Structural and Functional Characterisation of the Cold-Inducible RNA-Binding Protein CIRP and its Application to Enhanced Recombinant Protein Production

2014

Tilbe Creigh-Pulatmen

A thesis submitted to the University of Kent for the degree of Doctor of Philosophy

University of Kent
School of Biosciences

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

Tilbe Creigh-Pulatmen

12th September 2014

Acknowledgements

Firstly I would like to thank my PhD supervisors Dr. Richard Williamson, Professor Mark Smales and Dr Mark Howard for their endless support, encouragement and guidance throughout my PhD. I have been very lucky to have all three of you on board.

I would also like to thank Dr Michelle Rowe for her assistance and expertise with NMR spectroscopy and also Dr Anne Roobol and Jo Roobol for their help with the mammalian studies. A big thanks to Professor Anne Willis, Dr Helen King and Dr Amandine Bastide at the University of Leicester for welcoming me to their laboratory and their help with the radiolabelled shift assays.

Thank you to all my lovely friends in the lab for all the laughs and for always being available for support, cheering me up on difficult days and the many enjoyable chats over cups of tea.

Finally a very big thanks to my mother for helping me come so far in my academic career and sister for their endless support and encouragement, always putting a smile on my face when I most needed it. Last but not least I would like to thank my husband for always being by my side, keeping me going and pretending to understand what I was talking about, turning my tears to laughter! I could not have done it without you.

Contents

Abbreviations

A/CDk2- Cyclin A/ Cyclin Dependant Kinase 2

AMPK- Adenosine Monophosphate Activated Protein Kinase

ARIA- Ambiguous Restraints for Iterative Assignment

ATP- Adenosine Triphopshate

ATR- Ataxia telangiectasia and Rad3 related

BSA- Bovine Serum Albumin

C-CIRP- C-terminal Domain of CIRP

CCPN- Collaborative Computing Project for NMR

CHO- Chinese Hamster Ovary

CIRP- Cold Inducible RNA Binding Protein

CMV- Cytomegalovirus

CNS- Cyrstollography and NMR System

Co-IP- Co-immunoprecipitation

CSD- Cold Shock Domain

CSP- Cold Shock Protein

CV- Coloumn Volume

D_2O- Deuterium Oxide

DAMP- Damage Associated Molecular Patterns

DANGLE- Dihedral Angles from Global Likelihood Estimate

DAPI- 4',6-diamidino-2-phenylindole

dATP- Deoxyadenosine Triphosphate

dCTP- Deoxycytidine Triphosphate

dGTP- Deoxyguanosine Triphosphate

dNTP- Deoxyribonucleotide Triphosphate

dTTP- Deoxythymidine Triphosphate

Dyrk1b- Dual Specificity Tyrosine-Phosphorylation-Regulated Kinase 1B

ECL- Enhanced Chemiluminescence

EDTA- Ethylenediaminetetraacetic acid

eIF- Eukaryotic Translation Initiation Factor

EMSA- Electro Mobility Shift Assay

ERK- Extracellular Signal Regulated Kinase

ESI- Electrospray Ionization

ExPASY- Experimental Protein Analysis System

FITC- Fluorescein isothiocyanate

FL-CIRP- Full-length CIRP

FPLC- Fast Protein Liquid Chromatography

HEK- Human Embryonic Kidney

HMGB1- High Mobility Group Box 1

hnRNPA1- Heterogeneous Nuclear Ribonucleoprotein A1

hnRNP a18- Heterogeneous Ribonucleoprotein Particle a18

HSQC- Heteronuclear Single Quantum Coherence

IDP- Intrinsically Disordered Protein

IFN-γ- Interferon γ

IgG4- Immunoglobulin G4

IPTG- Isopropyl β-D-1-thiogalactopyranoside

IRES- Internal Ribosome Entry Site

Kb- Kilobase

lac1- Lactose Operon 1

LB- Louria-Bertani

MCRE- Mild Cold Responsive Element

MD2- Myeloid Differentiation Factor 2

Mr- Relative Molecular Mass

mTOR- Mammalian Target of Rapamycin

MWCO- Molecular Weight Cut-off

N-CIRP- N-terminal Domain of CIRP

NF-Y- Nuclear Transcription Factor Y

NMR- Nuclear Magnetic Resonance

NOE- Nucler Overhauser Effect

NOESY- Nuclear Overhauser Effect Spectroscopy

NSR- Nuclear Spin Relaxation

PABP- Poly Adenine Binding Protein

PAS- Polyadenylation Site

PBS- Phosphate Buffered Saline

PDB- Protein Data Bank

Pi- Phosphate

ppm- Part per Million

PRMT1- Protein Arginine N-Methyltransferase 1

PTB- Polypyrimidine Tract Binding Protein

Rbm3- RNA Binding Motif

RMSD- Root Mean Square Deviation

RNP- Ribonucleoprotein

RPA2- Replication Protein A2

rpm- Revolutions per Minute

RRM- RNA Recognition Motif

SDS-PAGE- Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SG- Stress GranuleseIF2α- Eukaryotic Translation Initiation Factor 2α

SOB- Super Optimal Broth

TEMED- Tetramethylethylenediamine

TLR4- Toll Like Receptor 4

TNFα- Tumour Necrosis Factor α

TOCSY- Total Correlation Spectroscopy

TRITC- Tetramethylrhodamine

TRX- Thioredoxin

TSS- Transcription Start Site

UTR- Untranslated Region

UV- Ultra Violet

v/v- Volume per Volume

V- Volts

VEGF- Vascular Endothelial Growth Factor

w/v- Weight per Volume

WT- Wild Type

YB1- Y Box Binding Protein 1

Abstract

Expression of the RNA binding protein CIRP is up regulated in mammalian cells upon perception of mild cold shock (27-32°C), conditions that can result in enhanced recombinant protein yields from mammalian cells and improved protein folding and activity. CIRP also binds to key proteins involved in the control of mRNA translation initiation, potentially acting as a bridge between the RNA and protein synthesis machinery. CIRP has two domains, an N-terminal RNA binding domain and an arginine/glycine rich C-terminal domain that is natively disordered. The N-terminal domain includes two RNA-binding sites, RNP1 and RNP2 that are conserved across many RNA binding proteins. Here, the RNA binding of the N-terminal domain of CIRP was investigated by introducing mutations into the RNP1 and RNP2 RNA binding sites and monitoring subsequent RNA binding using electromobility shift assays and NMR. These studies show that the F49 and F9 residues in these regions are important for RNA interactions. Further, NMR dynamics studies showed that the region (β2-β3 loop) just before the RNP1 sequence that includes the F49 residue has increased motion compared to the rest of the protein. Chemical shift analysis was used to map those residues in CIRP involved in RNA binding that mapped onto the RNP1 and RNP2 sites. Mutation of the F49 and F9 sites to Ala residues disrupted RNA binding as shown by NMR studies. Mutation of each residue resulted in some conformational change in the structure of the domain as determined by HSQC-NMR, particularly for the F49 residue in RNP1. The mutation of one of the phenylalanine residues affected the chemical shift of the other, confirming their proximity in space. The C-terminal is a natively disordered domain but the studies presented here suggest this plays a role in RNA binding and ligand specificity. Cell lines stably expressing CIRP were also generated to further investigate the function of CIRP and to determine binding partners. CIRP was found to interact with the translation initiation factor eIF4G, both the full length CIRP and N-CIRP molecule, suggesting that binding to 4G occurs through the N-terminal domain. The over-expression of CIRP enhanced recombinant firefly luciferase expression when the luciferase mRNA contained a CIRP 3'UTR binding sequence. CIRP over-expression also enhanced growth of Chinese hamster ovary cells at 37°C but not at reduced temperature. From these studies it is hypothesised that CIRP can bind specific mRNAs and enhance their expression through its interaction with both the specific mRNA and the translation machinery via eIF4G leading to increased protein expression of the target mRNA under conditions of mild cold-stress (32°C).

Chapter 1

Introduction

1.1 Introduction

The first recombinant protein produced by mammalian cells and used as a therapeutic was human tissue plasminogen activator approved for clinical use in 1986 (Becerra, Berrios, Osses, & Altamirano, 2012). The recombinant protein biopharmaceutical product market has ever since been rapidly growing with sales reaching 138 billion US dollars in 2010 (Vergara et al., 2014). As there is an ever-increasing demand for recombinant protein biopharmaceuticals, much research has being undertaken to improve recombinant protein production and quality of the protein product in mammalian cells. Around 60-70% of recombinant protein therapeutics are produced in mammalian cells where the CHO (Chinese hamster ovary) cell expression host is most commonly used (Kim, Kim, & Lee, 2012; Kumar, Gammell, Meleady, Henry, & Clynes, 2008).

Mammalian cells are the most commonly used expression system for biotherapeutic recombinant protein production destined for use in the clinic as they have the ability to undertake the required human-like assembly, folding and post-translational modifications of such proteins for their activity and to avoid antigenicity issues (Lim et al., 2010). Mammalian production systems have improved in terms of the viable cell concentration that can be achieved in the bioreactor (from 1-2 to 10-15 million cells/ml) as well as the recombinant protein productivity that can be obtained (from 0.05-0.1 to 1-5+ g/l) in the past 20 years (Vergara et al., 2014). Productivity of cells is proportional to the number of viable cells and culture viability and longevity. To date, researchers have investigated and adopted a number of strategies to increase viable cell concentration, extending culture longevity by inhibiting apoptosis and by application of environmental stress (Kou et al., 2011). In particular there is much interest in the area of reduced temperature cultivation to prolong culture viability and hence the productive lifetime of a culture.

Mild cold-shock (27-32°C) perception by mammalian cells has been reported to result in improved recombinant protein yields and to facilitate improved protein folding and product quality (Masterton et al 2010, Al-Fageeh et al., 2006). A study by Fox et al (2005) reported that enhanced interferon gamma expression from CHO cells at reduced temperatures was the result of elevated mRNA levels compared to those

observed at 37°C. The cold stress response(s) of mammalian cells have not only been taken advantage of in the biopharmaceutical industry, but have also been implicated and applied in adaptive thermogenesis, cold tolerance, storage of tissue, organ and cells and in the treatment of brain damage in the medical setting (Al-Fageeh, Marchant, Carden, & Smales, 2005).

The response of cells to elevated temperatures (heat-shock) have been extensively studied in prokaryotes and eukaryotes, this response involves the coordinated expression of a large number of heat shock proteins (HSPs) (Li & Srivastava, 2004). However, the mechanism(s) by which mammalian cells respond to cold-stress are much less well defined and how these relate to improved recombinant protein production under mild hypothermic conditions is not well understood. It is thought that the cellular responses involve coordination and modulation of transcription, translation, the cell cycle and cell cytoskeleton organization (Robool et al, 2009). Under conditions of mild hypothermia (32°C), mammalian cells can survive and continue to divide (Fujita, 1999) and culturing of the cells at these temperatures can result in prolonged generation time, maintenance of cell viability, supressed release of waste products, delayed apoptosis and improved tolerance against shear stress and reduced proteolytic degradation of the recombinant protein (Yoon et al, 2003). There have been an increasing number of reports where cultivation of mammalian cells at low temperatures has been used to improve recombinant protein production (Al-Fageeh & Smales, 2006; Fogolín, Wagner, Etcheverrigaray, & Kratje, 2004; Fox et al., 2005; Huang et al., 2010; Roobol, Carden, Newsam, & Smales, 2009; Wagstaff, Masterton, Povey, Smales, & Howard, 2013;Yoon, Song, & Lee, 2003). The cold-shock response(s) are discussed in more detail below.

1.2 The Cold Shock Response in Mammalian Cells

Unlike the heat shock response, which is largely conserved across eukaryotes, the mammalian cell response to sub-physiological temperatures has been studied in only a few organisms and in less detail and is therefore less well understood (Al-Fageeh et al., 2005). During therapeutic recombinant protein manufacturing, mammalian cells are normally cultured at 37°C to replicate *in vivo* conditions, however there have been an increasing number of reports which use lower temperature (28-34°C) culturing to improve recombinant protein yields (Masterton & Smales, 2014). At very low temperatures proliferation is compromised but at mild hypothermic conditions e.g. 32°C cells are able to survive and proliferate (Al-Fageeh et al., 2005).

Cold stress can potentially result in two different stresses being imposed upon mammalian cells; one related to reduced temperature and one that is related to changes in available oxygen concentration due to low temperature (Ohsaka, Ohgiya, Hoshino, & Ishizaki, 2002). Cells respond to sub-physiological conditions by the up-regulation of a small group of proteins called cold shock proteins (CSPs). Both prokaryotes and eukaryotes respond by synthesising cold shock proteins but there is little conservation between prokaryotes and eukaryotes in the nature of these proteins other than the fact most are involved in binding RNA. This is discussed in more detail below.

The cellular processes that are reported to be affected as a result of the responses of mammalian cells to cold stress are summarised in figure 1.1. Expression of cold shock proteins has been shown to ensure accurate and enhanced translation of mRNA and at mild hypothermic temperatures mRNA levels are generally increased (due to increased stability/reduced turnover) and degradation of RNA decreased (Marchant, Al-Fageeh, Underhill, Racher, & Smales, 2008; Sahara, Goda, & Ohgiya, 2002; Sonna, Fujita, Gaffin, & Lilly, 2002). Early experiments investigating temperature manipulation in cell culture date back to 1965, where it was shown that an accumulation of cells in mitosis occurs at sub physiological temperatures (Rao & Engelberg, 1965).

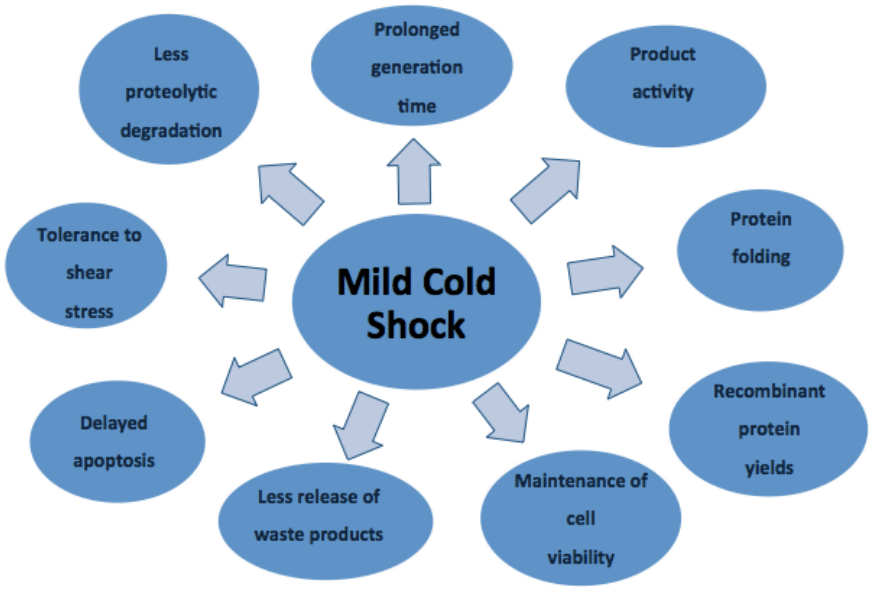

Figure 1.1 Summary of the reported cellular processes that are modified and cell responses observed in mammalian cells upon exposure to cold stress.

It has been suggested that mammalian cells respond to cold stress by attenuating global cap dependant mRNA translation (Ermolenko & Makhatadze, 2002), and that instead translation of specific transcripts can occur via a cap independent mechanism where an internal RNA structure within the 5'UTR of the mRNA called an IRES is involved in recruiting the ribosome (Holcik & Sonenberg, 2005). IRESs are sequences which allow translation initiation of an mRNA independent of the 5' cap recognition by initiation factors (Baranick et al., 2008). The cell membrane has also been shown to be remodelled upon perception of cold stress conditions where the rigidity is increased, this remodelling mitigating against this increased rigidity. The actual cellular architecture is maintained under mild hypothermia conditions and is only disrupted at more drastic low temperatures such as 4°C. As such, for some mammalian cells the cellular architecture at 37°C and at mild cold conditions (32°C) are indistinguishable (Roobol et al., 2009).

Another well characterised response to cold-shock is the arrest of the cell cycle at the G1 phase and this phase has been correlated with increased cellular productivity in mammalian cells (Fussenegger, 2001). Mammalian cells are able to proliferate at 32°C but when the temperature drops below approximately 22-25°C, proliferation ceases, even though it can be restored upon return to 37°C (Roobol et al., 2009). As a

result, the positive effects observed in terms of recombinant protein expression from mammalian cells at reduced temperatures are achieved under mild cold-shock conditions (27-35°C) rather then extreme temperature reductions (Roobol et al., 2009). The reported cell cycle arrest of mammalian cells at the G1 phase is due to the activation of the p53 signalling pathway, p53 being activated as a result of activation of ataxia telangiectasia mutated and Rad3-related kinase (ATR)-p53-p21 pathway. The activation of ATR is thought to be due to changes in the lipid composition of the cell membrane at reduced temperature (Roobol et al., 2011). At reduced temperatures, the lipid composition of the cell membrane changes leading to increased rigidity and reduction in cellular membrane associated functions (Los & Murata, 2004; Panadero, Pallotti, Rodríguez-Vargas, Randez-Gil, & Prieto, 2006; Roobol et al., 2009; Zhang, Zhao, & Ma, 2007). The levels of polyunsaturated fatty acids are increased under cold conditions to maintain fluidity and this activates ATR and its down stream effectors, the p53-p21 signalling pathway, leading to cell cycle arrest (Roobol et al., 2011).

With regard to mRNA translation in mammalian cells upon cold-shock, one of the mammalian cold shock induced proteins, CIRP (discussed in detail in section 1.4) interacts with the 3'UTR mRNA of ATR and it has been proposed that this interaction results in enhanced translation of ATR mRNA (Yang et al., 2010). It has been suggested that under cold conditions an increase in ATR protein is required for the activation of cell cycle arrest. Through this interaction with the 3'UTR and the interaction between CIRP and eIF4G , CIRP may act as a bridge to facilitate the translation of ATR mRNA under cold stress in a similar way to the function of the polyA-binding protein (Yang et al., 2010). This is discussed in more detail later in this chapter (section 1.4.3).

The cold shock of mammalian cells has also been linked to the formation of cytoplasmic structure called stress granules (SGs) (Hofmann, Cherkasova, Bankhead, Bukau, & Stoecklin, 2012). These structures contain stalled translation pre-initiation complexes (Hofmann et al., 2012). The attenuation of global protein synthesis at the initiation stage has been shown to be controlled by the phosphorylation of the translation initiation factor eIF2α (Holcik & Pestova, 2007; Joshi, Kulkarni, & Pal,

2013; Maquat, Tarn, & Isken, 2010). Further, there has been recent data that has suggested that activation of AMP-activated protein kinase (AMPK) is initiated under cold-shock conditions as well as inhibition of mTOR signalling and reduced mitochondrial function in parallel with attenuation of protein synthesis (Hofmann et al., 2012). Another response of mammalian cells to cold shock that is involved in translational control is that some microRNA (miR) expression is elevated or reduced at sub physiological temperatures. microRNAs act as negative regulators of translation of specific targets (Barron et al., 2011). With respect to cold, upon temperature shift to 31°C, Barron and colleagues reported that six microRNAs were up-regulated (miR-518d, miR-30e, miR-345, miR-219, miR-489 and miR219) and four were down-regulated (miR-7, miR-101, miR-320, miR-199) where miR-7 is specifically associated with cell cycle arrest (Barron et al., 2011).

The responses of mammalian cells to temperature downshift therefore generally involves suppression of protein synthesis and it has been suggested that there are 5 general mechanisms by which these responses are elicited – some passive and others active responses of the cells. These are; (1) increased transcription of specific target genes via elements in the promoter region of such genes, (2) alternative pre-mRNA splicing (3) cold shock specific IRESs (internal ribosome entry segments) in mRNAs resulting in the preferential and enhanced translation of such mRNAs upon cold shock, (4) reduction of RNA degradation, (5) a general reduction in transcription and translation (Sonna et al, 2002). The supressed protein synthesis is associated with growth arrest, but paradoxically, the up-regulation of specific cold inducible proteins. There has been little characterisation of those proteins whose expression is up-regulated upon cold-shock in mammalian cells, however there are two that are considered cold-shock inducible proteins (although they are induced by other stresses as well), Rbm3 (RNA binding motif protein 3, NCBI: AAH06825.1) and CIRP (cold inducible RNA binding protein or a18 hnRNP, NCBI: AAC04895.1), which are up regulated upon mild hypothermia but also by hypoxia and other stresses too (Lleonart, 2010).

Although the cold-shock response in mammalian cells is poorly characterised compared to the heat shock response, the cold shock response has been widely

investigated in a number of prokaryotes where the responses to such temperature fluctuations are better defined. Bacterial cold shock proteins (CSPs) are highly conserved, although the *E.coli* cold response is the most characterised (Al-Fageeh & Smales, 2006). A quick change in temperature from 37°C to 10°C leads to the expression of approximately 27 CSPs in *E.coli* while global transcription and translation is down regulated (Gualerzi, Maria Giuliodori, & Pon, 2003). These CSPs are required for mRNA stability and bind single stranded RNA to enhance translation (Bae, Jones, & Inouye, 1997; Giuliodori, Brandi, Gualerzi, & Pon, 2004), in a manner that may be similar to that proposed for, but not yet confirmed, the two mammalian CSPs.

The major CSPs in *E.coli* consist of 9 homologous genes and have been named in alphabetical order from CspA to CspI. The synthesis of CspA, B, G and I is very sensitive to temperature and their expression is dramatically up-regulated upon cold perception (Horn, Hofweber, Kremer, & Kalbitzer, 2007). Synthesis of CspA reaches approximately 10-13% of total protein synthesis under cold stress at 10°C. CspA is a 7.4 kDa protein, and similar to the mammalian RNA binding proteins, contains the RNP1 and RNP2 (see section 1.3 for a detailed description of these) sequences that bind RNA. This again provides further evidence that a key response to cold-shock is the up-regulation of proteins involved in RNA binding/chaperoning.

CspA is the most well characterised cold-shock protein. For example, promoters have been identified which drive cold shock protein *CspA* gene expression in *E.coli* and has led to the development of a cold-inducible expression vector which contains the CspA gene promoter and the CspA 5'UTR sequence and has been shown to improve the expression of recombinant proteins in *E.coli* under sub-physiological conditions (Takaya, Nomura, & Kato, 2002). However, even though the bacterial CSPs seem to have a similar function of RNA binding, and use a similar mechanism to mammalian CSPs, no close homology is found between the bacterial and mammalian CSPs outside of the presence of the RNP1 and RNP2 RNA binding sequences. Figure 1.2(a) below shows a sequence and secondary structure alignment of the *E.coli* CspA and mouse CIRP sequences which confirms little similarity between the two (approx. 10% as determined by the Clustal alignment tool). The PDB structure of CspA and the

RNA binding domain of human CIRP also confirm that there is no homology between them and the two proteins have different folds (figure 1.2 (b)). The RNA binding domain of CIRP is composed of two α-helix and four β-strands in an order of β1α1β2β3α2β4 whilst CspA has a secondary structure composition of β1α1β2β3α2β4β5. The RNA binding motifs RNP1 and RNP2 of CIRP are located on strands β3 and β1 but are located on strands β2 and β3 in CspA.

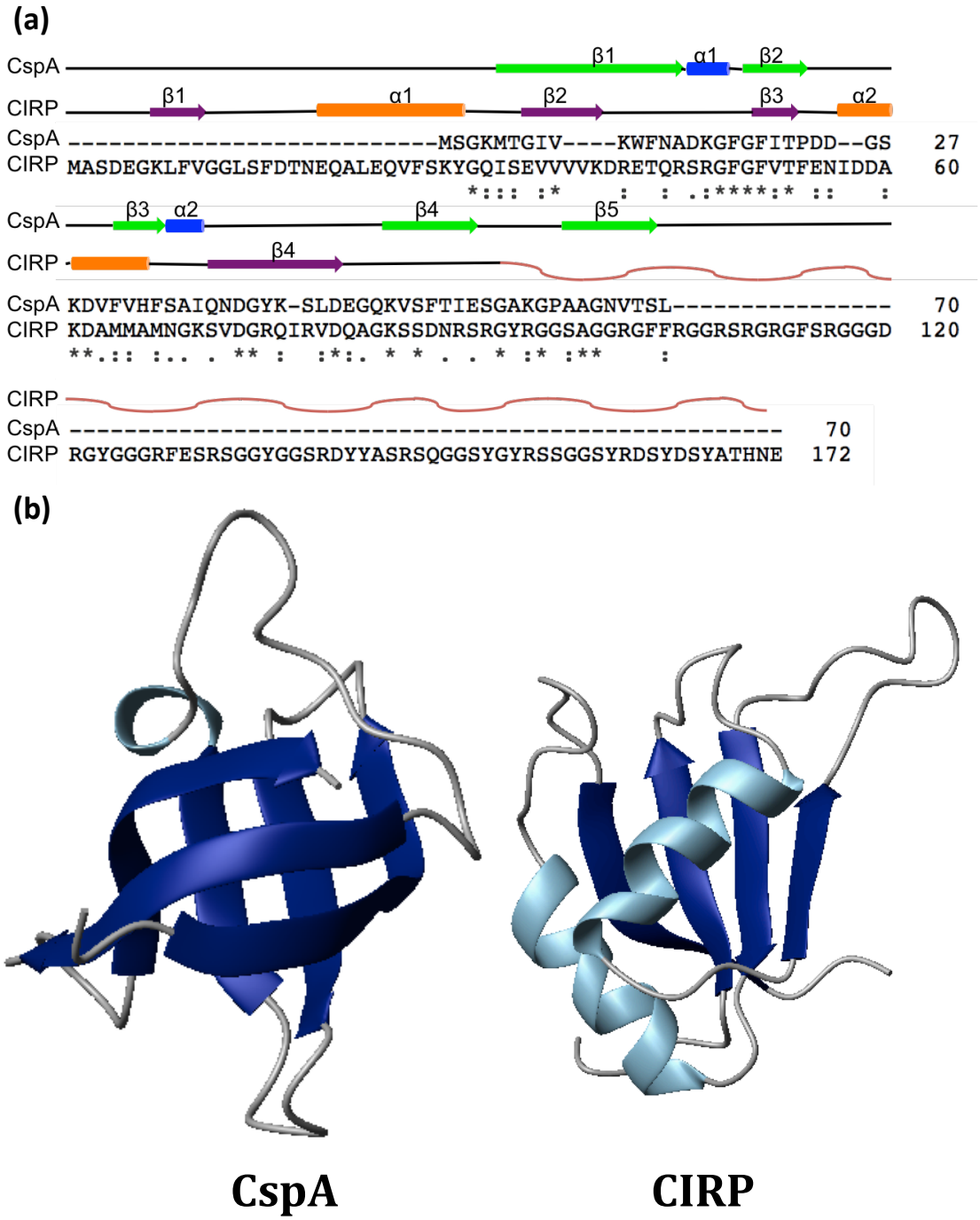

Figure 1.2 Sequence and structural comparison of the bacterial cold shock protein CspA and mammalian cold shock protein CIRP (a) Amino acid sequence and secondary structure alignment of CspA and mouse CIRP. (b) Structural comparison of CspA PDB structure 1MJC and human CIRP PDB structure 1X5S

1.3 Cold Shock Proteins

As described in section 1.2, prokaryotic and eukaryotic cells generally respond to sub physiological temperature by up regulating a small group of proteins called cold shock proteins (CSP) (Al-Fageeh & Smales, 2006). Further, these CSPs are involved in the modulation of a number of cellular processes including transcription, translation and DNA-dependent processes like recombination and repair by their mechanism of binding to single stranded nucleic acids (Chaikam & Karlson, 2010). Whilst CSPs are induced upon a reduction in temperature and are involved in adaptation to low temperatures, some are also functional under normal temperature growth conditions (Ermolenko & Makhatadze, 2002; Horn et al., 2007). Cold shock proteins are characterised by the presence of a cold shock domain with two consensus RNA binding motifs which are also present in RNA recognition motif (RRM) type RNA binding proteins and as such is referred to as an RNA binding domain (Manival, Ghisolfi-Nieto, Joseph, Bouvet, & Erard, 2001).

The cold shock domain (CSD) is one of the oldest and most well-conserved nucleic acid binding domains from bacteria to higher animals to plants (Chaikam & Karlson, 2010). Most functions associated with cold shock proteins are through their RNA binding activity, where upon hypothermia translocation of the CSP from the nucleus to cytoplasm occurs (Chaikam & Karlson, 2010). There are two RNA binding motifs that are conserved across the majority of RNA binding proteins and these are termed the RNP1 and RNP2 sequence. RNP1 and RNP2 are generally located on the β1 and β3 strands and are located adjacent to each other in the structure (Mihailovich, Militti, Gabaldón, & Gebauer, 2010) (figure 1.3).

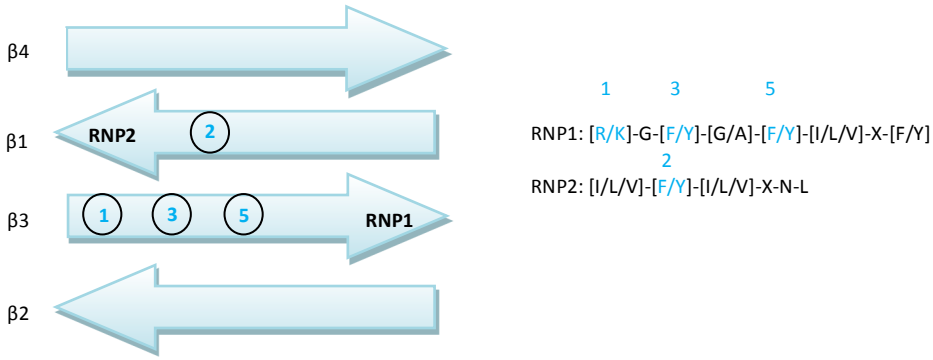

Figure 1.3 Model showing RNP1, RNP2 motifs and residues essential for RNA binding.

The CSD was initially identified in bacterial proteins. In prokaryotes, CSDs are contained within small (approximately 7 kDa) proteins that have only a single cold shock domain (Chaikam & Karlson, 2010). In animal and plant systems C-terminal auxiliary domains can be present as well as an N-terminal CSD (Chaikam & Karlson, 2010). The C-terminal domains of most of these eukaryotic CSD containing proteins have unknown functions but it is proposed this domain has non-specific affinity for RNA and to be docking sites for protein ligands (Evdokimova & Ovchinnikov, 1999; Wolffe, 1994). The 3D protein structure of a number of prokaryotic and eukaryotic CSPs have now been solved, although no 3D structures of plant CSPs have yet been reported. Proteins structures described as cold shock proteins in the protein data bank are listed below in table 1.1

Protein Name	**Organism**	**PDB ID**
CSPB	*B. Subtilis*	1CSP
KIAA0885	Human	1WQF
CSPA	*E.Coli*	1MJC
TmCsp	*T. Maritima*	1G6P
Cold Shock Like Protein	*R. Ricketsii*	2LSS
Enoyl-ACP Reductase	*B. Subtilis*	3O1F
Yfia	*E.Coli*	1N36
YB1	Human	1H95
Lm-CSPa	*U. Monocytogenes*	2LXJ
Cold Shock Protein 1	*T. Thermophilus*	3AOJ
Ta-Csp	*T. aquatics*	2MO1

Table 1.1 Cold shock proteins, which have 3D structures solved as listed in the PDB to date.

As described in section 1.2, CSPs are generally thought to act as RNA chaperones and to be involved in the regulation of gene expression under cold-shock conditions via modulation of global or specific RNA transcription, translation and RNA turnover. Reports suggest that CSPs bind single stranded nucleic acids with a K_d value in the micro to nano molar range (Mihailovich et al., 2010). The aromatic residues found in the RNP1 and RNP2 sequences are essential for RNA binding by enabling base stacking (Maris, Dominguez, & Allain, 2005) (figure 1.3). Bacterial CSPs are also

reported to be involved in the maintenance of chromosome structure and DNA replication as well as their RNA binding function (Gaudreault, Guay, & Lebel, 2004; Ise et al., 1999). The involvement of CSPs in modulation of mRNA translation is via binding to mRNA and stabilising these RNA ligands. For example, CspA destabilises RNA secondary structures that form in the cold and that inhibit ribosome scanning and thus mRNA translation (Jiang, Hou, & Inouye, 1997), whilst CspE melts partially double stranded and hairpin structures and binds polyA tails (Phadtare & Severinov, 2005).

Eukaryotic CSPs have also been reported to participate in DNA maintenance. For example, the human YB1 protein which is a member of the Y-box protein family, has affinity towards secondary structure in damaged DNA and aids DNA repair. (Gaudreault et al., 2004). Y-box binding proteins are DNA and RNA binding proteins functioning as nucleic acid chaperones and also have functions in proliferation, differentiation and stress responses (Eliseeva, Kim, Guryanov, Ovchinnikov, & Lyabin, 2011). Eukaryotic CSPs are also involved in translation regulation specifically during the initiation stage, for example the YB1 protein promotes the 43S pre-initiation complex to scan the 5'UTR of mRNA (Evdokimova et al., 1998; Nekrasov et al., 2003). Although there are now a large number of processes that CSPs are reported to be involved with modulating upon cold perception, little is known about their RNA and protein binding mechanisms and ligands and how the binding of these works at a mechanistic level to elicit the desired cellular response(s) to the cold.

1.3.1 The Mammalian CSPs, CIRP and Rbm3

There are two mammalian cold shock proteins that have been characterised, CIRP (Cold inducible RNA binding protein) and Rbm3 (RNA binding motif protein 3) (Derry, Kerns, & Francke, 1995; Sheikh, 1997). CIRP was initially identified in a screen for transcripts that are induced upon DNA damage due to UV stress (Fornace, Alamo, & Hollander, 1988; Sheikh, 1997) and subsequently CIRP was the first cold induced protein identified in mammalian cells (Nishiyama, Itoh, et al., 1997). Rbm3 was initially discovered in cancer cell lines upon cold shock stress (Danno et al., 1997). These proteins are also up-regulated upon other stresses such as hypoxia and UV irradiation (Lleonart, 2010).

CIRP and Rbm3 belong to highly conserved glycine rich RNA binding family of proteins, which contain an RNA binding N-terminal cold shock domain containing the RNP1 and RNP2 sequences and a glycine rich C-terminal domain (Al-Fageeh & Smales, 2009). The two proteins are highly homologous (Liu et al., 2013) (figure 1.4), they share a 61% identity at the amino acid level over the entire sequence and an 80% identity in the N-terminal RNA binding domain. However, as stated above, they show little homology to the CSPs induced upon cold-shock in bacterial systems (Roobol et al., 2009).

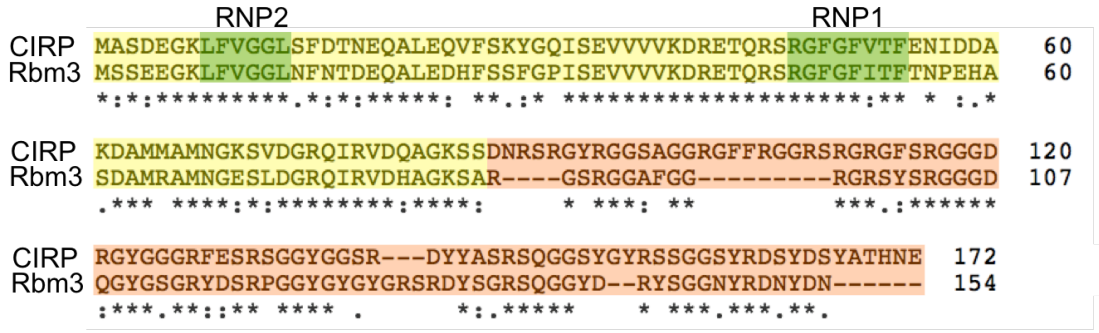

Figure 1.4 Amino acid sequence alignment of human CIRP and human Rbm3 proteins. The yellow region is the N-terminal RNA binding domain and the orange region is the glycine C-terminal region. The RNA recognition motifs RNP1 and RNP2 are highlighted in green.

The overexpression of CIRP and RBM3 in mammalian cells can lead to an increase in general protein synthesis (Dresios et al., 2005; Sakurai et al., 2006; Smart et al., 2007). Both proteins are thought to act as RNA chaperones and modulate transcription and translation under cold stress (Liu et al., 2013) by interacting with components of the transaltional machinery to stimulate translation. The meraction of CIRP with the translational machinery is discussed in more detail in section 1.4.3. The cellular control of the response to cold stress and recovery have also been suggested to be regulated or partially regulated by p53 and eIF3i levels (Roobol et al., 2009) and it is higly likely that CIRP and RBM3 have cruicial roles in this process, possibly through interaction with their mRNA to stabilise them and upregulate their protein synthesis.

CIRP has been associated with modulation of both cap-dependant and cap-independent mRNA translation, but it has been shown that RBM3 does not posess the sequence required for IRES sites for cap-independent translation (Baranick et al.,

2008). It has been suggested that Rbm3 is involved in regulation of global protein synthesis at 37°C and 32°C via the alteration of microRNA containing complexes (Dresios et al., 2005) and that CIRP plays a key role in the activation of the extra cellular signal regulated kinase pathway, protecting the cells from apoptosis by induction of tumor necrosis factor TNFα (Sakurai et al., 2006). CIRP and Rbm3 translocate from the nucleus to the cytoplasm upon cold stress and the localisation is reported to be regulated by methylation of arginine residues (De Leeuw et al., 2007). Rbm3 methylation has been shown to take place by stimulation of primary cortical neurons with a glutamate receptor agonist (Pilotte, Cunningham, Edelman, & Vanderklish, 2009). As well as their proposed role in protein synthesis through interaction with the translation machinery and specific mRNAs under cold stress, the two proteins have also been shown to be proliferative and proto oncogenic proteins. Their inhbition has been shown to decrease cell proliferation (Zeng, Kulkarni, Inoue, & Getzenberg, 2009). The possible cellular pathways in which CIRP and Rbm3 are suggested to be involved, and hence through which they can elicit their responses, are summarised in figure 1.5 below (Lleonart, 2010).

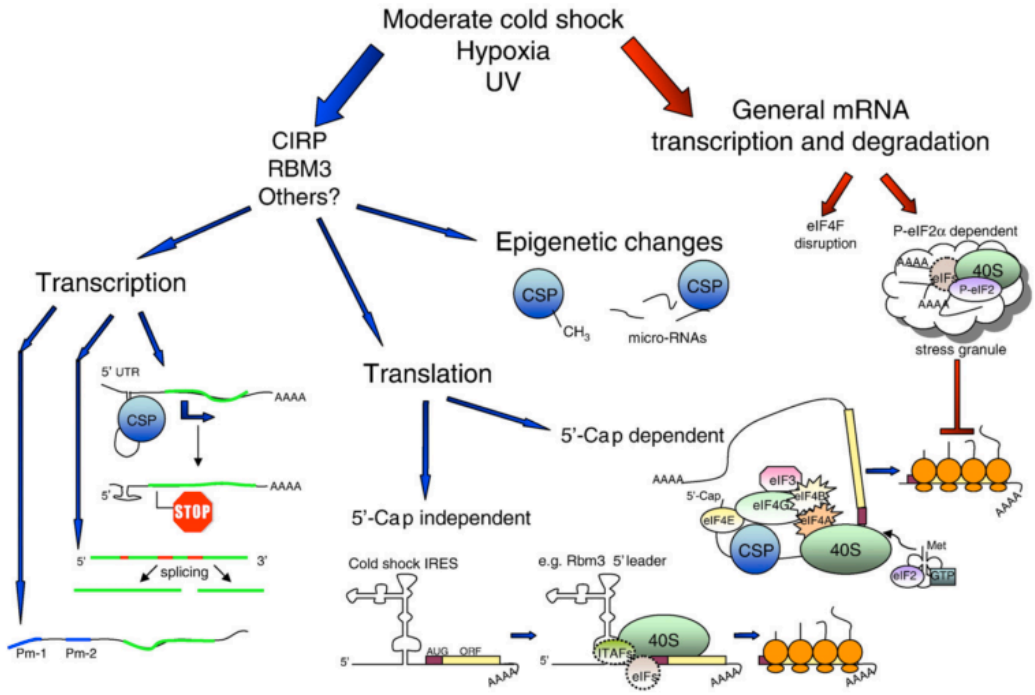

Figure 1.5 Pathways that CIRP and RBM3 are suggested to interact with, leading to modulation of transcription and translation under cold stress conditions in mammalian cells. (Lleonart, 2010)

1.3.1.1 Rbm3

The gene for the Rbm3 protein is found on the Xp11.2 region of the human chromosome and encodes for a 157 amino acid, 17 kDa protein closely related to other human RNA-binding proteins and shares 94% identity with the 18 kDa mouse protein (Derry et al., 1995). Rbm3 is expressed in the pancreas, adrenal gland, cerebellum, lung, spleen, intestine, uterus and testis and Rbm3 mRNA levels have been found to be increased in HepG2, NC65, HeLa, T24, K562 and TAMA26 cultured cells upon a temperature shift from 37°C to 32°C with maximal expression between 6-12 hours after temperature shift (Danno, Itoh, Matsuda, & Fujita, 2000; Y. Liu et al., 2013; Lleonart, 2010). During cold shock, oxygen levels in cells are reduced (hypoxia) and consistent with this Rbm3 up-regulation also occurs in response to hypoxia (Wellmann et al., 2004). There is also evidence to suggest that Rbm3 alters global translational under cold stress by altering microRNA levels, counteracting their effect (Dresios et al., 2005). microRNAs either induce the cleavage of their target mRNAs or inhibit their translation by interacting with the mRNAs through the RNA-inducing silencing complex (White-Grindley & Si, 2006). In this way, Rbm3 removes translation repression of microRNAs by removal of the microRNA itself. With regard to translation, over expression of Rbm3 under mild hypothermia was reported to result in increased levels of 80S monosomes and polysomes suggesting increased association of ribosomal subunits during translation initiation and direct interaction of Rbm3 with 60S ribosomal subunits has been reported (Dresios et al., 2005).

As well as a reported role in modulating translation under cold stress conditions, Rbm3 has been reported to also be involved in cytokine dependant proliferation (Baghdoyan, Dubreuil, Eberlé, & Gomez, 2000), poxvirus replication (Wright, Oswald, & Dellis, 2001), development of cancer (Baldi et al., 2003) and apoptosis (Kita, 2002). It is considered to be a proto-oncogene as it enhances the stability and translation of mRNAs for cyclooxygenase 2, VEGF and interleukin 8 and its overexpression has been shown to induce NIH3T3 cells to grow in a cancerous type manner (Cok, Acton, Sexton, & Morrison, 2004; Sureban et al., 2008; Zeng et al., 2009). Its expression in sertolli cells of mice during testis development may also suggest a developmental function (Danno et al., 2000)

1.4 Cold inducible RNA binding protein (CIRP)

As mentioned previously in this chapter, expression of CIRP (Cold-Inducible-RNA-Binding-Protein) is up-regulated in mammalian cells upon perception of mild cold-shock (27-32°C) (Nishiyama, Higashitsuji, et al., 1997). CIRP has also been suggested to bind to key proteins involved in the control of mRNA translation initiation, potentially acting as a bridge between the mRNA and the protein synthesis machinery (Yang, Weber, & Carrier, 2006). RNA binds to the RNP1 and RNP2 sequence present on the N-terminal domain (figure 1.6) of CIRP as described in Chapter 3 (section 3.3.3). CIRP also has an arginine/glycine rich C-terminal domain (figure 1.6), which is natively disordered with an unknown function (Hseih, 2010).

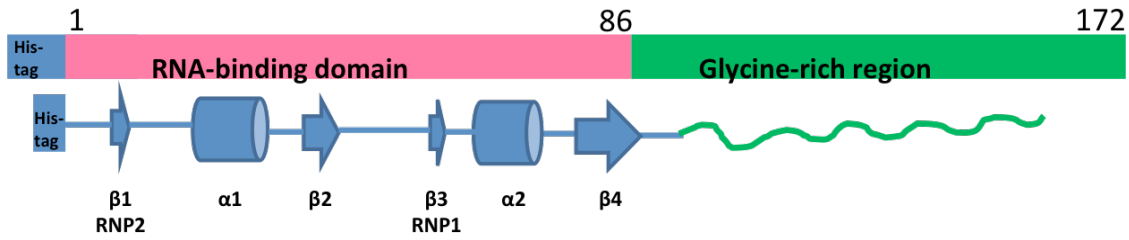

Figure 1.6 Schematic representation of the N-CIRP and full-length CIRP constructs.

CIRP transcription is up regulated in response to mild cold shock (Fajita, 1999). The promoter regions responsible for this up-regulation have also been mapped (Al-Fageeh & Smales, 2013). Human CIRP is usually nuclear but translocates to the cytoplasm upon perception of mild cold stress. The *CIRP* gene is located on the chromosomal locus 19p13.3 and encodes a 172 amino acid, 18 kDa protein with 95.3% identity to mouse CIRP (Nishiyama, Higashitsuji, et al., 1997). CIRP is thought to function as an RNA chaperone, i.e. protect and restore RNA conformations thus enhancing translation of its target RNA, preventing formation of mRNA secondary structure and it has been suggested that CIRP gene expression is controlled by cold responsive elements within its promoter (Al-Fageeh & Smales, 2013; Fujita, 1999).

CIRP has been shown to have a role in adapting cells to cold shock and in supressing cell proliferation. This was determined by overexpression of CIRP at 37ºC which caused extended doubling time and cells accumulating in the G1 phase of the cell cycle (Nishiyama, Itoh, et al., 1997). Up regulation of CIRP at physiological

temperature has also been shown to enhance recombinant protein production; 40% for gamma interferon (Tan, Lee, Yap, & Wang, 2008). CIRP presumably exerts its effects through its RNA binding activity but it is not well defined as to how it does this. CIRP has been reported to bind to a limited number of RNAs (discussed further in section 1.4.2, see table 1.2), one being thioredoxin mRNA (Yang et al., 2006), interacting specifically at the 3' untranslated region of the RNA, where it may be involved in modulating the efficiency of translation and RNA stability. RNA constructs that contain the CIRP binding region have enhanced resistance to degradation by RNases and cells transfected with an antisense CIRP vector have diminished response to stress (Yang et al, 2001).

1.4.1 Induction and Regulation of CIRP under Mild Hypothermia

The CIRP gene was initially isolated from mouse testis by cDNA cloning of an RNA recognition motif containing RNA binding protein (Nishiyama, Itoh, et al., 1997). As mentioned in the previous section, CIRP gene expression is transcriptionally controlled during mild cold shock. Al-Fageeh et al characterised the 5'UTR of the transcript encoding CIRP at varying temperatures and times (Al-Fageeh & Smales, 2009). The study showed the generation of three major CIRP transcripts with different transcription start sites that as a result had 5'UTRs of varied lengths, which were regulated in response to temperature and time. The longer of the 5'UTRs possesses IRES-like activity, which is associated with cap-independent translation. Upon temperature downshift two CIRP 5'UTRs with different transcription sites are generated, the levels of these transcripts control CIRP protein expression. Further, the stability of the long transcript up regulated at 32°C is enhanced. The major shorter transcript found at 37°C does not encode for full-length CIRP or may be inactive through being a truncated version (Al-Fageeh & Smales, 2009). This type of gene expression control via alternative 5'UTRs has been previously reported in a number of systems (Hughes, 2006; Tan, Mohandas, & Conboy, 2006).

Work by these investigators on the *CIRP* 5'UTRs showed that the expression levels of the longer 5'UTR transcripts are regulated over time at 32°C where their stability increases. During the first 6 hours of low temperature, levels of all 5'UTR transcripts are up-regulated and beyond this time the levels decrease except for the longer 5'UTR

whose levels and stability are increased. The IRES activity of the longer 5'UTR results in a several fold increase in translation. More recently, studies by Al-Fageeh et al discovered a core promoter localised between -264 and +1 relative to the transcription start site (TSS), which controls the basal transcription level of CIRP (Al-Fageeh & Smales, 2013). This promoter contains a TATA box at -37 and is a single peak promoter, which means that it has a compactly defined transcription start site and is involved in tightly regulated gene expression (Müller, Demény, & Tora, 2007). Further, a CAAT box at -92 and an inverted CAAT box at -187 were identified suggesting specific regulation of the CIRP gene and the deletion of these CAAT motifs led to a loss of CIRP promoter activity. The NF-Y CAAT motif binding protein binds to the CAAT box which is estimated to be present in 30% of eukaryotic gene promoters and is required for expression of many genes (Mantovani, 1998).

The longer CIRP 5'UTR mRNA was found to be transcribed by a different promoter upstream of the core promoter characterised as a broad peak promoter which is associated with ubiquitously expressed genes and does not contain a TATA box (Müller et al., 2007). Electro mobility shift assays (EMSA) showed that the two CIRP promoters bind to factors in a nuclear extract yielding different sized complexes, suggesting that the two promoter bind different factors. Luciferase reporter assays showed expression from the core promoter but little from the upstream promoter at 37°C. These findings suggest that when the global transcription efficiency is decreased upon cold stress, the CIRP alternative promoter, which is weakly active at 37°C, is induced and generates the CIRP mRNA containing IRES which leads to more efficient translation at these temperatures (Al-Fageeh & Smales, 2013).

Transcriptional regulation is thus key for gene expression of CIRP at mild hypothermic temperatures. Studies have found that the half life of CIRP mRNA is not highly influenced by cold stress (Al-Fageeh & Smales, 2009; Nishiyama, Itoh, et al., 1997). Further support for the transcriptional regulation of CIRP expression was the identification of a mild cold responsive element (MCRE) in the 5' flanking region of the mouse *CIRP* gene which was shown to enhance gene expression at 32°C in cultured mammalian cells (Sumitomo et al., 2012). The Sp1 transcription factor binds to this MCRE to induce CIRP expression and Sp1 down-regulation supressed CIRP

induction. It was also shown that the levels of Sp1 protein in the nucleus was increased at 32°C for binding to the MCRE region on the CIRP gene. This enhancer element increases the gene expression of CIRP 3- to 7-fold in various cell lines at 32°C.

RNA binding protein regulation can also be mediated by post-translational modifications (Yang et al., 2006). As described previously, CIRP is localised to nucleus and translocates into the cytoplasm upon mild cold shock. This process is thought to be regulated by methylation. CIRP has been identified as an arginine methylated protein by Bedford et al (Bedford, 2007). There has been an increased interest in the regulation of protein function through methylation and many RNA binding proteins have been shown to undergo methylation (Liu & Dreyfuss, 1995; McBride & Silver, 2001). Arginine residues are methylated by arginine methyl transferase I, which catalyses asymetric dimethylation (Tang et al., 2000). The recognition motif for the transferase PRMT1 has been identified as (F/G)GGRGG(G/F) (Kim et al., 1997). Many RNA binding proteins have a C-terminal RGG domain including CIRP. Aoki et al identified the Xenopus homolog of PRMT1 as a binding partner of Xenopus CIRP (XCIRP2), showing that the methylation of XCIRP2 leads to accumulation of XCIRP2 in the cytoplasm and confirmed the interaction of XCIRP2 with PRMT1 *in vitro* (Aoki, Ishii, Matsumoto, & Tsujimoto, 2002). The study identified a 43 amino acid region at the C-terminal of XCIRP2 as the methylation site named RG4 which when mutated led to accumulation of XCIRP in the nucleus and that upon over expression of PRMT1, cytoplasmic XCIRP2 levels increased.

There is a high amino acid sequence homology between XCIRP2 and mammalian CIRP showing 74% identity to mouse CIRP (Aoki et al., 2002). Leeuw et al also showed that methylation of arginine residues in the RGG motif on the C-terminal of human CIRP is required for nuclear exit of the protein in mammalian cells (De Leeuw et al., 2007). The nucleocytoplasmic distribution of CIRP is determined by the RGG motifs of the glycine rich C-terminal domain and the exit of CIRP from the nucleus is in a methylation dependant manner. The capacity of CIRP to stimulate translation

and cell survival under stress conditions is related to this translocation in stress playing a protective role as discussed in the next section.

1.4.2 RNA Binding of CIRP

It is not known how the up-regulation of CIRP leads to improved recombinant protein production but it has been described as an RNA chaperone in several investigations (Nishiyama, Itoh, et al., 1997; Sonna et al., 2002). To date some mRNAs have been identified which specifically CIRP binds to (Yang & Carrier, 2001) but a sequence specifity has not been identified. However Yang et al (2001) identified a 51 base oligonucleotide common RNA motif which was found 11 times more in the 3'UTR of the sequences tested than in the 3'UTR of the whole gene database. This sequence was also found to be present in the 3'UTR of the identified targets (table 1.2). The over expression of CIRP has been shown to stabilise mRNAs by binding to them and it has been also been shown that CIRP also binds mRNA of ribosomal proteins which could lead to an increase in their translation (Yang & Carrier, 2001).

RNA binding proteins have an important function in the response of cells to cold stress (Liu et al., 2013). The effect of temperature on mRNA structure is not well understood but under mild hypothermia cells display increased recombinant mRNA levels and improved mRNA half life suggesting increased mRNA stability (Fox et al., 2005). Cold inducible RNA binding proteins are thought to ensure and improve translation of specific mRNAs at low temperatures (Becerra et al., 2012). The translation and stability of mRNAs is regulated by RNA binding proteins and small RNAs as stable ribonucleoprotein complexes (RNP) (Nishiyama, Higashitsuji, et al., 1997). CIRP is a member of a family of RNA binding proteins that consist of an N-terminal RNA binding domain and a glycine rich terminal domain (Nishiyama, Itoh, et al., 1997). The RNA binding domain consists of approximately 90 amino acids and includes two highly conserved RNA binding sequence RNP1 and RNP2 and a number of other conserved amino acids, which are mostly hydrophobic (figure 1.7) (Sheikh, 1997).

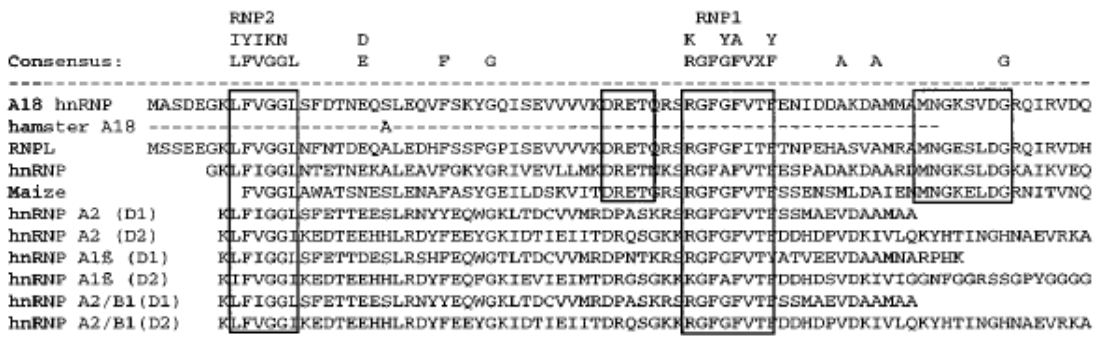

Figure 1.7 Sequence alignment indicating the conserved RNP1 and RNP2 and other regions in several mammalian cold shock proteins. (Sheikh et al, 1997)

As described above (section 1.4.1), CIRP (hnRNP A18) translocates from to nucleus to the cytoplasm under cold/UV stress or hypoxia and binds to stress responsive 3'UTR of specific mRNAs increasing their translation (Yang & Carrier, 2001). Yang et al (2001) identified 46 mRNA ligands, which CIRP binds to. More than 20 of these were identified as stress-responsive mRNA targets (table 1.2). Two of these mRNAs that CIRP has been reported to specifically bind to are the 3'UTR mRNA of thioredoxin (TRX) and replication protein A (RPA2), which are both stress responsive transcripts. Their translation and stability was reported to increase upon binding by CIRP (Yang & Carrier, 2001).

The levels of TRX protein correlate with CIRP levels. TRX plays an important role during oxidative stress through its up-regulation in stress conditions. It has a regulatory role in cell signalling and gene transcription. Its ability to bind DNA and activate of transcription factors is increased through its redox activity and it also protects cells under stress conditions by quenching reactive oxygen species which increase under cellular stress and are damaging to the cells (Yang et al., 2006). The TRX 3'UTR is reportedly able to bind both the RNA binding domain and the C-terminal glycine rich domain when expressed independently, but maximal binding is observed when both domains are present suggesting that the C-terminal domain must have a function in RNA binding (Yang et al., 2006). More recently, the ATR 3'UTR mRNA was also identified as an mRNA target of CIRP and overexpression of CIRP resulted in increased ATR protein levels (Yang et al., 2010). The significance of this is discussed later in this chapter (section 1.4.3).

Ribosomal Proteins	Stress-Responsive Proteins	Other
L3	Replication protein A2	18 S rRNA
L	Thioredoxin	Novel sequence 1
L10	Ferritin L chain	Sequence of unknown function
L13A	Nucleophosmin	Histone
L15	Glutathione transferase Ω	USF-1 gene
L17	Translational elongation factor 1α	cctd
L19	TIM 22 homolog	HSPC023 mRNA
L29	Cyclophilin B	Novel sequence 2
L30	NADH-ubiquinone oxidoreductase	Hypothetical protein FLJ20420
L32	Human Wilm's tumor related	Tubulin-α
S3a	Myosin regulatory light-chain	Cystein-glycine-rich protein
S5	Laminin-binding protein	
S6	Human G10 homolog edg-2	
S8	Translational elongation factor1/β2	
S11	Protein kinase C inhibitor I	
S12	Putative C-myc responsive (rcl)	
S13	ATPase	
	TAXREB107	

Table 1.2 mRNA ligands reported to bind CIRP. (Yang & Carrier, 2001)

CIRP has a typical RNA binding protein topology with the RNP1 and RNP2 motifs located on the β3 and β1 strands respectively. There are highly conserved aromatic residues exposed on the surface of these β sheets which interact with the RNA (Maris et al., 2005). An RNA binding model shows that residues in position 3 and 5 on RNP1 (β3) and the residue in position 2 on RNP2 (β1) are the essential residues for RNA binding (Maris et al., 2005) (figure 1.8). The interaction of these CIRP residues with RNA ligand was previously confirmed by NMR chemical shift studies undertaken on the N-terminal domain of CIRP, referred to as N-CIRP (Hsieh, 2010). Upon mixing with RNA ligand, the aromatic side chains in position 2 (RNP2) and 5 (RNP1) stack with adjacent RNA bases. The aromatic residue in position 3 (RNP1) interacts in a hydrophobic manner with adjacent sugar rings in the RNA ligand. The positively charged residue (Arg) in position 1 (RNP1) forms a salt bridge with the phosphate backbone of the RNA ligand (Maris et al., 2005) (figure 1.8).

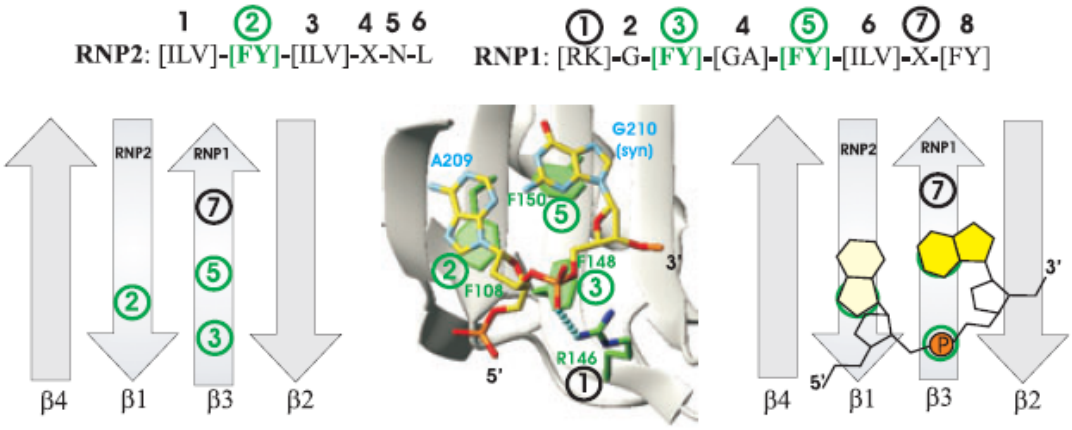

Figure 1.8 Model of RNA binding showing residues essential for RNA binding in RNP1 and RNP2 and their interaction with RNA (Maris et al., 2005)

As mentioned earlier, previous studies have reported that the 3'UTR of the thioredoxin (TRX) mRNA binds to CIRP (Yang et al., 2006) and these findings were further confirmed by Hseih (2010). Human thioredoxin 3'UTR mRNA is 166 nucleotides long excluding the poly-adenine (polyA) tail and the mouse TRX 3'UTR mRNA is 532 nucleotide long, however the first 127 nucleotides of the TRX 3'UTR from the two species is 69% identical. Electromobility shift assay studies by Hseih showed that full length CIRP and N-CIRP bind to full length human TRX 3'UTR mRNA, the first 127 nucleotides of human TRX 3'UR mRNA and also the first 113 nucleotides of mouse TRX 3'UTR mRNA (Hseih, 2010).

Binding studies of N-CIRP with mRNA by NMR also showed binding of N-CIRP to the TRX 3'UTR and also to a 10mer polyA RNA fragment which showed the largest chemical shift changes on minimal chemical shift maps (Hseih, 2010). As described above, it has also been shown that both the RNA binding domain and the glycine rich C-terminal domain are required for maximal RNA binding of CIRP (Yang et al., 2006). The exact nature of these interactions and the subsequent effect of CIRP binding on cellular responses remain unclear. However, reports have now suggested that the increased translation of specific mRNAs upon CIRP up-regulation is mediated by interaction with the initiation factor eIF4G of the translational machinery (Yang et al., 2010). A model whereby the 3'UTR RNA binding proteins can regulate translation initiation at the 5'UTR end has been proposed, where 3'UTR and 5'UTR of mRNA are bridged by the interaction of CIRP with eIF4G. This proposed model

results in circularization of the mRNA where the 3' and 5' end form a 'kissing' loop (Yang et al., 2006). The interaction of CIRP with the translational machinery is further discussed in section 1.4.3.

1.4.3 Interaction of CIRP with the Translational Machinery

As described above, CIRP binds to the 3'UTR (untranslated region) of particular mRNAs resulting in an increase in their translation (Yang & Carrier, 2001; Yang et al., 2010) and one of these is the 3'UTR of the ATR protein (ataxia telangiectasia mutated and Rad3 related kinase) resulting in increased ATR protein expression (Yang et al., 2010). ATR belongs to a family of protein serine threonine kinases and the ATR signalling pathway is associated with cell cycle arrest due to compromised DNA replication (Abraham, 2001). It has been suggested that the binding of CIRP to the 3'UTR of specific mRNAs results in an increase in their translation and is mediated by an interaction between CIRP, the mRNA and the translational machinery (Yang et al., 2006). This proposal is supported by the studies of Yang et al who reported that CIRP binds to the eukaryotic initiation factor 4G (eIF4G). From these observations it is suggested that through the interaction with 3'UTRs and eIF4G, CIRP acts as a bridge between 3' and 5'UTRs to increase translation.

Translation initiation is the assembly of the ribosome on an mRNA for protein synthesis and involves the binding of the 40S ribosomal subunit to the eIF4F complex that forms at the capped 5' end of an mRNA. This is a complicated event with two major stages; formation of the 48S pre-initiation complex followed by the recruitment of the large 60S ribosomal subunit once the AUG start codon is found to assemble the 80S intact ribosome (figure 1.9) (Alekhina & Vassilenko, 2012).

The binding partners eIF3 (bound to eIF4G), eIF2 and eIF1A subsequently interact with the 40S subunit forging a link between mRNA and ribosome. mRNA circularisation is aided by the eIF4G interaction of eIF4A and polyA binding protein (PABP) which are bound to the 3' and 5' of the mRNA respectively where eIF4G acts as the central protein to bring together the 5' and 3' termini of the mRNA (Lamphear, Kirchweger, Skern, & Rhoads, 1995). Indeed, during this project the hypothesis that the eIF4G-CIRP-3'UTR RNA interaction can also circularise the mRNA under mild hypothermic conditions when CIRP is expressed and facilitate the translation of

specific mRNAs that contain CIRP binding 3'-UTRs was investigated using the ATR 3'-UTR mRNA, as this has been shown to bind to CIRP as described above.

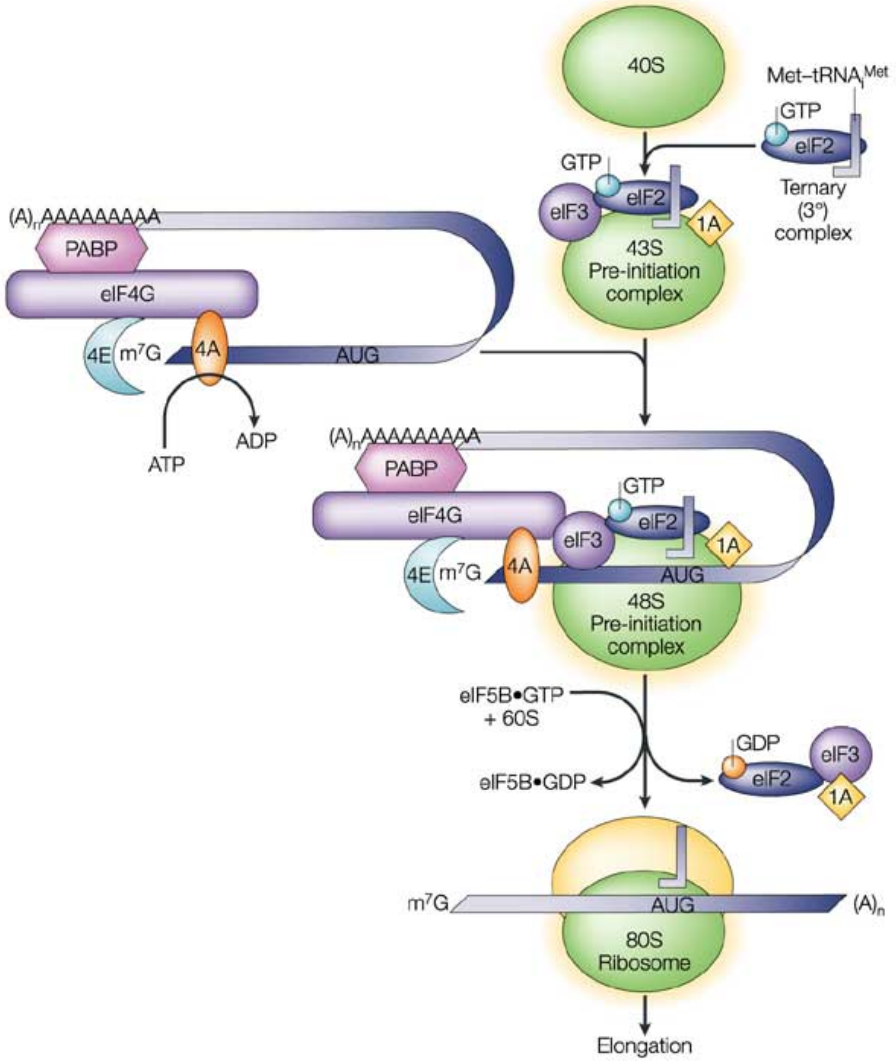

Figure 1.9 Pathway of translation initiation in eukaryotes (Klann & Dever, 2004)**.** This event requires multiple initiation factors including eIF4E, eIF4A, eIF3, and PABP, which in turn aids binding of mRNA to the 40S ribosomal to form the 48S pre-initiation complex (Korneeva, Lamphear, Hennigan, & Rhoads, 2000). eIF4G is the central protein for the initiation factors involved in mRNA recruitment to the ribosome (Keiper, Gan, & Rhoads, 1999). eIF4G serves as a scaffold protein for the binding of the initiation factors eIF4E, eIF4A and PABP at the 5' cap of an mRNA to promote the recruitment of RNA by ribosome (Hinton, Coldwell, Carpenter, Morley, & Pain, 2007). The central region (figure 1.10) of eIF4G binds ATP dependent RNA helicase eIF4A, the 40S binding factor eIF3 and RNA (Korneeva et al., 2000).

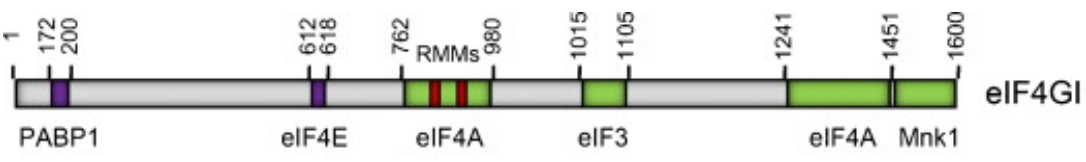

Figure 1.10 Schematic figure of the eIF4G initiation factor, with the binding domains indicated. Figure adapted from Yángüez & Nieto, 2011.

1.4.4 Structure and Function of CIRP

CIRP consists of a well-characterised N-terminal RNA binding domain and a glycine rich C-terminal domain that is natively disordered. Although the function of the unstructured C-terminal domain remains unknown, it has been shown previously that both the RNA binding domain and the glycine rich domain is required for maximal RNA binding (Yang et al., 2006). In recent years interest in natively disordered proteins has increased and NMR spectroscopy is recognised as a good method to study intrinsically disordered proteins (IDPs) (Cavanagh, Fairbrother, Palmer, Skelton, & Rance, 2007). Structured proteins give resonance peaks that are spread out due to the difference in the local environment of the nuclei but IDP spectra peaks are closely spaced or overlapping as the nuclei environments are indistinct. The C-terminal glycine rich domain of CIRP has been shown to be natively disordered by NMR studies (fig 1.11) (Hsieh, 2010).

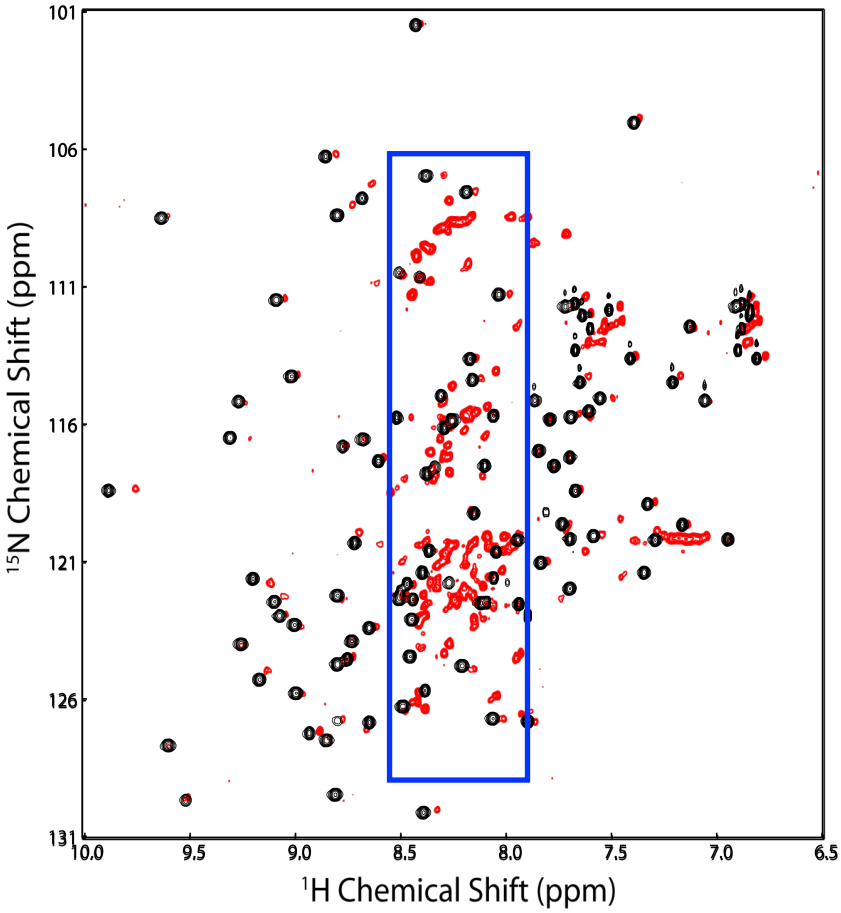

Figure 1.11 ^{15}N/^{1}H HSQC spectra of full length CIRP (red) and N-CIRP (black). ^{15}N-labelled N-CIRP gave well-resolved NMR data typical of a stable folded protein domain. The peaks originating from the C-terminal region (red) were found to predominate in the central region of the spectrum (shown by the blue box). Backbone amide ^{1}H resonances in this region are characteristic of a random coil structure suggesting that the C-terminal region of CIRP is unstructured. Data from PhD thesis of Hsieh (2010), University of Kent.

The structure and function of recombinant full-length CIRP has not been studied in detail and the contribution of the C-terminal region towards the function of CIRP is not known. Unstructured regions of proteins have been identified to act as decision points for specificity in binding and function (Pancsa & Fuxreiter, 2012). Intrinsically disordered regions of proteins lack a stable structure but still exhibit biological activities (Dunker et al., 2001). Due to their amino acid sequence, in the case of CIRP being rich in glycine residues, they have flexible characteristics and are known to be involved in functions such as recognition of proteins and nucleic acids and facilitate interactions and chemical reactions, accommodate post translational modification, alternative splicing, protein fusions and insertions or deletions (Oldfield & Dunker, 2014). IDPs may lack structure in the absence of a ligand but can fold when bound to a ligand and allow for the binding to multiple partners. For example the IDP human cell-cycle control protein p21, becomes structured when bound to its ligand cyclin (Kriwacki, Hengst, Tennant, Reed, & Wright, 1996).

A number of proteins with natively disordered regions have been shown to be important in cell cycle regulation (Follis, Galea, & Kriwacki, 2012). For example, the disordered region of p27 causes cell cycle arrest by binding and closing of the cycling A/CDk2 complex (Galea et al., 2008). These findings are fitting with CIRP studies showing that over expression of CIRP in cultured cells results in cell cycle arrest in the G1 phase (Nishiyama, Itoh, et al., 1997). Studies show that IDPs have weak binding affinities for their binding partners and fast on/off binding rates (Shammas, Travis, & Clarke, 2013). Also consistent with CIRP function, IDPs and IDP regions are also associated with having function as RNA chaperones (Tompa & Csermely, 2004). All these findings strongly suggest that the C-terminal domain of CIRP has important functional roles that are currently ill defined.

On-the-other-hand, the N-terminal region of CIRP has a clear RNA binding function (Manival et al., 2001; Maris et al., 2005; Nishiyama et al., 1998; Yang & Carrier, 2001;Yang et al., 2006). The N-terminal binding domain has a β1α1β2β3α2β4 topology and consists of two RNA binding consensus sequences on strands β1 and β3; these are called RNP1 and RNP2 as described earlier (fig 1.6) (Sheikh et al, 1997). The RNA binding of this domain is discussed in detail in section 1.4.2. Studies have identified RNP1 and RNP2 as the specific binding regions for RNA with some

residues being more important than others. The highly conserved aromatic residues exposed on the surface of the β sheets 1 and 3 are thought to interact closely with the RNA (Manival et al, 2001). Fitting with this model, CIRP possesses these conserved sequences RNP1 and RNP2 on the β sheets 1 and 3 and the essential phenylalanine aromatic residues in position 2 of RNP2 and positions 3 and 5 in RNP1. The RNA binding domain of human CIRP has been deposited in the protein data bank (1X5S) (figure 1.2) but not been published whilst there is no structural data publicly available for the mouse protein studied in this project.

Aside from CIRP being involved in RNA binding and stabilising leading to increased translation of specific RNAs, the specific function of CIRP is still largely unknown. The RNA binding function and interaction of CIRP with the translational machinery is discussed in detail in previous sections in this chapter (sections 1.4.2. and 1.4.3). The main focus to date and of this project has been upon the observation that the over expression of CIRP leads to an increase in recombinant protein production under cold stress conditions via unknown mechanisms. The current understanding is therefore that CIRP stabilises and facilitates the translation of specific mRNAs necessary for the cells survival under stress conditions (e.g. cold-shock) and several additional interactions of CIRP to those discussed above have now been reported. For example, Qiang et al (2013) have identified CIRP as a DAMP (damage associated molecular pattern) where CIRP levels were reported to be elevated in the blood of haemorrhagic shock patients. Haemorrhagic shock is associated with an inflammatory response. CIRP was found to be up-regulated in the heart and liver and released into the blood stream by translocating from nucleus to cytoplasm in the macrophages. Release of tumour necrosis factor α (TNFα) and HMGB1 from macrophages and induction of inflammatory response and tissue injury was stimulated by recombinant CIRP. This effect of TNFα and HMGB1 was reduced in CIRP deficient mice. It was also reported human CIRP residues 106-125 bind toll-like receptor 4 (TLR4)-myeloid differentiate factor 2 (MD2) complex to mediate this extracellular activity and as such CIRP is thought to be associated with promotion of inflammatory responses in shock and sepsis (Qiang et al., 2013). A further study by Sakurai et al (2006) reported that under mild hypothermic conditions CIRP protects cells via the activation of the ERK (extracellular signal regulated kinase) pathway, which in turns inhibits TNFα induced apoptosis. It was discovered that in BALB/3T3 cells which had been treated with

tumour necrosis factor α, reduction of temperature lead to an increase in CIRP expression and levels of phosphorylated ERK, leading to inhibition of apoptosis (Sakurai et al., 2006).

Another mechanism that CIRP may be involved in modulating upon its expression was reported in a study by Masuda et al (2012) where proliferation of undifferentiated spermatogonia was stimulated by the interaction of CIRP with Dyrk1b (Masuda et al., 2012). CIRP is expressed in the mouse testis' germ cells. It was found that CIRP knockout mice lead to a reduction in undifferentiated spermatogonia. The study showed that proliferation of spermatogonia is stimulated by CIRP confirming its proliferative role. It was found that CIRP supressed the kinase activity of the Dyrk1b protein by interacting with it, which stops its kinase activity to phosphorylate p27. When active, p27 has a role toward preventing cell cycle progression at G1 by inhibiting cyclin E/CDK2. Also, p27 phosphorylation by Drkb1 was increased in the absence of CIRP. CIRP increases cell proliferation by leading to a reduction in p27 levels. This supports the onco-protein role, which has been linked to CIRP (Masuda et al., 2012).

A number of studies have implicated up-regulation of CIRP in cancer cells. As mentioned previously, CIRP is up-regulated in response to hypoxia and in the environment of tumour cancer cells, there is a reduction in oxygen levels as tumours grow (Harris, 2002) and initiation and progression of cancer has been linked to oxidative stress. The proteins TRX and RPA2 have both been associated as being prognostic indicators in certain cancer and potential therapeutic targets (Givalos et al., 2007; Noike, Miwa, Soeda, Kobayashi, & Miyagawa, 2008) and as described above, CIRP specifically binds to the 3'UTR mRNA of both these proteins. Further, CIRP's role in protecting cells against apoptosis under stress conditions supports this (Sakurai et al., 2006). CIRP has been found to be overexpressed in various human tumours, for example in a study involving 193 patients, CIRP was overexpressed in 33% of colon and 45% of breast carcinomas providing further support of CIRP having a role in tumourignesis (Lleonart, 2010). The potential of CIRP as a target in cancer therapy has also been suggested as Zeng et al (2009) showed that the inhibition of CIRP reduced cell proliferation in prostate cell lines and as well as impairing cell survival,

CIRP knock down also enhances chemosensitivity to cisplatin and adriamycin (Zeng et al., 2009).

Liu et al (2013) have also reported that CIRP regulates circadian gene expression by controlling alternative polyadenylation (APA), a further link to the translational apparatus. These authors found that the 3'UTR binding sites for CIRP were enriched near the polyadenylation sites of mRNA (PASs), they identified common RNA binding protein binding motifs on target 3'UTR mRNAs using computational methods and also showed that reduced levels of CIRP lead to shortening of the 3'UTR and up-regulation of CIRP lengthened the 3'UTR. APA is an important mechanism of post-transcriptional regulation, many mammalian genes have tandem UTRs generated by multiple PASs and so without changing the mRNA code, the length of 3'UTRs can be altered which changes the accessibility of RNA binding proteins and has an effect on mRNA stability, translation and localisation (Liu et al., 2013). For example, 3'UTR are shortened during T lymphocyte activation but lengthened in embryonic development (Ji, Lee, Pan, Jiang, & Tian, 2009; Mayr & Bartel, 2009).

1.5 Cold Shock and Recombinant Protein Production

Under conditions of mild hypothermia (32°C), mammalian cells can continue to survive and divide (Fujita, 1999) and culturing of the cells at these temperatures can result in enhanced recombinant protein (Yoon, Song, et al., 2003). In the biopharmaceutical manufacturing industry, mammalian cells are used routinely for the production of recombinant proteins used as therapeutics in the clinic. Normally mammalian cells are cultured at 36.5-37°C to mirror the conditions found in the body (Yoon, Song, et al., 2003). Cell proliferations takes place during exponential growth of the culture at which time recombinant protein synthesis is not optimal as the cell's resources are focused on processes necessary for doubling of cell biomass (Masterton & Smales, 2014). Recombinant protein production from mammalian cells is optimal during the stationary phase as there is less competition between cellular processes and protein synthesis (Gammell, Barron, Kumar, & Clynes, 2007). During late phase of culture, mammalian cell viability rapidly declines due to apoptosis limiting production (Goswami, Sinskey, Steller, Stephanopoulos, & Wang, 1999). However,

reduced temperature can be used to prolong the culture lifetime before cell death to extend production.

Low temperature culturing of mammalian cells for the synthesis of therapeutic recombinant proteins is widely used (Masterton & Smales, 2014). The temperatures described as mild hypothermia in this sense are 27-32°C, below these temperatures translation is halted and protein production will not increase (Roobol et al., 2009). One of the first studies to develop a process of temperature manipulation was reported by Jenkins and Hovey (Jenkins & Hovey, 1993), where they changed culture temperature between 39°C and 34°C and the temperatures less than 37^{o}C resulted in maintenance of a viable culture for longer and a 35% increase in protein yield.

At mild hypothermic conditions growth rate of the cells is reduced and cell doubling time increases (Marchant et al., 2008). As a result of culturing cells at lowered temperatures, metabolite depletion and build up of toxic by-products is also decreased. As the metabolism is reduced, apoptosis is also delayed as cells arrest at the G1 phase of the cell cycle, cells are not apoptotic and protein synthesis is active. Delayed apoptosis is advantageous in the sense that cell viability is elongated and cells are productive for longer periods of time (Hwang, Yoon, Koh, & Lee, 2011; Marchant et al., 2008). However, this does mean running cultures for extended periods of time, therefore a biphasic approach was developed where cells are cultured at 37°C during the first phase to achieve the desired biomass when the cells rapidly grow until stationary phase of growth is achieved. This is followed by the second phase where the temperature is reduced (27-35°C) for protein production (Sunley, Tharmalingam, & Butler, 2008; Yoon, Kim, Song, & Lee, 2006). This biphasic strategy was applied to enhance the synthesis of recombinant human IFN-γ from CHO cells (Fox, Patel, Yap, & Wang, 2004).

Most reports suggest that the cells major response to reduced temperature is the decrease in the rates of global protein synthesis (Roobol et al., 2009) which conflicts with the general observations that recombinant protein production is increased at these temperatures. This observation can be explained by the fact that under these conditions there is reduced metabolic burden on the cell and the mRNA levels are

increased and mRNA stabilised. Therefore, as well as the increased recombinant mRNA levels, there is less competition from the translation of endogenous mRNA and protein folding leading to enhanced recombinant protein yields (Masterton & Smales, 2014).

The effects achieved from culturing mammalian cells at reduced temperatures is cell line and product dependent. For example, the synthesis of recombinant human growth hormone from CHO cells was improved at reduced temperature but leads to decreased cell viability (Rezaei, Zarkesh-Esfahani, & Gharagozloo, 2013). It has also been reported that culturing cells at mild hypothermic temperatures results in no change or a decrease in productivity. For example Yoon et al (2003) found that expressing anti-4-1BB antibody in CHO cells at sub physiological temperature did not have an effect on yield at 33°C and lead to an 3.9 fold decrease at 30°C (Yoon, Song, et al., 2003). Marchant et al (2008) showed that expression of the IgG4 monoclonal antibody (cB72.3) is host cell specific. Temperature shift from 37°C to 32°C in CHO cells lead to 50% increase in productivity whereas a negative effect on production was observed in NS0 murine myeloma cells (Marchant et al., 2008). This difference between two cell lines was attributed to metabolite utilization, cell growth and recombinant mRNA level differences between the cells lines. Further, it has been suggested that improved recombinant protein products at lower temperature cultivation is not only due to growth arrest of the cells but also because of the elevated levels of recombinant mRNA at sub physiological temperatures as a result of increased mRNA stability through secondary structure formation of the mRNA and reduction in the rate of mRNA degradation (Fox et al., 2005; Kou et al., 2011; Marchant et al., 2008; Roobol et al., 2009). For example increased mRNA levels of recombinant IFN-γ mRNA at sub physiological levels was associated with elevated IFN-γ protein levels from CHO cells (Fox et al., 2005).

As well as improving product levels, culturing mammalian cells at reduced temperatures have been reported to improve product quality i.e. more accurate protein folding (Masterton et al., 2010) and more accurate post-translational modifications, specifically glycosylation (Borys et al., 2010). This in turn leads to reduced protein

aggregation. Studies have shown that there is no difference between the biological activity of protein produced at 37°C and at 32°C (Becerra et al., 2012).

1.6 Project Aims

Mammalian cells respond to culturing at sub-physiological temperatures through activation of a number of mechanisms as described in the previous sections of this introduction. However, the exact cellular mechanisms which are activated and the relationships between these mechanisms are not well understood. Gaining a higher level of understanding as to how the mechanisms are initiated and how this coordinated response leads to the improved recombinant protein production and quality is crucial if new strategies, cell lines and vector systems to exploit these responses to improve recombinant protein product yield and quality at such temperatures are to be devised. Further, a better understanding of the mechanisms activated during cold stress could also potentially help develop strategies for diseases which involve cold shock proteins such as male infertility and cancer.

In this study the structure and function of the cold-shock inducible protein CIRP were investigated. CIRP plays a crucial role in the response to cold shock in mammalian cells and is involved in the modulation of many of the cellular responses to cold-shock and cold-stress. Extending our knowledge of this protein and how its up-regulation is involved in modulating these responses will aid in defining the role of CIRP in these responses and how CIRP and cold-shock responses may be manipulated in the cell. Specifically here the following were investigated;

- RNA binding of full-length CIRP, N-CIRP and C-CIRP domains and investigating the essential residues for the RNA binding of CIRP.
- Function of the glycine rich C-terminal domain.
- The structure of the RNA binding domain of mouse CIRP and its dynamic properties in space.
- Mechanisms by which CIRP expression might lead to an increase in recombinant protein production from cultured Chinese hamster ovary (CHO) cells.

Chapter 2

Mutagenesis, Expression and Purification of CIRP and N-CIRP

2.1 Introduction

2.1.1 Recombinant Protein Expression and Purification

To investigate the structure and ligand binding activity of the N-terminal domain (N-CIRP), NMR was chosen as a suitable method. N-CIRP is a good candidate for NMR studies as it is of an appropriate size, soluble in the necessary buffers to at least 1 mM and it is well structured and stable. For ligand binding studies, ^{15}N-^{1}H HSQC experiments were used and a minimum protein concentration of 0.1 mM is required. For structural studies the protein concentration must be 0.5-1 mM for which 6-12 mg/ml of CIRP is required. To achieve such yields using mammalian cell systems is expensive and requires the production of stable cell lines, which can take a long time. Another point to be considered is that for NMR experiments, it is required for the protein to be isotopically labelled. This is achieved by enriching the growth media with ^{15}N or ^{13}C or both. The cost of isotopic labelling is much cheaper in *E.coli* than mammalian cells. Full-length protein can also be characterised using similar NMR experiments as used for N-CIRP and also similar RNA binding studies carried out to investigate any changes in binding when both the N- and C- regions are present.

To achieve as high a yield as possible, the proteins were expressed in the *E.coli* system because of its many advantages such as short culture time, low cost of media and high expression yields (Baneyx 1999; Sorensen & Mortensen 2005a). Depending on the culture size and the media used, a suitable yield can be obtained within 24 h after the cells have been transformed. For the purposes of this project, *E.coli* expression was found suitable as we believe that CIRP does not possess any essential post translation modifications that would require the use of eukaryotic systems for expression.

For NMR experiments, protein samples are required to be of high purity (>95%). The pET vector system (Novagen) encoding CIRP/N-CIRP was used (figure 2.1). The pET28a vector has T7 promoter, lacI repressor for the induction of expression and kanamycin resistance and results in high expression yields due to the strong T7 promoter (Rosenberg et al. 1987; Studier & Moffatt 1986; Studier et al. 1990). The pET system codes for a number of tags to aid purification, we used the His-tag to enable nickel affinity purification. pET28a(+) with the CIRP/N-CIRP gene insert was kindly provided by Shu-Ju Hsieh.

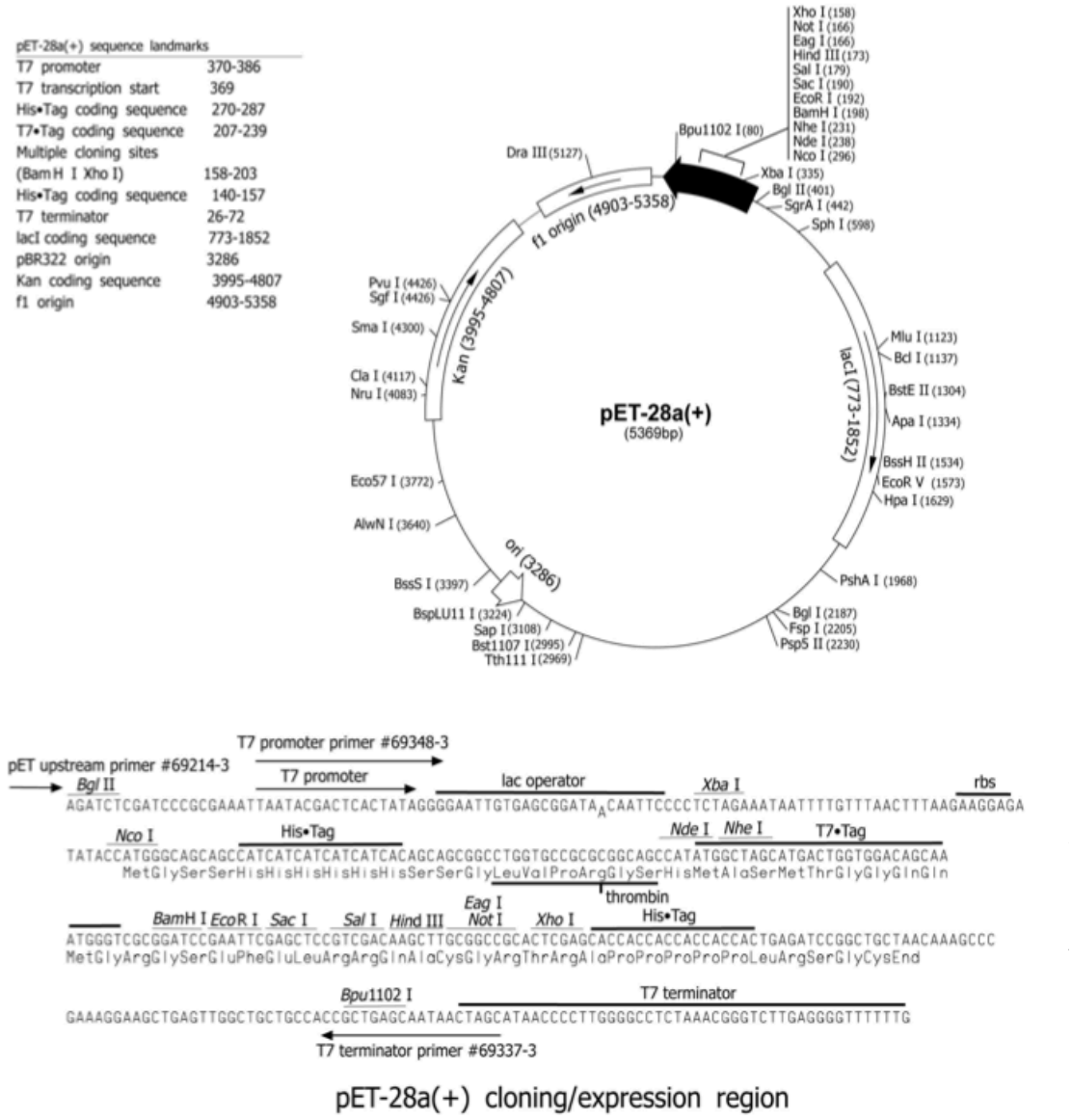

Figure 2.1 Schematic representation of the pET28a (+) vector map. (Novagen pET System Manual)

Protein is loaded on to the nickel affinity chromatography column, the His tagged protein binds to the matrix. Once the tagged protein is bound it can be eluted by introducing a competing agent such as imidazole which competes with the His-tag (Simpson 2009; Janson & Ryden 1998). Ion exchange chromatography was chosen to further purify N-CIRP after nickel affinity chromatography to remove further contaminants. Ion exchange chromatography relies on charged interactions between the protein and charges immobilised on the matrix (Gräslund et al. 2008). The bound protein is eluted using a salt gradient (Simpson 2009; Millner 1999).

2.1.2 Mutagenesis

NMR experiments can be used to study protein-ligand interactions by looking at chemical shift changes of the protein in the presence and absence of ligand. N-CIRP consists of two RNA binding consensus sequences on strands β1 and β3; these are called RNP2 and RNP1 (figure 1.6) (Sheikh 1997). Studies have identified RNP1 and RNP2 as the specific binding regions for RNA with some residues being more important than others (Chapter 1, section 1.4.2). The highly conserved aromatic residues exposed on the surface of the β strands 1 and 3 are thought to interact closely with the RNA (Manival et al. 2001). To confirm the ligand-binding site in the N-terminal domain of CIRP and the role of the N-terminal domain in full-length CIRP, the highly conserved aromatic residues F9 (RNP2, position 3) and F49 (RNP1, position 2) were mutated. The importance of these residues could be assessed by their effect on the binding activity of N-CIRP and the structure and stability of the protein.

Site directed mutagenesis by overlap extension PCR is a convenient method to introduce mutations into a known sequence (Ling & Robinson 1997). Using PCR, complementary primers are used to generate two DNA fragments having overlapping ends (figure 2.2). These fragments are combined in a subsequent overlap reaction in which the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand (Ho et al. 1989; Aiyar et al. 1996).

The overlapped product is then amplified further by PCR using a pair of primers i.e. the two end primers. Mutations to the sequence are introduced by incorporating changes into the overlapping primers. pET28a vector coding N-CIRP was used as the template DNA to generate these mutant fragments, which were then cloned back into the pET28a vector for protein expression in *E.coli*.

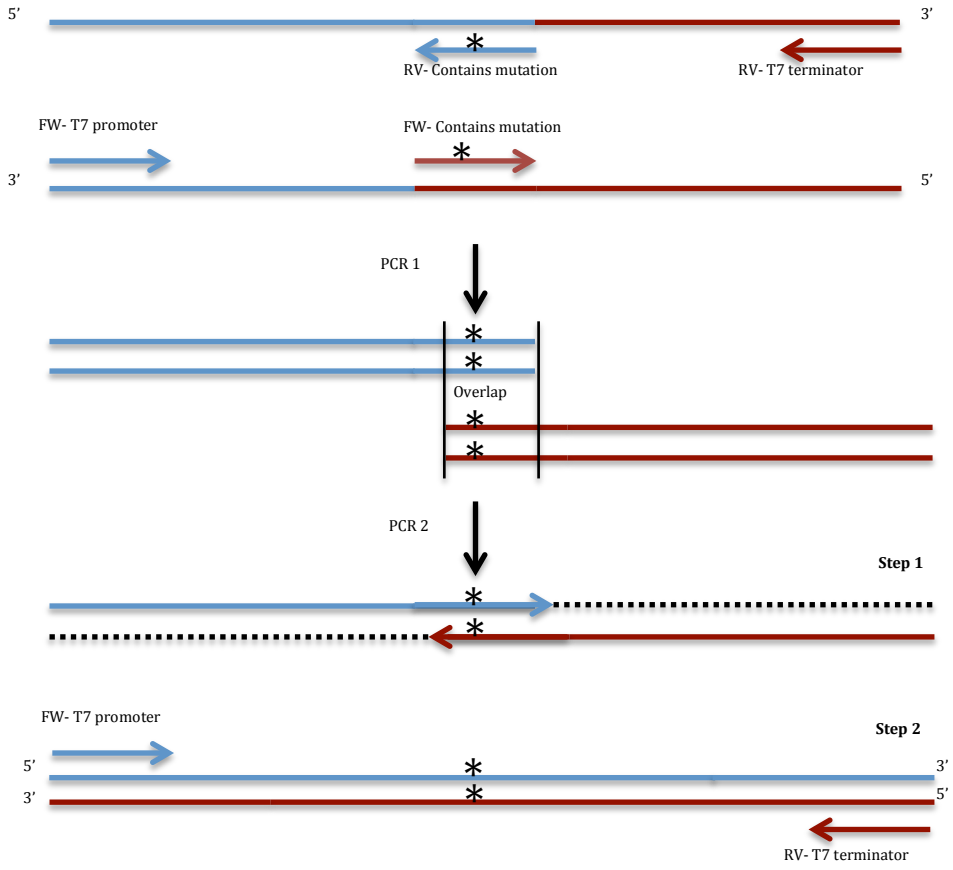

Figure 2.2 Schematic diagram of site-directed mutagenesis by overlap extension. Site of mutagenesis is indicated by the small black star.

2.1.3 Protein Solubility

There are several factors that can contribute to protein aggregation; pH, protein hydrophobicity (Luan et al. 2004) and hydrogen bonding (Pace et al. 1996; Myers & Pace 1996) N-terminal and C-terminal amino acid composition (Sati et al. 2002, Doray et al. 2001; Oswald et al. 1994), and high protein concentration (Idicula-Thomas & Balaji 2005). Studies by Shu-Ju Hseih showed that the C-terminal region of CIRP is natively unstructured presumably due to the high glycine and arginine residue content in this region (figure 2.3), whereas the N-terminal region is stable and soluble on its own. The studies also showed that full length CIRP is unstable and has poor solubility after purification.

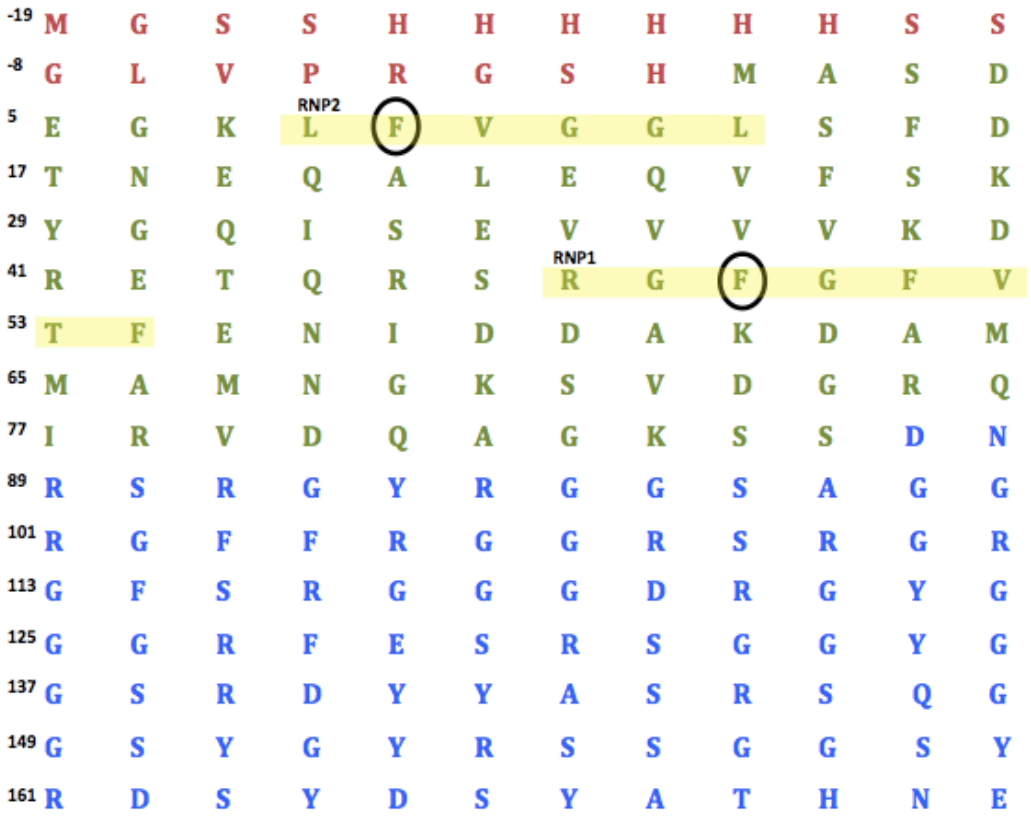

Figure 2.3 The amino acid sequence of mouse CIRP. The hexaHis tag, N-terminal region and C-terminal region are shown in red, green and blue respectively. Highlighted regions indicate the two RNA binding sequences and amino acid residues chosen for mutagenesis are circled.

Various techniques were tested in this chapter to increase final soluble CIRP protein yields by looking at protein which was soluble on expression and also protein which was soluble on expression but formed precipitate *in vitro*. There are many approaches to improve protein solubility on expression such as the use of tags or fusions at the N- or C-terminus of the protein, modification of growth conditions, such as temperature, medium or induction time (Jana & Deb 2005; Smith 2007; Sorensen & Mortensen 2005b). Culture temperature can be dropped from 37°C to 20°C following the addition of the induction molecule IPTG and cells can be induced overnight at this lowered temperature instead of the regular 2-3 hour induction time at 37°C. Protein expression at reduced temperatures can increase the solubility of aggregation-prone recombinant proteins and limit their degradation by proteases (Vasina & Baneyx 1996). However, around 50% of proteins over expressed in *E.coli* are still insoluble despite the approaches mentioned above, these can be recovered using a denaturation/refolding technique (Maxwell et al. 2003).

The aggregates are generally captured by centrifugation. A popular method of solubilisation of aggregates is by the addition of concentrated chemical denaturants such as guanidine or urea and a reducing agent like DTT to reduce any incorrect disulphide bonds(Middelberg 2002). These solubilised aggregated proteins can be further purified using chromatography techniques before refolding as contaminants can interfere with refolding (Maxwell et al. 2003). Refolding of the protein is initiated by the reduction of denaturant concentration by rapid dilution or dialysis with uniform mixing. Efficiency of refolding depends on the competition between aggregation and correct folding. For maximum refolding and to minimize aggregation, protein concentration should be low around 10-100 μg/ml (Jungbauer & Kaar 2007). Renatured protein should be obtained using a minimally complex method with as few steps as possible. Chemical denaturation and refolding by dilution or dialysis is a simple and repeatable method to obtain soluble protein. The refolded proteins should be analysed to check for correct activity and structure.

Another solubility problem which was investigated can be the formation of inclusion bodies on expression and the method of obtaining soluble protein is by inclusion body washing. Inclusion bodies are densely packed protein molecules formed from partially folded protein intermediates (Singh & Panda 2005). These are found insoluble in the pellet following cell lysis. Inclusion body washing helps to isolate and purify the aggregates and once isolated the protein can be solubilized and refolded. Protein recovery steps include, isolation of inclusion bodies, solubilisation of inclusion bodies, refolding, and purification (Ramón et al. 2014; Burgess 2009; Basu et al. 2011). It is important to wash the inclusion bodies before solubilisation to remove any contaminants as these can interfere with refolding. Pellets are washed with a buffer containing detergent by resuspending the pellet and homogenizing followed by a centrifugation step. This wash can be repeated several times. The washed inclusion bodies can then be solubilized using chemical denaturants such as urea or guanidine HCl and refolded by dilution or dialysis methods (Basu et al. 2011).

As well as insolubility on expression, *in vitro* solubility can be a problem i.e. the protein precipitating after expression and purification. Many methods have been proposed to recover insoluble proteins (Marston & Hartley 1990; Tsumoto et al.

2010). Studies by Shu-Ju Hseih showed that CIRP could be partially re-solubilized in sodium acetate buffer at pH 4.8 when the precipitate was washed with this buffer as CIRP precipitated *in vitro* after purification. Additions of some amino acids can further improve solubility and stability of protein formulations (Arakawa et al. 2007). Addition of the two amino acids arginine and glutamic acid together in particular have been shown to enhance the solubility of proteins with poor *in vitro* solubility (Golovanov et al. 2004). The presence of these molecules leads to enhanced crowding which suppresses protein-protein association reducing intermolecular interactions between proteins (Shukla & Trout 2011).

2.2 Materials and Methods

2.2.1 Site-directed Mutagenesis by Overlap Method

To generate the F9A, F49A and F9AF49A mutants, site directed mutagenesis by the overlap method was used, where two set of primers are used in two seperate PCR reactions which generate overlapping DNA fragments (figure 2.2). pET28a(+) plasmid containing wildtype N-CIRP between the Nde1 and EcoR1 and restriction sites (figure 2.1) was used as the template DNA for the PCR reactions. Forward and reverse primers (table 2.1) were designed to contain the mutation sequence complementary to the region of N-CIRP where the mutation was to be introduced. These were paired with T7 promoter forward and T7 terminator reverse primers in two separate PCR reactions. Two fragments were produced with overlapping sequences containing the desired mutation. The two fragments were mixed and used as the template in a PCR reaction using the T7 promoter and terminator primers to produce the final fragment (figure 2.2.). To make the double mutant F9AF49A, the F9A mutant DNA was used as the template and the F49A mutation was introduced using the same overlap method described above using the primers designed for the F49A mutation.

Primer Name	Sequence (5'>3')	T_m (°C)
T7 Promoter Forward	TAATACGACTCACTATAGGG	53.2
T7 Terminator Reverse	CTAGTTATTGCTCAGCGGT	54.5
F9A Forward	AGCTTGCCGTGGGAG	53.3
F9A Reverse	CTC CCA CGG CAA GCT TG	57.6
F49A Forward	GAT CCC GAG GCG CTG GGT TTG TC	67.8
F49A Reverse	GAC AAA CCC SGC GCC TCG GGA TC	67.8

Table 2.1 Primer sequences used for site directed mutagenesis of N-CIRP and their melting temperatures.

2.2.2 Polymerase Chain Reaction

For the PCR reactions Pfu DNA polymerase (Promega, 3U/μl) or Taq DNA polymerase (Roche, 5U/μl) was used. PCR reactions were run in a thermo cycler with a heated lid (Techne TC-3000). To produce the two overlapping fragments, Pfu was used in the reactions. The primer pairs and DNA template used are described above (section 2.2.1). A total reaction volume of 50 μl was used in the PCR tube. A final

concentration of 0.5 μM of each primer was added from a 5 μM stock. For the dNTPs (Promega 100 mM) a 10 mM stock of dTTP, dCTP, dATP and dGTP was used and a final concentration of 20 μM. A maximum of 250 ng of template was added to each reaction and 1 unit of Pfu polymerase. Two master mixes were made for each reaction and these were mixed together just before placing the tube in the instrument. Master mix 1 consisted of the DNTPs, primers and template DNA made up to a total volume of 25 μl with sterile water. Master mix 2 contained the polymerase and the 10x polymerase buffer, also made up to a total volume of 25 μl. The PCR program was as follows: initial denaturation at 95°C for 2 min followed by 25 cycles of; denaturation at 95°C for 1 min, annealing for 1 min at 5°C below the melting temperature of the primer with the lowest melting point extension at 72°C for 2 min. After the 25 cycles, there was a final extension step at 72°C for 5 min. PCR products were analysed on 1% agarose gels (section 2.2.3) and visualised with ethidium bromide to confirm that the correct sized product was present. PCR products were purified using a PCR clean up (QIAGEN, QIAquick PCR Purification Kit) or gel extraction kit (QIAGEN, QIAquick Gel Extraction Kit) following the kit protocols.

To produce the full fragment from the two overlapping fragments, Taq polymerase was used. Taq polymerase was used for this step as the enzyme adds A tails to the 3' end of the fragments and this was required for the cloning of the fragment into pGEM T-easy vector. For the PCR reaction, the T7 promoter and terminator primer pair was used. The same protocol as the Pfu polymerase described above was used to set up the PCR reaction. For the template DNA, 5 μl of each purified fragment was added into master mix 1. The extension time during the 25 cycles was increased to 3 min due to size of the final product being longer. Products were analysed on 1% agarose gels (section 2.2.3).

2.2.3 DNA Gel Electrophoresis

1% agarose gels were prepared in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.5). 0.5 μg/ml ethidium bromide was added to the gel mixture before pouring. Wells were loaded with samples prepared with 6x sample buffer (0.4% orange G, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 15% Ficoll 400, 10mM

Tris, 50 mM EDTA pH 8.0) to a final concentration of 1x. The gels were run in TAE buffer at 100 V for 45 min and visualized under UV light.

2.2.4 Cloning of PCR Fragments into the pGEM T-easy Vector System

The pGEM T-easy vector system (Promega) was used as a convenient system for cloning of PCR products (figure 2.4). The vector is linearised with 3' terminal thymidine at both ends to aid the ligation of A tailed products. It makes the restriction digestion of the products to give them the correct sticky ends to clone into final plasmid easier as restriction digest of small PCR products can be difficult. The vector contains T7 and SP6 RNA polymerase promoters within the alpha-peptide coding region of the enzyme beta-galactosidase. A successful insert inactivates the alpha-peptide allowing blue/white screening of colonies.

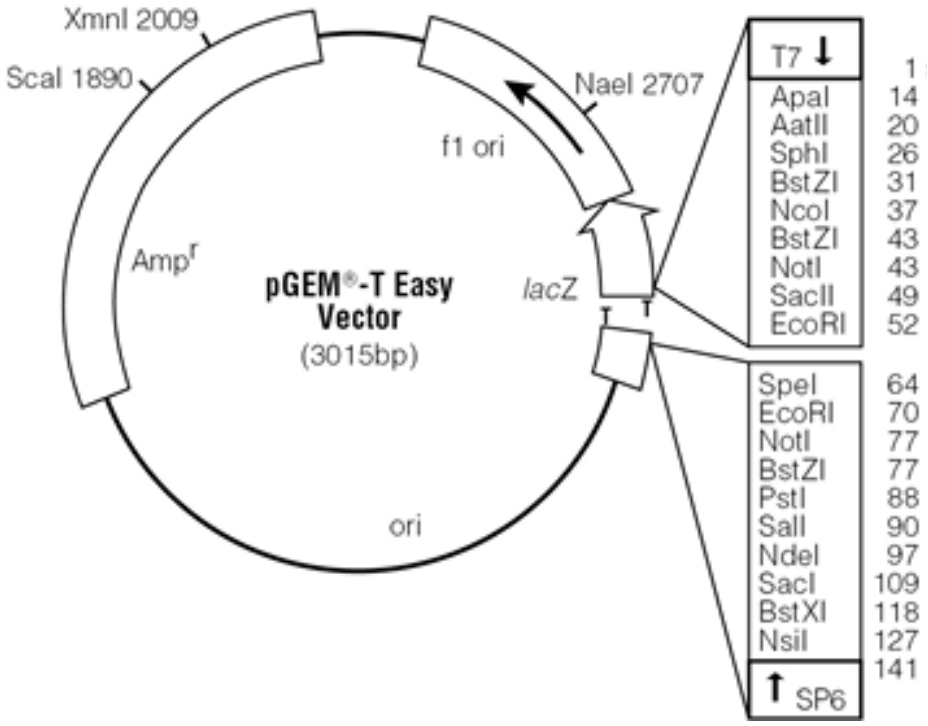

Figure 2.4 Schematic representation of the pGEM T easy vector map (www.promega.co.uk).

Final DNA fragments after the overlap PCR reaction were analysed on agarose gels and purified either using a gel extraction or PCR cleanup kit. As Taq polymerase was used, the A tailed fragments were ligated directly into the pGEM T-easy vector.

Ligation mixes were set up according to the pGEM protocol in a 10 µl volume with 50 ng of pGEM vector, 3 units of T4 DNA ligase (Promega, 3 U/µl), 2x rapid ligation buffer and insert DNA. DNA was quantified using Nanodrop equipment. Various ratios of insert:vector were tested; 5:1, 2:1, 1:1. The amount of insert DNA to be added was calculated using the equation 2.1 below.

Equation 2.1:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert: vector molar ratio} = \text{ng of insert}$$

Two different controls were used. One ligation was carried out using the control insert DNA provided with the pGEM system as a positive control. One ligation was setup using the vector but with no insert added as a negative control to see the amount of background self-ligation. Ligation mixtures were incubated for 1.5-2 h at room temperature and 3-4 µl were transformed into competent DH5α cells. 100 µl of transformed cells were spread onto LB agar plates containing 100 µg/ml ampicillin and incubated overnight at 37 °C. For the blue white/screening, the plates were spread with 100 mM IPTG and 50 mg/ml Xgal and incubated for 30 min at 37°C before use. A transformation control was also set up by transforming DH5α cells with 1 µl of plasmid DNA. Any white colonies were checked for the correct size insert by PCR screening using the T7 promoter and T7 terminator forward and reverse primers and running products on 1% agarose gels.

2.2.5 PCR Screening of Colonies

For PCR screening, colonies were picked with a sterile 200 µl pipette tip and dipped into 50 µl of water and pipetted up and down 2-3 times. The samples were boiled for 2 min and centrifuged at 13 000 rpm for 2 min. 33 µl of the supernatant was then used for the PCR reaction using appropriate primers and Taq polymerase. Cycle and temperature settings were set up as described in section 2.2.2. Size of the PCR products were then analysed by running on 1% agarose gels. Colonies corresponding to the PCR products with the correct size were grown overnight at 200 rpm, 37 °C in 5 ml/ 10 ml (high copy number pGEM/low copy number pET28a) of LB containing

100 μg/ml ampicillin (pGEM) or 50 μg/ml kanamycin (pET28a) for DNA extraction using the miniprep kit (QIAGEN, QIAPrep Spin Miniprep Kit). The extracted DNA was then sent off for sequencing for further confirmation (Beckman Coulter Genomics).

2.2.6 Restriction Enzyme Digestion

To isolate the N-CIRP mutant DNA fragment from the pGEM T-easy vector, restriction enzyme digestion was used with Nde1 and EcoR1 (Roche) enzymes. 1 μg of DNA was cut using 1 unit of each enzyme in a 10-20 μl reaction volume in buffer H. The reaction was incubated at 37°C for 1 h as recommended by the supplier. To purify the excised fragment, the entire reaction volume was run on a 1% agarose gel. The gel piece containing the fragment was cut out and purified using a gel extraction kit.

2.2.7 Ligation of DNA Fragments into pET28a (+) Vector

For the expression of the N-CIRP mutants, the fragments were cut out from the pGEM T-easy vector and ligated into the pET28a (+). Prior to ligation, the pET28a (+) vector was linearized by digesting with the EcoR1 and Nde1 enzymes and purified using the methods described above (section 2.26). A ratio of 15:1 (insert: vector) was used as suggested in the pET manual in 10 μl reaction (vector quantity = 50 ng). 1 μl of T4 DNA ligase was added per reaction and the samples incubated for 1.5-2 h at room temperature. After incubation, 3-4 μl of the ligation mix was transformed into competent DH5α cells. Negative (no insert) and positive (1 μl of plasmid DNA) controls were also included. 100 μl of transformed cells were spread on LB agar plates containing 50 μg/ml kanamycin and incubated overnight at 37°C. Colonies were checked by PCR screening using the T7 promoter and T7 terminator forward and reverse primers as described above (2.2.5). Constructs with correct sequences were transformed into BL21 (DE3) cells for protein expression and purification.

2.2.8 Competent Cells

DH5α competent cells were generated for the purpose of DNA extraction from cells for cloning and DNA amplification. BL21 (DE3) competent cells were generated for protein expression. LB agar plates not containing antibiotics were streaked with BL21

or DH5α glycerol stocks and incubated at 37°C overnight. 50 ml of super optimal broth (SOB) medium was inoculated with a single colony and shaken at 150 rpm, 37°C until an A_{600} of 0.4-0.5 was reached. Cells were harvested by centrifuging at 3000 rpm for 15 min. The pellet was resuspended in 25 ml of ice cold 50 mM $CaCl_2$ and incubated on ice for 30 min. Cells were then pelleted by centrifuging at 3000 rpm for 15 min and resuspended in 2 ml of ice cold 50 mM $CaCl_2$. 50% (v/v) sterile glycerol was added to a final concentration of 10% and cells were divided into 100 µl aliquots in microfuge tubes. The aliquots were flash frozen in liquid nitrogen and stored at -80°C.

2.2.9 Transformation

Aliquots of competent cells in microfuge tubes were thawed on ice and 1 µl of DNA were added to 30 µl of cells and gently mixed. The cells were incubated on ice for 10 min followed by a heat shock treatment in a 42°C water bath for 45 s. Cells were returned to ice for 5 min, 200 µl of LB medium was added and the cells shaken at 150 rpm and 37°C for 45 min to 1 h. Several dilutions of transformed cells were spread on to LB agar plates containing 50 µg/ml kanamycin (Sigma) and incubated in a static incubator overnight at 37 °C.

2.2.10 Glycerol Stocks

For the generation of glycerol stocks 50 ml LB medium containing 50 µg/ml kanamycin was inoculated with a single colony of the transformed BL21 cells. The cells were incubated with shaking at 200 rpm, 37°C until an A_{600} of 0.5-0.6 was reached. 500 µl of the culture was removed and placed into a cryo tube, 50% (v/v) sterile glycerol was added to a final concentration of 10%. Cells were flash frozen in liquid nitrogen and stored at –80 °C.

2.2.11 Protein expression

LB agar plates containing 50 µg/l kanamycin were streaked with glycerol stocks of BL21 cells transformed with pET28a (+) plasmid containing the full-length CIRP, N-CIRP or N-CIRP mutant gene. Plates were incubated overnight at 37°C. For N-CIRP and N-CIRP mutant expression, cells were grown in minimal media (appendix A) containing $(^{15}NH_4)_2SO_4$. For all full-length CIRP and unlabelled preparations of N-

CIRP and N-CIRP mutants, the same procedures were followed using LB media. 50 ml of minimal media containing 50 µg/l kanamycin and 0.6 g/l $(^{15}NH_4)_2SO_4$ was inoculated with a single colony from the LB plates grown overnight. The culture was incubated by shaking at 200 rpm and 37°C until A_{600} reached 0.6-1.0. Cells were pelleted by centrifuging at 3000 rpm for 15 min and the pellet resuspended in 1ml of fresh minimal medium. The resuspended cells were used to inoculate a 1 l main culture of minimal medium containing 50 µg/l kanamycin and 0.6 g/l $(^{15}NH4)_2SO4$ The culture was incubated by shaking at 200 rpm, 37°C, until A_{600} reached 0.6. When an A_{600} of 0.6 was reached, 1 mM IPTG was added to induce protein expression and the culture incubated further for 2.5-3 h. Before the addition of IPTG, a 500 µl pre-induction sample was taken and the cells pelleted by centrifuging at 13000 rpm for 5 min. Pellets was stored at -20°C for analysis by SDS-PAGE. Following induction, a 500 µl post-induction sample was taken from the culture. Cells were harvested by centrifuging at 10000 rpm for 15 min at 4°C (Beckman Coulter Avanti J-25, JA10 rotor). The pellet was resuspended in 25 ml of 20 mM sodium phosphate buffer containing 0.5 M sodium chloride, pH 7.3 (lysis buffer), and frozen at -20°C for lysis.

For small-scale expression of full-length CIRP, a main culture of 50-200 ml LB containing kanamycin was used without a starter culture and induced at A_{600} of 0.6. Amount of lysis buffer to take up the harvested cells was adjusted in proportion to culture size.

2.2.12 Cell Lysis

Lysis was carried out to recover soluble proteins from the cells for the purification of full-length CIRP, N-CIRP or N-CIRP mutants. For lysis, the frozen, resuspended pellet was thawed at room temperature. Lysozyme was added to a final concentration of 100µg/ml from a 10mg/ml stock in lysis buffer and 5% (v/v) Triton X-100 in lysis buffer was added to a final concentration of 0.1%. The lysate was mixed and incubated at room temperature for 20-30 min. At the end of the incubation, 2 M $MgCl_2$ was added to a final concentration of 10 mM and 2 mg/ml DNase1 to a final concentration of 20 µg/ml. The lysate was incubated for a further 20-30 min at room temperature until the viscosity was reduced. A 500 µl total lysate sample was taken for analysis by SDS-PAGE. To separate the soluble fraction, the lysate was

centrifuged at 13000 rpm, 4°C for 15 min (Beckman Coulter, Avanti J-25, JA25.50 rotor). The supernatant was stored on ice for further purification. The pellet was resuspended in 25 ml lysis buffer. 500 µl samples were taken from both the pellet and the supernatant for SDS PAGE analysis.

2.2.13 Nickel Affinity Chromatography Purification

Full-length CIRP, N-CIRP and N-CIRP mutants were purified by nickel affinity chromatography using the His tag derived from the pET28a (+) vector. Nickel-bound chelating sepharose was used to purify hexaHis-CIRP and hexaHis-N-CIRP. A 3 ml column was prepared from 6 ml of 50% slurry of chelating sepharose fast flow (GE Healthcare UK Ltd.). The column was allowed to flow by gravity. Before use, 5 column volumes (CV) of water was passed through to remove the storage solution (20% ethanol). The column was charged with 2 CV of charging solution (0.2 M nickel sulphate). 3 CV of water was then passed through to wash off unbound Ni^{2+} ions, followed by 5 CV of acidic buffer (20 mM sodium acetate, 0.5 M NaCl, pH 4). The column was then equilibrated with 5 CV of equilibration buffer (20 mM sodium phosphate, 0.5 M NaCl, pH 7.4) and the lysate was loaded on to the column. 5 CV of equilibration buffer was passed through following loading followed by 8 CV of wash buffer (20 mM sodium phosphate buffer, 0.5 M sodium chloride, 60 mM imidazole, pH7.4) to remove loosely bound proteins. To elute N-CIRP or N-CIRP mutants, 8 CV of elution buffer (20 mM sodium phosphate, 0.5 M sodium chloride, 150 mM imidazole, pH 7.4) was used and 2 ml fractions were collected. When purifying full-length CIRP, 8 CV of EDTA buffer (20 mM sodium phosphate, 0.5 M sodium chloride, 20 mM EDTA pH 7.4) was used.

After use, column was then stripped with 3 CV of strip buffer (20 mM sodium phosphate, 0.5 M sodium chloride ,20 mM EDTA), was washed with 10 CV water and stored in 20% ethanol. The collected fractions and the flow through from the loading of the lysate were analysed by SDS-PAGE. The fractions containing CIRP protein were identified and pooled for further purification by ion exchange chromotography.

2.2.14 Urea Nickel Affinity Chromatography Purification

To purify full length CIRP in the presence of urea, the same procedure was followed as described in section 2.2.13 with the change that the acidic, wash, equilibration and the EDTA elution buffers was made with the addition of 6M urea and 1 ml fractions were collected of the elution.

2.2.15 Dialysis

Dialysis tubing (Medicell Membranes Ltd, MWCO = 12-14000 Daltons) was prepared by boiling for 10 min in 500 ml of soaking buffer (1 % w/v sodium bicarbonate, 10 mM EDTA, pH 8.0). The tubing was then rinsed with water and boiled in water for 5 min. Tubing was stored in 20% ethanol at 4°C. Dialysis was carried out with stirring overnight at 4°C against 2 l of buffer unless stated otherwise.

2.2.16 Inclusion Body Wash

For the inclusion body washing procedure, 50 mM Tris-HCl, 10 mM EDTA, 0.5 % Triton X100, pH 8.0 was used as the wash buffer. The lysis pellet which had been resuspended in lysis buffer following centrifugation and frozen, was thawed at room temperature. It was centrifuged at 13000 rpm for 15 minutes and the supernatant removed. The pellet was resuspended in 12.5 ml of inclusion body wash buffer and homogonised with a glass homogeniser for 5 minutes and centrifuged at 12000 rpm for 5 minutes. The supernatant was removed and stored on ice. The wash procedure was repeated once with wash buffer and once with water. The final pellet of inclusion bodies was resuspended in 2.5 ml water. Samples were taken from the pellets and supernatants at the end of each step for SDS-PAGE analysis.

2.2.17 Protein Refolding

After solubilising CIRP in urea, it was refolded by diluting the concentration of the urea present. This was done by either dilution or dialysis. For dilution, the protein sample in 6 M urea was added dropwise into refolding buffer. The volumes used were dependant on the protein concentration and are described in the results section. For the dilution method the protein was added drop wise while stirring into refolding buffer 50 mM Tris, 100 mM NaCl, 50 mM Arg, 50 mM Glu pH 7.4, to reduce the concentration of urea from 6 M to 200 mM. This was stirred overnight at 4°C. The

total volume was then concentrated down to 2ml using spin concentrator and 20 ml of refolding buffer was added to further reduce the urea concentration to 2 mM. The total volume was then concentrated to 2 ml again. To remove the urea with dialysis, the purified sample was dialysed in two steps. First step reduced urea concentration to 200 mM and the second down to 6.7 mM. Each step was an 8 hours dialysis at 4 °C with stirring. The diluted sample was stirred overnight at 4°C to allow the protein to refold. The sample was then concentrated using spin concentrators and then buffer exchanged by adding the same volume of refolding buffer as the first step. The refolded protein was then concentrated to 0.5 mg/ml (Vivaspin 20 concentrator, 10,000 MWCO, Sartorius).

2.2.18 Anion Exchange Chromotography Purification

Anion exchange was carried out on an AKTA FPLC purification system (Amersham Pharmacia Biotech). A 3 ml anion exchange column was prepared with quaternary ammonium (Q) Sepharose Fast Flow (GE Healthcare UK Ltd.) media. A flow rate of 2 ml/min was used throughout the purification process. All buffers were degassed and filtered through a 0.45 μm nitrocellulose membrane. The column was washed with water initially to remove the 20% ethanol storage solution. It was then equilibrated with 5 CV of equilibration buffer (20 mM Tris-HCl, pH 8.0). The sample was filtered (0.45 μm membrane) and loaded via a 50 ml superloop. 2 ml fractions of the flow through were collected. Following loading, the column was washed with 5 CV equilibration buffer. The elution buffer used was 20 mM Tris-HCl, 0.5 M sodium chloride pH 8.0. To elute the protein, a linear gradient of 0-0.5M NaCl was used over 20 CV, followed by a further 5 CV of the high salt solution. The elution was collected in 1 ml fractions. Protein was detected by absorbance at 280 nm and fractions containing protein were analysed by SDS-PAGE.

2.2.19 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples were analysed using a discontinuous gel system containing a resolving (12.5% or 15% acrylamide, 375 mM Tris-HCl, pH 8.8, 0.1% SDS, 0.1% ammonium persulfate and 0.02% TEMED) and stacking (4% acrylamide, 125 mM Tris pH 6.8, 0.1% SDS, 0.1% ammonium persulfate and 0.02% TEMED) gel. For

full-length CIRP samples, 12.5% gels were used, and 15% for N-CIRP and mutants. Samples were prepared in reducing sample buffer (50 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol) and boiled for 3 min before loading. Gels were run using Tris-glycine running buffer (25 mM Tris-HCl, 250 mM glycine, 0.1% SDS, pH 8.3) at 100 V and stained with coomassie brilliant blue R-250 (0.2% coomassie R-250, 50% methanol, 10% acetic acid) and destained with a solution of 10% methanol and 10% acetic acid. Gels were stored in water.

2.2.20 Quantifying Protein Concentration

To quantify the amount of purified protein (N-CIRP or N-CIRP mutants) band intensities on SDS-PAGE were compared with intensities of known standards (lysozyme). Varying amount of lysozyme was loaded on the gel from 0.5 μg to 4 ug and different amounts of N-CIRP was also run on the same gel. After the gels were stained, they were scanned and the intensity of all bands was estimated using the software ImageJ. A standard curve was constructed from lysozyme samples, which was used to estimate N-CIRP amounts.

2.3 Results

2.3.1 Site-directed Mutagenesis

For RNA binding experiments of N-CIRP, three mutant constructs were generated; F9A, F49A and F9AF49A. These residues were selected for mutation as they were identified as the key residues for binding in the RNA binding sequences RNP1 and RNP2 (figure 2.3). They were changed to an alanine removing the aromatic ring, which contributes to the binding of the RNA ligands to N-CIRP.

2.3.1.1 Generation of Mutant DNA Fragments by Overlap PCR

6 PCR reactions were run, to produce two overlapping fragments for the mutants F9A, F49A and F9AF49A. Two fragments with overlapping 5' and 3' ends of expected size were amplified successfully by PCR for all three mutants. The overlapping upstream and downstream DNA fragments were purified. The expected full-length fragment for all 3 mutants was 540 bp, which was the length of DNA to be amplified between the T7 promoter and T7 terminator. The full-length fragments were successfully generated and analysis of the PCR products on a 2% agarose gel showed a band between the 500 and 600 bp markers matching the expected size. Figure 2.5 shows the overlapping fragments generated for the F9A mutants and the final product from the overlap PCR reaction as an example.

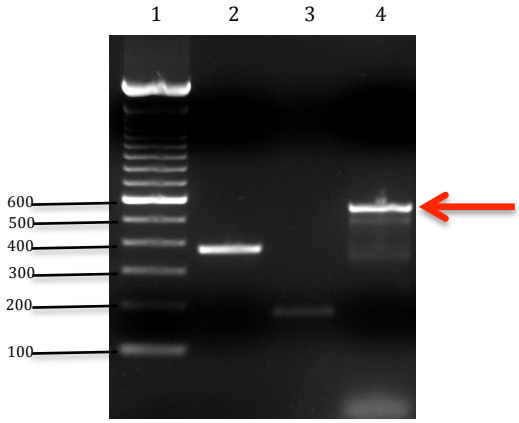

Figure 2.5 2% agarose gel showing the overlapping PCR products and the full-length fragment for the F9A mutant. Lane 1: 100 bp ladder, lane 2: downstream fragment (expected size: 369 bp), lane 3: upstream fragment (expected size: 184 bp), lane 4: PCR product from the overlap reaction (expected size: 540 bp) indicated by red arrow.

2.3.1.2 Cloning Mutant DNA Fragments into pET28a(+) vector via the pGEM T-easy system

The pGEM T-easy system was used as a method in between the generation of DNA fragments by PCR and cloning of the fragments into the final plasmid pET28a(+).The ligation reaction set up with a 5:1 molar ratio (insert:vector) was found to be successful and this ratio was used for subsequent ligations of all 3 mutants. White colonies were observed on the transformation plates and these were checked for the correct inserts by PCR colony screening.

PCR products from many of the screened colonies showed bands between the 500 and 600 bp marker around the correct expected size of 540 bp for all 3 mutants. An example PCR colony screen for the F9AF49A mutant is shown by figure 2.6(a) and similar results were seen for the other two mutants. DNA from the positive colonies was taken for double restriction digest with the Nde1 and EcoR1 enzymes to isolate the N-CIRP DNA for cloning into the pET28a (+) vector. An example gel is shown in figure 2.6(b) for the digestion of pGEM T-easy vector with the F9AF49A mutant insert. The analysis of the digested plasmid on agarose gel showed a main band for the positive clones at around 260 bp (figure 2.6 (b)), which was as expected as the size of the N-CIRP DNA between the two restriction sites was calculated to be 263 bp. Purified vector DNA was sequenced and the correct sequences of all 3 mutants was confirmed.

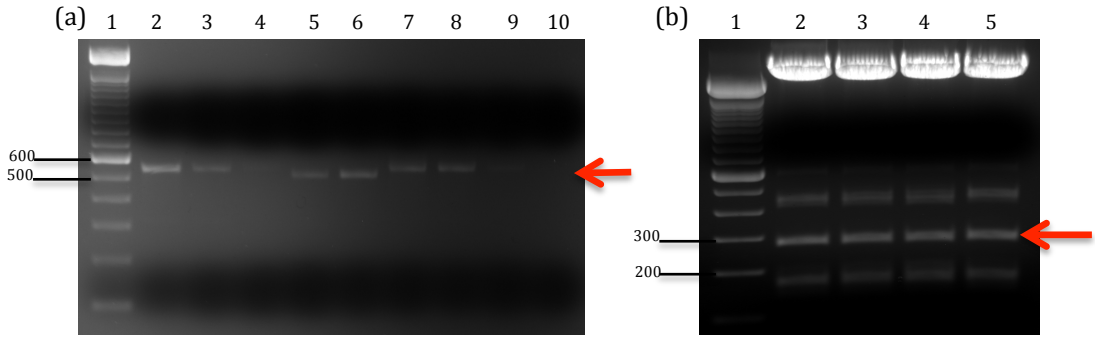

Figure 2.6 2% agarose gels. (a) PCR products from colony screen of the F9AF49A mutant fragment ligations into pGEM T-easy plasmid. Arrow indicates the position of the correct sized inserts. Lane 1: 100 bp ladder, lane 2: positive control, lanes 3-10: PCR product from screen of colonies 1-8. **(b) Restriction digest of the pGEM T-easy vector containing F9AF49A mutant DNA insert with Nde1 and EcoR1 enzymes.** Bands belonging to N-CIRP fragments are shown by the red arrow. Lane 1: 100 bp ladder, lane 2-5: clones 1-4.

2.3.2 Expression and purification of WT N-CIRP and Mutants

2.3.2.1 Growth, Induction and Cell Lysis

Once the sequence of the mutants in the pET28a (+) vector had been confirmed, plasmids were transformed into BL21 (DE3) *E.coli* cells for protein expression. In the wild-type pre-induction sample (figure 2.7 (a)), N-CIRP expression was absent but in the post-induction sample which was taken from the culture after 2.5-3 h of induction, a band was observed at the expected size of N-CIRP (11.6 kDa). Supernatant and pellet samples taken after centrifugation showed that there was very little N-CIRP in the pellet and a band similar to intensity of the band seen in total lysate sample was observed in the supernatant suggesting that all recombinant N-CIRP is soluble. This supernatant fraction was then taken forward to isolate N-CIRP from the rest of the soluble proteins. Post induction expression levels and behaviour of the mutants was similar to that of the wild type although a small amount of each (<10%) remained in the pellet fraction.

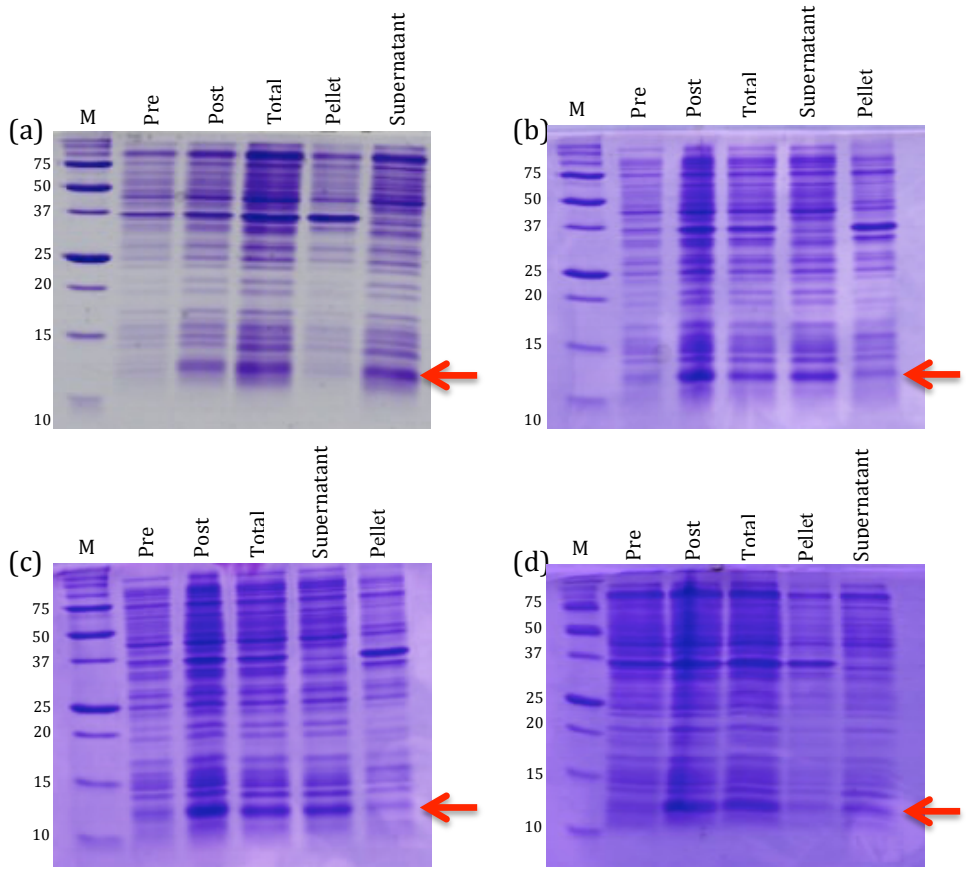

Figure 2.7 15% Reducing SDS PAGE analysis of pre/post induction, total lysate, lysis supernatant and lysis pellet samples of wild type N-CIRP and N-CIRP mutants (a) Wild type (b) F9A mutant (c) F49A mutant (d) F9AF49A mutant. Red arrow indicates the N-CIRP band.

2.3.2.2 Nickel Affinity Chromatography Purification

WT and mutant N-CIRP was purified from the lysis supernatants by nickel affinity chromatography. The nickel affinity chromatography profile of the mutant F9AF49A is shown as an example in figure 2.8.

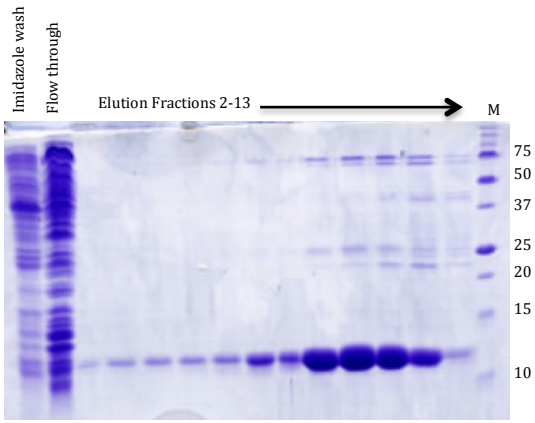

Figure 2.8 15% Reducing SDS PAGE analysis of nickel affinity purification of F9AF49A N-CIRP mutant.

All the protein loaded on to the nickel affinity column had bound as no N-CIRP protein was seen in the flow through sample on the gel (figure 2.8). There were a large number of proteins in the 60 mM Imidazole wash but no N-CIRP was seen here as the purpose of the low concentration imidazole wash is to remove loosely bound proteins. To elute N-CIRP, 150 mM Imidazole was used, the elution fractions on the gel showed that all of the protein eluted in 12 fractions. N-CIRP bands at the correct size were observed as well as some high MW contaminants. The WT protein and all 3 mutants showed a very similar elution profile with bands of similar intensity i.e. a similar amount of protein (figure 2.8). Fractions containing wild type or mutant N-CIRP protein were pooled for further purification by ion exchange chromatography.

2.3.2.3 Ion Exchange Chromatography Purification

The pooled fractions from nickel affinity chromatography of wild type N-CIRP and all three mutants were buffer exchanged by dialysis. The elution profile of the F9A mutant and SDS-PAGE analysis of the fractions is shown in figure 2.9 as an example.

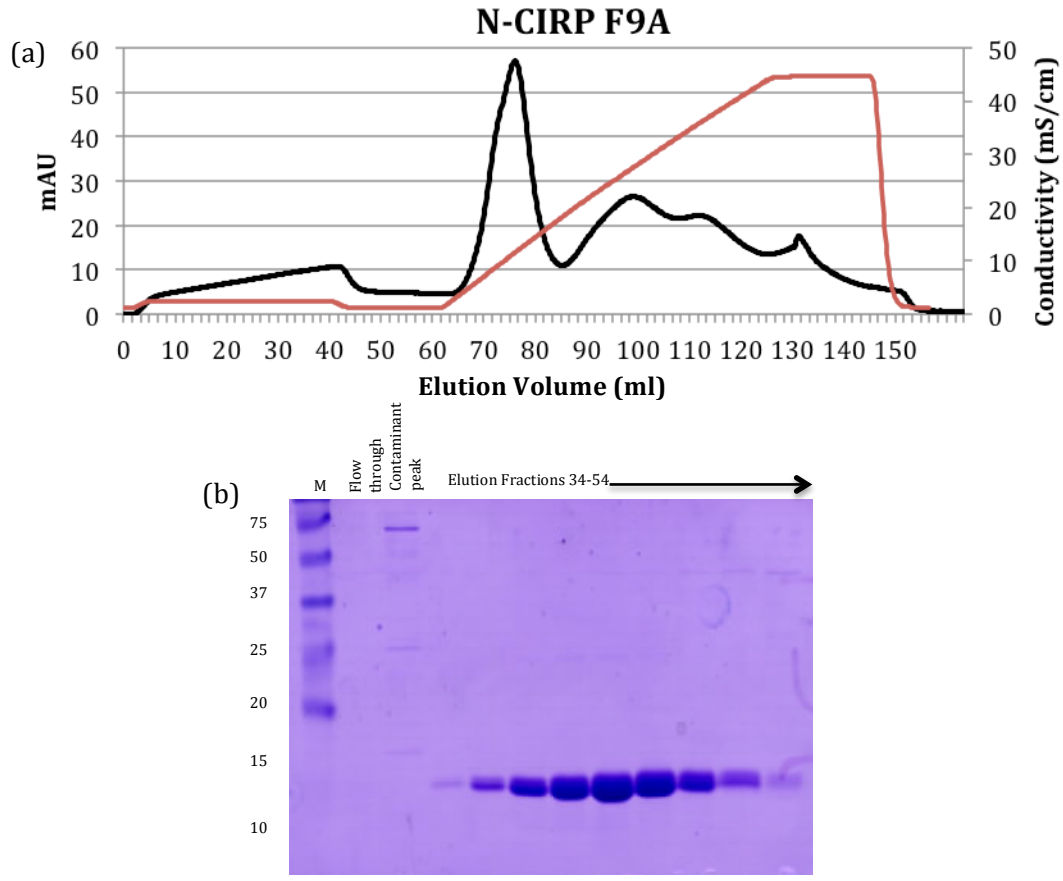

Figure 2.9 (a) Ion exchange chromatogram and (b) 15% reducing SDS-PAGE analysis of individual fractions taken from the purification of N-CIRP mutant F9A. Elution fractions were taken between elution volumes 65 and 85 ml as shown on chromatogram.

The elution profile of the F9A protein showed that N-CIRP eluted in a single peak (figure 2.9 (a)). A low concentration of salt was adequate to elute the protein from the IEX matrix. There was a small increase in the absorbance during loading, suggesting some protein was present in the flow through, however SDS-PAGE analysis showed no presence of N-CIRP in this fraction. A broad peak with a lower absorbance than N-CIRP appeared later on in the gradient. All 3 mutants and wild type showed a very similar profile. The proteins were eluted at the same point of the gradient as the wild type over 18-20 fractions (figure 2.9 (a)). The SDS-PAGE analysis showed a purified

N-CIRP band for the wild type and the mutants at the correct molecular mass (figure 2.9 (b)). The fractions containing N-CIRP or mutants were pooled (18-20 ml).

2.3.2.4 Mass spectrometry analysis of WT N-CIRP and mutants

The mass of the purified wild type and mutant proteins were determined by electrospray mass spectrometry. The theoretical mass was calculated using the ExPASy ProtParam tool. The expected average mass for the wild type was 11488.6 Da with the N-terminal methionine removed. For the F9A and F49A mutants, the calculated value was 11412.6 Da, 76 Da less than the wild type as this is the difference between the mass of a phenyalanine and alanine residue. For the F9AF49A mutants the expected mass was 11336.6, 152 Da less than the wild type as two phenylalanines are changed to alanines.

Two major peaks were seen on all of the spectra (figure 2.10). The major peak being the expected product for the wild type and the mutants. The second peak was 178 Da bigger than the main product consistent with the His-tag modification (alpha-N-6-phosphogluconoylation) previously observed in *E.coli* (Geoghegan et al. 1999), this peak was not observed on the double mutant spectra.

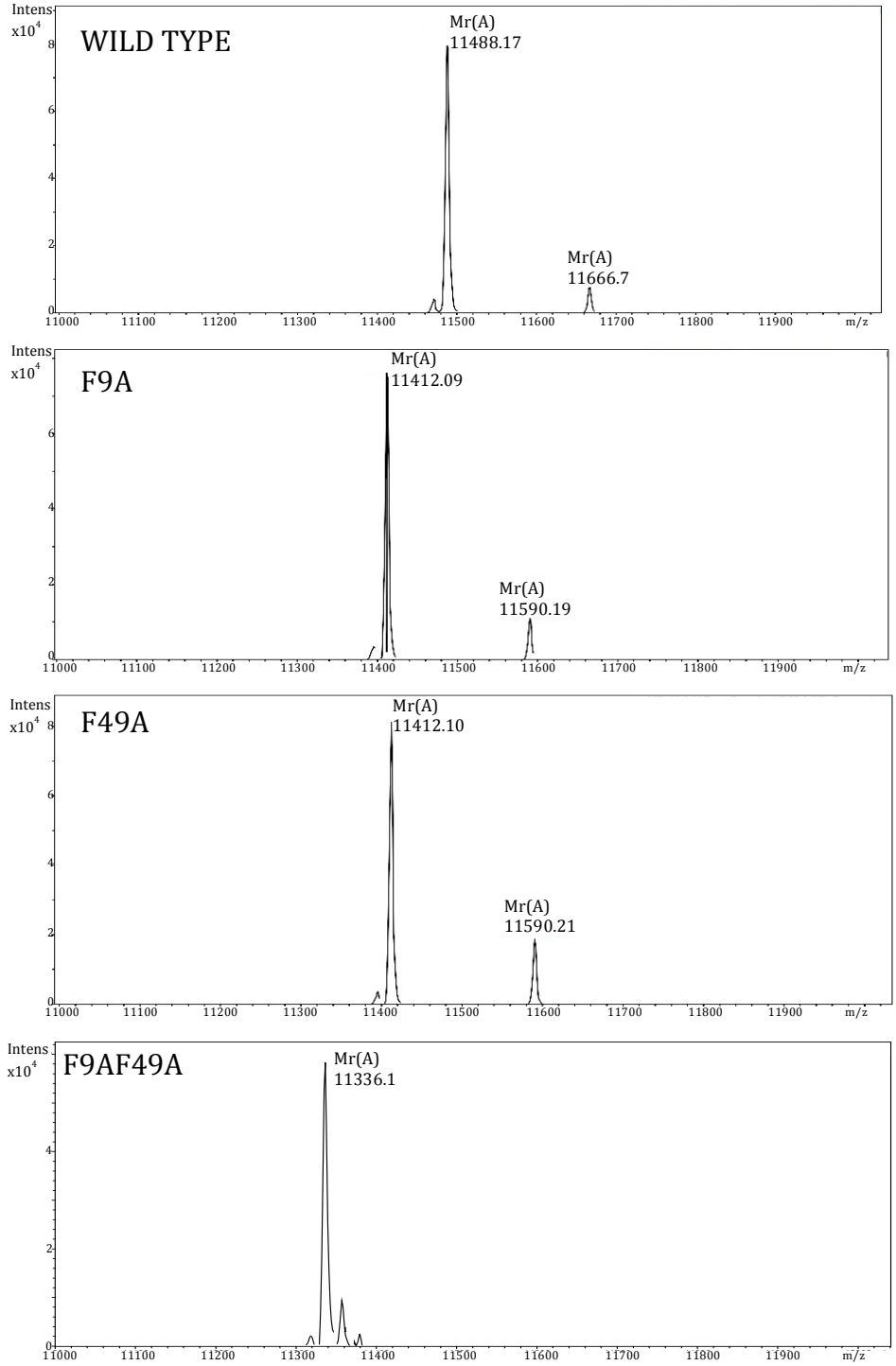

Figure 2.10 ESI-mass spectra of purified wild-type and mutant N-CIRPs.

2.3.2.5 Protein Concentration Estimation

To estimate the amount of protein expressed at the end of each prep, increasing amounts of protein from 2.5 µl to 10 µl was run on a SDS PAGE gel alongside some known protein standards (figure 2.11(a)). The amount of protein had to be estimated in this way as N-CIRP does not possess any tryptophan residues, so estimation by absorbance could not be carried out as this is the usual method.

Image J analysis allowed quantification of the staining density of each band, which was used to construct a standard curve from the know amounts of lysozyme loaded on the gel (figure 2.11 (b)). R^2 values for the fits were 0.97-0.99 in every calculation. The equation of the line was used to estimate the amounts of N-CIRP. A 1 l culture yielded around 10-15 mg of protein after purification. This was the same for all 3 of the mutants.

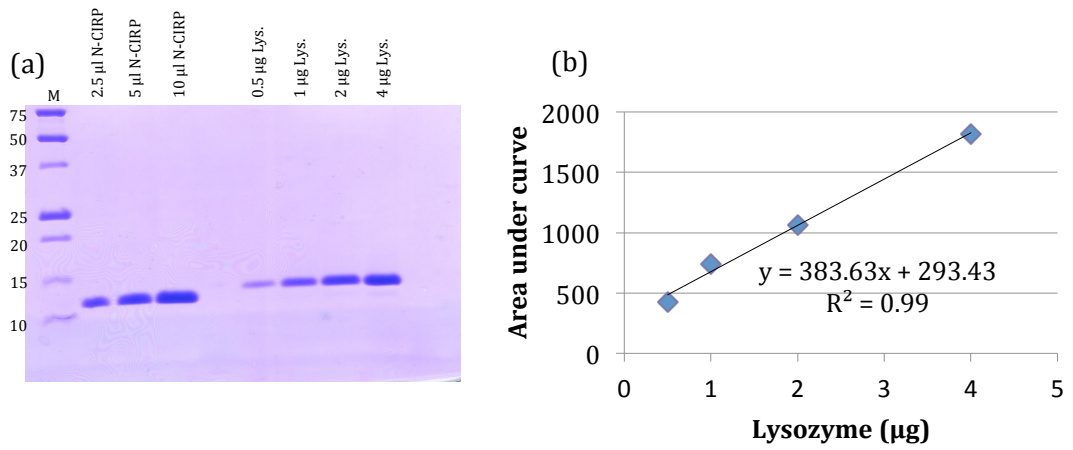

Figure 2.11 (a) 15% Reducing SDS-PAGE for the estimation of the amount of N-CIRP protein expressed and purified using lysozyme standards. (b) Lysozyme standard curve.

2.3.3 Expression and Purification of Full-Length CIRP

2.3.3.1 Small Scale expression and Purification

A small test expression of full-length CIRP was carried out in 50 ml LB media.

The calculated theoretical mass of CIRP is 20797.3 Da. In the pre-induction sample, there were no bands at the expected size for CIRP. The post induction sample showed that CIRP was expressed. A band just below the 20 kDa marker (figure 2.12) was seen running at a slightly smaller size than expected. After lysis CIRP was observed in the total lysate sample. The intensity of the CIRP band was reduced in the lysate supernatant and a small amount of CIRP was remaining in the lysis pellet. The lysate supernatant sample was purified by nickel affinity chromatography on a 1 ml small-scale purification column and 1 ml fractions of the elution was collected. Out of the 15 elution fractions collected, 3 of them contained small amounts of CIRP. Due to the low yields, SDS-PAGE gels are not shown here. These fractions from nickel affinity purification was analysed by mass spectrometry (figure 2.12). A peak was present at the correct size (20796.58 Da) for CIRP. Various other peaks were also present indicating the presence of other proteins in the sample. All of the other peak masses were identified as CIRP fragments by using the ExPASy FindPept tool confirming that these are degradation products of the full-length protein.

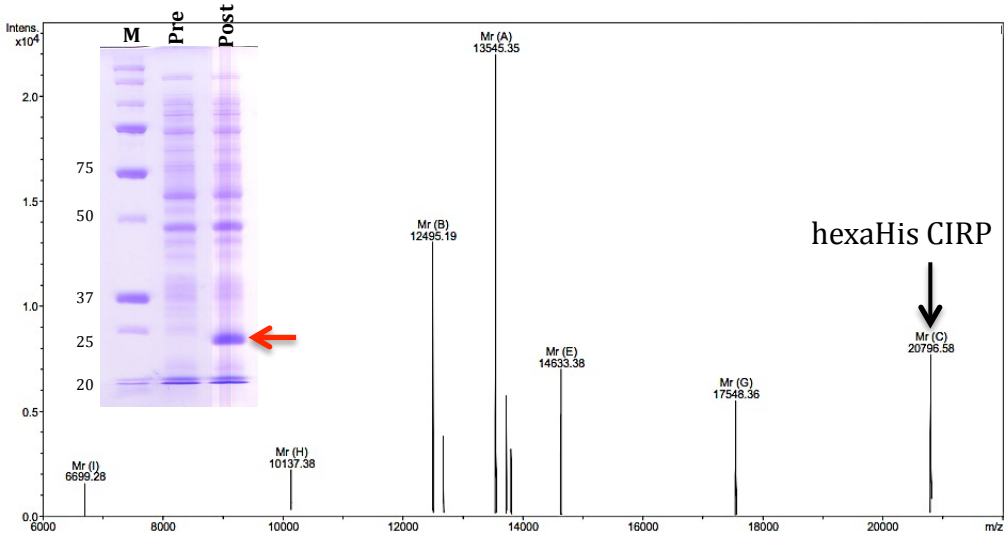

Figure 2.12 ESI-mass spectrum of nickel affinity purified full-length CIRP and 12.5% SDS PAGE analysis of the small-scale expression of full-length CIRP, pre and post induction. CIRP band indicated by red arrow.

2.3.3.2 Effect of protease inhibitors, expression temperature and sonication on CIRP

To inhibit the degradation of CIRP and to increase the amount of CIRP in the soluble fraction, expression at lower temperatures was tested and lysis was carried in the presence of protease inhibitors. The effect of sonicating the lysate to aid cell lysis was also investigated. A small-scale expression was carried out in 200 ml LB media. For lower temperature expression, the culture temperature was lowered to 20°C following the addition of IPTG and cells were induced overnight instead of for 2.5-3 hours. For the additional sonication step, the original lysis procedure was followed as normal, but prior to the centrifugation step to separate the soluble and insoluble fractions of the lysate, the lysate was sonicated on ice for 2 min. To test the effect of protease inhibitors, at the end of induction at 37°C, the culture was divided into two and harvested as normal. Each harvested cell pellet was taken up in 5 ml of lysis buffer with/without protease inhibitors (Roche- Complete Mini Protease inhibitor cocktail tablets).

The total lysate sample without and with inhibitors (figure 2.13(a)) contains a similar amount of CIRP protein. When the lysis supernatant samples were compared, the sample which includes the inhibitors had a larger amount of CIRP compared to the sample without inhibitors. The same conclusion was also made for the lysis pellet samples. Sample with the inhibitors showed more total protein as well as CIRP protein. Some of the CIRP protein expressed remained insoluble in the pellet. As the culture size increased, the ratio of soluble: insoluble CIRP decreased. Induction at 20°C lead to an increase in the amount of CIRP protein expressed as well as total protein in the cells (figure 2.13(c)). When the amount of CIRP in the lysis pellet compared to the lysis supernatant was compared between 37°C (figure 2.13(b)) and 20°C, a higher proportion of CIRP was seen in the pellet vs. the supernatant for the 20°C expression compared to 37°C. About 10% of total CIRP was in soluble in the lysis pellet when expressed at 20°C whereas this number was about 30% at 37°C expression. Sonication of the total lysate lead to small increase in the amount of CIRP that remained soluble in the lysis supernatant, this was increased in the 37°C expression sample.

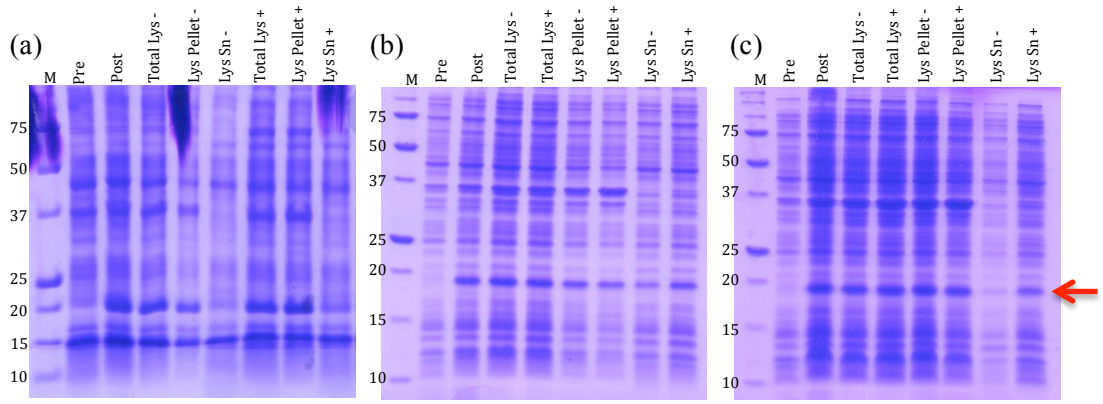

Figure 2.13 12.5% SDS PAGE analysis of the effect of low temperature expression, sonication and protease inhibitors on CIRP degradation. (a) Effect of protease inhibitors, the -/+ notations indicate the absence and presence of protease inhibitors (b) 37°C expression (c) 20°C expression, the -/+ notations indicate not sonicated and sonicated samples. CIRP band position is indicated by red arrow.

2.3.3.3 Large Scale Expression and Purification

To provide material for further characterization studies on full-length CIRP, large-scale expressions were carried out in 1-liter of LB media.

Full length CIRP was successfully expressed in a 1-litre culture (Figure 2.14(a)). The total lysate sample showed a band for CIRP similar to the post induction sample. A larger amount of the protein was seen in the lysis pellet compared to in the supernatant.

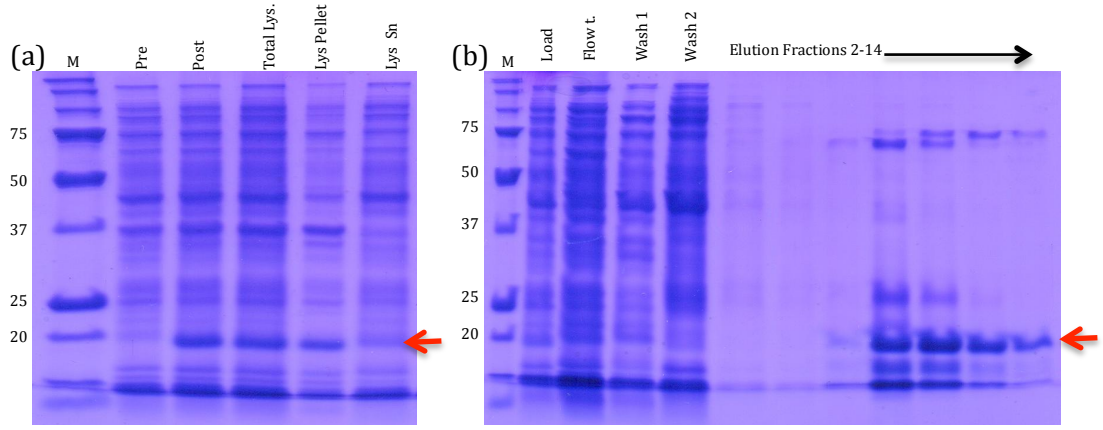

Figure 2.14 12.5 % SDS PAGE analysis of expression and lysis (panel a) and purification (panel b) of large scale CIRP expression. CIRP band position is indicated by red arrow.

The small amount of remaining soluble CIRP in the supernatant was purified. SDS PAGE analysis of the nickel affinity fractions showed that there was some CIRP in

the flow through of the load so not all of it bound to the column and some was also washed out in the wash fractions (figure 2.14(b). The protein that was bound to the column was eluted over 8 fractions. 5 out of the 8 fractions, which were the fractions that contained the most protein as seen from the SDS PAGE analysis, contained some precipitate upon short storage (approx. 1 h at 4°C). The fractions containing the aggregates were centrifuged and the supernatants were taken off and pooled together. The pellets were resuspended in the same volume of buffer as the original fraction (1 ml) and pooled together. The soluble fractions were then dialysed overnight into 50 mM sodium acetate pH 4.8 buffer as previous studies by Shu-Ju Hsieh had shown that CIRP is more stable and soluble in this buffer condition. At the end of the dialysis process more precipitates had formed, the precipitates were pelleted by centrifugation and the supernatant was removed. The insoluble and soluble fractions before and after dialysis were analysed by SDS PAGE.

The pooled soluble fractions of CIRP (figure 2.15(a)) contained some CIRP protein but also contaminating proteins as well. After dialysis, the majority of this protein remained soluble in the sodium acetate buffer. The analysis of the precipitate seen after dialysis showed that a very small amount of CIRP protein was present in this precipitate but this was very small compared to the fraction that remained soluble. Analysis of the precipitate observed in the elution fractions after nickel affinity showed that it contained a large proportion of the CIRP protein, this insoluble protein did not contain any contaminants and was quite pure (figure 2.15 (a)). Approximately 60% of CIRP precipitate was soluble after resuspending in the sodium acetate buffer (figure 2.15(b)). The total amount of soluble CIRP protein was estimated to be 800 μg at a concentration of 400 μg/ml. To test the optimal storage conditions for the protein, it was divided into two. One half was stored at 4°C and the other half at -20 °C for a week (figure 2.15(c)). CIRP which was stored at 4°C showed degradation products on the gel. Storage at -20°C did not show any degradation, however, the amount of protein was estimated to be 200 μg/ml compared to the 400 μg/ml prior to storage.

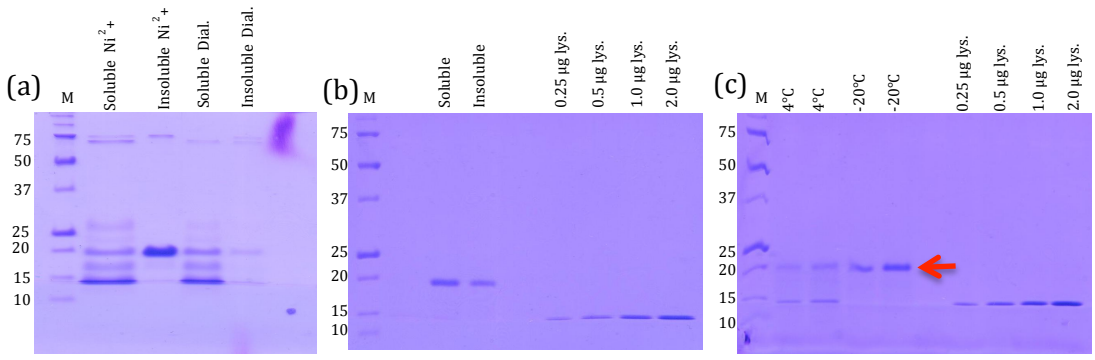

Figure 2.15 12.5% SDS PAGE analysis of soluble and insoluble CIRP and re-solubilisation. (a) Soluble and insoluble fraction analysis following nickel affinity purification, followed by dialysis in sodium acetate. (b) Resolubilisation of insoluble protein following nickel affinity in sodium acetate. (c) Resolublised N-CIRP storage at 4°C and -20°C. CIRP band position is indicated by red arrow.

It was noted that one of the degradation products seen on the gel was at 13-14 kDa. This is 1-2 kDa larger than the size of the N-terminal region. As the C-terminal region is natively disordered it was thought that it could be this region undergoing proteolysis. To confirm this, the CIRP sample that was stored at 4°C which showed degradation (figure 2.15(c)) was run through a small nickel affinity column. As CIRP contains an N-terminal His tag, the degradation product would bind to the column if it were the N-terminal region of the protein.

SDS PAGE analysis of the nickel affinity purification showed that all of the 13-14 kDa product bound to and could be eluted from the Ni^{2+} affinity column (figure 2.16). This behaviour suggests that this fragment is the N-terminal region of CIRP. By the time this experiment was carried out, all full-length protein had degraded; therefore the samples on gel do not contain any full length CIRP. Fractions were analysed by mass spectrometry (figure 2.16). Various peptides between 13 and 15 kD were present and no full length CIRP was present. The position of these peptides in the CIRP amino acid sequence were estimated by using the ExPASy FindPept tool, the results showed that most the peptides found in the sample were consistent with the N-terminal of N-CIRP including the His tag except for one (MW=13879.91)(table 2.2).

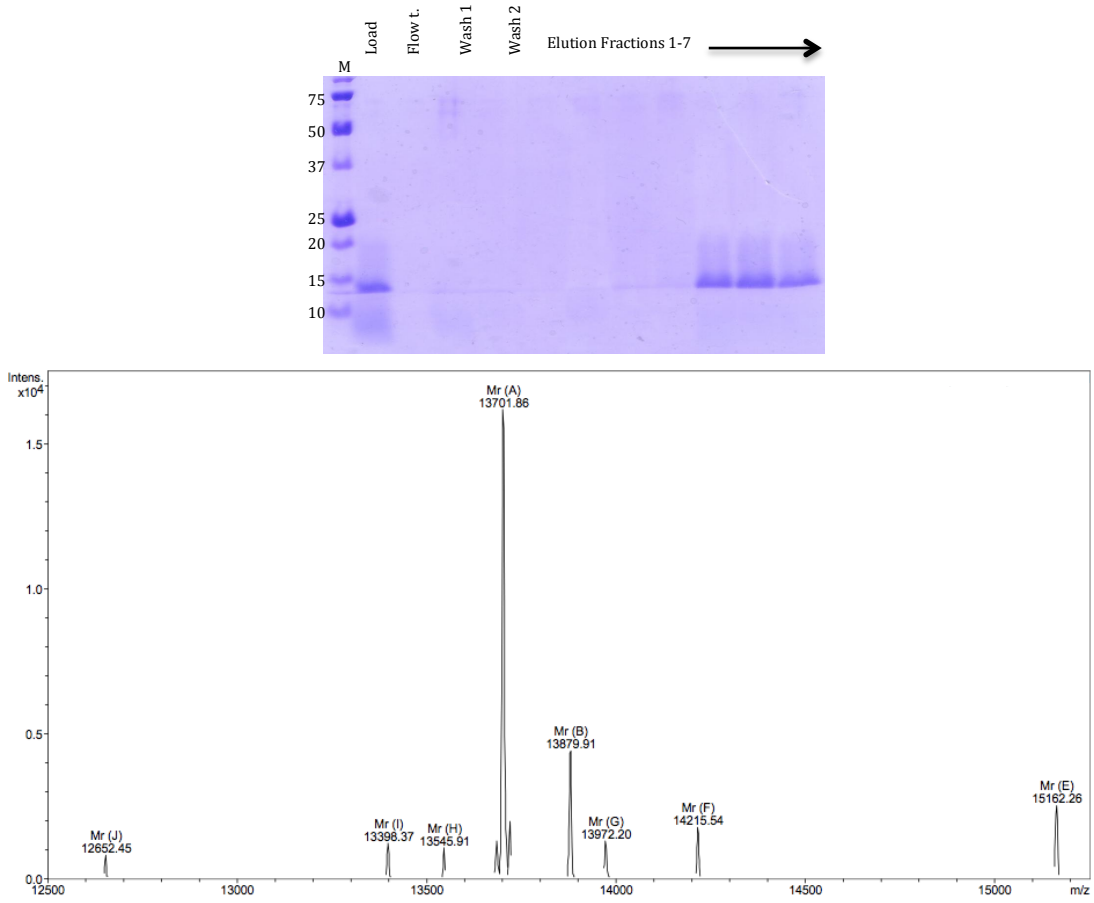

Figure 2.16 Analysis of the purified CIRP degradation products by nickel affinity chromatography and ESI mass spectrometry.

Mass (Da)	ΔMass (Da)	Sequence Position
12562.450	0.437	1-115
	0.437	2-116
	0.515	27-143
13398.320	1.297	1-124
13545.910	1.468	62-187
	0.064	23-148
	0.888	42-166
	0.934	1-125
13701.860	1.234	60-186
	0.184	19-145
	1.172	1-126
	1.172	2-127
13879.910	0.674	26-153
	0.658	15-143
	0.108	62-190
	0.409	10-138
13972.200	0.136	19-148
	1.123	1-129
14215.540	0.104	6-136
	1.049	1-131
	1.049	2-132
15162.260	0.130	39-179
	1.306	1-141

Table 2.2 Peptides identified in CIRP degradation products and their position in the protein sequence. The masses which include the His tag are shown in red.

2.3.3.4 Effect of Arginine and Glutamic Acid on Solubility of CIRP

As described in the previous section 2.3.3.3, CIRP precipitated in the fractions eluted from Ni-affinity chromatography and a proportion of this precipitate could be re-solubilised in sodium acetate buffer. The amino acids arginine and glutamic acid (50 mM of each) were added to the sodium acetate buffer to investigate their affect on CIRP solubility.

When the buffer containing Arg and Glu was compared with the buffer without Arg and Glu, there was only a small difference between the amounts of protein that re-solubilised (figure 2.17). The buffer containing Arg and Glu contained a little more soluble CIRP and similarly more CIRP remained insoluble in the buffer without Arg and Glu.

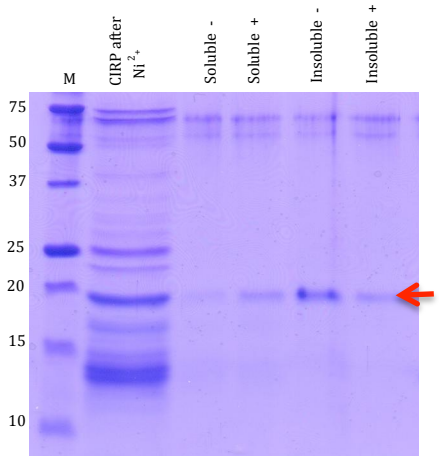

Figure 2.17 The effect of arginine and glutamic acid on CIRP solubility. CIRP was resolubilised in sodium acetate buffer with (+) or without (-) Arg and Glu following precipitation after nickel affinity purification. CIRP band position is indicated by red arrow.

2.3.3.5 Inclusion Body Washing of Lysis Pellet To Recover CIRP

Following lysis, a large proportion of the CIRP protein was found to be in the lysis pellet. The amount of protein which could be purified from the lysis supernatant was not enough for further experiments and hence experiments were carried out to try and recover the CIRP from the lysis pellet.

The CIRP protein in the lysis pellet was unexpectedly solubilized by the inclusion body wash buffer (figure 2.18(a)). After each wash, approximately 50% of the CIRP protein in the pellet was solubilised. After wash 2, the majority of the protein had been solubilised and after wash 3 almost no CIRP was left in the pellet.

The supernatants of the inclusion body washes 1 and 2 were pooled together and dialysed overnight into the nickel affinity equilibration buffer for purification. Majority of the CIRP from the inclusion body wash remained soluble through dialysis (figure 2.18(b)). There was a small amount of CIRP in the aggregate seen in the sample after dialysis. The soluble fraction after dialysis was purified by nickel affinity chromatography. The elution fractions had some aggregates in them. Each fraction was centrifuged and the pelleted aggregates were resuspended in 1ml sodium acetate buffer as described previously. However, the aggregates were not successfully re-solubilised and all of the CIRP protein that was purified was found in the insoluble fraction.

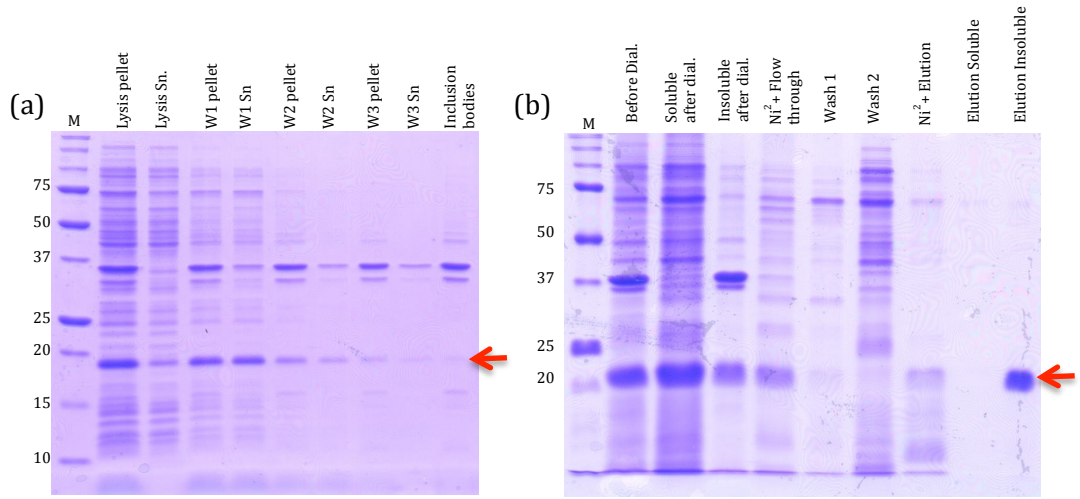

Figure 2.18 12.5 % SDS PAGE analysis of CIRP lysis pellet inclusion body washes and purification of solubilised CIRP. (a) Inclusion body washing of CIRP lysis pellet, W= wash (b) Dialysis, purification and solubilisation of CIRP recovered by inclusion body washing. CIRP band position is indicated by red arrow.

2.3.3.6 CIRP Refolding Experiments

Another approach taken to recover soluble CIRP was by denaturation of the inclusion bodies with urea and refolding by dilution. Urea was added to the wash buffer solubilized lysis pellet and the CIRP protein purified in the presence of the denaturant to avoid multiple precipitation and re-solubilisations of the protein and therefore also minimizing the amount of protein loss throughout the different steps. Once the protein was purified, it could then be refolded by diluting the denaturant present.

Similar to previous purifications without urea, some CIRP protein was seen in the flow through of the load, which had not bound to the column (figure 2.19(a)) The protein was eluted with EDTA and CIRP was found to have eluted over 8 fractions similar to previous purifications. Two methods of refolding were tested, these were dilution and dialysis. At the end of both techniques the final refolded sample had a cloudy appearance and a pellet was formed when centrifuged.

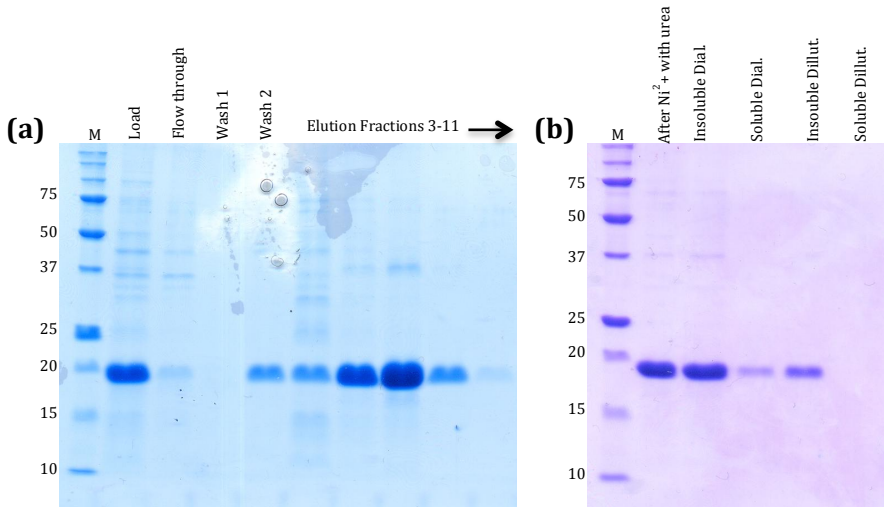

Figure 2.19 12.5% SDS PAGE analysis of the 6 M urea nickel affinity purification and refolding. (a) Nickel affinity purification in the presence of urea fractions (b) Refolding of CIRP by removing urea using the dialysis or the dilution method.

Both methods of refolding showed formation of large amounts of aggregated protein upon the dilution of the urea concentration. CIRP protein did not remain soluble when the urea was removed. The dialysis refolding method showed that a very small amount of CIRP was still soluble (figure 2.19(b)) but most of the protein precipitated. The dilution refolding method showed no soluble CIRP protein was present when the

urea was removed. It was also noticed that the amount of precipitated CIRP was much less than the initial total protein so some protein was lost. This could be due to use of the spin concentrators where protein can stick to the concentrator membrane.

2.4 Discussion

In this chapter, the generation of the N-CIRP mutants and the process by which the wild type N-CIRP and mutants were expressed and purified was described. To generate the mutants, a PCR-based site-directed mutagenesis method was used. The residues F9 and F49 were both changed to alanine, removing the aromatic ring, which is a key element for binding (chapter 1, section 1.4.2). A double mutant was also made containing both mutations. The DNA overlap fragments upstream/downstream of the mutation were made successfully for all mutants after some primer optimization. The full-length fragment was successfully generated for the F9A fragment. However, initially no product was made from the overlap reaction for the F49A mutants. The annealing temperature required for the overlap regions must be taken into consideration when using this method. A new step to the PCR program was introduced where 5 cycles would run prior to 25 main cycles and these 5 cycles were set to have a low annealing temperature of 43°C compared to the 48°C of the main cycles. The overlap primers used had a T_m of 67.8°C which corresponds with the T_m of the region of overlap so the need for a temperature of 43°C for annealing of this region was unexpected. However this approach successfully produced the full fragment for the F49A mutant. Full length fragments were successfully cloned into the pET28a(+) plasmid successfully via the pGEM T-easy system. The constructs were confirmed by PCR screening and DNA sequencing.

Wild type N-CIRP and all mutants were expressed successfully. ^{15}N was added to the expression media as the sole nitrogen source for NMR experiments. Post induction samples showed a band for N-CIRP and mutants at the expected size of 11.6 kDa, therefore the mutations had no effect on the expression of the protein. Wild type N-CIRP and mutants remained stable throughout the lysis procedure and there was no evidence to suggest that mutant proteins were more susceptible to proteolysis. The mutants behaved in the same way as the wild type in terms of binding and elution from the nickel affinity and ion exchange columns. Furthermore, there was not a difference between protein yields for the wild type and the mutants after purification. Wild type and mutants N-CIRP were therefore successfully generated, expressed and purified for further structure and binding activity analysis by NMR as described in chapters 3 and 4. The structure of N-CIRP mutants was confirmed by mass

spectrometry and the mutants were found to be as soluble and as stable as the wild type upon storage. A 1 liter culture yielded around 10-15 mg of protein after purification. This was the same for all 3 of the mutants.

Various methods were tested to express and purify soluble full-length CIRP protein. It was necessary to express the full-length protein to fully characterise CIRP function as a whole and to investigate the function of the C-terminal region. This disordered region may have made the recombinant expression and purification of the protein challenging by reducing the solubility of the protein. The unstructured nature of the C-terminal region also makes the protein prone to degradation and yields were higher when the lysis step was carried out in the presence of protease inhibitors. Mass characterisation of a degraded sample of CIRP (figure 2.16) confirmed that it is C-terminal region that is degraded first whereas the folded N-terminal domain is relatively resistant to proteolysis. Full-length CIRP was successfully expressed in *E.coli* with a yield of approximately 5 mg/l. However, the final yield of soluble protein was significantly less as the protein precipitated after Ni^{2+}affinity purification, on dialysis and concentration. Interestingly, intact CIRP runs at a slightly smaller size than expected on SDS-PAGE. It is not clear why this so. It may be due to degradation and the presence of degradation products were seen on the gel. Mass spec did show however the presence of full-length CIRP (figure 2.16). When cells are lysed, proteases which normally have regulatory roles are released in to the lysate and all proteins are exposed to proteolysis (Ryan 2011). These degradation products were not observed when expressing the N-terminal domain of the protein and as the C-terminal region is unstructured it is likely to be prone to proteolysis. The lysis without protease inhibitors showed a reduced amount of CIRP compared to lysis with inhibitors. Studies have shown that unstructured regions of proteins are highly susceptible to degradation especially *in vitro* and in the absence of ligands (Mittag et al. 2010; Tompa et al. 2008). Although some degradation products were still observed on the gel when protease inhibitors were present, this experiment confirmed that the CIRP is prone degradation by proteolysis during the lysis procedure.

It was noted that the larger the scale of expression, the larger proportion of CIRP protein present the in the lysis pellet. High local concentration of proteins susceptible to aggregation can result in the protein favouring the aggregate form rather than

folding correctly (Idicula-Thomas & Balaji 2005). Lysis of cells from a 1 litre expression showed that around 2/3 of the total protein was in the insoluble lysis pellet. This suggests that the aggregation of CIRP could be concentration dependent and a useful future experiment would be to use higher lysis buffer volumes to keep the CIRP concentration lower. Reports suggest that induction of recombinant protein expression at lower temperatures can improve protein folding and quality of the final product (Jana & Deb 2005; Smith 2007; Sorensen & Mortensen 2005b). CIRP was induced at 20°C overnight; reduced temperature induction for 16 h increased the amount of protein expressed by the cells but the majority of the expressed protein was still found in the insoluble lysis pellet. This finding further confirms that the poor solubility of CIRP causes it to precipitate at higher concentrations.

Purification of the soluble CIRP in the lysis supernatant resulted in precipitation of CIRP in the purified elution fractions, precipitation was seen in the fractions containing the highest concentration of CIRP, once again suggesting that solubility is influenced by protein concentration. ¾ of the purified protein precipitated, the remaining soluble protein was not very pure as it contained contaminating proteins and degradation products whereas the precipitated protein sample was much purer. Previous studies by Shu-Ju Hsieh had shown that CIRP is more soluble and stable in 50 mM sodium acetate buffer, pH 4.8. Resuspending the CIRP precipitate in this buffer lead to around 60% of the precipitate becoming soluble. This re-solubilisation is beneficial as this solubilised protein is higher in purity than the nickel affinity eluted fractions. This precipitation step can be seen as an extra purification, which also yields the protein at relatively high concentration.

Mass spec analysis of Ni^{2+} affinity purified CIRP samples confirmed that the degradation of CIRP is due to the unstructured C-terminal region. The N-terminal region of CIRP on its own is stable; whereas the unstructured C-terminal region is prone to proteolysis. Aggregation of CIRP upon purification suggests that taking it out of the cellular environment causes the protein to precipitate. The addition of the amino acids arginine and glutamic acid made a small improvement to the amount of CIRP that could be re-solubilised. Molecular crowding of protein in a cellular environment can help enhance their solubility which the amino acids assist to achieve

(Shukla & Trout 2011). Addition of amino acids have increased the solubility of protein in many studies, arginine and glutamic acid in particular help with this (Arakawa et al. 2007; Golovanov et al. 2004).

Inclusion body washes were performed on the lysis pellet and it was expected that this process would be unable to re-solubilise the CIRP protein. Inclusion bodies are typically composed of densely packed denatured protein molecules that are insoluble in aqueous buffers (Singh & Panda 2005). The solubilisation of the CIRP protein could be due to the buffer components EDTA and Triton X-100 and also due to the fact that the amount of CIRP is diluted by the addition of the buffer after every wash and so allowing more CIRP to dissolve. If CIRP was forming true inclusion bodies it would not be expected to become soluble in the inclusion body wash buffer. The CIRP protein solubilised from the lysis pellet was less soluble after Ni^{2+}affinity purification and none of the precipitate could be re-solubilised in the sodium acetate buffer. This suggests that the CIRP protein aggregates from the lysis pellet may be a different form compared to the CIRP in the lysis supernatant, as it is more insoluble. CIRP was purified in the presence of urea successfully with no loss of protein through precipitation, however, the refolding process was unsuccessful and the aggregated protein could not be recovered by resuspending in sodium acetate buffer.

The studies described in this chapter suggest that there are various different factors that can lead to the precipitation of CIRP, such as protein concentration, buffer conditions, presence of other proteins and culture conditions. The difference between how much of the protein precipitate could be solubilised form the lysis pellet or the aggregated protein following nickel affinity purification suggests that there may be more than one form of the protein; 1) soluble, 2) protein prone to aggregation which can be re-solubilised 3) insoluble protein. The fact that some precipitation occurs after purification suggests that to be soluble, full-length CIRP requires to be in a cellular environment and even in the presence of its ligands The greatest amount of purified soluble protein (400 μg per 1 litre culture) was achieved from re-solubilisation of the precipitates following nickel affinity chromatography in the sodium acetate buffer; this was also the most pure CIRP sample that was achieved. CIRP protein in the lysis pellet could not be salvaged. To maximise the amount of

soluble refolded protein, there are many conditions that should be taken into consideration, such as buffer type and pH, protein concentration, temperature, storage conditions, and there are many additives available that can be included to assist with protein solubility (Jungbauer & Kaar 2007). These could be further optimised to maximise future CIRP yields.

Chapter 3

RNA Binding Studies of CIRP and N-CIRP

3.1 Introduction

Nuclear Magnetic Resonance is an ideal tool to investigate structure-function relationships of proteins and N-CIRP is well suited to such analysis as it is a well-behaved, soluble, small molecule. NMR has been developed for many decades for both small and large biomolecules (Bloch, Hansen, & Packard, 1946; Bloembergen, Purcell, & Pound, 1947; Rollin & Hatton, 1947). NMR is based on the fundamental property of nuclear spin. For example ^{1}H, ^{13}C and ^{15}N isotopes have nuclei with spin ½ and give rise to well resolved high-resolution resonances, therefore they are deemed as NMR active isotopes. Isotopes with a spin of 0 do not give rise to NMR spectra and are NMR inactive, for example, the naturally occurring ^{12}C isotope. Naturally occurring ^{14}N, with a spin of 1, gives broad line shapes so it cannot be used for NMR experiments either. ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P are all common NMR active nuclei used in NMR analysis, although some have a low natural abundance; e.g. ^{13}C is 1.1% and ^{15}N is 0.37%. To utilise these nuclei the protein must be labelled with ^{13}C and/or ^{15}N isotopes usually by supplementing growth media with ^{15}N-ammonium sulphate and/or ^{13}C-glucose to enable acquisition of NMR spectra from these isotopes together with ^{1}H. The chemical shift of a nucleus depends on the chemical group that it belongs to and its microenvironment, therefore chemical shift changes of a resonance indicate a change in its microenvironment and hence a chemical or structural change. NMR can also deal with folded and unfolded proteins and can be used to identify unfolded regions of proteins. Regions of proteins, which are of a disordered nature or where there is degradation, give resonances in the random coil region of the spectrum (Jacobsen, 2007).

One of the aims of this project was to characterise the contribution of the N- and C-terminal regions of CIRP to its function. This chapter describes the investigation of the N-terminal domain of CIRP and its binding to RNA by the use of NMR spectroscopy. N-CIRP has a β1α1β2β3α2β4 topology which contains two RNA binding sequences RNP1 and RNP2 on strands β1 and β3 (Sheikh, 1997) (figure 1.6). Some residues on these strands have been identified as being more involved in RNA binding than others and the aromatic residues are highly conserved(Manival, Ghisolfi-Nieto, Joseph, Bouvet, & Erard, 2001). Residues in positions 3 and 5 on RNP1 and position 2 on RNP2 are essential for RNA binding (Maris, Dominguez, & Allain, 2005) as described in detail in section 1.4.2. As described in chapter 2, two of these

residues (position 3 of RNP1 and position 2 of RNP2) were chosen for mutagenesis to determine the effect on RNA binding. Interaction of these residues with RNA ligands was previously confirmed by NMR chemical shift perturbation studies on N-CIRP by Shu-Ju Hseih (2010). Further structural characterisation of N-CIRP will help to reveal the importance of the structure-function relationship of this protein in the mammalian response to mild cold shock. In this chapter, N-CIRP analysis by simple ^{15}N-^{1}H HSQC experiments to study the effects of mutations, and the RNA binding of full-length CIRP, N-CIRP and mutants were investigated. Further structural analysis through structure determination and dynamics experiments of N-CIRP is described in Chapter 4.

3.1.1 ^{15}N-^{1}H Heteronuclear Single Quantum Coherence NMR

^{15}N labelled protein is used to record two-dimensional HSQC NMR spectra. ^{15}N-^{1}H HSQC experiments are usually first to be performed on labelled protein as it is an ideal experiment to check whether the protein is folded and stable by looking at the dispersion and resolution of the spectrum (McIntosh & Dahlquist, 1990). ^{15}N-^{1}H HSQC data can be used to measure chemical shift changes and these provide detailed information about hydrogen bonding interactions, oxidation states, the ring current influence of aromatic residues and nature of hydrogen exchange dynamics (Mielke & Krishnan, 2009). Each residue of the protein (except proline) has an amide proton attached to a nitrogen in the peptide bond. The HSQC informs on the correlation between the nitrogen and amide proton (figure 3.1) and each amide yields an observable peak in the spectra with the exception of proline, which lack the amide proton.

Changes in chemical shifts provide indications of structural changes. Each residue has a separate backbone resonance for ^{1}H and ^{15}N; ^{15}N-^{1}H HSQC experiments are therefore a simple way to determine whether a protein conformationally or structurally changes under different conditions by the chemical shift changes observed. To follow changes in backbone NH chemical shifts, the control 2D ^{15}N-^{1}H HSQC spectra should have been assigned. Backbone assignments of N-CIRP were completed by Shu-Ju Hseih previously.

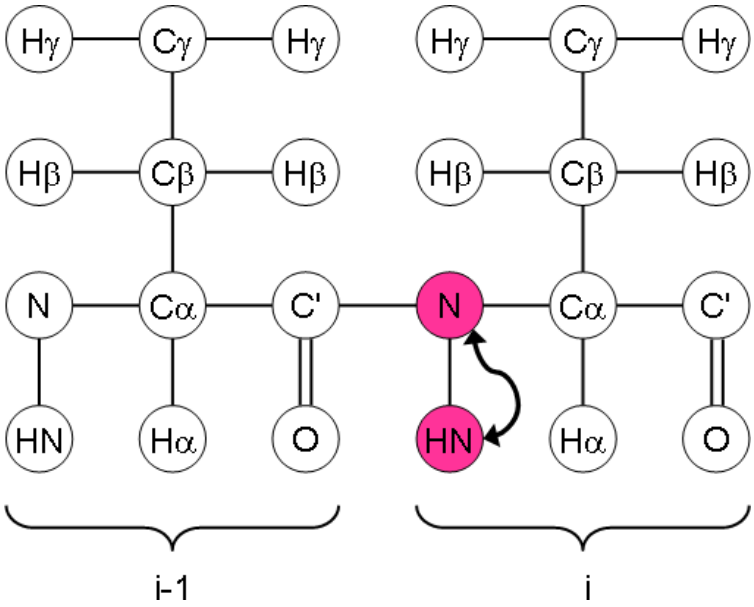

Figure 3.1 Magnetisation is transferred from hydrogen to the ^{15}N labelled nitrogen and transferred back to the hydrogen for detection. http://www.protein-nmr.org.uk

3.1.2 RNA Binding Studies

One of the aims of this project was to study the RNA binding of CIRP to RNA ligands and to investigate the role of the N- and the C- terminal domains in binding. As described above, N-CIRP has two RNA binding consensus sequences, which have been found to be conserved across many RNA binding proteins. These sequences are the RNP1 and the RNP2 sequences on the β1 and β3 strands of the N-terminal domain (figure 1.6). Previous studies have reported that the 3'UTR of the thioredoxin (TRX) mRNA binds to CIRP (Yang, Weber, & Carrier, 2006) and these findings were further confirmed by Shu-Ju Hseih (2010). Human thioredoxin 3'UTR mRNA is 166 nucleotides long excluding the poly-adenine (polyA) tail and the mouse TRX 3'UTR mRNA is 532 nucleotide long, however the first 127 nucleotides TRX 3'UTR from the two species is 69% identical. Electro mobility shift assays studies by Shu-Ju Hseih showed that CIRP and N-CIRP binds to full length human TRX 3'UTR mRNA, the first 127 nucleotides of human TRX 3'UR mRNA and also the first 113 nucleotides of mouse TRX 3'UTR mRNA. Binding studies of N-CIRP by NMR also showed binding to these RNAs and also to a 10mer polyA RNA fragment which showed the largest chemical shift changes on minimal chemical shift maps. The identified RNA ligands from these studies were used here to test the RNA binding of CIRP, N-CIRP and N-CIRP mutants. If the mutations successfully inactivate the RNA binding of the

N-terminal domain, this could be exploited to investigate the function of the C-terminal of region CIRP.

For the investigation of RNA binding by NMR, 10mer polyA RNA was used as this showed the largest chemical shifts in Shu-Ju Hseih's findings. Nuclear magnetic resonance can have many applications, it can be used to detect intermolecular interactions and to study protein-ligand interaction (Coles, Heller, & Kessler, 2003; Maurer, 2005; Takeuchi & Wagner, 2006; Wishart & Nip, 1998). ^{15}N-^{1}H HSQC experiments are a good way to study structural changes of a protein under different conditions i.e. in the presence or absence of ligands by changes in the chemical shifts in the spectra of the backbone amide resonances. It is only necessary to label the protein or the ligand, changes of the labeled protein can be detected upon binding to unlabeled ligand. The changes in the chemical shifts can then be presented as chemical shift maps. Chemical shift mapping is a popular method for the investigation of protein-ligand interactions. As well as being able to confirm ligand binding, the location of binding site can also be discovered by this method by looking at the chemical shift changes of individual resonances (Clarkson & Campbell, 2003; Lee et al., 1997; Rajagopal, Waygood, Reizer, Saier, & Klevit, 1997; Yuan, Davydova, Conte, Curry, & Matthews, 2002). If the ^{15}N-^{1}H spectrum of the protein is assigned, the spectra in the presence and absence of ligand can be compared and chemical shifts changes of individual amino acid resonances calculated. Therefore those residues that experience an environmental change can be identified.

^{5}N-^{1}H HSQC spectra were collected to identify any structural change as a consequence of the mutations made to N-CIRP RNP1 and RNP2 sites and the addition of RNA to N-CIRP and N-CIRP mutants, perturbations between wild type N-CIRP and N-CIRP mutant spectra were followed. HSQC spectra were also used to confirm the folded nature of N-CIRP and mutants.

Electro mobility shift assays (EMSA) are a sensitive method of studying RNA binding of proteins and do not require high concentration of protein and ligand. The radiolabelled RNA can be detected at levels as low as nanograms in gels. *In vitro* synthesized and radiolabelled RNA is used to investigate the interaction between the mRNA and target protein. The recombinant radiolabelled RNA is usually synthesized

using *in vitro* transcription systems containing a vector including a T7, SP6 or T3 promoter and labeled with ^{32}P, ^{33}P, ^{35}S or ^{3}H (Krieg & Melton, 1984; Schenborn & Mierendorf, 1985). Binding assay can then be used to study the binding once the RNA and protein have been generated. EMSA is a popular and powerful method for detecting protein ligand interactions (Rio, 2014). If a protein-RNA complex is present, it migrates more slowly than the free RNA when ran on a polyacrylamide or agarose gel. If binding of the RNA to protein is present, shifts in the migration of the radiolabelled RNA band will be observed and an interaction can be confirmed. If binding is absent, the migration position of the RNA will remain unchanged in the presence of protein.

Using this method the binding between mouse CIRP, N-CIRP and N-CIRP mutants to the mouse and human TRX 3'UTR RNAs described above was investigated. Plasmids containing the TRX 3'UTR were generated and provided by Shu-Ju Hsieh. RNA transcription and radiolabelling was carried out by Dr Helen King at the University of Leicester. All EMSA experiments were carried out in collaboration with Prof Anne Willis' lab at the University of Leicester.

3.2 Materials and Methods

3.2.1 Preparation of NMR Samples

^{15}N labelled N-CIRP was expressed and purified as described in Chapter 2. For the analysis of N-CIRP and N-CIRP mutants by ^{15}N-^{1}H HSQC experiments, the ^{15}N labelled purified proteins were prepared by dialysis and freeze drying in the same way. However, the proteins were in the buffer; 20 mM Tris-HCl pH 7.9, 3 mM $MgCl_2$, 5 mM DTT, 5 mM NaCl, 1 mM spermidine, RNase inhibitors (RNasin Plus, Promega) (1 unit/μl of protein) instead. For ^{15}N-^{1}H HSQC experiments, samples of 100 μM protein concentration were prepared. For RNA binding experiments, 25 μM of 10mer polyA RNA (MWG) was added to the samples.

All samples for NMR analysis were prepared in 10 % (v/v) deuterium oxide (D_2O) (Goss Scientific Ltd.) in final sample volumes of 330 μl and transferred into 5 mm Shigemi NMR tubes BMS-005V for Varian and BMS-005B for Bruker spectrometers (Goss Scientific).

3.2.2 ^{15}N-^{1}H Heteronuclear Single quantum Coherence (HSQC) Experiment

^{15}N-^{1}H HSQC experiments were run with 2048 data points in the F2 dimension and 256 points in the F1 dimension at 25°C with a 1.5 hour duration with the assistance of Dr Michelle Rowe on a Bruker Avanace III, 600 MHz spectrometer using Top Spin 3.0 software with the ^{1}H centred at 4.766 ppm and ^{15}N at 115.532 ppm. Chemical shift referencing was based on the position of the water resonance with the exact value being related to the known relationship of the water resonance with temperature (Wishart & Sykes, 1994). Experiments were solvent supressed using WATERGATE (Piotto, Saudek, & Sklen, 1992). The data was processed using NMRpipe on Linux PCs and the NMR spectra were assigned using CcpNmr Analysis version 2.0.

3.2.3 Minimal Chemical Shift Mapping

Backbone resonances of the N-CIRP mutants were compared to the wild type with the assumption of minimal chemical shift. Minimal chemical shift mapping for backbone resonances was determined by assuming that the closest peaks in wild type and mutant spectra share the same assignment. The shortest distance was calculated with

an adjustment to account for the larger chemical shift range of the ^{15}N dimension compared to ^{1}H by using equation 3.1 below.

Equation 3.1:

$$\text{Minimal chemical shift value} = \sqrt{((\Delta^{1}H)^2 + (\Delta^{15}N/7)^2)}$$

$\Delta^{1}H$ and $\Delta^{15}N$ are the difference in the proton chemical shift and the nitrogen chemical shift respectively. The $\Delta^{15}N$ was divided by 7 to allow for the difference in window size between ^{15}N and ^{1}H. The number 7 being the factor difference between the two ppm axes.

For the RNA binding of N-CIRP, backbone resonances of the N-CIRP mutants with RNA were compared to the wild type with the assumption of minimal chemical shift. To study the effect of the Phe mutations on RNA binding, minimal chemical shift maps were constructed using the wild type chemical shifts and the chemical shifts of the mutants in the presence of RNA. The difference between the minimal chemical shifts with and without RNA were calculated and plotted as bar charts as shown by equation 3.2 below.

Equation 3.2:

Mutant RNA Shifts =
[Minimal map (WT vs. Mutant) – Minimal map (WT vs. Mutant + RNA)]

For the binding of wild type N-CIRP to RNA, minimal chemical shift maps were constructed by comparing the chemical shift of wild type N-CIRP with and without RNA.

3.2.4 Electro Mobility Shift Assay

^{32}P radiolabelled RNA was produced using Ribomax T7 transcription kit (Promega) by Dr. Helen King at the University of Leicester. RNase contamination was minimised using RNase ZAP and RNase free solutions and equipment. Disintegrations per unit time (dpm) of each RNA were counted per μl and 20 000 dpm of each RNA (0.004 μM) was incubated with different ratios of protein

(RNA:Protein; 1:0-1:2560). The proteins tested were full-length CIRP and N-CIRP. The production of these are described in chapter 2. Total reaction volume used was 20 µl, reactions were set up to contain 1x transcription buffer, 1 mg/ml yeast tRNA, RNAse inhibitor. The reactions were set up in 96 well plates and incubated at room temperature for 20 min. At the end of incubation time, 5x EMSA loading dye (0.1 M EDTA, 50% sucrose, 0.2% bromophenol blue) was added to each well to a final concentration of 1x. The samples were loaded onto TBE gels (6% bispolyacrylamide, 0.1x TBE, 2.5% glycerol, 0.02% TEMED, 0.005% ammonium persulfate) (1xTBE= 90 mM tris, 90 mM boric acid, 2 mM EDTA). The gels were run in 0.5x TBE buffer at 100 V until the dye front reached the bottom of the gel. Loading dye was loaded again and the run continued until the second dye front reached the end of the gel. Dye was loaded again into the left lane of the gel and allowed to run into the gel to help keep track of the orientation of the gel. The gels were dried on to Whatman paper using a heated gel dryer under vacuum. The dried gels were exposed to a phosphoimager screen overnight.

3.3 Results

3.3.1 ^{15}N-^{1}H HSQC NMR Spectra for N-CIRP

The N-CIRP NMR sample was soluble and gave a well-resolved NMR spectrum (figure 3.2). Backbone NH resonance assignments were provided by Shu-Ju Hsieh, which had been obtained by triple resonance sequential assignments carried out previously.

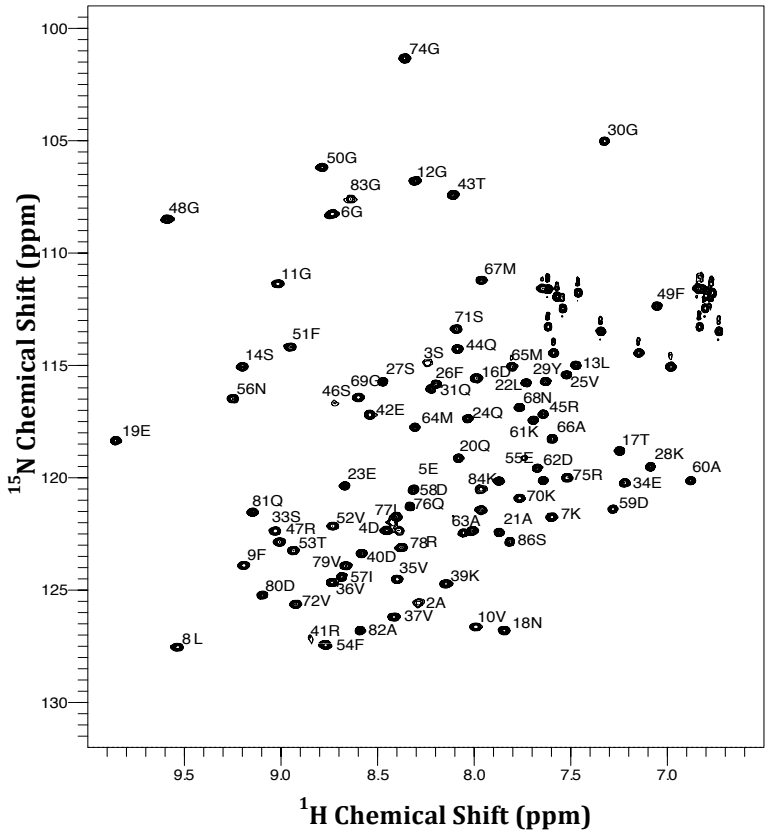

Figure 3.2 ^{15}N-^{1}H HSQC NMR spectra of ^{15}N labelled His tagged N-CIRP

3.3.2 ^{15}N-^{1}H HSQC NMR Spectra for N-CIRP Mutants

^{15}N-^{1}H HSQC NMR experiments were used to analyse the N-CIRP mutants F9A, F49A and F9AF49A described in Chapter 2. HSQC experiments are suitable to determine whether a protein is folded or unstructured and if it is soluble and stable.

Figure 3.3 shows ^{15}N-^{1}H HSQC spectra of the N-CIRP mutants and the wild type spectrum from section 3.3.1 as a reference (figure 3.2). Similarly to the wild type protein, the mutant F9A produced a well-resolved spectrum with dispersed peaks typical of a folded protein. When the peak dispersion of the F49A and F9AF49A spectra was compared with the wild type spectra, some broadening of peaks was observed in the central region, this suggest that these mutants are in conformational exchange.

As some peaks showed quite significant shifts, the backbone assignments from the wild type spectra could not be transferred across to the mutant spectra. Specifically large shifts of the peaks from residues that were mutated were observed. The F9 and F49 peaks are indicated by the red arrows in figure 3.3. When these are compared with the mutants' spectra, these peaks can no longer be seen for the corresponding mutants.

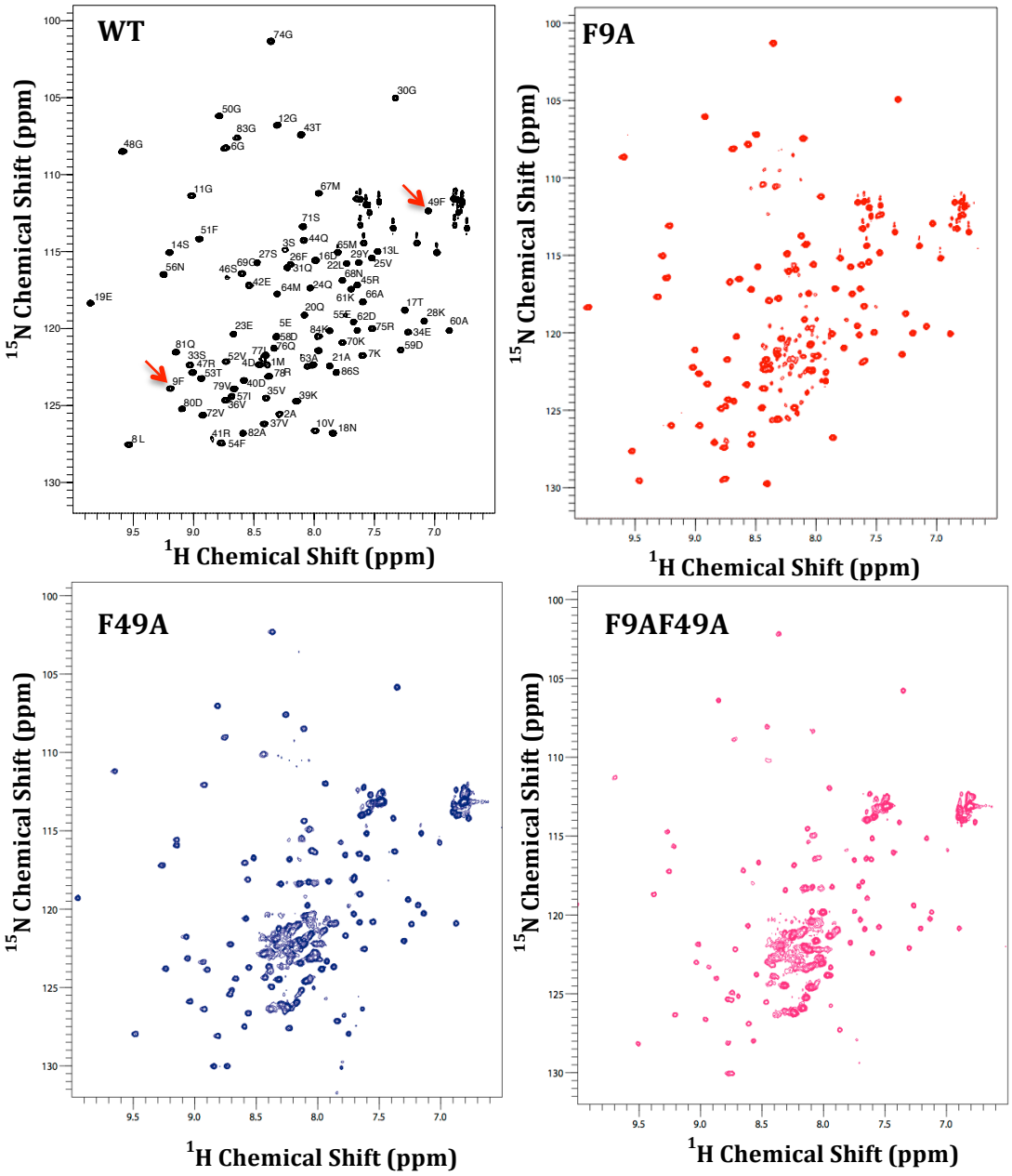

Figure 3.3 ^{15}N-^{1}H HSQC NMR spectra of ^{15}N labelled His tagged N-CIRP and mutants.

Minimal chemical shift maps were created to assess the chemical shifts changes between wild type N-CIRP and mutants. Figure 3.4 shows wild type HSQC spectra overlaid with the F9AF49A mutant spectra. The overlaid spectra shows that most peaks have shifted, this was also true for the overlays of the other two mutants with wild type N-CIRP spectra. The extent of these shifts was identified calculating minimal chemical shifts by using a 'nearest peak approach'.

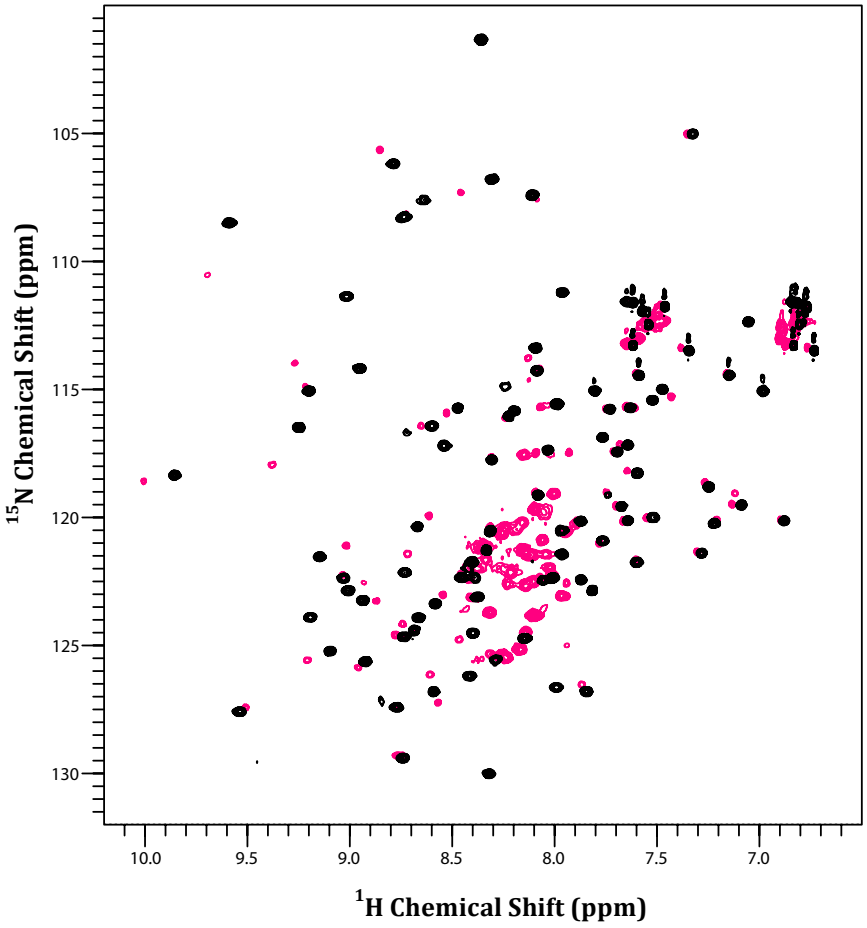

Figure 3.4 ^{15}N-^{1}H HSQC NMR spectra of ^{15}N labelled His tagged wild type N-CIRP (black) and F9AF49A mutant (pink) overlaid.

To calculate the minimal chemical shifts, chemical shift values of the two spectra were compared. As there were no assignments for the mutant peaks, the peak with the chemical shift closest to the known peak on the wild type spectra was taken to calculate minimal chemical shift. These shift values are represented as a bar chart in figure 3.5.

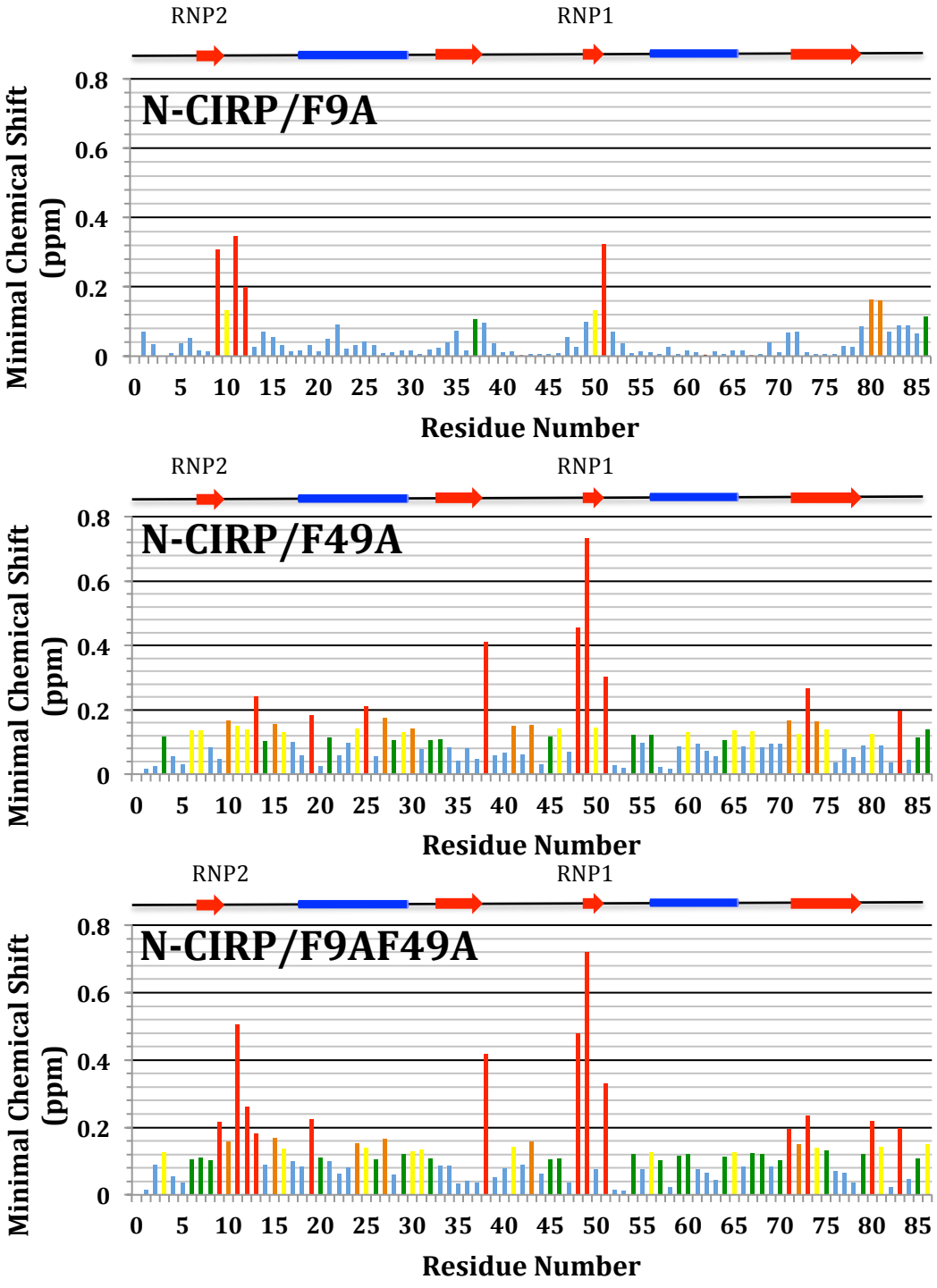

Figure 3.5 Minimal chemical shift values for the comparison of wild type N-CIRP protein to N-CIRP mutants. <0.1 ppm = Blue, 0.100-0.125 ppm =Green, 0.125-0.150 ppm = Yellow, 0.150-0.175 ppm =Orange, >0.175 ppm =Red

The minimal chemical shift maps (figure 3.5) show that shifts are smaller for the F9A mutant compared to F49A and F9AF49A. This is in agreement with the HSQC spectrum where F49A and F9AF49A mutants showed some broadened peaks suggesting an effect on the conformation of the protein in these mutants. The F9A

map showed that the largest shift is at the site of mutation (F9) and the residues on either side are also affected. This region is the RNP2 binding site. An interesting finding was that the next largest shift was the peaks belonging to residues 49-53, those being the residues belonging to the RNP1 sequence. The significance of this is discussed later. An average background shift of 0.032 ppm was calculated by taking the shift values less than 0.175 ppm for the F9A mutant. The F49A mutant showed larger shifts at the site of mutation and residues either site. Shifts across the whole protein were larger as can be observed by the colour coding with an average background shift of 0.092 ppm, 2.4 times larger that of the F9A mutant therefore the whole protein conformation is effected by this mutation. Similar to the F9A mutant shift map, F49A shift map showed large shifts around residues 9-14, which is the RNP2 binding site. The double mutant F9AF49A mutant map also showed larger shifts across the whole protein with a similar background shift value to the F49A mutant of 0.093 ppm. As this value has not grown larger due to the double mutation it was assumed that the F49A mutation is causing the conformational effect on the protein. The largest shifts were seen at the mutation sites of F9 and F49 and residues either sides of these.

3.3.3 RNA Binding of N-CIRP by Minimal Chemical Shift Mapping

The effect of the mutation on the F9 and F49 residues were determined by comparing the backbone resonances of the wild type and the mutant proteins in the presence of RNA to the backbone resonance of the wild type protein without any RNA and minimal chemical shift changes calculated. Minimal chemical shift changes of the mutants with no RNA compared to the wild type were subtracted from the minimal chemical shift changes of the mutants with RNA compared to the wild type.

RNA binding was followed by observing the 1D ^{31}P spectra of the 10mer polyA RNA in the presence and the absence of N-CIRP (figure 3.6). Upon the addition of protein to the RNA sample, the RNA peak broadened indicating the binding event.

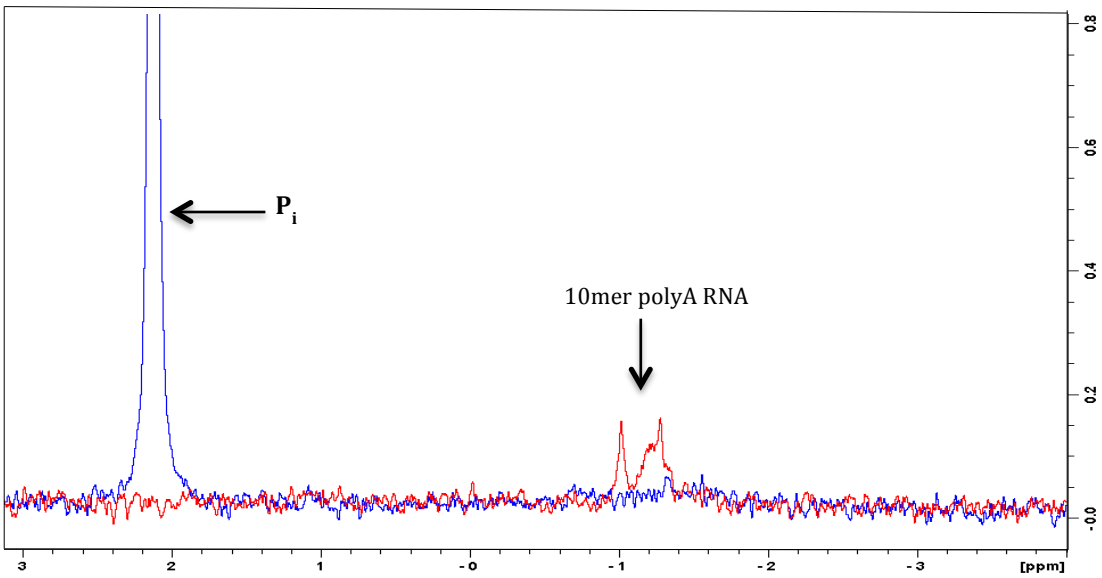

Figure 3.6 1D ^{31}P NMR spectra of the 10mer polyA RNA with (blue) and without (red) N-CIRP.

The minimal chemical shift maps for the RNA binding to wild type N-CIRP showed that there are small chemical shift changes all below 0.01 ppm across the protein (figure 3.7). Larger changes were observed around the RNP1 and RNP2 sequences suggesting that the polyA RNA is binding to these areas. Larger shifts compared to the rest of the protein were also at the N- and C-termini of the protein.

When the shift changes of the RNP1 and RNP2 sequences seen for the wild type were compared to those for the mutants, it was found that the values are were very small in

these regions at almost 0 (figure 3.7). For the F9A mutant some small general shift changes are present across the protein similar to the wild type but less suggesting that the RNAs effect on the conformation of this protein is little. However, for the F49A and F9AF49A mutants the same background was not observed i.e. the resonances of the protein with and without RNA are very similar so RNA is not having any effect on the conformation of the protein at all. The unexpected large values seen on the charts for the mutants for example the large value for the F49 peak on the F49A mutant chart is due to these peaks being absent on the mutant spectra and as a nearest peak approach is used to calculate the chemical shift, a value for the next nearest peak is produced as the peak for the original residue is no longer present due to the mutation.

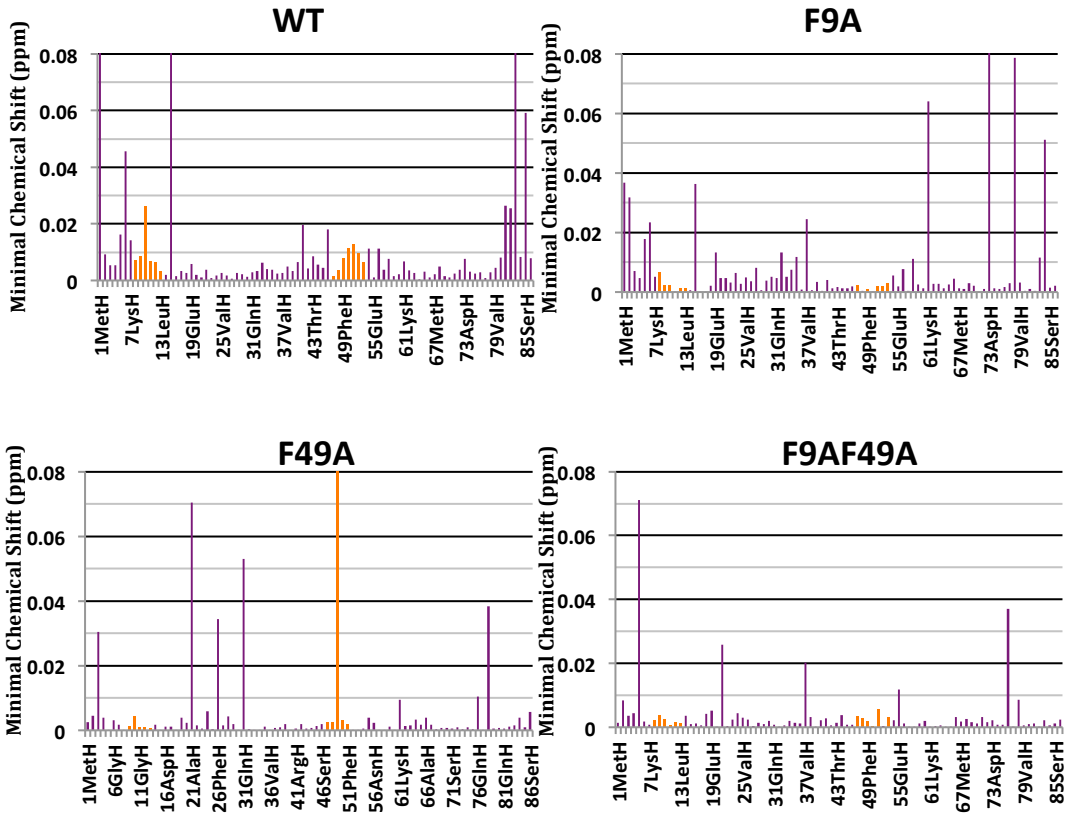

Figure 3.7 Minimal chemical shift values for interaction of with 25-µM 10mer polyA RNA. Bars for the residues belonging to RNP1 and RNP2 interaction sites are coloured in orange

3.3.4 RNA Binding of CIRP and N-CIRP by Electromobility Shift Assay (EMSA)

Binding of CIRP and N-CIRP to 3' UTR of human and mouse thioredoxin RNAs was tested by EMSA as this method does not require high concentration of RNA or protein. Different ratios of RNA:Protein tested are shown in table 3.1.

	control		CIRP				N-CIRP			
Lanes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
RNA (µM) 20000 dpm	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Protein (µM)	0.00	10.24	0.00	2.56	5.12	10.24	0.00	2.56	5.12	10.24
Ratio	0	2560	0	640	1280	2560	0	640	1280	2560

Table 3.1 Concentration and ratios of RNA and CIRP/N-CIRP protein in gels for EMSA. Red box indicates the ratios at which shifts were observed.

RNA:Protein ratios of 1:10, 1:20, 1:40, 1:80, 1:160,1:320 were also tested but binding was not observed at these ratios. Polypyrimidine-tract-binding protein (PTB) which is an RNA binding protein (Sawicka, Bushell, Spriggs, & Willis, 2008) provided by Professor Anne Willis' labaratory was used as a positive control at the highest ratio.

No binding was seen for N-CIRP at any of the ratios tested to any of the RNAs (figure 3.8,lanes 7-10). Band shifts seen at ratios 1:1280 and 1:2560 for full-length human thioredoxin 3'UTR and the first 127 nt of human trx 3'UTR and full-length CIRP (figure 3.8 (a) and (b) lanes 5 and 6). The shifts could be seen more clearly for the first 127 nt of human trx 3'UTR compared to the full length RNA. Shifts of the first 113 nt of mouse trx 3'UTR RNA bands was seen at lower ratio of 1:640 and also at 1:1280,1:2560 with full-length CIRP. Even though band shifts were seen, it was not easy to differentiate RNA only and RNA-Protein complex bands on the gels. Band shifts could not be seen for the positive control PTB RNA for the human trx 3'UTR RNA but a small shift was seen for the mouse trx 3'UTR RNA. Even though shifts were not observed for N-CIRP protein with any of the RNAs, a decrease in the intensity of the bands for mouse trx 3'UTR free RNA was seen for the highest ratios

which may suggest the presence of binding (figure 3.8 (c) lanes 9 and 10). As binding was not confirmed for N-CIRP the assays were not carried out for the mutants.

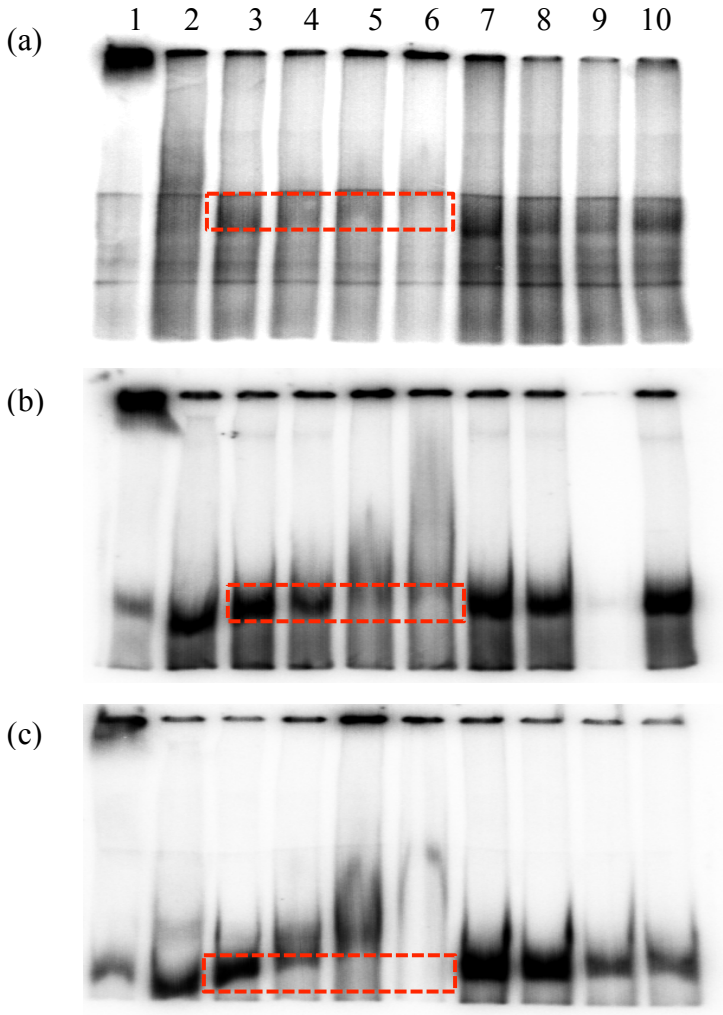

Figure 3.8 Electromobility shift analysis (EMSA) of recombinant ^{32}P radio labelled RNA binding to PTB (positive control, lane 2), full-length CIRP (lanes 3-6) and N-CIRP (lanes 7-10). (a) Full length human thioredoxin (trx) 3' UTR, (b) the first 127 nt of human trx 3' UTR, (c) the first 113 nt of mouse trx 3' UTR. Samples with shifted bands are shown by red boxes.

3.4 Discussion

This chapter describes investigation of N-CIRP binding to RNA by the use of NMR and gel shift techniques. ^{15}N labelled wild type and mutant proteins were produced as described in Chapter 2. NMR samples were prepared by freeze drying the proteins in the appropriate buffer as concentrating the protein by the use of spin concentrators was found to be problematic where the protein was adhering to the concentrator membrane.

The ^{15}N-^{1}H HSQC NMR spectra were collected to determine whether the proteins are folded, soluble and stable. Wild type N-CIRP sample remained soluble and gave a well resolved HSQC spectrum with dispersed peaks, indication of a folded and stable protein. Two-dimensional ^{15}N-^{1}H HSQC experiments were collected for recombinant N-CIRP mutant proteins F9A, F49A and F9AF49A. All proteins appeared soluble and remained soluble through out the experiment. The F9A mutant gave well-resolved spectra with dispersed peaks similar to the wild type protein. Some broadened peaks were observed in the random coil region (Jacobsen, 2007) on the F49A and F9AF49A spectra, this suggests some conformational disorder in the structure or some degradation. This could be seen more clearly when the spectra was overlaid with the wild type spectra (figure 3.4). This was not observed for the F9A mutant; therefore it was the F49A mutation that was having this effect on the N-CIRP structure. As discussed later in Chapter 4, some motion was also detected in the loop region before RNP1 containing the F49 residue in the wild-type protein as changes in the S^2 and τ_e parameters were seen. Peaks for all residues were present in the mutant spectra as in the wild type with large shifts observed. When mutant spectra were overlaid with each other it was seen that all of the peaks had shifted, there were no peaks that overlaid over one another.

The F9A chemical shift perturbation map showed that there were generally small shifts (average 0.032 ppm for background shifts) across all residues of F9A N-CIRP. Large shifts (>0.175 ppm) were seen for the F9 residue and the residues either side. Interestingly large shifts were seen in the RNP1 region of β3, suggesting that the RNA binding sequences RNP1 and RNP2 are close in space. The F49A mutants shift maps showed a larger general shift (0.092 ppm average background shift) for all residues compared to F9A. The greatest shifts were at F49 and either side of it and

similar to the F9A shift map, increased shifts were seen in the RNP2 region too. The general larger shift values across the whole protein for the F49A mutant confirms that this mutation has a larger structural effect on N-CIRP also agreeing with HSQC spectrum and the dynamics data (Chapter 4). This is also shown in figure 3.9, the surface representation of the chemical shift changes shown for each mutant shows that the F49A and F9AF49A are experiencing much larger structural changes compared to the F9A mutant. This could also be confirmed when looking at the shift map of the F9AF49A mutant, which also showed larger shift values across the protein compared to the F9A mutant. The largest shift were seen at the mutation sites of F9 and F49 as expected.

The mutation of the F9 residue of the RNP2 RNA binding sequence to alanine does not have an effect on the folding, solubility and stability of the protein. However the mutation of F49 into alanine in the RNP1 sequence does have a conformation effect on N-CIRP giving data which suggests that both folded and unfolded protein is present and the loop before this region appears more dynamic at the pico to millisecond time scale compared to the rest of the protein confirming that it is of functional and structural importance as discussed in chapter 4 (section 4.3.5). To confirm this hypothesis, dynamic studies would need to be carried out on the specific mutants to look into the dynamic properties of these regions.

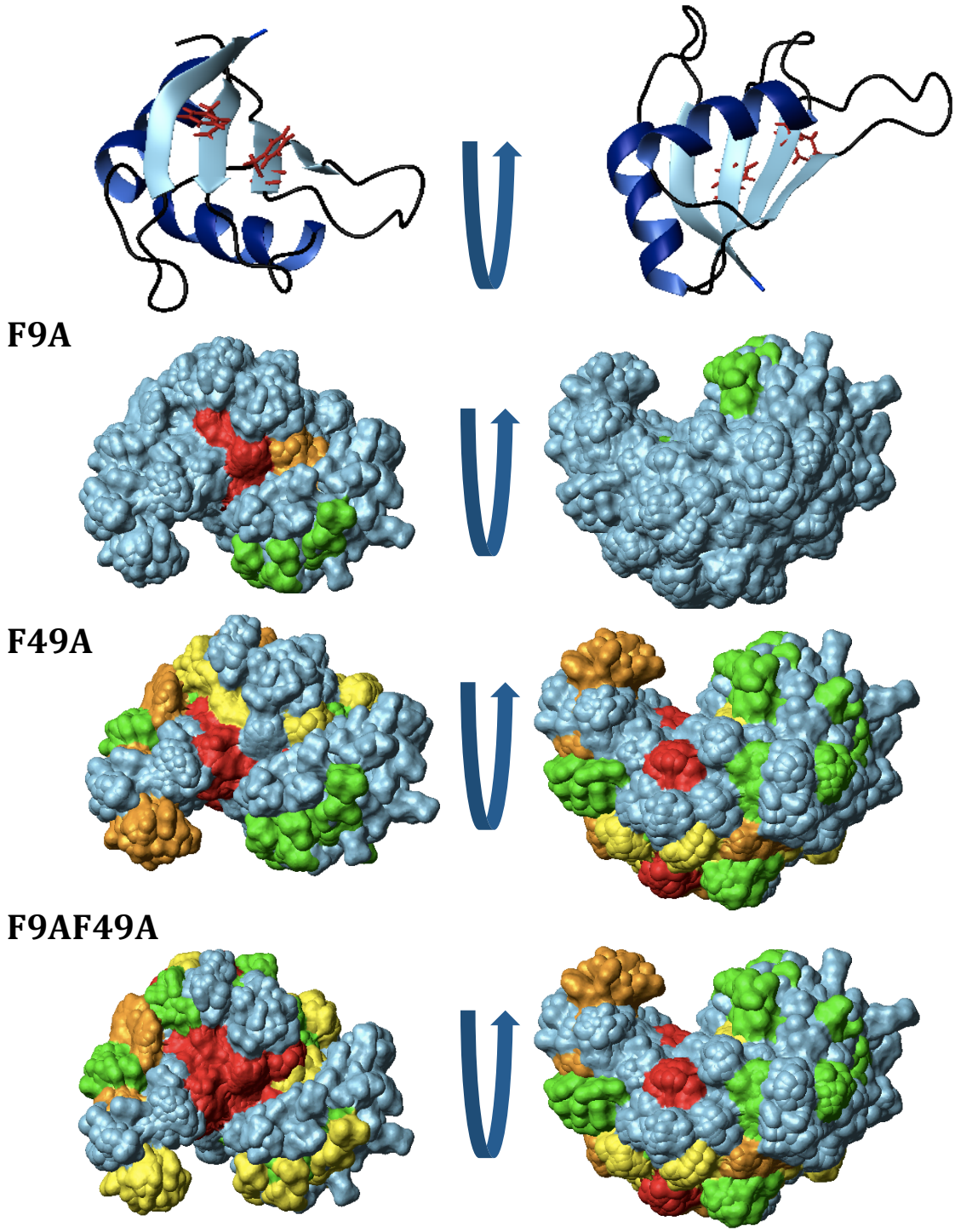

Figure 3.9 N-CIRP is shown as a surface representation with residues coloured according to the chemical shift difference on mutation. The top images show ribbon structure with the F9 and F49 residues in line style. N-CIRP structure shown is the PDB structure 1x5S. <0.1 ppm = Blue, 0.100-0.125 ppm = Green, 0.125-0.150 ppm = Yellow, 0.150-0.175 ppm = Orange, >0.175 ppm = Red

NMR RNA binding experiments of N-CIRP confirmed that the RNP1 and RNP2 regions identified in the N-CIRP sequence are the binding sites for the 10mer polyA RNA. Changes in chemical shift values were largest in these two regions when the resonances of the wild-type spectra were compared to the wild-type spectra with RNA. Initially the binding of the 10mer polyA RNA was confirmed by observing the 1D ^{31}P spectra of the RNA as a proof of concept before pursuing the detailed study of RNA binding using NMR.

Larger shifts were also seen at both the termini, this was likely due to structural rigidity being forced on the termini upon binding as the termini are less ordered compared the rest of the protein when in the free state. The changes in chemical shift were smaller than expected. This may be due to weak binding or rapid on and off exchange between bound and unbound forms. The minimal chemical shift maps also confirmed that the F9 and F49 are important residues for RNA binding. When the chemical shifts of the mutants with and without RNA were compared, chemical shifts changes in the RNP1 and RNP2 regions were small. The F9A mutants showed that there was no change in the chemical shift of the RNP2 resonance in the presence of RNA therefore the addition of RNA does not have any conformational effect on this region of N-CIRP i.e. no binding. Binding was depleted for the RNP1 sequence too, this meaning that both the F9 and F49 residues have to be present for binding to take place. Some small changes in shift were still observed particularly for residues 20-40 of the protein and the N-termini suggesting some conformational effect. The shift map for the F49A showed that no changes were present between the mutant protein with or without the presence of RNA so the RNA did not have any effect on the protein conformation and binding to both RNP1 and RNP2 region was absent. Unlike for the F9A mutants there were no other shift changes across the protein so the addition of RNA to this mutant has no effect on the conformation of the protein. This was also the same for the double mutant F9AF49A with no changes seen between the RNA and no RNA spectra.

The data confirms that F49 is more key for binding compared to F9 as the mutation of the F49 lead to larger depletion in RNA binding. This fits with the findings from the dynamics data for wild-type N-CIRP (Chapter 4,) and ^{15}N-^{1}H HSQC mutants spectra and chemical shift map comparison of the mutants to the wild-type protein where it

was concluded that F49 mutation has a larger structural effect on the protein (figure 3.3) and the region around this residue posseses more internal motion compared to the rest of the protein (figure 4.24).

Electromobility shift assays revealed that full-length CIRP bound to all three RNA ligands tested (3'UTR of full-length and first 127 nt of human trx and 3'UTR of mouse TRX) at very high ratios of protein (1:640-1:2560). A clear band for the RNA-Protein complex could not be seen but a clear shift of the free RNA band could be seen on the gel in particular for first 127 nt of human TRX 3'UTR and first 113 nt of mouse TRX 3'UTR (figure 3.8). The band shifts for the mouse TRX 3'UTR RNA were the strongest out of the 3 with no free RNA band observed at the highest ratio. Also, even though a shift was not seen there was reduction in the density of the free RNA band for N-CIRP at the two highest ratios with the mouse RNA. It is apparent that the binding is preferential toward the mouse RNA. This is expected as the CIRP protein is also of the mouse species. K_d could not calculated for any of the binding events, as the free RNA and RNA-Protein complex bands could not be clearly differentiated from one another.

This data suggests that the CIRP binding to RNA is stronger when the protein is in its full-length form rather than the N-terminal domain on it own. This agrees with the NMR binding data, as chemical shift changes seen were small with the N-CIRP protein binding to the polyA RNA. The C-terminal region must therefore have a function towards the binding of the protein to the RNA. Study by Yang et al (2006) has also shown that full length CIRP has a greater interaction with RNA ligands than the N-terminal domain alone. The concentration of protein had to be taken to 5 μM to see any binding. Even though a K_d was not calculated, the results from the EMSA and NMR experiment suggest that the RNAs tested do not bind to CIRP and N-CIRP very strongly therefore the identification of a specific ligand is required. The weak binding could also be due to rapid on and off exchange as mentioned above although rapid 'off rate' is synonymous with high K_d (low affinity). RNA binding proteins can bind to diverse RNA sequences with different levels of affinities (Dreyfuss, Matunis, Piñol-Roma, & Burd, 1993) and they have preferred substrates although this does not mean exclusivity, but the determinants towards these are not clear (Maris et al., 2005).

Chapter 4

Structural and Dynamics Analysis of N-CIRP

4.1 Introduction

This chapter describes the process of the structural and dynamics analysis of N-CIRP. Earlier work has provided information on N-CIRP binding to RNA ligands and also some insight into the C-terminal domain (chapter 3). As the C-terminal is natively disordered and does not produce a resolved NMR spectra, it could not be included in structure determination experiments (discussed further in section 1.4.4). Further functional studies of CIRP are described in Chapter 5. To further extend our knowledge of CIRP function and how this may be contributing towards improving recombinant protein production, a structure-function based knowledge is necessary. Even though there is a structure deposited in the protein data bank of the RNA recognition domain of CIRP (1X5S) (figure 1.2), this is of the human protein and it has not been subject to peer review. This project is based on the mouse protein. The two sequences are 95% identical with 8 residues being different. Only one of these residues is in the N-terminal domain (figure 4.1) and is a change seen frequently in protein evolution (Ala to Ser) with a score of 1 in the PAM250 matrix.

```
CIRBP_MOUSE MASDEGKLFVGGLSFDTNEQALEQVFSKYGQISEVVVVKDRETQRSRGFGFVTFENIDDA 60
CIRBP_HUMAN MASDEGKLFVGGLSFDTNEQSLEQVFSKYGQISEVVVVKDRETQRSRGFGFVTFENIDDA 60
            ********************:************************************

CIRBP_MOUSE KDAMMAMNGKSVDGRQIRVDQAGKSSDNRSRGYRGGSAGGRGFFRGGRSRGRGFSRGGGD 120
CIRBP_HUMAN KDAMMAMNGKSVDGRQIRVDQAGKSSDNRSRGYRGGSAGGRGFFRGGRGRGRGFSRGGGD 120
            ************************************************.***********

CIRBP_MOUSE RGYGGGRFESRSGGYGGSRDYYASRSQGGSYGYRSSGGSYRDSYDSYATHNE 172
CIRBP_HUMAN RGYGGNRFESRSGGYGGSRDYYSSRSQSGGYSDRSSGGSYRDSYDSYATHNE 172
            ***** ****************:****.*.*. ******************
```

Figure 4.1 Amino acid sequence alignment of the human and mouse CIRP proteins. Perfectly matched residues are indicated by an asterisk (*), conserved residues by a colon (:), and less conserved residues by a period (.). The yellow region is the N-terminal RNA binding domain and the orange region is the glycine C-terminal region.

NMR can be used to determine structure of proteins up to 50 kDa and beyond (Cavanagh, Fairbrother, Palmer, Skelton, & Rance, 2007). There are two main steps to structure determination of proteins; resonance assignment and determination of distance constraints for structure calculation. Before the structure of a protein can be determined, there are multiple NMR experiments that need to be carried out. These experiments allow the assignment of resonances, which belong to the backbone and side chain elements of each residue in the protein and it is important to assign as many

resonances as possible. Dynamic properties of N-CIRP were studied through nuclear spin relaxation experiments. Biological molecules are flexible and undergo conformational changes on different timescales (Mittermaier & Kay, 2009). Proteins do not simply occupy one structure but they are equilibrating as a set of time dependant structures and these dynamic properties of proteins are essential for describing their function (Kleckner & Foster, 2011). Dynamic characteristics of proteins affect a wide a range of functions such as signalling and regulation (Smock & Gierasch, 2009), rate and pathway of protein folding (Neudecker, Lundstrom, & Kay, 2009) and ligand binding (Boehr, Nussinov, & Wright, 2009). NMR experiments are run using isotopically enriched protein. Nuclear magnetic resonance and NMR active isotopes are described in Chapter 3 (section 3.1).

4.1.1 Resonance Assignment

NMR experiments which correlate ^{1}H, ^{13}C and ^{15}N in an amino acid can be used for the assignment of resonances (Cavanagh et al., 2007; Griesinger & Sattler, 1999; Kay, 2005). There are many different experiments that study the connectivity between different nuclei, which allow the identification of NMR frequencies of NMR active nuclei. The first step is to assign the protein backbone. Backbone resonances are ^{15}N, ^{1}H and $^{13}C\alpha$. A common strategy used for backbone assignment is the through triple resonance spectra by sequential assignment (Griesinger & Sattler, 1999; Whitehead, Craven, & Waltho, 1997). Triple resonance experiments produce 3D data, which correlate ^{15}N, ^{1}H, $^{13}C\alpha$ and $^{13}C\beta$ of a residue (Cavanagh et al., 2007), which allow for rapid assignment of proteins as large as 700 amino acids by the use of NMR experiments which show specific resonances (Kay, 2005). Sequential assignment matches distinctive spin systems to the amino acids to which it belongs (Wüthrich, 1986). The triple resonance spectra collected are CBCANH and CBCA(CO)NH (Grzesiekt & Bax, 1992a, 1992b). The CBCA(CO)NH correlates $^{1}H^{N}$ and ^{15}N resonance of residue (*i*) with $^{13}C\alpha$ and $^{13}C\beta$ resonance of the previous residue (*i-1*). The CBCANH spectrum makes the correlation $^{1}H^{N}$ and ^{15}N resonance with $^{13}C\alpha$ and $^{13}C\beta$ resonance within the same residue. Reference chemical shifts can be used as guidelines to identify particular residue types (Wishart, Bigam, Holm, Hodges, & Sykes, 1995) to help make assignments. The average chemical shifts for $^{13}C\alpha$ and $^{13}C\beta$ are shown in figure 4.2.

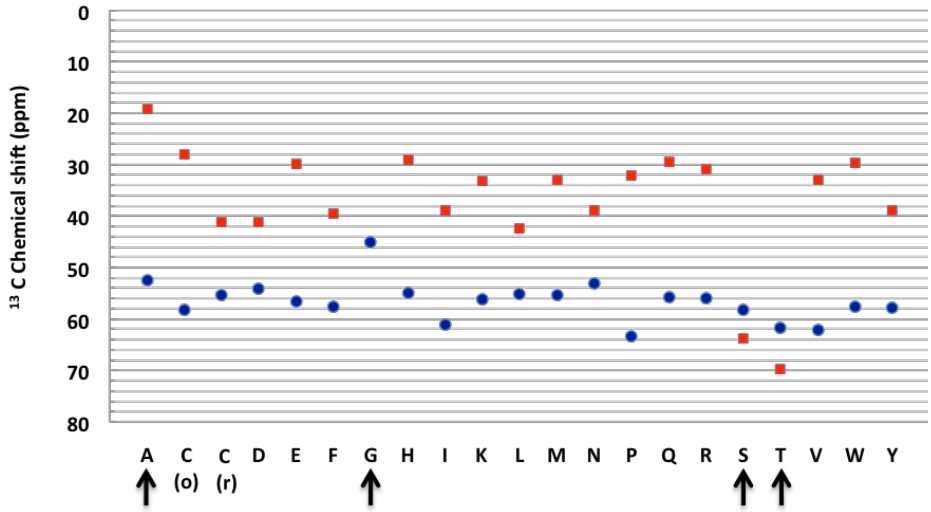

Figure 4.2 Average random coil backbone amino acid resonances of $^{13}C\alpha$ and $^{13}C\beta$ (Wishart et al., 1995) Blue dots indicate Cα and red squares indicate Cβ. Residues with distinctive shifts are indicated by the arrows.

The amino acids serine, threonine, alanine and glycine in particular have distinct chemical shifts that are easily distinguished from the rest as shown by the arrows in figure 4.2. Serine and threonine have especially high $^{13}C\beta$ shifts, whereas alanine's $^{13}C\beta$ shift is especially low. Glycine is easily distinguishable as it lacks $^{13}C\beta$. Therefore, once these have been identified in the protein spectra, they can be used as starting points and one can sequentially move through the spectra and assign the backbone resonances on either side of A, G, S and T amino acids. As proline lacks an NH, it cannot be correlated with its previous residue (*i-1*). The assignment of the spectra was made using CcpNmr analysis software and how this process is carried out is described in further detail in the results section 4.3.1.

Backbone assignments can then be used to aid assignment of ^{1}H, ^{13}C side chain resonances. Side chain assignments are needed for structure determination. Assignment of ^{1}H-^{13}C side chain resonances is usually made using HCCH-TOCSY spectra (Bax & Clore, 1990) which provides three dimensional data correlating ^{1}H and ^{13}C nuclei through ^{13}C resonances (Cavanagh et al., 2007). The spectra has ^{1}H axis and ^{13}C axis; the ^{13}C axis is used to make further ^{13}C side chain assignments that could not be made with the CBCANH and CBCA(CO)NH spectra. The ^{1}H axis provides side chain information by creating identifiable patterns for ^{1}H side chain resonances. Figure 4.3 shows residue specific patters for both ^{13}C and ^{1}H resonance of each amino acid which can be used as guidelines to identify particular residue types.

This experiment gives a spectrum where there is a strip for each carbon frequency in which (in principle) all side chain hydrogen resonances are visible with a diagonal which crosses at the proton resonance to which carbon it is bound to (figure 4.8). Cα, Cβ, Hα and Hβ resonances (already assigned from previous experiments) form the basic starting points for the side chain assignment. Hα and Hβ resonances can be identified from cross peaks at that residue's Cα and Cβ resonance position and if the residues contains further ^{1}H and ^{13}C nuclei, these will be correlated to the α and β resonances and can be identified using reference chemical shift patterns shown in figure 4.3. The assignments are essential for the full interpretation of NOESY data to determine sufficient distance constraints for three dimensional structure calculations.

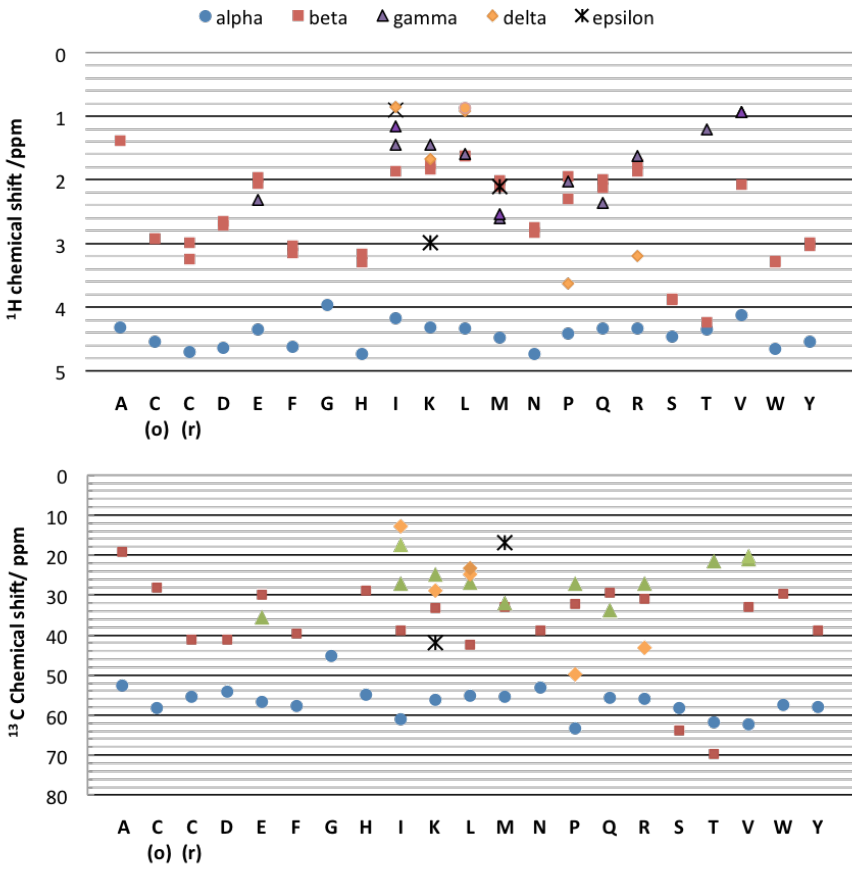

Figure 4.3 Average random coil amino acid resonances of ^{13}C and ^{1}H (Wishart et al., 1995)**.**

4.1.2 Secondary Structure Prediction

There are several secondary structure prediction programs which have been developed to date such as TALOS and TALOS+ (Cornilescu, Delaglio, & Bax, 1999; Shen, Delaglio, Cornilescu, & Bax, 2009), DANGLE (Cheung, Maguire, Stevens, & Broadhurst, 2010), PREDITOR (Berjanskii, Neal, & Wishart, 2006) and SimShiftDB (Ginzinger & Coles, 2009).

Secondary structure is determined by the dihedral angles of the protein backbone. Phi (Φ) is the angle that rotates around the N-C_α bond and psi (Ψ) is the angle that rotates around the C_α-C' bond (figure 4.4). A Ramachandran plot shows the possible combinations of Φ and Ψ angles and the regions of the plot that give rise to α-helices or β-strands. Chemical shifts of ^{1}H and ^{13}C can inform on a helical or beta strand conformation (Wishart, Sykes, & Richards, 1992) (figure 4.4). Known chemical shifts of helices and strand were used to develop the chemical shift index and secondary structure can be determined using ^{1}H and ^{13}C chemical shifts.

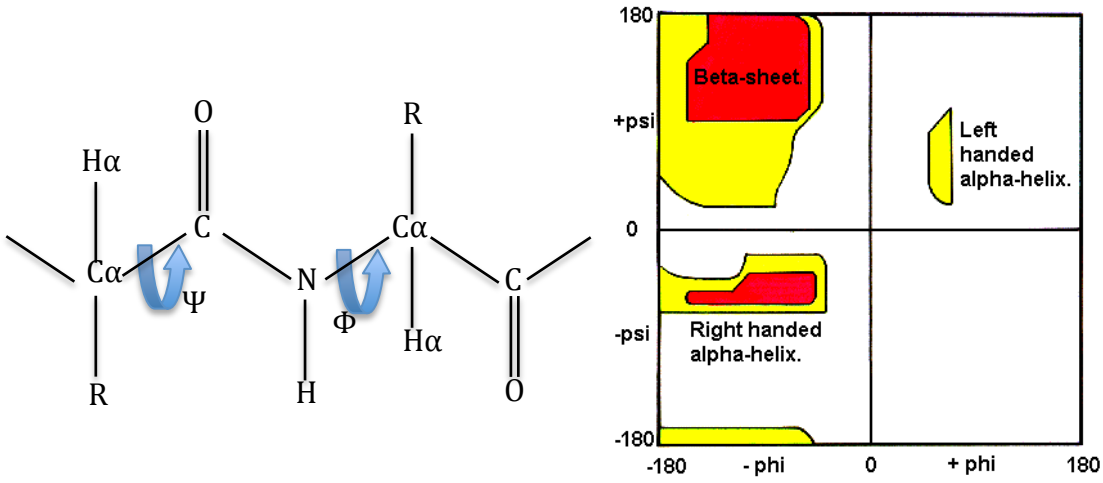

Figure 4.4 An amino acid unit indicating the angles phi and psi (Φ and Ψ) (left) and Ramachnadran Plot (right). In the Ramachadran plot, white areas represent sterically disallowed conformations, the red regions are where there are no steric clashes i.e. regions for α-helices and β-strand conformations. Yellow areas represent conformations allowed if shorted van der Waals radii are used for the calculation i.e atoms come closer together.

DANGLE (Dihedral Angles from Global Likelihood Estimate) is the most recent algorithm to estimate Φ and Ψ backbone dihedral angles for each residue (and also takes into account neighbouring amino acids) to predict a secondary structure of a protein using chemical shift measurements and a database of known structures and

shifts (Cheung, Maguire, Stevens, & Broadhurst, 2010). The package is provided by CcpNmr Analysis program.

A global likelihood estimate (GLE) diagram is assembled from the B (Φ and Ψ) values of each bin in Ramachandran space. The island that contains the largest score in the diagram, the Bmax value, is termed the principal island. DANGLE determines the angular means of Φ and Ψ within the principal island and these are the prediction of the backbone conformation (Cheung et al., 2010). In this chapter ^{1}H, ^{15}N and ^{13}C chemical shift values were used to predict the secondary structure of N-CIRP with DANGLE.

4.1.3 Structure Calculation

The process which ends with the structure determination of a protein can be broken down into four parts; (1) collection of experimental data, (2) processing the data, (3) determination of structural constraints and (4) calculations of structures which satisfy the constraints (Sutcliffe, 1993). Once as many assignments as possible are known, these can be used along with additional data, e.g. through space distance restraints and dihedral angles, for structure calculation. An iterative process is used to refine the calculated structures as the initial calculations invariably produce a list of restraint violations which do not satisfy the model. The iterative process involves the analysis of the violations before re-calculating the structure and eventually reducing or eliminating the list of violations.

4.1.3.1 Distance Restraints

Experimental data which constrains the geometry of the structure is necessary to refine the protein structure. The nuclear Overhauser effect experiments provide restraints for protein structures through NOE cross peaks. An NOE cross peak is due to a transfer of magnetisation between two ^{1}H nuclei coupled by dipole-dipole interaction (Guntert, 1997). These experiments allow through space correlations between pairs of ^{1}H nuclei, which may be far away in protein sequence but close in three-dimensional space. Therefore these ^{1}H-^{1}H correlations provide distance measurements, which can be used as constraints in a protein structure calculation. For most well behaved proteins, NOE cross peaks define distances up to 6Å between ^{1}H nuclei (Barsukov & Lian, 1993). NOE experimental data is collected through NOESY

experiments, where the intensity of the NOE cross peak is related to the distance between the two nuclei. Intensity decreases as distance gets larger and cannot be observed more than approximately 6Å as mentioned earlier. There are other experimental data, which can be used for structural information; these include dihedral angles, hydrogen-bond restraints and residual dipolar couplings (RDCs) (Barsukov & Lian, 1993; Bax, Kontaxis, & Tjandra, 2001; Cavanagh et al., 2007; Guntert, 1997; Wüthrich, 1986).

4.1.3.2 Automated Structure Determination

Determination of protein structure by NMR is a complex and time-consuming process. There are a number of methods for the calculation of protein structures using resonance assignment data and distance restraints which produce an ensemble of closely related structures. Various programs are available which have been developed over the years to automate part of or the entire process. Some examples are FLYA (López-Méndez & Güntert, 2006), Auremol (Gronwald et al., 2004), ARIA (Nilges, 1995), CYANA (Güntert, 2004) and CLOUDS (Grishaev & Llinás, 2005). The most time consuming aspect of the process is the correct and unambiguous assignment of NOE cross peaks which is also prone to errors, this step is considered the bottle neck in the process of structure calculation (Wishart, 2005). Overlapping chemical shifts increase ambiguity and this gets higher as the molecular weight of the protein increases.

The ARIA program can use ambiguous NOEs (Linge, O'Donoghue, & Nilges, 2001), it uses the concept of ambiguous distance restraint where an NOE cross-peak with several assignment possibilities can be converted into a single distance restraint. This single restraint is the sum of the restraints from all of the assignment possibilities which is correct as long as one of the assignments is accurate. The data can then be fed into CNS (crystallography and NMR system) for automated structure calculations using these ambiguous NOEs. There are other automated methods available but these rely on complete and accurate chemical shift list for every resonance to make unambiguous NOE assignments (Güntert, 2003), for example DYANA and CYANA (Oezguen, Adamian, Xu, Rajarathnam, & Braun, 2002). After the calculations, structures are validated using computational methods such as RMSD (root mean square deviation) calculations of the ensemble backbone, analysis of violation lists,

number of restraints, comparison to other similar structures (Spronk, Nabuurs, Krieger, Vriend, & Vuister, 2004).

4.1.4 NMR Relaxation Dynamics

Relaxation is the process where over time magnetisation returns to it equilibrium position (Keeler, 2005). Measuring relaxation parameters can give information about both internal and global motions. These motions occur at different timescales and are measured by different NMR experiments (figure 4.5). For example loops and termini regions of proteins tend to have fast motion measured on the nanosecond to picosecond timescale.

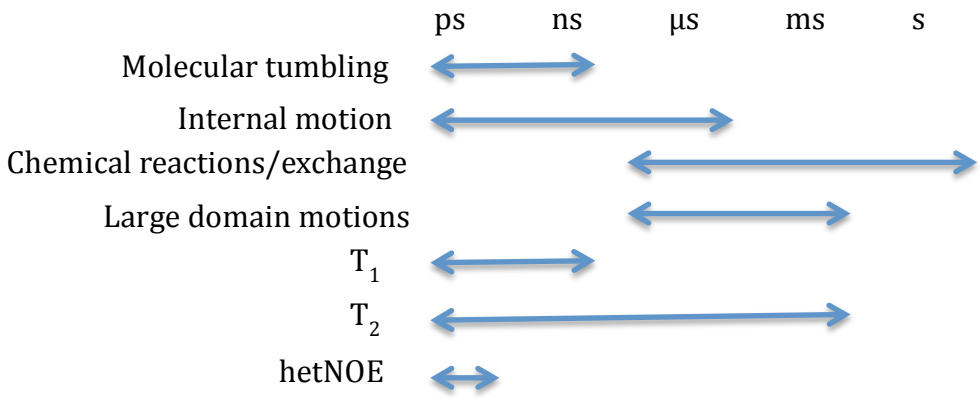

Figure 4.5 Different timescales that can be viewed by different NMR experiments. Based on figure from (Henzler-Wildman & Kern, 2007)

The nuclear spin relaxation (NSR) time of decay yields T_1, T_2 relaxation times and heteronuclear Nuclear Overhauser Effect (hnNOE) relaxation rates (Igumenova, Frederick, & Wand, 2006; Jarymowycz & Stone, 2006; Kempf & Loria, 2003). T_1 and T_2 provide fast movement information on a nano to picosecond timescale. T_1 is the longitudinal relaxation time and conveys information about the overall size and shape of the protein, for example, the more the T_1 varies across the protein, the more asymmetric the protein is (Palmer, 2001). T_2 is the transverse relaxation time and also reflects size and shape of protein but this value also fluctuates due to the process of chemical exchange as T_2 relaxation can extend up to millisecond timescales (Palmer, 1997). The fundamental time constants produced by NMR relaxation (T_1 and T_2) become relaxation rates (R_1 and R_2) in reciprocal form (ie $1/T_1$ and $1/T_2$). Rate of

relaxation is dependent on the physical environment and motions the molecule is experiencing.

hnNOE is obtained by comparing signal intensities in the presence of and absence of heteronuclear dipolar relaxation which is a through space interaction between pairs of nuclear spins that affects the magnetic field experienced by each spin (Kleckner & Foster, 2011). hnNOE effects are observed for the amide ^{1}H and ^{15}N as a difference in intensity when saturation of the proton is switched on and switched off. The observed hnNOEs are typically negative and become more negative as motion becomes faster on a picosecond timescale (Brüschweiler, 2003; Palmer, 2001). hnNOE is determined by the equation: hnNOE= $(I-I_0)/I_0$, where I is the intensity of the proton saturated resonance and I_0 is the intensity of the non-saturated resonance (Noggle & Schirmer, 1971). If the protein is rigid and tumbles isotropically, the T_1, T_2 and hnNOE values will be similar at each site (Kleckner & Foster, 2011). Therefore these parameters can give a complete picture of the dynamic properties of a protein.

Spectral density can be used to describe T_1 and T_2. Spectral density describes motion that occurs at specific frequencies and is field strength dependent on the NMR spectrometer used (Peng & Wagner, 1992a, 1992b) . Following equations define spectral density:

Equation 4.1:

$$\frac{1}{T_1} = \frac{d^2}{4}[J(\omega_h - \omega_N) + 3J(\omega_N) + 6J(\omega_h + \omega_N)] + \frac{c^2 J(\omega_N)}{3}$$

Equation 4.2:

$$\frac{1}{T_2} = \frac{d^2}{8}[4J(0) + J(\omega_h - \omega_N) + 3J(\omega_N) + 6J(\omega_h) + 6J(\omega_h + \omega_N)] + \frac{c^2(3J(\omega_N) + 4J(0))}{18}$$

Equation 4.3:

$$d^2 = \left(\frac{\gamma_H \gamma_N \hbar \frac{\mu_0}{4\pi}}{r_{NH}^3}\right)$$

Equation 4.4:

$$c^2 = (\omega_N \Delta_N)^2$$

Where J(ω) is shown in equation 4.5. γ_N(-2.71×10^{-7} $rads^{-1}T^{-1}$) and γ_H (2.68×10^{-8} $rads^{-1}T^{-1}$) are the gyromagnetic rations of ^{15}N and ^{1}H, μ_0 (1.26×10^{-6} NA^2) is the permeability of free space. r_{NH} (1.02×10^{-10} m) is the length of the amide bond, ħ (6.63×10^{-34} Js) is Planck's constant divided by 2π, ω_N (600 MHz = -3.82×10^{8} $rads^{-1}$) and ω_H (600 MHz = 3.77×10^{9} $rads^{-1}$) are the Larmor frequencies of ^{15}N and ^{1}H. Δ_N (-0.00016 s^{-1}) is the difference between the perpendicular components of the chemical shift tensor (Broadhurst, Hardman, Thomas, & Laue, 1995).

ModelFree analysis (Lipari & Szabo, 1982) is the most commonly used approach for the interpretation of T_1, T_2 and hnNOE (Fischer, Majumdar, & Zuiderweg, 1998; Palmer, 1997). This approach assumes that internal motion is independent and faster than molecular rotation (Chen & Tjandra, 2008) and no structural model is required. Spectral density function equation (Abragam, 1961; Mcconell, 1987; Wagner, 1993) cannot be used to describe individual nuclei in the protein (equation 4.5). To break down spectral density functions into different motional properties the ModelFree formalism was proposed (Lipari & Szabo, 1982) by adding additional parameters to the equation (equation 4.6).

Equation 4.5:

$$J(\omega) = \frac{2}{5} \frac{\tau_m}{1 + (\omega\tau_m)^2}$$

Equation 4.6:

$$J(\omega) = \frac{2}{5} \left(\frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S^2)\tau'}{1 + (\omega\tau')^2} \right)$$

$1/T_1$, $1/T_2$ and hnNOE are proportional to J(ω). J(ω) is spectral density, S^2 is the order parameter and τ_m is the global molecule correlation time. $\tau'^{-1} = \tau_m^{-1} + \tau_e^{-1}$ where τ_e is internal motion correlation time which indicates the timescale of bond vector reorientation i.e. internal motion (Andrec, Montelione, & Levy, 2000). The left half

inside the brackets represents global motion and right half represents internal motion. If the protein is rigid, has no internal motion, the right spectral density expression equals 0, causing global motion to dominate. As protein becomes less rigid and S^2 reduces from 1.0 to 0.0, influence of internal motion τ_e, i.e. the influence of the right hand side, increases. S^2 is a measure of the magnitude of the angular function of a chemical bond vector and so it reflects the flexibility of the polypeptide chain at these sites (Ishima & Torchia, 2000). S^2 can take a value between 0 and 1, where 1 defines a rigid system and it decreases as flexibility increases, for example termini typically have reduced S^2 values compared to structured regions.

Global correlation time, τ_m is the time taken for molecule to rotate through one radian (Abragam, 1961). τ_e is the effective correlation time and gives an indication of fast and slow internal motions. τ_m should be estimated and it can be obtained by determining T_1/T_2. τ_m is dependent on the size of protein and can be estimated using equation 4.7 where N is the number of residues present i.e. the larger the protein the larger the τ_m. T is the temperature in Kelvin (Daragan, Ilyina, Fields, Fields, & Mayo, 1997).

Equation 4.7:

$$\tau_m = \frac{9.18 \text{ x } 10^{-3}}{T} \exp\left(\frac{2416}{T}\right) N^{0.93}$$

The interpretation of T_1, T_2 and hnNOE can be approached in several ways. Plots of T_1, T_2 and hnNOE vs residue number provide information on the dynamic properties of proteins. T_1 vs T_2 plots inform on tumbling characteristics of the molecule including isotropy/anisotropy or indication of chemical or conformational exchange by adding ModelFree parameters as contour lines.

4.2 Materials and Methods

4.2.1 Preparation of NMR Samples

Double labelled ^{13}C and ^{15}N N-CIRP for resonance assignment and ^{15}N N-CIRP for dynamics experiments was expressed and purified as described in Chapter 2. To concentrate the protein for NMR analysis, it was dialysed into 2 litres of sodium phosphate buffer, pH 6.5 containing sodium chloride and freeze-dried. The dried samples were then taken up in a 330 μl volume to make a 150 μM ^{13}C and ^{15}N double labelled and 0.6 mM ^{15}N labelled sample. The double labelled N-CIRP was resuspended in 100% deuterium oxide (D_2O) (Goss Scientific Ltd) and the ^{15}N labelled N-CIRP was prepared in 10% (v/v) D_2O. The concentration of the phosphate and salt dialysis buffer was prepared so that when the freeze-dried material was resuspended, the final concentration would be 20 mM sodium phosphate and 50 mM sodium chloride. The sample was transferred into a 5 mm Shigemi NMR tube BMS-005B (Goss Scientific) for Bruker spectrometers.

4.2.2 Data Acquisition and Processing

All spectra were collected at 25°C with the aid of Dr. Michelle Rowe. The data was processed using NMRpipe and NMRDraw on Linux PCs and the NMR spectra were assigned using CcpNmr Analysis version 2.0. Side chain resonance assignments were obtained from a 0.1 mM $^{13}C/^{15}N$ labelled N-CIRP sample using HCCH-TOCSY experiments (Kay, Xu, Singer, Muhandiram, & Forman-Kay, 1993). Experiments were performed with the sample exchanged in D_2O. Each experiment was acquired with 2048 points (9014 Hz) in the direct F3 dimension (^1H) and 64 points (3244 Hz) in F2 (^{13}C) and 256 points (9014 Hz) in F3 (^1H) indirect dimensions. The ^{13}C dimension was centred on 35.734 ppm with limits at 10.00 and 80.00 ppm such that resonances outside these regions were aliased into the spectrum in order to optimise resolution. The HCCH-TOCSY ^{13}C mixing period was set to 11.3 ms. The number of transients and relaxation delay was set to provide 109 hours experimental time for the HCCH-TOCSY.

The ^{13}C-edited NOESY was acquired on a Varian UnityINOVA 800 MHz NMR system operating at 18.8 Tesla at the MRC National Institute for Medical Research (NIMR), Mill Hill, London. The ^{13}C-edited NOESY used a mixing time of 100 ms

was acquired with 2048 points (12000 Hz) in the direct F3 dimension (^{1}H) and 128 points (8000 Hz) in F2 (^{13}C) and 360 points (10500 Hz) in F1 (^{1}H) indirect dimensions. The ^{13}C dimension of this edited NOESY was centred on 41.066 ppm with limits at (10.00 and 80.00 ppm) such that resonances outside these regions were folded into the spectrum in order to optimise resolution. The number of transients and relaxation delay was set to provide 68 hours experimental time.

Carrier points were defined as 4.766 ppm for all ^{1}H, 41.109 ppm for ^{13}C and 117.436 ppm for ^{15}N where applicable.

3D ^{15}N TOCSY-HSQC and NOESY-HSQC experiments were run with 2048 points (8000 Hz) in the 1H F3 dimension, 192 points (8000 Hz) in the 1H F2 dimension and 64 points (2150 Hz) in the ^{15}N F1 dimension. NOESY experiments had a mixing time of 100 ms, and TOCSY experiments a 60 ms mixing time and both experiments were collected with 4 transients.

4.2.3 Amino acid sequential backbone resonance walkthrough

N-CIRP CBCANH and CBCA(CO)NH 3D spectra were previously acquired by Dr. Michelle Rowe. The triple resonance experiment CBCANH and CBCA(CO)NH were collected with a double labeled 1 mM N-CIRP sample with the N-terminal hexaHis-tag removed. CBCANH and CBCA(CO)NH experiments were run with 1024 points (8000 Hz) in the 1H F2 dimension, 64 points (2150 Hz) in the 13C F1 dimension and 120 points (2000 Hz) in the ^{15}N F3 dimension over 32 transients. 2 X linear prediction was used to make 3 X data in ^{15}N resonance. These spectra had been partially assigned previously by Shu-Ju Hseih. A sequential walkthrough was carried out using the spectra to confirm the existing assignments and additional backbone Cα, Cβ and amide N and H resonance assignments were made as described further in the results section 4.3.1.1.

4.2.4 Amino acid side chain NMR resonance assignments

The assignments made from the triple resonance spectra were used as starting points to make side chain assignments. HCCH-TOCSY is a 3D experiment with two ^{1}H axes and a ^{13}C axis. It correlates the backbone and side chain ^{13}C and ^{1}H with each other (section 4.1.1). As the backbone resonances had already been assigned through

the triple resonances experiments, the side chain resonances which correlate with these could be identified. The spectrum was acquired using double labelled N-CIRP as described above. The spectrum was used to complete the ^{13}C side chain assignments: it also provided ^{1}H side chain information and these were assigned using residue specific identifiable patterns of the proton resonances (figure 4.3). This is further described in the results section 4.3.1.2.

4.2.5 Secondary Structure Prediction

Secondary structure prediction was completed using DANGLE (dihedral angles from global likelihood estimates) software (Cheung et al., 2010). DANGLE uses Bayesian methods to estimate Φ and Ψ backbone dihedral angles for each residue and secondary structure using chemical shift measurements and a database of structures with known chemical shifts (section 4.1.2). Maximum number of islands was set to 2 when running DANGLE.

4.2.6 Structure Calculation

N-CIRP structure calculations were carried out using the automated structure determination software ARIA (Ambiguous Restraint for Iterative Assignment) and CNS running on a Linux platform. For each round of calculations 9 iterations were performed. For the first 8 iterations, 20 structures were calculated and after each step the 7 lowest energy structures were used for the calculation of next of 20 structures. For the final iteration, 20 structures were calculated and the 10 with the lowest energy were analysed. The assignments and peak information imported into the ARIA software is described in detail in section 4.2.7. The parameter settings used in ARIA and CNS for structure calculation is detailed below (table 4.1 and 4.2).

Iteration	Violation Analysis		Partial Assignment	
	Violation tolerance (Å)	Violation threshold (Å)	Ambiguity cut-off	Maximum number of contributions
0	1000.0	0.5	1.0	20
1	5.0	0.5	0.9999	20
2	3.0	0.5	0.999	20
3	1.0	0.5	0.99	20
4	1.0	0.5	0.98	20
5	1.0	0.5	0.96	20
6	0.1	0.5	0.93	20
7	0.1	0.5	0.9	20
8	0.1	0.5	0.85	20

Table 4.1 Violation analysis and partial assignment settings used in ARIA for structure calculation of N-CIRP.

Default settings were used with CNS unless specified otherwise. All setting for CNS are described below.

(a) (b)

Ambiguous and unambiguous restraints	
Force constant high temp	10.0
Initial force constant cool 1	10.0
Final force constant cool 1	50.0
Final constant cool 2	50.0
Flat bottom harmonic well	
Left switch-point high temp	0.5
Right switch point high temp	0.5
Left asymptote high temp	-1.0
Right asymptote high temp	1.0
Left switch-point cool 1	0.5
Right switch-point cool 1	0.5
Left switch-point cool 2	0.5
Right switch-point cool 2	0.5
Right asymptote cool 2	0.1
Hydrogen bond restraints	
Use H-bonds starting at iteration	0
Force constant high temp	10.0
Initial force constant cool 1	10.0
Initial force constant cool 1	50.0
Force constant cool 2	50.0
Dihedral angle restraints	
Force constant high temp	5.0
Force constant cool 1	25.0
Force constant cool 2	200.0

Parameter	Value
Type	Torsion angle
Random seed	Increased by 1 for each
TAD high temperature	run
TAD time-step factor	10000.0
Cartesian high	9.0
temperature	2000.0
Cartesian first iteration	0
Time-step	0.003
Cool1 final temperature	1000.0
Cool2 final temperature	50.0
High-temp steps	10000
Cool1 steps	10000
Cool2 steps	8000

Table 4.2 (a) Annealing parameters used for structure calculation with CNS. (b) Dynamics settings used for structure calculation with CNS.

4.2.7 Analysis and Refinement of Calculated N-CIRP Structures

Structure calculations initially used ^{13}C carbon and ^{1}H side chain assignments made using the HCCH TOCSY spectra described above for preliminary calculation to optimise the calculation method. 10 structures were calculated and to determine the quality of the resultant structure ensemble, they were viewed using MOLMOL (Koradi, Billeter, & Wüthrich, 1996) and RMSD (root mean square deviation) values were calculated. Subsequent calculations included ^{15}N backbone assignment information as well as peak information from a 3D ^{15}N NOESY spectrum. The ^{15}N NOESY peaks were assigned for the NH backbone resonance and also the side chain protons. The NOESY spectrum also provides proton NOE peaks which give inter-residue information (section 4.1.3.1), this peak data was also included in the calculations to help determine the proximity of amino acids to each other. 10 structures were calculated again and RMSD values calculated.

To improve the calculated structures further, the HCCH TOCSY spectra assignment and peak information was replaced by information from a 3D ^{13}C NOESY spectra which contained all the information that is included in an HCCH-TOCSY spectra but it also provides additional inter-residue information like the ^{15}N NOESY spectra with extra NOE peaks. This is further described in the results sections. The assignments made on the HCCH TOCSY spectra were transferred onto the ^{13}C NOESY spectra and the extra NOE peaks were also picked and included in the calculation along with the ^{15}N NOESY information and the RMSD values for the ensemble of 10 structures were analysed. This was followed by further analysis of both the ^{15}N and ^{13}C NOESY spectra and making assignments of the inter-residue NOE peaks to define contacts between helices and strand residues and a further 10 structure calculations.

The final refinement step taken was by analysis of any restraint violations at the end of the calculations. Every violation was analysed and if necessary it was corrected by reassigning or being removed. A new round of calculations was carried out each time after violations had been analysed, improving the RMSD value each time.

4.2.8 RMSD Value Calculation

The RMSD values were calculated by opening the PDB structure ensemble of 10 in the MOLMOL software (Koradi et al., 1996). Calculation was made by using the option "fit➔to first". RMSD was calculated for all atoms of all residues, the backbone atoms only, all secondary structure element atoms or backbone atoms of secondary structure elements.

4.2.9 ^{15}N NMR Relaxation Experiments

^{15}N T_1 experiments were run with 1820 data points in the F2 dimension and 256 points in the F1 dimension. The delay times used were 64, 128, 256, 385, 513, 641 and 769 ms. ^{15}N T_2 experiments were run with 2048 data points in the F2 dimension and 256 points in the F1 dimension. The delay times used were 20, 40, 60, 80, 100, 120, 140 and 160 ms. {^{1}H, ^{15}N} NOE experiments were run with 2048 data points in the F2 dimension and 256 points in the F1 dimension. Experiments were run with {^{1}H,^{15}N} NOE ON and OFF with a two second saturation. hnNOES were calculated using $(I-I_0)/I_0$ (section 4.1.4).

All experiments were run with proton centred at 4.766 ppm and nitrogen at 116.579 ppm at 25°C by Dr Michelle Rowe on a Varian Inova, 600 MHz spectrometer. The data was processed using NMRpipe on Linux PCs and the NMR spectra were assigned using the software CcpNmr Analysis 2.0. ^{15}N T_1 and T_2 values were obtained by measuring the change in ^{15}N-^{1}H peak volume intensities over time. Each data set per amino acid was fitted to equation 4.8 with in the CcpNmr Analysis software.

Equation 4.8:

$I= I_0\exp(-kt)$ where k is R_1 or R_2.

Where, I is the ^{15}N-^{1}H peak height intensity at a given time, t, from the start of the experiment, k is exponential decay rate constant

4.3 Results

4.3.1 Resonance assignment of N-CIRP

The backbone and side chain resonance assignment of N-CIRP ^{1}H, ^{13}C and ^{15}N nuclei are listed in appendix B. Percentage assignments are listed in table 4.3 below.

Assignment Type	Nucleus	% Assigned
Backbone	$^{1}H^{N}$	100
	^{15}N	100
	$^{13}C\alpha$	97
	$^{13}C\beta$	97
Side-Chain	$^{1}H\alpha$	95
	$^{1}H\beta$	90
	$^{1}H\gamma$	94
	$^{1}H\delta$	67
	$^{1}H\varepsilon$	24
	$^{13}C\gamma$	71
	$^{13}C\delta$	48
	$^{13}C\varepsilon$	28

Table 4.3 Extent of resonance assignments for N-CIRP.

4.3.1.1 Conformation and further assignment of backbone resonances

The majority of the backbone resonance assignments for N-CIRP were achieved by Shu-Ju Hseih using a sequential assignment strategy with the triple resonance experiments, CBCA(CO)NH and CBCANH collected as described in section 4.2.3. The $^{1}H_{N}$ and ^{15}N assignments were mapped to the peaks in a ^{15}N-^{1}H HSQC spectrum (figure 3.2). To confirm the assignments already made, a sequential walkthrough of the spectra was carried out as shown in figure 4.6. The CBCA(CO)NH correlates ^{1}H and ^{15}N resonance of residue (*i*) with $^{13}C\alpha$ and $^{13}C\beta$ resonance of the previous residue (*i-1*). Using these rules further assignments which were absent were completed with the aid of the chemical shift pattern for amino acids in random coil conformation used a reference (figure 4.2) (Wishart & Sykes, 1994). The assignment process for residues 49-52 is shown in figure 4.6. For example, if 50 Gly is *i*, then i-1is for 49 Phe. The CBCA(CO)NH spectrum shows the $^{13}C\alpha$ and $^{13}C\beta$ resonance peak of 49 Phe at the amide ^{1}H,^{15}N resonance of 50 Gly as shown in the left-hand side 50 Gly strip pair in figure 4.6. The CBCANH spectrum (right-hand side) shows the $^{13}C\alpha$ resonance peak

of 50 Gly and ^{13}Cα and ^{13}Cβ resonance peaks of 49 Phe at the amide ^{1}H,^{15}N resonance of 50 Gly. The chemical shifts were linked both forward and backward looking at these patterns and matching the peaks at the same ^{13}C chemical shifts. If these matches could not be made, residues which have unique patterns as described earlier in the chapter (section 4.1.1) were searched for and used as a starting point to assign further previous or proceeding residues.

The ^{1}H, ^{15}N, ^{13}Cα and ^{13}Cβ resonances assignment were confirmed and were assigned to 100% using the CBCANH and CBCA(CO)NH spectra. The chemical shift values are shown in appendix B.

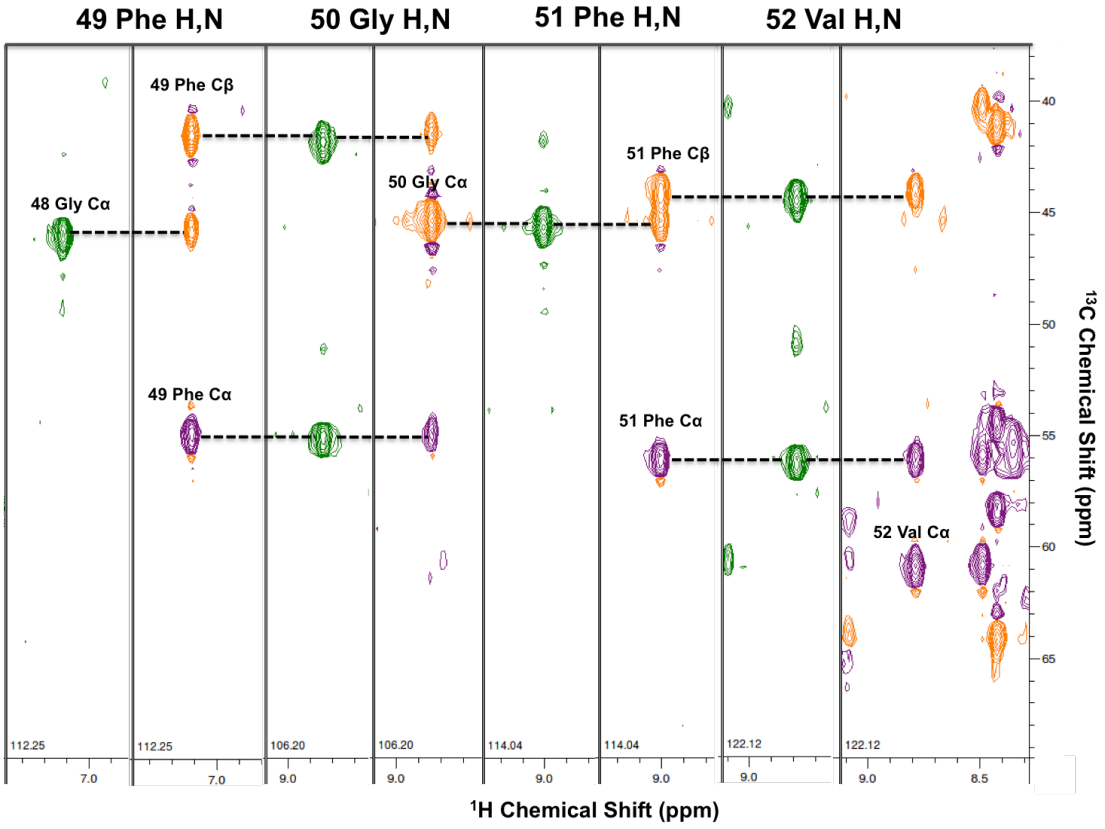

Figure 4.6 An example of backbone triple resonance assignment of N-CIRP. The CBCA(CO)NH strips are shown on the left and the CBCANH strips are shown on the right for each residue. The ^{15}N chemical shift position is shown at the bottom of each strip. Sequential matches are indicated by a dotted line. CBCA(CO)NH peaks are in green, positive Cα CBCANH peaks are purple and negative Cβ CBCANH peaks are orange with the exception of Gly Cα peaks which is negative.

4.3.1.2 Side Chain Assignment

The backbone chemical shifts assignments were used as starting points for the assignment of ^{1}H and ^{13}C side chain resonances. A 3D ^{15}N-edited TOCSY experiment which couples the backbone ^{15}N, ^{1}H backbone to the side chain ^{1}H side chain resonances was used to begin the assignment process for each residue. Residue specific patterns as described in section 4.1.1 were used to aid the process. Figure 4.7 shows examples of side chain resonances assigned using the ^{15}N-edited TOCSY spectra. Peak on the diagonal is the chemical shift of the hydrogen which is bonded to the nitrogen at the ^{15}N chemical shift indicated. The correlated ^{1}H resonances are found at this ^{1}H chemical shift.

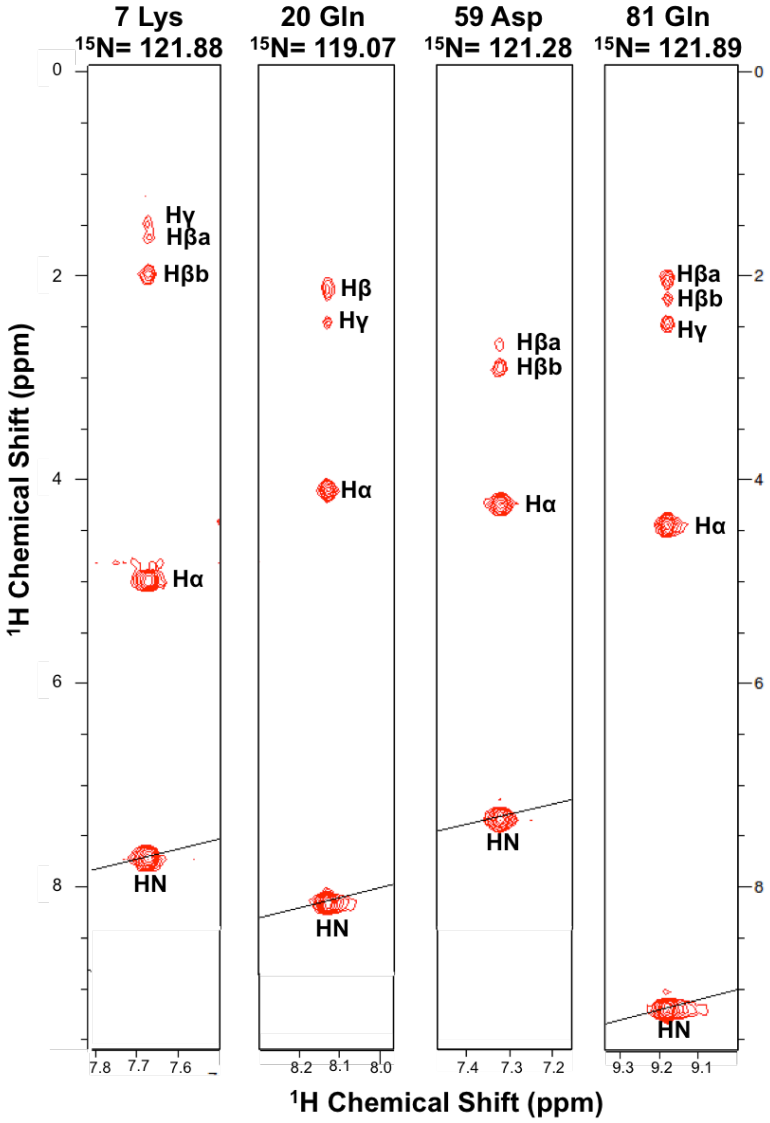

Figure 4.7 ^{15}N-edited TOCSY NMR ^{1}H-^{1}H strips showing example assignments for ^{1}H side chain resonance of residues 7, 20, 59 and 81. The ^{15}N chemical shift position of each strip is indicated above each strip.

The majority of the side chain assignments were made using the HCCH-TOCSY spectrum. The side chain hydrogen resonance assignments made on the ^{15}N-edited TOCSY spectra aided this process. The reference chemical shift patterns shown in figure 4.3 were also used to identify resonances for specific residues. Resonances were identified by moving through ^{13}C chemical shift planes. The spectrum shows a strip of side chain ^{1}H resonances which are bonded to the side chain carbons for every residue. A separate strip is seen for every carbon in the side chain of the amino acid at the chemical shift plane for that carbon. The peak on the diagonal belongs the hydrogen which is bonded to the carbon at that that chemical shift plane. For example, at the chemical shift plane of Cα, the hydrogen resonance on the diagonal will belong to Hα. The below figures (figure 4.8-4.11) show example assignments of ^{13}C and ^{1}H side chain resonances for different amino acids as specified in the figure legend.

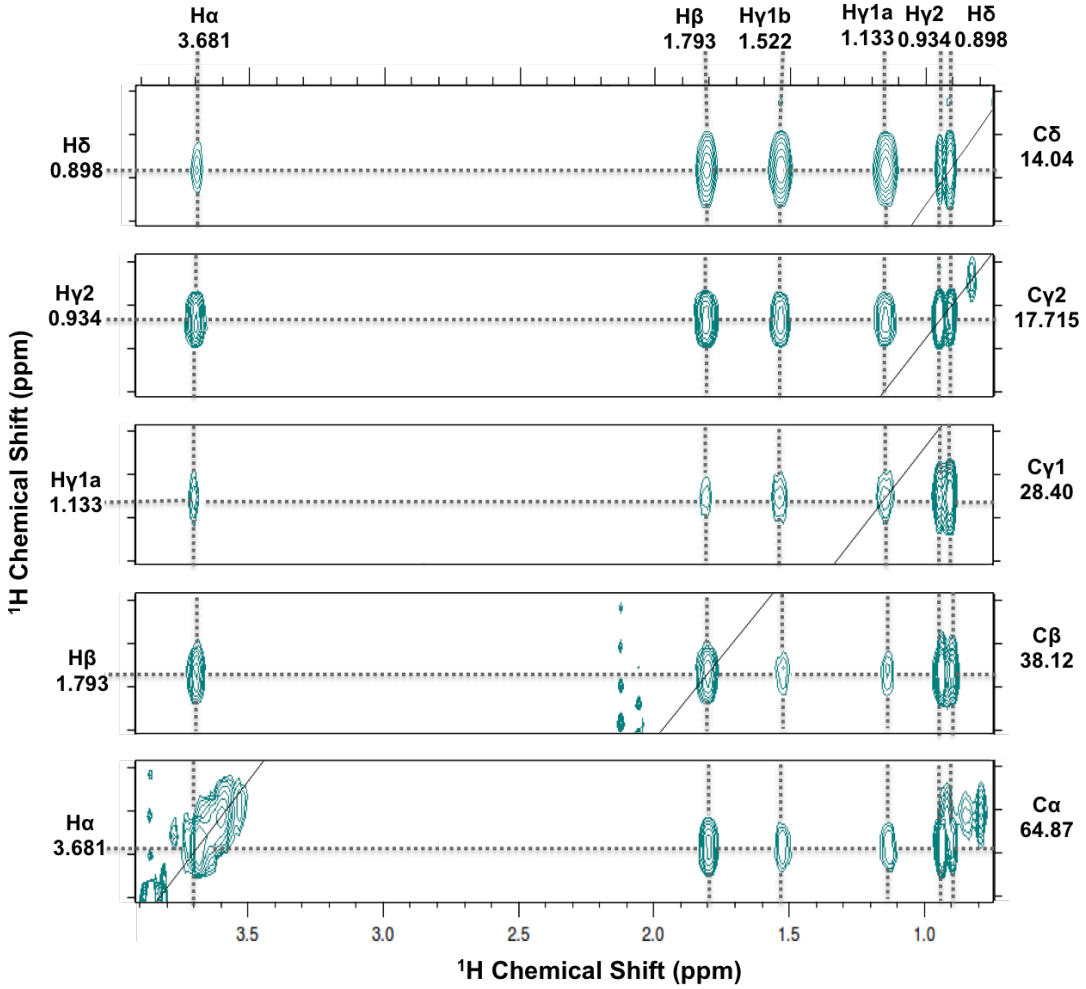

Figure 4.8 HCCH TOCSY ^{1}H-^{1}H strips for residue 57 Ile showing the assignment of the side chain ^{1}H and ^{13}C resonances. The ^{13}C chemical shift of each plane is indicated to the right, which is the carbon to which the hydrogen on the diagonal is bonded. The assignment and ^{1}H chemical shift of the hydrogen on the diagonal is indicated on the left. The assignment and the ^{1}H chemical shift of the side chain hydrogens are indicated on top.

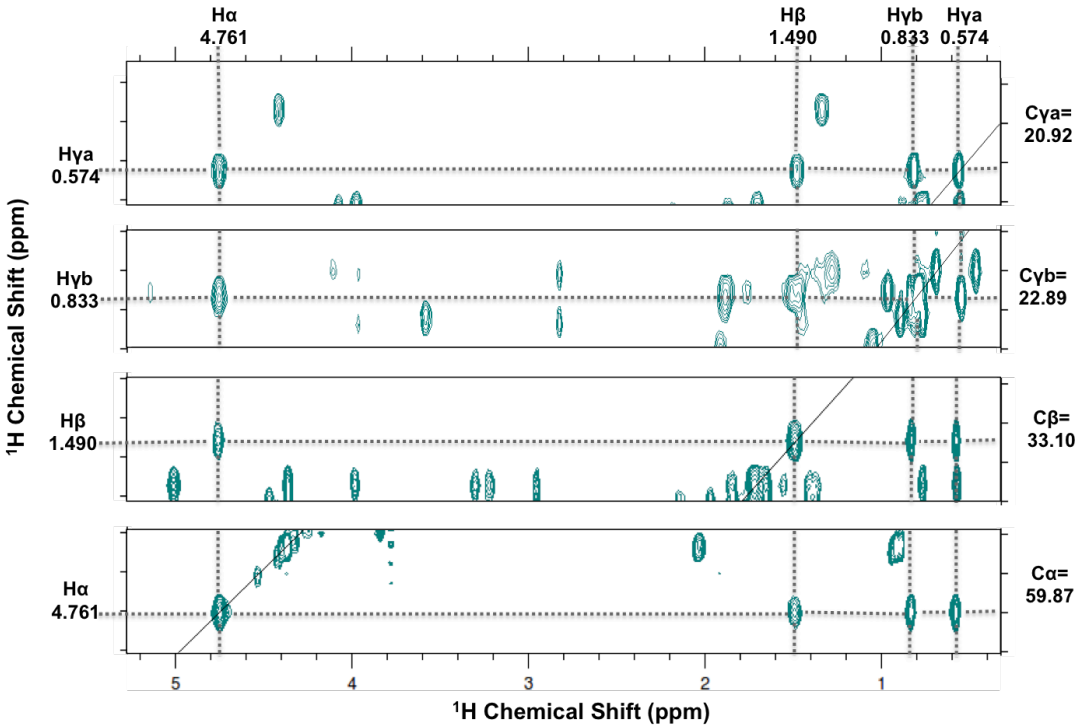

Figure 4.9 HCCH TOCSY ^{1}H-^{1}H strips for residue 10 Val showing the assignment of the side chain ^{1}H and ^{13}C resonances. The ^{13}C chemical shift of each plane is indicated to the right, which is the carbon to which the hydrogen on the diagonal is bonded. The assignment and ^{1}H chemical shift of the hydrogen on the diagonal is indicated on the left. The assignment and the ^{1}H chemical shift of the side chain hydrogens are indicated on top.

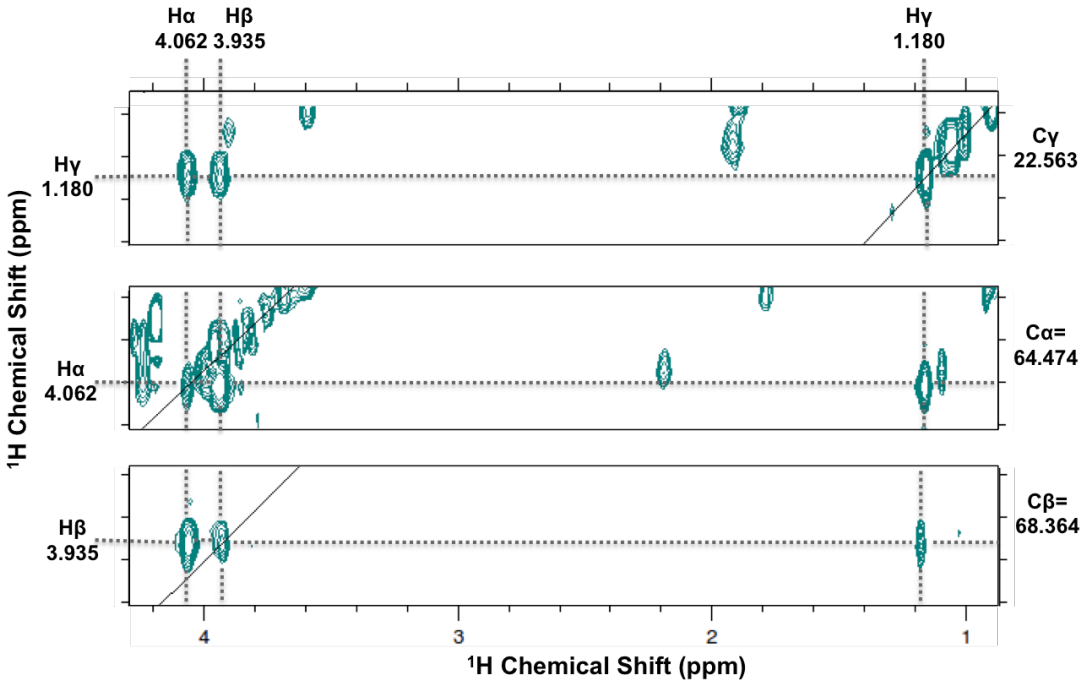

Figure 4.10 HCCH TOCSY ^{1}H-^{1}H strips for residue 17 Thr showing the assignment of the side chain ^{1}H and ^{13}C resonances. The ^{13}C chemical shift of each plane is indicated to the right, which is the carbon to which the hydrogen on the diagonal is bonded. The assignment and ^{1}H chemical shift of the hydrogen on the diagonal is indicated on the left. The assignment and the ^{1}H chemical shift of the side chain hydrogens are indicated on top.

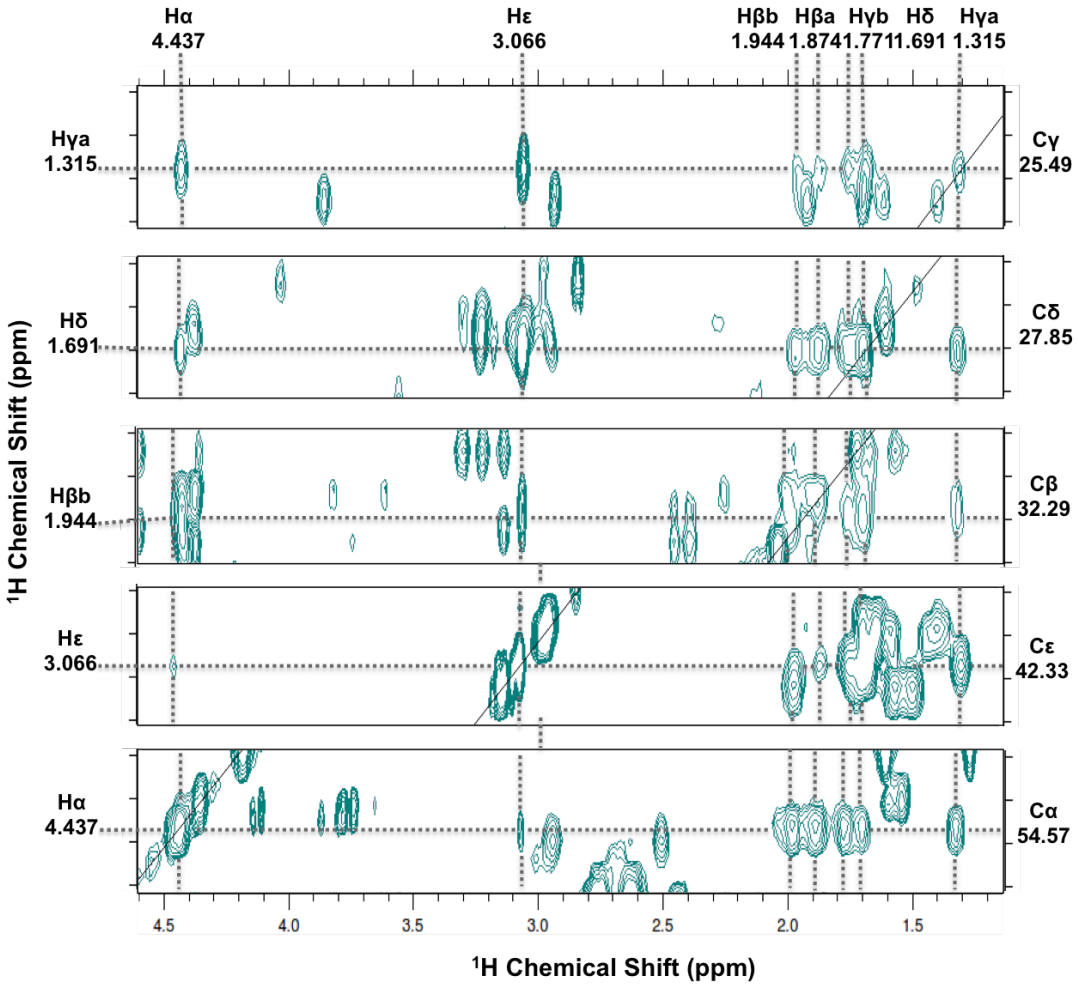

Figure 4.11 HCCH TOCSY ^{1}H-^{1}H strips for residue 70 Lys showing the assignment of the side chain ^{1}H and ^{13}C resonances. The ^{13}C chemical shift of each plane is indicated to the right, which is the carbon to which the hydrogen on the diagonal is bonded. The assignment and ^{1}H chemical shift of the hydrogen on the diagonal is indicated on the left. The assignment and the ^{1}H chemical shift of the side chain hydrogens are indicated on top.

4.3.2 Secondary Structure Determination

Several programs are available for predicting protein secondary structure using chemical shifts of ^{1}H, ^{15}N and ^{13}C. DANGLE was used to compare the predicted N-CIRP secondary structure with the N-CIRP (1X5S) secondary structure available in the protein data bank (PDB). Chemical shifts of ^{1}H, ^{15}N and ^{13}C were provided by Shu-Ju Hseih.

Figure 4.12 shows examples of Ramachandran plots for determining secondary structures using the DANGLE prediction program. The middle window shows an ambiguous result (residues 16 Asp) indicating this residue is placed in a loop region.

The left window shows β-strand prediction (residue 9 Phe). Residue 60Ala in the right window shows an α-helix prediction. Ramachandran plots were analyzed for each residue in this way to determine secondary structure.

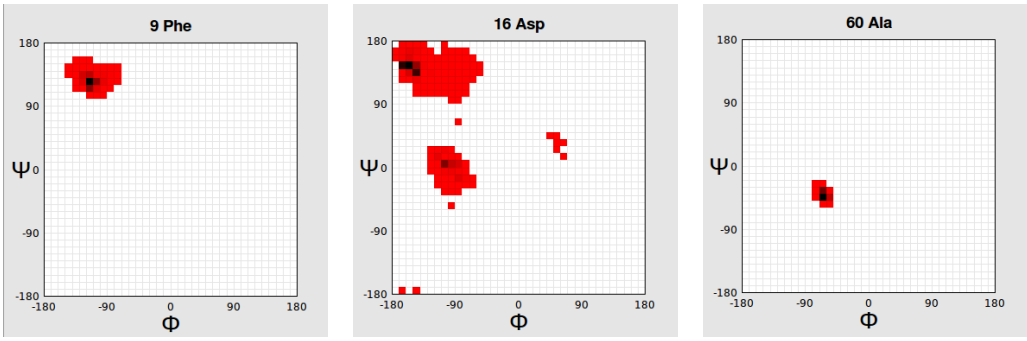

Figure 4.12 Representative global likelihood estimate (GLE) diagrams of Ramachandran space distribution plots from DANGLE for residues 9,16 and 60 of CIRP that were predicted to be a section of β-sheet, loop and α-helix). Bins with B-scores below the threshold value are in white; other bins are shaded from red (threshold value) to black (Bmax)

The comparison between the 1X5S structure and the secondary structure predicted by DANGLE is shown in figure 4.13. The predicted secondary structure is largely similar to that of the structure 1X5S. The α- helices are especially similar being almost identical in length and in the same region of the sequence. The DANGLE prediction for α1 being identical to the 1X5S structure between residues 19-29 and α2 was only one residue shorter in the DANGLE prediction between amino acids 57-66 instead of 57-67. The length of the β-strands are different, generally shorter than the 1X5S structure except for β2 being longer. DANGLE predicted β1 as 3 residues long (8-10) whereas the sheet was placed between residues 7-11 in 1X5S. Similarly DANGLE prediction for β3 was 3 residues long (51-53) but 49-53 in 1X5S. β2 however was predicted to be across 6 residues (34-39) by DANGLE but it was shorter in 1X5S, (34-37). Even though the length of the strands varied between DANGLE and 1X5S, they were predicted to be in the same region in both. β4 shows the most dissimilarity being more upstream in the DANGLE prediction (71-80) and also longer than the 1X5S strand (78-83).

The secondary structure arrangement of N-CIRP was also compared to the RNA binding protein hnRNP1A RRM1 whose structure has recently been solved (Barraud

& Allain, 2013) and this protein displays a typical RNA binding protein fold. It possessed two RNA binding domain similar to one another, figure 4.13 is showing the RRM1 domain. The secondary structure arrangement of hnRNPA1 RRM1 is similar to that of N-CIRP with the exception of an extra strand and each of the secondary structure elements being further downstream in the sequence. β1 and α2 are of a similar length as in N-CIRP. β4 and α1 are longer in N-CIRP than in hnRNP1A RRM1 and β2 and β3 are shorter in length. β strand and α sheet information is heavily defined by hydrogen bond information. No hydrogen bond data was collected for mouse N-CIRP which would improve the accuracy of the DANGLE prediction.

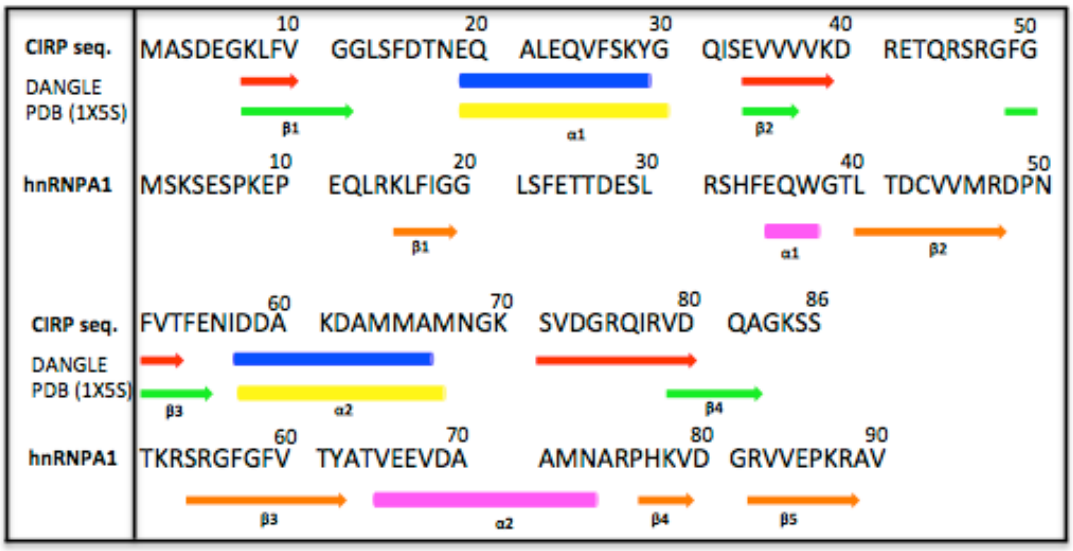

Figure 4.13 Schematic representation of the secondary structure of N-CIRP in the protein data bank (1X5S) and predicted secondary structure of N-CIRP using DANGLE. Also shown here is the sequence and secondary structure of the RNA binding protein hnRNPA1 RRM1 (2LYV) which has a similar structure to the human CIRP structure. Sequence numbering is indicated above the sequence and prediction is shown below the sequence; α-helices are represented by cylinders and β-strands are represented by arrows.

4.3.3 N-CIRP Structure Calculations

Many steps were carried out during the process of calculating the structure of N-CIRP. A summary of the process is shown in the table 4.4 below, where the qualities of the structures were assessed by the RMSD value.

	RMSD (Å)- Total Backbone	RMSD (Å)- Structured Regions Backbone
^{13}C NOESY data	3.96	2.23
^{13}C NOESY and ^{15}N NOESY data	3.45	2.10
+ Manual Assignments	2.82	2.2
Violation correction 1	3.75	1.22
Violation correction 2	2.14	1.16
Violation correction 3	3.03	1.56
Violation correction 4	3.51	2.70
Violation correction 5	4.12	3.32
Violation correction 6	2.77	1.01
Violation correction 7	2.17	1.03
Violation correction 8	1.63	0.89

Table 4.4 Summary of the N-CIRP structure calculation and violation analysis process. For the RMSD calculations backbone atoms C, Cα and N were considered and for structured regions, secondary structure elements as predicted by DANGLE were used.

4.3.3.1 NOESY spectra and Manual NOE Assignments

^{13}C HCCH-TOCSY and ^{15}N-edited TOCSY spectra were useful for the assignment of N-CIRP resonances. These spectra provide residue information but do not provide inter-residue data. 3D ^{13}C NOESY and ^{15}N NOESY spectra contain equivalent peaks to the TOCSY spectra but also have extra peaks which provide through space inter-residue information. Having this extra inter-residue information is essential for structural calculations. The peak assignments on the HCCH TOCSY spectra were transferred to the ^{13}C NOESY spectra as well as picking the extra inter-residue NOE peaks (figure 4.14). The same process was carried out as with the^{15}N-edited TOCSY assignments and ^{15}N NOESY spectra. The chemical shift information of the assigned N-CIRP resonances taken from the NOESY spectra and the unassigned NOE peaks were used for the structure calculations in ARIA (figure 4.14).

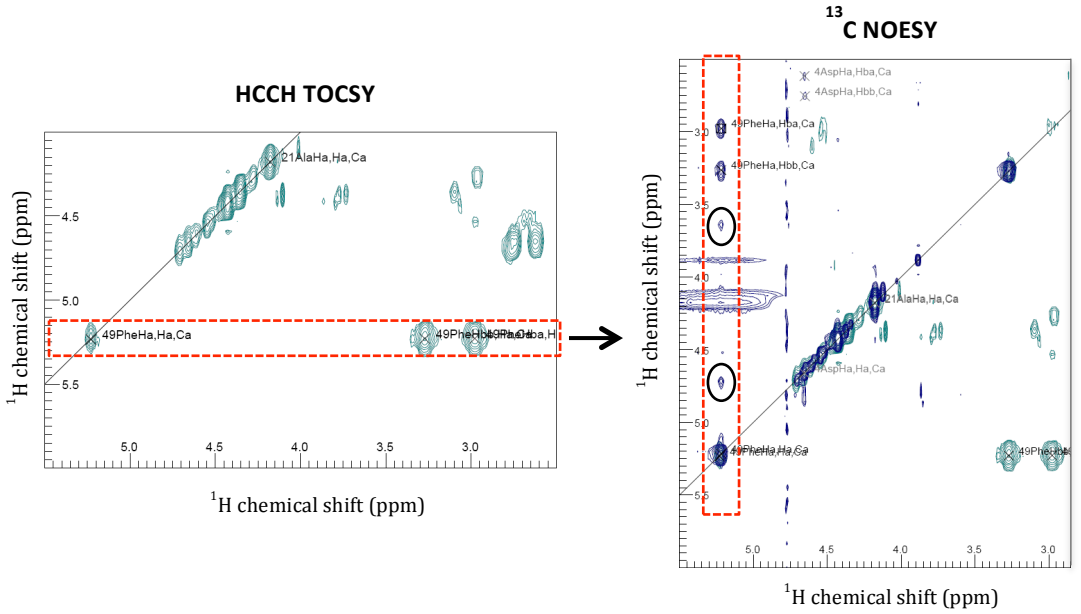

Figure 4.14 The process of assignment transfer from HCCH TOCSY spectra to the ^{13}C NOESY spectra and the identification of the NOE cross peaks which are indicated by circles.

Once the transfer of the assignments were complete and all NOE cross peaks were selected, this data was used to run an automated structure calculation of N-CIRP using the ARIA software. Initially calculations were carried out using the ^{13}C NOESY assignments and peak data only, i.e. without any ^{15}N-^{1}H assignment information. The calculated ensemble had a backbone RMSD of 3.96 Å and was showing initial formation of the secondary structure elements (figure 4.15). 3 strands could be seen out of a total of 4 determined with DANGLE (section 4.3.2) and out of the two helices none were observed.

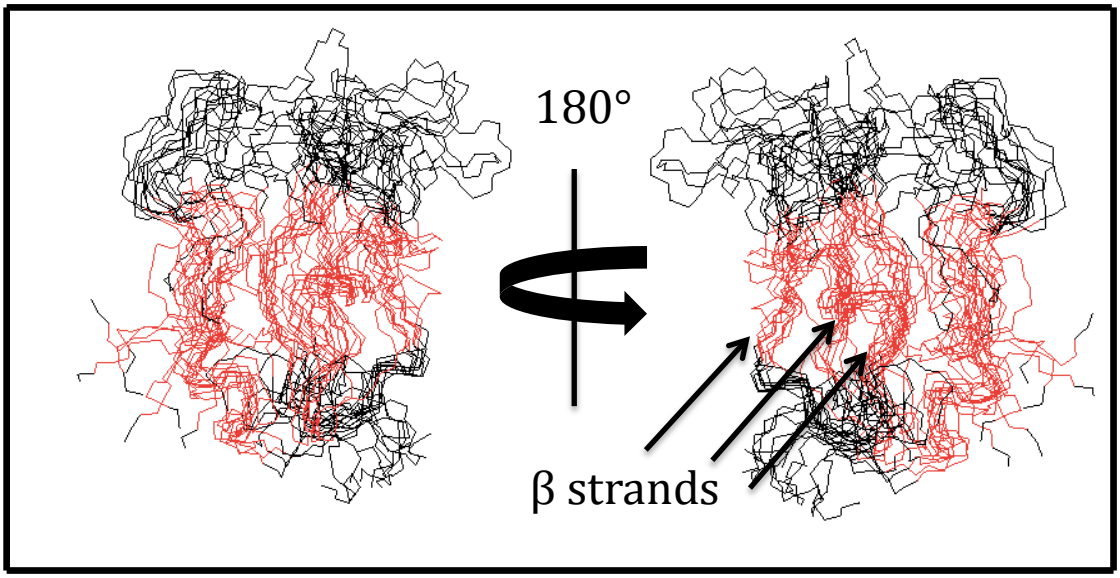

Figure 4.15 First N-CIRP structure ensemble as calculated by ARIA with only ^{13}C-^{1}H assignment data. The structure ensemble was produced by fitting al 10 structures together by the secondary structure backbone in MOLMOL.

After this, another calculation was carried out using both ^{13}C NOESY and ^{15}N NOESY assignments and peak information. The structure ensemble from this calculation had a lower backbone RMSD value 3.45 Å and the regions with the secondary structure elements looked more defined and tightly fitted together and one of the helices had started to show (figure 4.16).

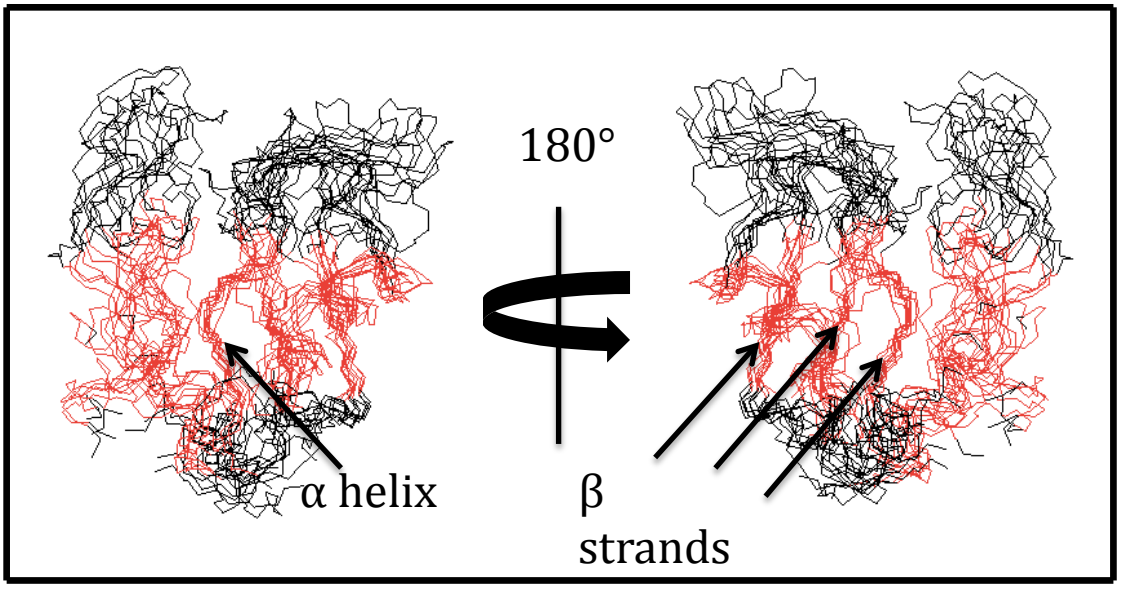

Figure 4.16 N-CIRP structure ensemble as calculated by ARIA after addition of ^{15}N-^{1}H assignment data. The structure ensemble was produced by fitting al 10 structures together by the secondary structure backbone in MOLMOL.

To increase structural precision, more data was included in the ARIA structure calculation. More data, which gives indication of how the protein is folded in space and which, add restraints to calculation. This was achieved by making manual NOE assignment using the CcpNmr analysis software. ARIA software makes automated ambiguous NOE assignments to calculate the structure of the protein, the addition of correct unambiguous manual assignments result in a more accurate structure. Parameters reported by Barsukov and Lian (table 4.5) were used to identify NOE cross peaks of resonances close in space. These parameters are helpful to define secondary structures, and short and medium range interaction between strands.

Some assignments were also made by examining the calculated structure (figure 4.16) and identifying potential NOEs between residues, which appeared in close proximity to each other on the structure.

Parameter	α-Helix	3_{10}-Helix	β-Anti-parallel	β-Parallel
$d_{\alpha N}(i, i)$	2.6	2.6	2.8	2.8
$d_{\alpha N}(i, i+1)$	3.5	3.4	**2.2**	**2.2**
$d_{\alpha N}(i, i+2)$	4.4	**3.8**		
$d_{\alpha N}(i, i+3)$	**3.4**	3.3		
$d_{\alpha N}(i, i+4)$	4.2	(>4.5)		
$d_{NN}(i, i+1)$	**2.8**	**2.6**	4.3	4.2
$d_{NN}(i, i+2)$	4.2	4.1		
$d_{\beta N}(i, i+1)$	2.5-4.1	2.9-4.4	3.2-4.5	3.7-4.7
$d_{\alpha\beta}(i, i+1)$	2.5-4.4	3.1-5.1		
$d_{\alpha\alpha}(i, j)$			**2.3**	4.8
$d_{\alpha N}(i, j)$			3.2	3.0
$d_{NN}(i, j)$			**3.3**	4.0
$^{3}J_{HN\text{-}H\alpha}$	(<4)	(<4)	(>9)	(>9)
NH exchange rate	**slow**	**slow**	**slow**	**slow**

Table 4.5 Parameters that are useful for secondary structure identification; short sequential and medium range ^{1}H-^{1}H distances, vicinal coupling constants and amide hydrogen exchange rates. The most readily available parameters are highlighted in bold adapted from Barsukov & Lian, 1993.

Approximately 260 cross peaks were identified and assigned, the majority of these were made using the ^{15}N NOESY spectra (220) (figure 4.18) and the remainder on ^{13}C NOESY (figure 4.17).

For example, the first ^{13}C NOESY strip on figure 4.17 shows that a cross peak on the 8 Val ^{13}Cα strip had been assigned to 52 Hα, this assignment was made through examination of the calculated structures and discovering that these two residues on strands β1 and β3 respectively are close in space and the presence of the cross peak confirmed that these two strands run alongside each other. Also seen on strip 3, the 52 Val Hα on β3 and 35 Val Hα on β2 are in close distances to each other suggesting that strands β3 and β2 are antiparallel to each other. From this we can conclude that the strands β1-3 are antiparallel in the order of β1, β3 and β2. The strips 2 and 4 shows the $d_{\alpha\beta}(i, i+3)$ parameter where 22 and 57 Hα are i and 25 and 60 Hα i+3. This suggests that these residues are in an alpha helix formation in the structure. This agrees with the DANGLE prediction (section 4.2) from where these residues are found on helices α1 and α2 respectively.

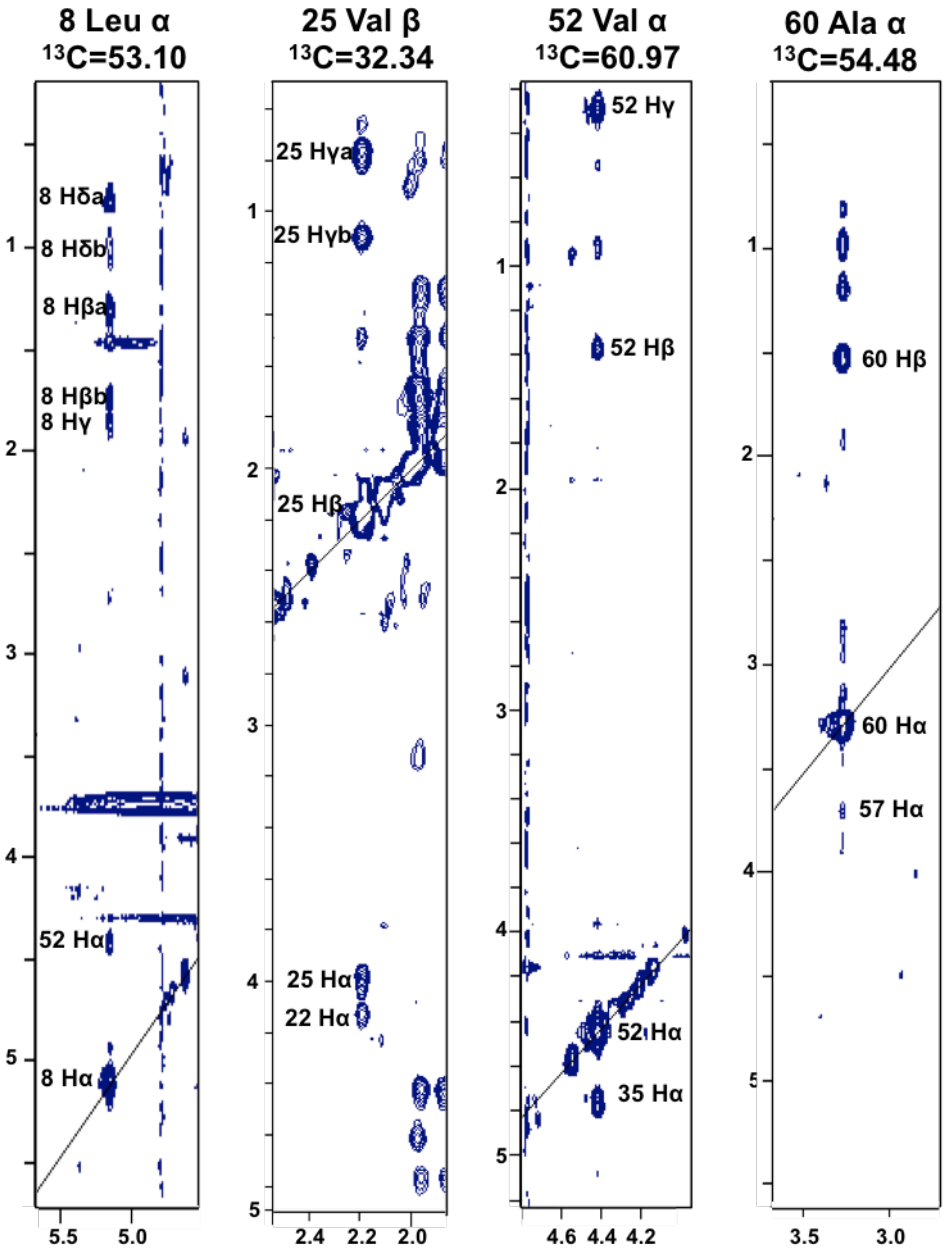

Figure 4.17 Example of ^{13}C NOESY strips showing cross peak assignments. The ^{13}C chemical shift position is indicated above each strip.

Similar assignments were also made on the ^{15}N NOESY spectra (figure 4.18). For example residues 40 Asp and 56 Asn shown in strips 2 and 3 correlate with the prior amino acids 39 and 55 NOE peaks, suggesting that these residues are in strand formation in the structure ($d_{\alpha N}$ (*i, i+1*), (table 4.5).

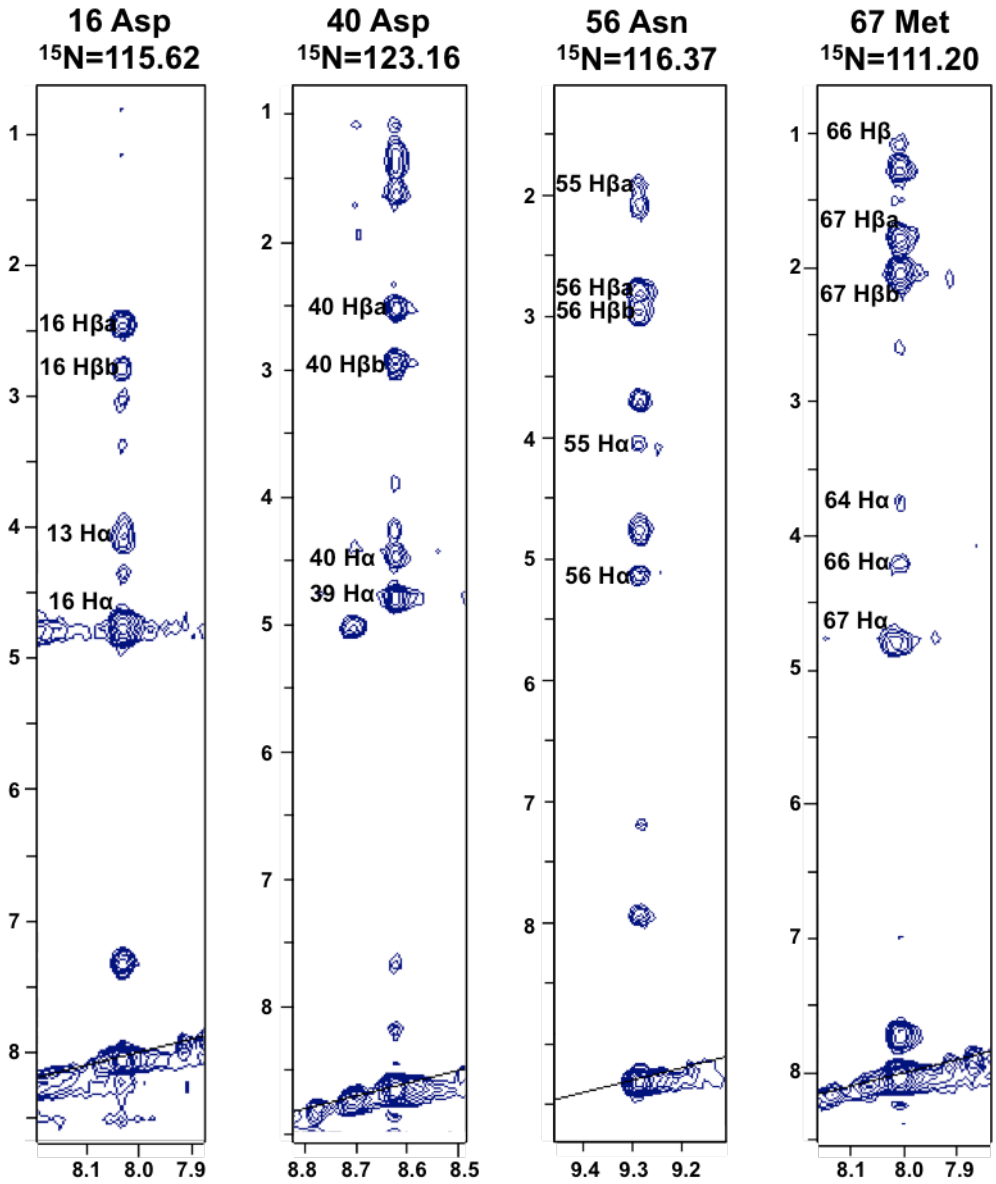

Figure 4.18 Example of ^{15}N NOESY strips showing cross peak assignments. The ^{15}N chemical shift position is indicated above each strip.

ARIA structure calculation was performed following the addition of the manual NOE assignments. The ensemble of the 10 calculated structures had a significantly improved backbone RMSD of 2.82 Å. The structures showed that one of the helices and two strands was very well defined, with the second helix and the other two strands also formed but not as tightly fitted together (figure 4.19). A higher agreement of the loop regions between the structures was also observed. The structure also looked more condensed with the strands and helices in better agreement with each other compared to the previous calculations where it looked more collapsed onto itself.

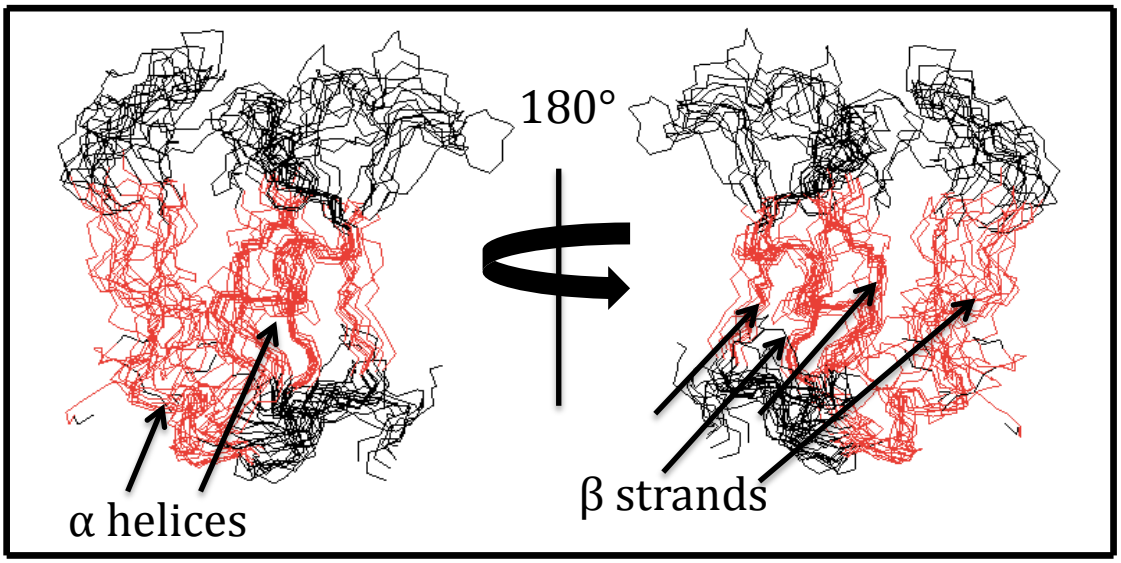

Figure 4.19 N-CIRP structure ensemble as calculated by ARIA after addition of manual NOE assignment data. The structure ensemble was produced by fitting al 10 structures together by the secondary structure backbone in MOLMOL.

4.3.3.2 Structure Refinement

All NOE data from ^{13}C NOESY spectra, ^{15}N NOESY spectra and the manual NOE cross peaks assignments were used for all proceeding calculations. To improve the structures after each calculation the NOE violation list was analysed. Each violation was checked on the appropriate spectra using CcpNmr analysis software and either corrected or if necessary removed. An improvement in the RMSD value was observed after each calculation and the checking of the violations list. 8 rounds of these corrections were performed as summarised in table 4.5. The most improved structure is shown in figure 4.20 below. This ensemble had a backbone RMSD value of 1.63 Å which is approximately half of the ensemble calculated initially and the RMSD across

structured regions was 0.89 Å. The 10 calculated structures appeared tightly fitted together and in agreement for both the loop and the structured regions. 4 well defined strands and 2 helices were easily identifiable with the strands in the order of β4, β1, β3, β2. The boundaries for the strands and helices were found to be in agreement with the DANGLE secondary structure determination in sections 4.3.2 and also with the human N-CIRP structure on PDB (1X5S).

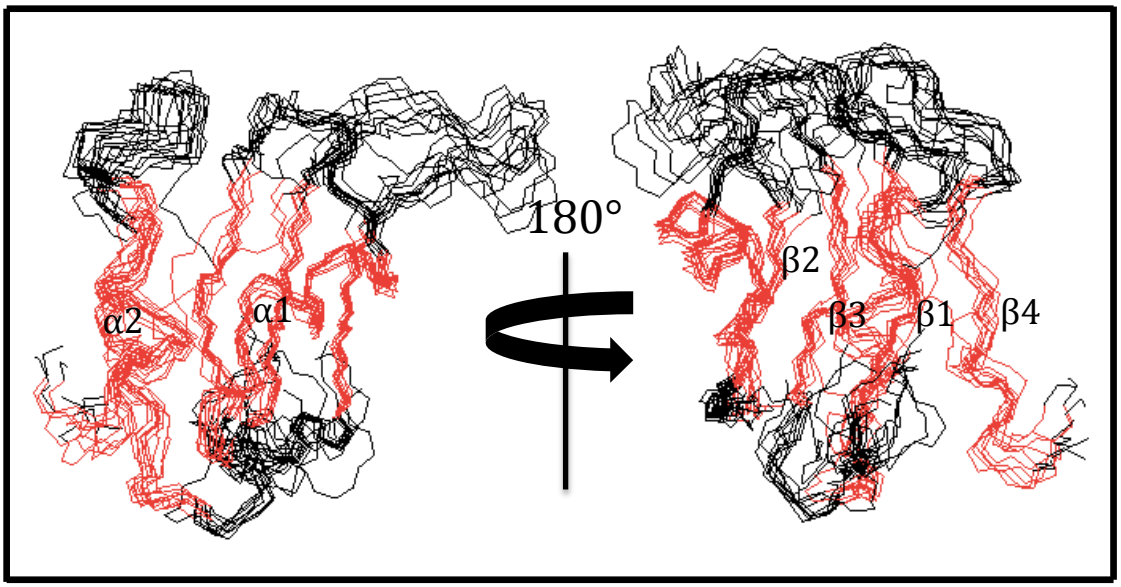

Figure 4.20 Final N-CIRP structure ensemble as calculated by ARIA after rounds of violation corrections. The fitted ensemble structure was produced by fitting all 10 structures together by the secondary structure backbone in MOLMOL.

The figure 4.20 shows the most refined version of the N-CIRP structure calculated during this project. A more precise structure can be achieved by the addition of dihedral angle and hydrogen-bonding data in future experiments.

4.3.4 ^{15}N T_1,T_2 and hnNOE Relaxation Experiments

To study N-CIRP dynamics, relaxation experiments were used. ^{15}N HSQC experiments were set up with a series of T_1 or T_2 relaxation delays. T_1 and T_2 values were obtained by measuring peak volumes after a range of relaxation delays (figure 4.21). T_1, T_2 and hnNOE values for each residue are listed in appendix C.

Figure 4.21 shows example relaxation decay curves for residues across the N-CIRP protein. These were generated in the CcpNmr Analysis software by following peak volume changes of each amino acid peak on the HSQC spectra collected with a range of relaxation delays. These were then used to produce T_1 and T_2 values (figure 4.22) for each residue by the CcpNmr Analysis software.

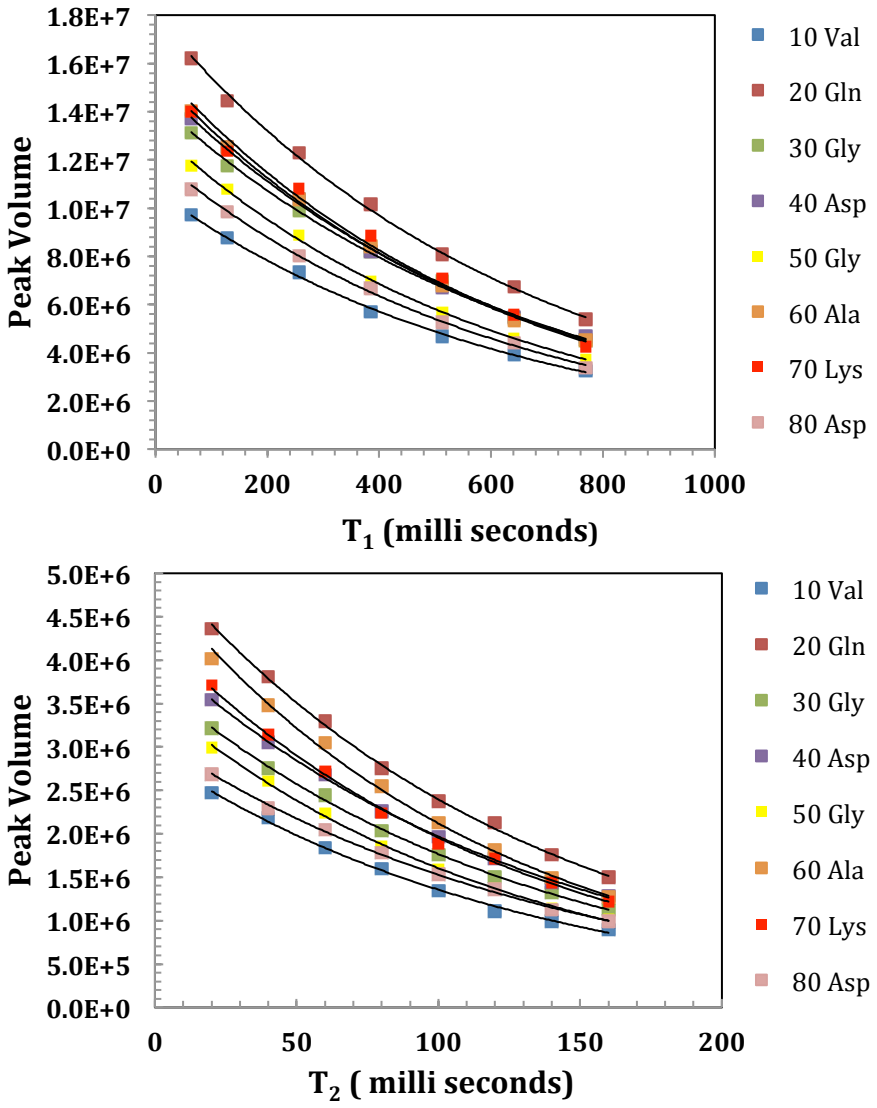

Figure 4.21 Example relaxation decay curves for the calculation of T_1 and T_2 values.

The relaxation times shown by the scatter graphs (figure 4.22) give an indication of protein motion. Relaxation times for all 3 parameters (T_1, T_2, hnNOE) did not change significantly across the central regions of the protein. A small increase in the T_1 values were observed at the N- and C-terminal residues with an average T_1 time of 0.64±0.06 s across the protein. Similarly the T_2 times also showed a small increase at the termini residues with an average T_2 of 0.15±0.04 s. The hnNOE values were also similar across the centre of the protein with a -0.23±0.33 average. The extreme residues at the N- and C-termini had values more negative hnNOE compared the rest of the protein which mirrored the T_1 and T_2 increases.

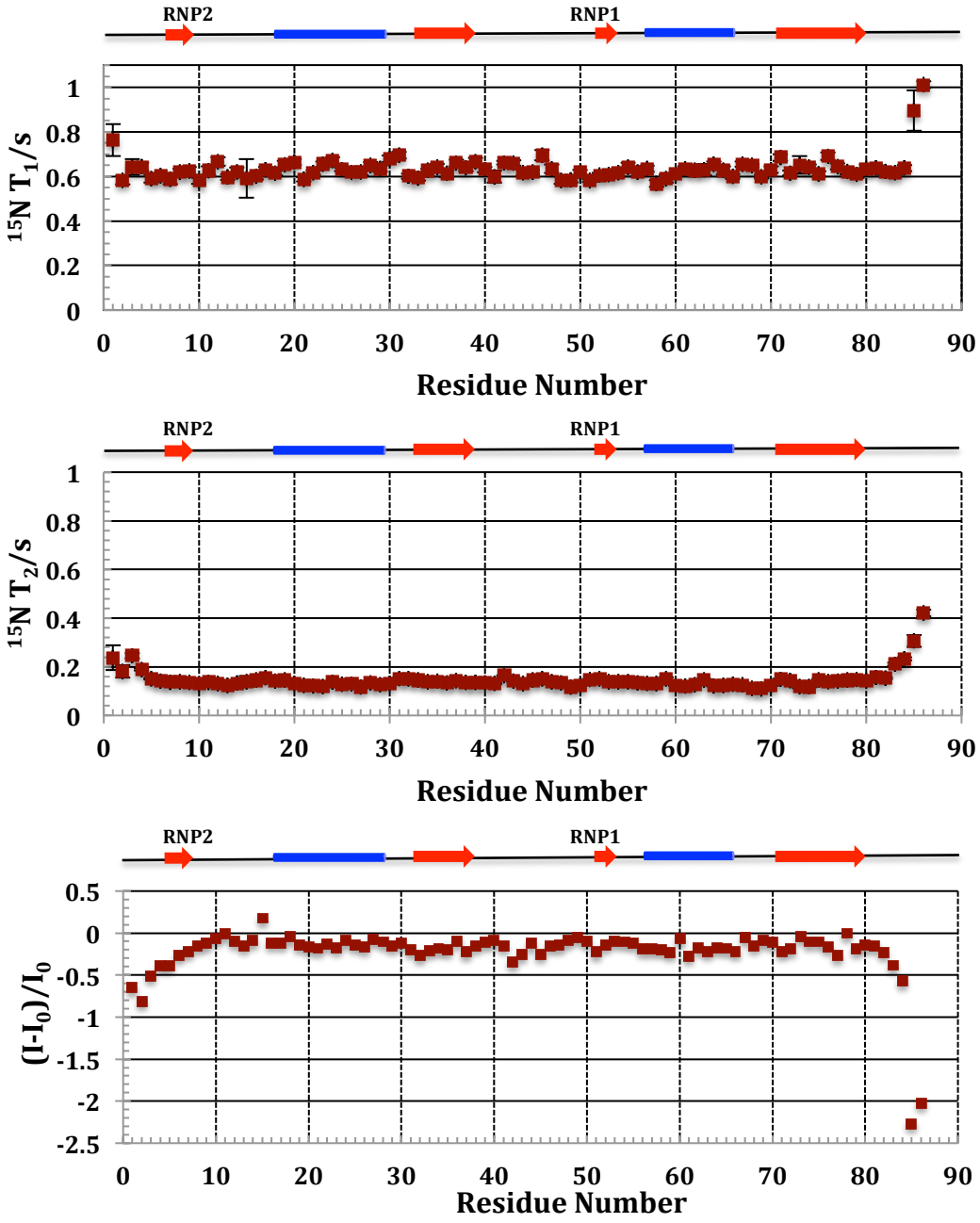

Figure 4.22 Plots of ^{15}N T_1, T_2 and heteronuclear NOE relaxation rates.

T_1 vs. T_2 plot (figure 4.23) shows the T_1 and T_2 values for all residues of N-CIRP. The plot includes dipolar relaxation curves (600 MHz) for the order parameters S^2 and τ_m provided by Mark Howard. The data points on the plot were packed tightly together in the same region, except for the termini residues, which were diffused across the graph as labelled on the plot. The average ^{15}N T_1 value was calculated as 0.63±0.03 s and the ^{15}N T_2 value was 0.14 ± 0.01 s. The average S^2 value was 0.77 ± 0.08 and the average τ_m was 6.18±0.21 ns.

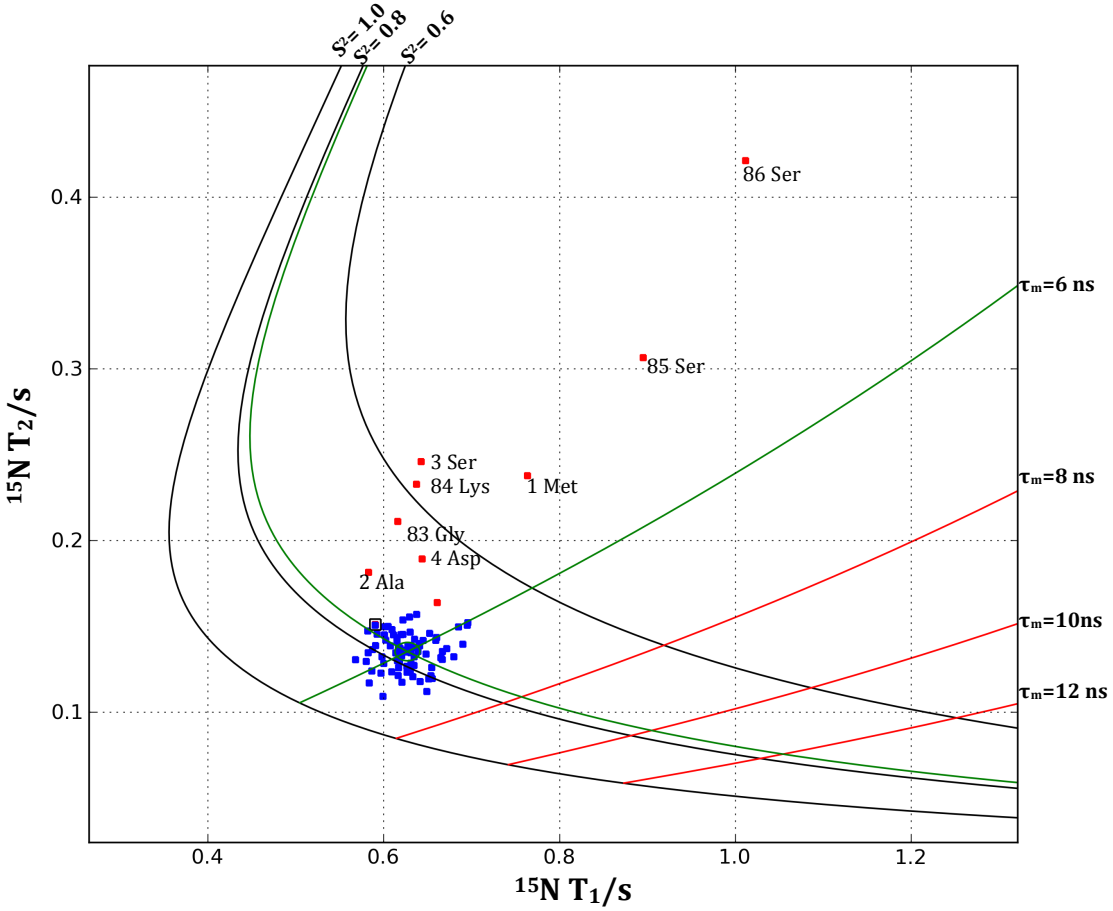

Figure 4.23 Plot of ^{15}N T_1 and T_2 of each amide of N-CIRP. Contour lines are shown for S^2 order parameter and τ_m.

4.3.5 ModelFree Calculated Order Parameters

The T_1, T_2 and NOE relaxation times were used for ModelFree analysis to better understand N-CIRP motion. S^2 and τ_e values were calculated for each residue (appendix C). S^2 is an order parameter and describes spacial restriction of internal motion, S^2=1.0 meaning complete rigidity. τ_e is a measure of the correlation time of exchange and has been used to describe timescale of internal motion.

S^2 values (figure 4.24) for each residue of N-CIRP revealed that the order parameter across the protein is similar for most residues being at around 0.8. This value agrees with the average S^2 calculated from the T_1 vs. T_2 plot contours (0.77 ± 0.08). S^2 was reduced closer to 0 for the residues at the termini. A similar pattern was observed for the τ_e values across N-CIRP, most being around 0.0 and an increase was seen for the termini residues. A small change was seen in the order parameter S^2 in loops between β2-β3 and α2-β4. The change in loop between β2-β3 was also present on the τ_e plot.

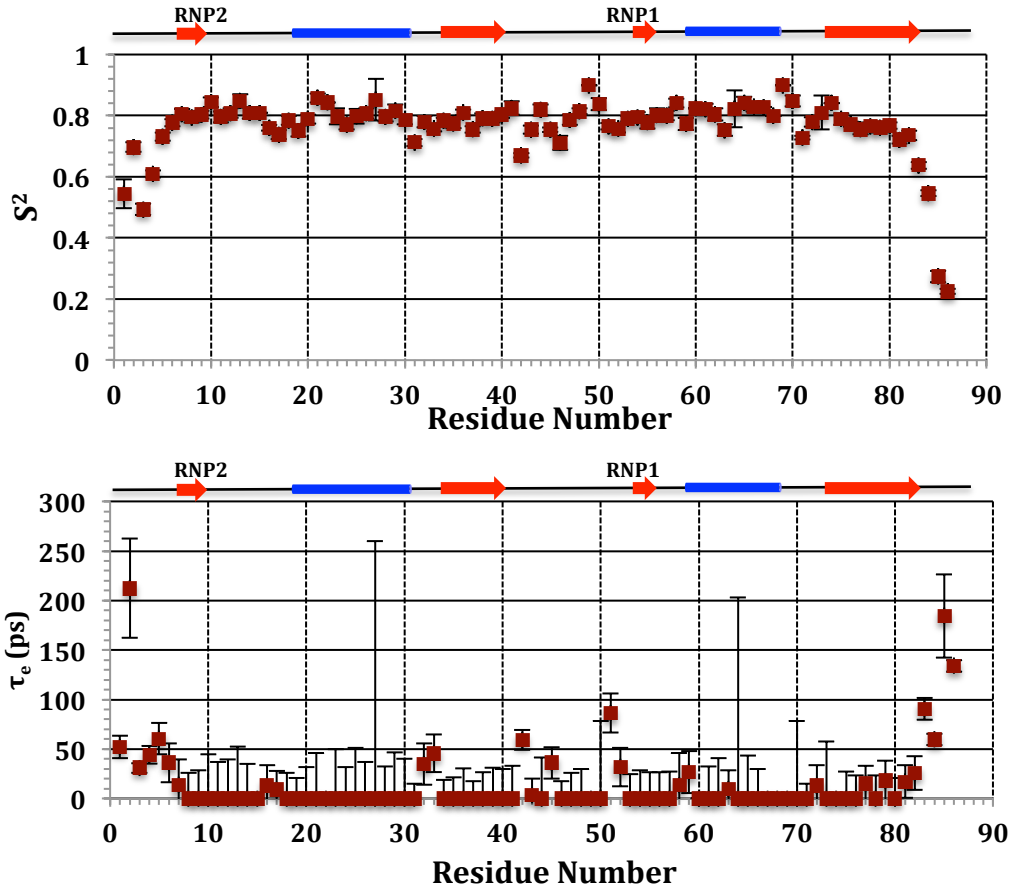

Figure 4.24 Plots of Model Free parameters S^2 and τ_e against N-CIRP residue number.

4.4 Discussion

In this chapter N-CIRP structure and dynamics was investigated to help further our understanding of this RNA binding domain by looking at mobility solution and its fold. The secondary structure prediction using DANGLE confirmed that the protein has topology of β1α1β2β3α2β4 (figure 4.1.3) and in agreement with the PDB structure (1X5S) and also similar to the RRM1 domain of RNA binding protein hnRNPA1 (Barraud & Allain, 2013). For structure calculations, resonance assignment of N-CIRP backbone and side chains were sufficient to run automated structure calculations. The backbone resonances ^{1}H and ^{15}N were completely assigned for each residue and the other components Cα, Cβ, Hα,Hβ and Hγ were all assigned for above 90% of the residues (table 4.3). All assignments can be found in appendix B.

Following the assignment, N-CIRP structures were calculated using an automated procedure. Manual NOE assignment can be a very time consuming process so ARIA software was chosen for the calculation process as this makes automated ambiguous assignment (Linge et al., 2001). NOE manual assignments were also included at later stages (figures 4.17 and 4.18). ARIA was found to operate well, as even after the very initial structure calculation only using ^{13}C and ^{1}H data, structured regions and a general fold of the protein was visible with an RMSD of 2.23 Å over regions with secondary structure. With the addition of extra data from the ^{15}N NOESY spectrum and some manual NOE assignments, the RMSD values did not reduce significantly specifically over the structured regions (table 4.4.). However, the analysis and correction of the violations improved the structure significantly where the final RMSD value was 1.63 Å over the whole protein backbone and 0.89 Å for the regions in secondary structure. Having an RMSD below 1.0 Å is very good at this stage of structure determination before the addition of extra information such as dihedral angles and hydrogen bonds which are what really condense and tighten the structure (Barsukov & Lian, 1993; Cavanagh et al., 2007).

Due to the large variety of studies carried out during this project to investigate N-CIRP function, time was not available to complete the structure with the dihedral angle and hydrogen bond information, which will have to be done in the future. However the final structure achieved here provides information on the topology of the protein and enables the RNA binding sites on β1 and β3 to be located (figure 4.25).

The two RNA binding motifs are on the strands 1 and 3 as predicted by the secondary structure analysis by DANGLE (Cheung, Maguire, Stevens, & Broadhurst, 2010). The phenylalanine residues which were reported to be essential for RNA binding by Maris et al's model (figure 1.8) are shown in figure 4.25 (Maris, Dominguez, & Allain, 2005). The positions of these agree with the model shown in Chapter 1 (section 1.4.2) where the aromatic rings in position 2 of RNP2 (in this case F9) and in position 5 of RNP1 (F51 in this case) are predicted to ring stack. Looking at the position and the close proximity of these two residues this is also the case for N-CIRP (figure 4.25). The residue F49 which is the position 3 of RNP1 is thought to interact with the RNA ribose sugar and phosphate backbone. This agrees with the finding in the previous chapter (3) where the mutation of this residue had the largest effect on RNA binding of N-CIRP as well as inducing a significant structural change (sections 3.3.2 and 3.3.3).

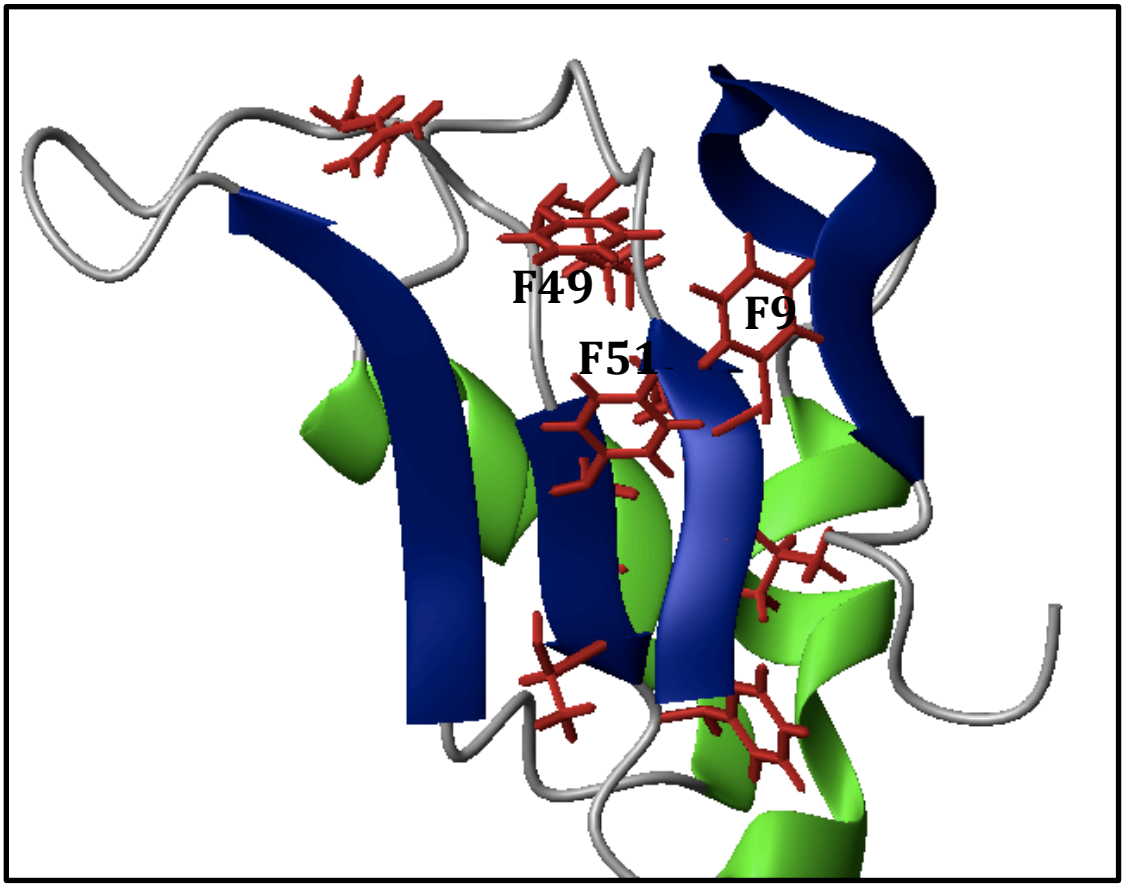

Figure 4.25 Ribbon structure of N-CIRP. RNP1 and RNP2 residues on β1 and β3 are shown in red.

N-CIRP possesses a fold typical of an RNA binding protein. The solved structure of human N-CIRP confirmed this (figure4.26 (b)) and having the typical β1α1β2β3α2β4 (Sheikh, 1997). The DANGLE software predicted mouse N-CIRP also has this topology and the structure determined has a very similar fold to the human protein (figure 4.26 (a)). To confirm the RNA binding fold the N-CIRP structure was compared to the structure of another typical RNA binding protein hnRNPA1 RRM1 (Barraud & Allain, 2013) (figure 4.26 (c)). Both human and mouse N-CIRP have a very similar fold to that of hnRNPA1 RRM1 except the absence of a strand, however, when looked at closely, mouse N-CIRP has a long β4 strand which curves around (figure 4.26 (a)) and hnRNPA1 RRM1 has a β4 strand very closely followed by β5 in a similar position to where the mouse N-CIRP β4 curves around. So it may be that CIRP also posseses an extra strand here that was not picked by DANGLE or even hnRNPA1 RRM1 strands 4 and 5 are meant to be one long strand together.

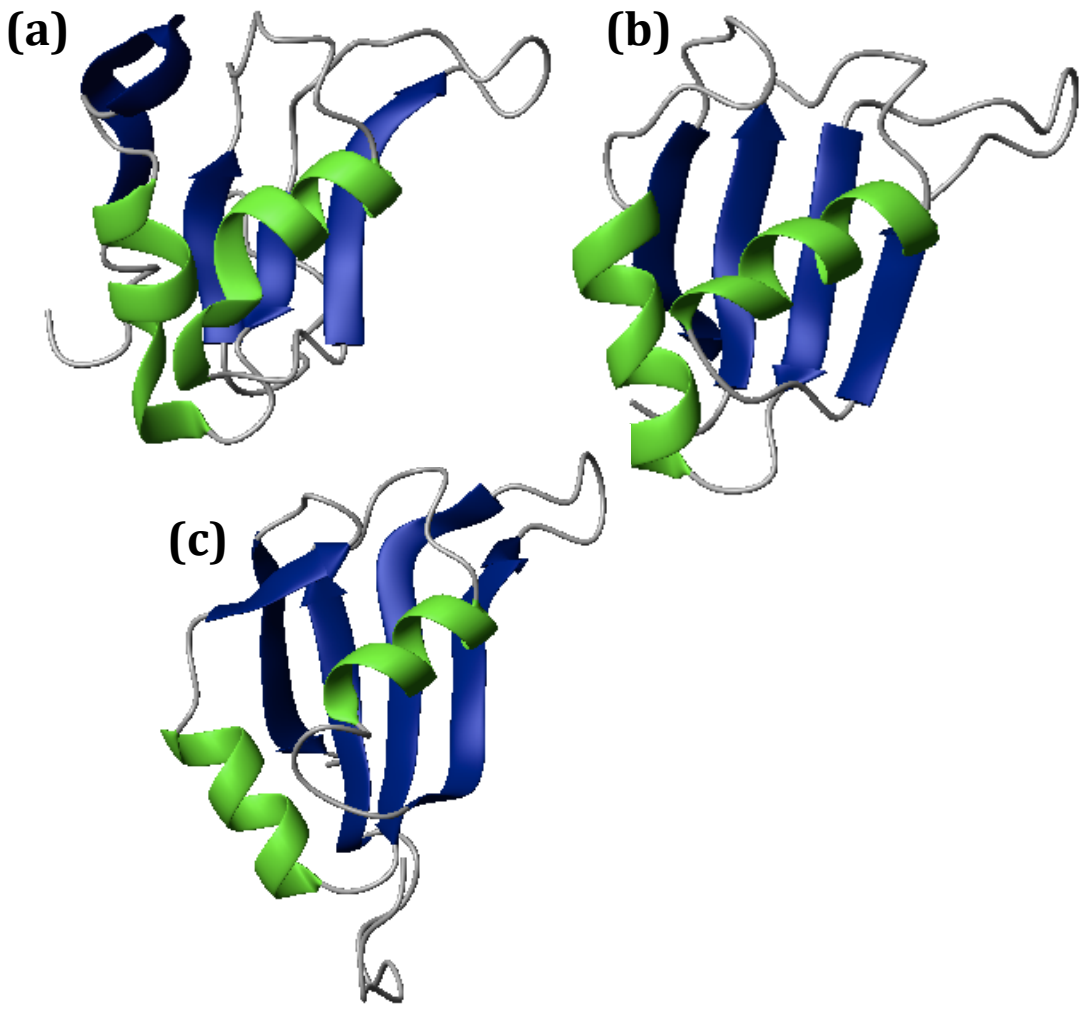

Figure 4.26 Ribbons structures of mouse N-CIRP (the structure closest to the mean) (a), human N-CIRP (PDB- 1X5S) (b) and hnRNPA1 RRM1 (PDB- 2LYV) (c).

The calculated structure was further analysed by superimposing with human N-CIRP as well as the structure of hnRNPA1 RRM1. This process was carried out using the web-based software DaliLite (Holm & Park, 2000). The superimposition between human N-CIRP and mouse N-CIRP gave an RMSD value of 2.6 Å over 75 residues, which were picked by the software as equivalent between the two structures. The RMSD value between mouse N-CIRP and hnRNPA1 RRM1 was 2.5 Å over 72 equivalent residues. The close similarity of the structure to both human N-CIRP and hnRNPA1 RRM1 confirms the RNA binding protein fold of the protein and it accuracy. The superpositions showed that two proteins aligned well with mouse N-CIRP on both cases specifically at the secondary structure regions. There is no doubt that mouse N-CIRP shares the same fold as human N-CIRP and the RMSD value will improve after further refinement has been carried out.

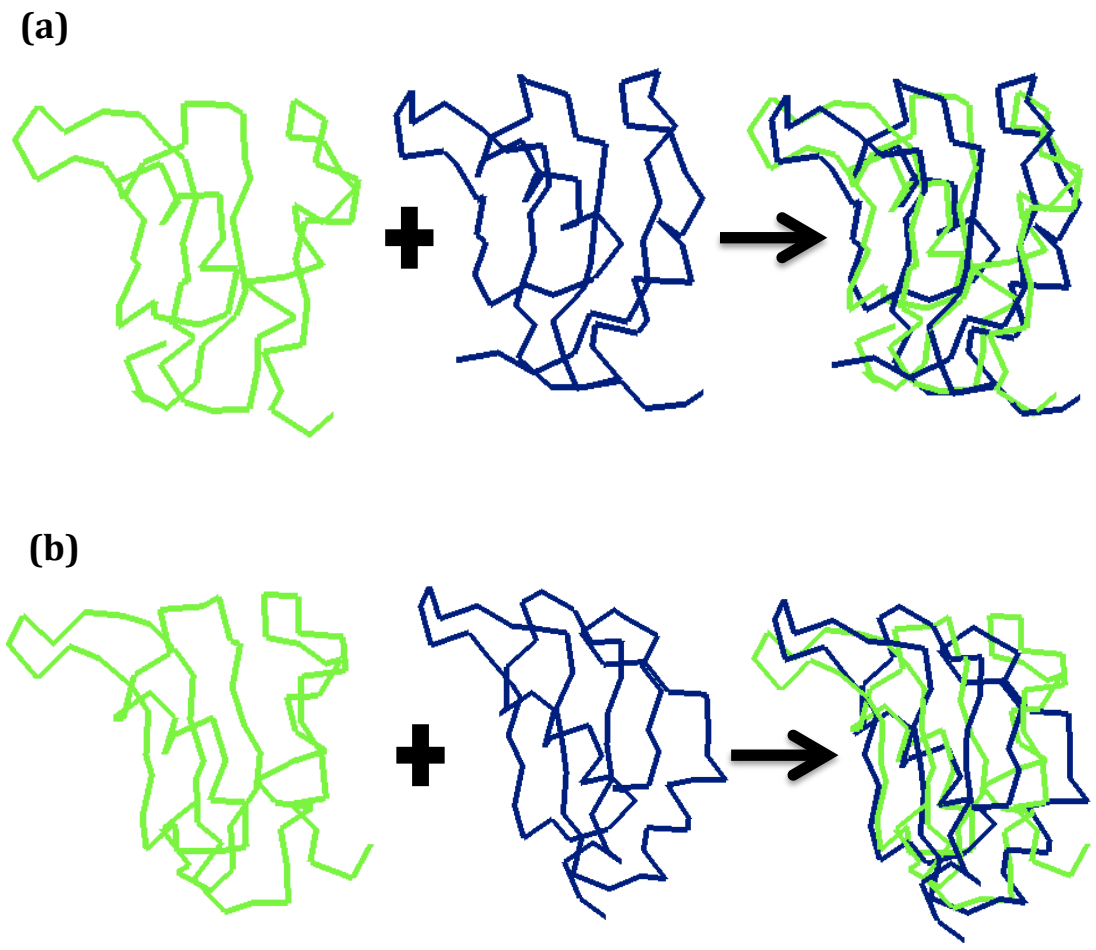

Figure 4.27 The superimposition of mouse N-CIRP structure with human N-CIRP or hnRNPA1 RRM1 achieved using DaliLite software (www.ebi.ac.uk/Tools/structure/dalilite). Mouse N-CIRP backbone is shown in green and human N-CIRP and hnRNPA1 RRM1 backbone in navy. (a) mouse N-CIRP and human N-CIRP, (b) mouse N-CIRP and hnRNPA1 RRM1.

The NMR dynamics data gained from the relaxation experiments revealed that N-CIRP is rigid except for some flexibility at both termini. If the protein is rigid and tumbles isotropically, the T_1, T_2 and hnNOE values will be the similar at each site (Kleckner & Foster, 2011). The data points on the T_1 scatter graph revealed that T_1 values were 0.6-0.7 s for most residues (figure 4.22). There were no extreme rises or dips except a small increase for the extreme residues at the N- and C-termini. This is expected as most proteins have increased motion at the termini (Jacob & Unger, 2007). The loop regions can be flexible regions of proteins however this was not observed here. Variation of the T_1 value across protein suggests anisotropy (Palmer, 2001), but the similarity of values across the protein show that N-CIRP is isotropic. The same was found from the T_2 data with most of the residues having a T_2 value of around 0.15 s (figure 4.22). The T_2 time parameter also gives and indication of the size of protein but also reduction of T_2 indicates chemical exchange (Palmer, 1997). This further confirms that WT N-CIRP is isotropic and not involved in any chemical exchange.

The hnNOE becomes more negative as picosecond motion increases (Brüschweiler, 2003; Palmer, 2001). hnNOE values across most residues were approximately -0.2, this became more negative for residues in the termini (figure 4.22). Termini residues are expected to have elevated motion, but the rest of N-CIRP displays little picosecond motion.

The isotropic nature of N-CIRP was confirmed from the T_1 vs T_2 scatter plot that provided concentrated focus of points (figure 4.23). In this plot, the termini residue points were more dispersed but this is as expected as it reflects faster motion in these regions. Dipolar relaxation contour lines for τ_m and S^2 were plotted on the T_1 vs. T_2 plot to estimate the S^2 and τ_m values for N-CIRP. As the majority of data points concentrate in one region of the plot, this suggests most residues have a similar S^2 and τ_m values i.e. the protein has a similar motion across all regions. The average S^2 value was estimated to be 0.77 ±0.08 from this plot and indicates that the protein backbone is fairly rigid with little internal motion. A value of 6.8±0.21 ns was estimated from the plot for the τ_m, this is a typical tumbling time of a protein of such molecular

weight (11.6 kDa), as the estimated value calculated using the Daragon & Mayo equation (Daragan et al., 1997) (equation 4.7) was 6.3 ns.

To further describe dynamic motion in N-CIRP, ModelFree analysis was used. This uses the T_1, T_2 and hnNOE times to produce S^2 order parameters for each residue and τ_e internal motion correlation times without the need of a structural model. T_1, T_2 and hnNOE values were imported into ModelFree to produce S^2 and τ_e values, which were plotted against residue number (figure 4.24). The S^2 values were similar for most residues with the exception of the residues at the N- and C-terminus, which have a lower order, as they are more flexible. The residues in the loop regions between sheets β2 and β3 and also between α2 and β4 showed a small change of 0.05 to 0.1 in the order parameter suggesting these regions have some increased motion and are less rigid compared to the rest of the protein. The change in S^2 on the β2-β3 loop is interesting as this the region just before the RNP1 sequence including F49. This agrees with the findings of Chapter 3 where the mutation of this residue was found to have an effect on N-CIRP structure (section 3.3.2). This region appears more dynamic on the pico to millisecond time scale compared to the rest of the protein. The remaining residues had a value of around 0.8, which agrees with the value calculated from the T_1 vs. T_2 plot suggesting a significant degree of backbone rigidity with little motion that is common with a compact structure. τ_e values for most residues were also very small, with some motion in the same loop regions, agreeing with the rest of the calculated parameters. N-CIRP is a rigid protein with the exception of some motion at the termini. The loop regions are expected to be more flexible and also the RNA binding sites of RNP1 and RNP2 at β1 and β3 may have been expected to have more motion but this was not observed at the fast timescales that were analysed (picoseconds to milliseconds) and this may have been due to the absence of a ligand. The mobility of the residues at the termini was as expected as these regions are often more flexible in a protein due to fewer structural constraints (Jacob & Unger, 2007).

In conclusion, this chapter gave an insight to the structure and dynamic properties of N-CIRP. N-CIRP adopts the expected RNA binding fold as seen in the human protein. This structural information can be used to make assumptions as to the RNA binding mechanisms that N-CIRP follows by utilising data already known about proteins which also possess this typical RNA binding protein structure. In the absence

of a ligand, N-CIRP does not display much motion at the picosecond to millisecond timescale and is relatively rigid with the exception of the N- and C-terminal residues as well as the protein being isotropic and mouse N-CIRP has typical RNA a binding protein structure very similar to human N-CIRP.

Chapter 5

Functional Studies of CIRP in Mammalian Cells

5.1 Introduction

As described in the introduction chapter of this thesis, expression of CIRP (Cold-Inducible-RNA-Binding-Protein) is up-regulated in mammalian cells upon perception of mild cold-shock (27-32°C) (Nishiyama, Higashitsuji, et al., 1997). Culturing of cells at these temperatures can also result in improved recombinant protein yields, improved protein folding and activity (Al-Fageeh, Marchant, Carden, & Smales, 2005; Masterton, Roobol, Al-Fageeh, Carden, & Smales, 2010). CIRP has also been suggested to bind to key proteins involved in the control of mRNA translation initiation, potentially acting as a bridge between the mRNA and the protein synthesis machinery. RNA binds to the RNP1 and RNP2 sequence present on the N-terminal domain of CIRP as described in Chapter 3 (section 3.3.3). CIRP also has an arginine/glycine rich C-terminal domain, which is natively disordered with an unknown function (Shu-Ju Hseih 2010). This chapter describes the generation of CIRP constructs to assess the influence of CIRP on recombinant protein production, and to investigate whether modulation of CIRP, N-CIRP and C-CIRP influences mRNA translation/protein synthesis, and to investigate potential binding partners of CIRP. Through such investigations, a better understanding of the function of CIRP and its involvement in mechanisms which upon mild hypothermia lead to improved recombinant protein production can be achieved.

CIRP has been reported to bind the 3'UTR of particular mRNAs resulting in an increase in their translation (Yang & Carrier, 2001; Yang et al., 2010). One of these is the 3'UTR of the ATR protein (ataxia telangiectasia mutated and Rad3 related kinase) and it has also been shown that overexpression of CIRP leads to increased ATR protein levels (Yang et al., 2010). ATR belongs to a family of protein serine threonine kinases and the ATR signalling pathway is associated with cell cycle arrest due to compromised DNA replication (Abraham, 2001). This mechanism is described in more detail in chapter 1 (section 1.4.3). It is possible that CIRP may play a role in this cold induced phenotype by, upon its up-regulation, binding to the 3'UTR mRNA of ATR leading to an increase in ATR protein amounts. The cell cycle arrest response to cold shock is thought to be a primary cold-shock response that contributes to increased protein production at such temperatures in mammalian cells (Kumar, Gammell, & Clynes, 2007).

It has been suggested that the binding of CIRP to the 3'UTR of specific mRNA results in an increase in their translation and is mediated by an interaction between CIRP, the mRNA and the translational machinery (Yang, Weber, & Carrier, 2006). This mechanism was supported by the studies of Yang et al who reported that CIRP binds to the eukaryotic initiation factor 4G (eIF4G). From these observations it is hypothesised that through the interaction with 3'UTRs and eIF4G, CIRP acts as a bridge between 3' and 5'UTRs to increase translation. The translation initiation process (figure 1.9) and the role of eIF4G is discussed in detail in chapter 1 (section 1.4.3).

As described in detail in chapter 1, CIRP consists of a well-characterised N-terminal domain and a natively disordered C-terminal domain. A further investigation undertaken here was to express each of the two domains on their own in CHO cells to determine if either has aspects of CIRP function. CIRP constructs of full-length, N-terminal and C-terminal were therefore generated in mammalian cell expression vectors to study the interaction of these regions with RNA and protein and their effect on recombinant protein production in Chinese hamster ovary (CHO) cells.

5.1.1 CIRP Expression in Mammalian Systems

To be able to investigate CIRP function and its effect on recombinant protein production in mammalian expression systems, the industrially relevant CHO (Chinese hamster ovary) cell line was used as a model system. Mammalian cells are used extensively for the production of recombinant proteins because of their ability to correctly fold and assemble polypeptides and proteins and are able to perform human-like post translation modifications (Andersen & Krummen, 2002; Baldi, Hacker, Adam, & Wurm, 2007; Chu & Robinson, 2001; Pham, Kamen, & Durocher, 2006). The dominance of this system is demonstrated by the fact that of the 58 biopharmaceuticals approved between 2006 and 2010, 32 were expressed in mammalian systems (Walsh, 2010). CHO cells are widely used for the production of recombinant proteins as they produce high yields rapidly, with good product quality and have regulatory authority approval (Ye et al., 2009). Nearly 70% of recombinant therapeutic proteins are made in CHO cells and this is expected to be maintained in the next decade (Jayapal, Wlaschin, Hu, & Yap, 2007; Yeon-Gu, 2012). The advantages of using CHO cells in an industrial setting include human-like post

translational modification capacity, low specific productivity can be overcome by gene amplification, CHO cells can produce recombinant protein with glycan forms that are compatible with, and bioactive in, humans and can be easily adapted to grow in the regulatory friendly serum-free suspension conditions for large scale production (Kim, Kim, & Lee, 2012). Recently use of temperature reduction from 37°C to 32-34°C during industrial fermentation to enhance recombinant protein production has been popular. As such, a better understanding of how CHO cells respond to reduced temperature, and whether CIRP induction can be linked to the use of specific 3'-UTR elements to improve recombinant protein production was investigated here.

In order to manipulate the expression of CIRP and the N- and C-terminal domains, the gene sequences for these were cloned into the commercially available pcDNA5/FRT vector to allow their over-expression in CHO cells. The pcDNA5/FRT mammalian expression vector system possesses a human CMV (cytomegalovirus) promoter for high level expression of the gene of interest (Andersson, Davis, Dahlbäck, Jörnvall, & Russell, 1989; Boshart et al., 1985; Nelson, Reynolds-Kohler, & Smith, 1987). It has 10 unique restrictions sites for easy cloning of the gene of interest and ampicillin resistance. The vector also contains a FLP Recombination Target (FRT) site which allows integration of the vectors into a specific site in the FlpIn host cell genome allowing for the generation of stable cell lines when transfected together with the Flp recombinase expressing pOG44 vector and the hygromycin resistance gene allows for the selection of stable cells lines (Gritz & Davies, 1983). The Flp recombinase allows a recombination between the FRT sites (Craig, 1988; O'Gorman, Fox, & Wahl, 1991; Sauer, 1994). When the FRT vector is inserted into the genome at the FRT site, the SV40 promoter and ATG initiation codon is in frame with the hygromycin resistance gene so cells with the FRT vector integrated can be selected (Figure 5.1). This makes the FRT plasmid a good expression vector for the expression of CIRP in CHO cells as it yields high protein expression through the generation of stable cell lines, or conversely, the vector can be used for transient expression whereby the vector is not incorporated into the genome.

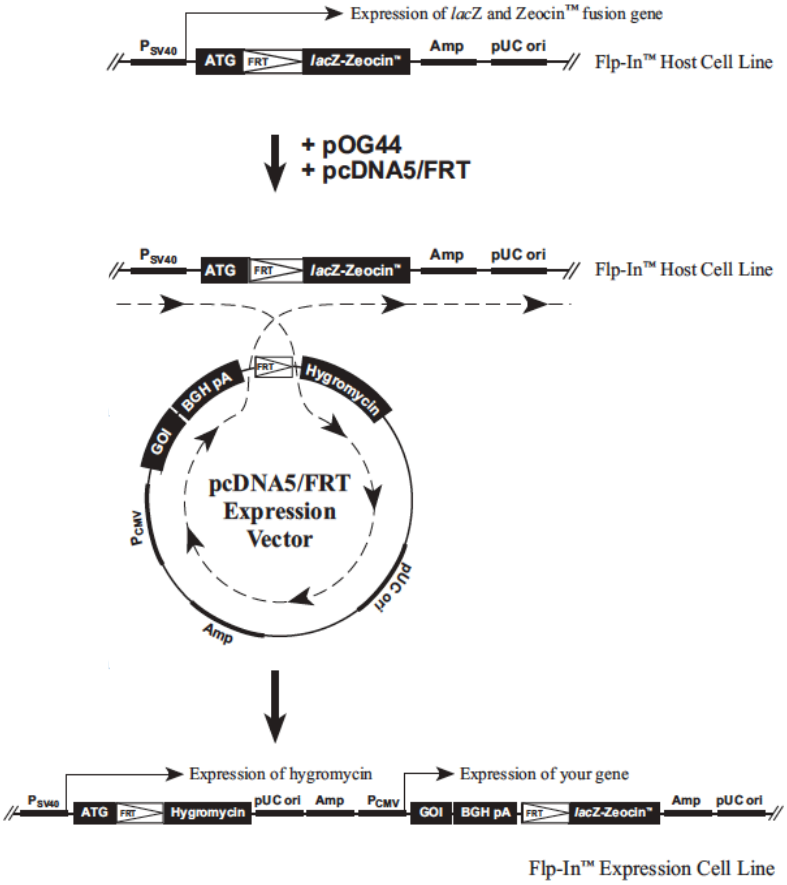

Figure 5.1 Schematic representation of the integration of the commercially available FRT expression plasmid into the host cell genome (Invitrogen Flp-In System Manual).

5.1.2 Protein-Protein Interactions

To study the interaction between eIF4G and CIRP it is necessary to investigate protein-protein interactions. There are various methods of studying protein-protein interactions. Co-immunoprecipitation (Co-IP) is one of the most commonly used methods to determine protein binding partners. This method uses the specific interaction between an antibody and the protein of interest given that the antibody-protein interaction does not disrupt the protein interactions of interest. It is one of the most convincing methods to provide evidence that two or more proteins interact *in vivo* (Monti, Orrù, Pagnozzi, & Pucci, 2005; Ren, Emery, Kaboord, Chang, & Qoronfleh, 2003). Co-IP is performed under conditions that preserves protein-protein interactions even when cells are lysed and protein solubilised. The antibody specific for the protein of interest is incubated with the cell lysate to allow the antibody and protein to form a complex. After the antibody is bound to the protein, the complex can be isolated from the cell lysate using protein A or protein G, which will bind to the

antibody bringing with it any protein that interacts with the target protein (Miernyk & Thelen, 2008). When this procedure is carried out with the appropriate controls, proteins that specifically interact with the protein of interest can be identified successfully. There are many advantages to using this method to study protein interactions; it allows high specificity, it is a relatively easy method to carry out and it is compatible with a number of subsequent analysis protocols (e.g. SDS-PAGE, western blotting) and it allows the study of protein-protein interactions *in vivo* (Miernyk & Thelen, 2008).

As CIRP binding has only been studied by NMR methods *in vitro*, this was a suitable method to study the interaction of CIRP, N-CIRP and C-CIRP *in vivo* with other proteins. This approach was simplified further here by the addition of a V5 tag to the CIRP constructs, as the highly specific V5 antibody could be used to carry out the Co-IP experiments. Importantly, studies in the laboratory at Kent have shown that the V5 antibody does not recognise any CHO cell proteins and hence will only result in the pull down of recombinant CIRP that is V5 tagged. A potential disadvantage of this method is that attaching a tag at the N- or C-terminal end of CIRP, N- and C-CIRP may interfere with protein-protein interactions.

5.1.3 The Effect of CIRP on Recombinant Protein Production

Overexpression of CIRP in mammalian cells has been shown to lead to increased recombinant protein expression (Tan, Lee, Yap, & Wang, 2008). Further, reduced temperature culturing of mammalian cells results in the up-regulation of CIRP expression and this is associated with enhanced recombinant protein production and improved protein quality (Al-Fageeh & Smales, 2006). Here, whether over-expression of CIRP and its interactions with eIF4G and specific mRNAs can be used/manipulated to further enhance recombinant protein production was investigated.

In order to determine the effect on recombinant protein production, there are a number of approaches to determine the amounts of recombinant protein expressed by the cell. This can be done by transfecting the cells with a reporter gene and determining its expression levels after the desired manipulation(s) have been made. Detection of protein expression in cells through reporter bioluminescence is a commonly used

technology and often utilises recombinant luciferase expression. Luciferases are oxidative enzymes that catalyse the oxidation of a substrate, luciferin, to form oxyluciferin with the emission of a photon (Thorne, Inglese, & Auld, 2010). The emission of the photon can be measured and quantified which is directly correlated to the amount of luciferase protein expressed by the cell. Therefore, this is a good method for studying how manipulations to the cellular machinery effect the amount of recombinant protein that is expressed, in this case the effect of CIRP manipulation on recombinant luciferase expression. Further advantages of using luciferase are (1) measurement of luminescence does not require excitation, which lowers the background signal resulting in a sensitive assay unlike other methods such as fluorescence (Simeonov et al., 2008), (2) measurement of dynamic changes in reporter expression can be determined (Corish & Tyler-Smith, 1999), and (3) the detection is simple using commercially available reagents containing the substrate and changes in protein expression can be detected by making comparisons to the appropriate controls.

Two commonly used luciferase assays are firefly and gaussia luciferase. Firefly is an intracellular protein, which requires the lysis of cells before the detection reagent can be added and gaussia is a secreted protein and can be detected in the culture media. Both these assays were used in this project as described below in the appropriate sections.

5.2 Materials and Methods

5.2.1 Cloning

For the expression of CIRP, N-CIRP and C-CIRP in mammalian systems, the gene sequences were cloned into the pcDNA5/FRT plasmid using the pcDNA3.1 plasmid as a template. The appropriate gene sequences were initially cloned into the pcDNA3.1 plasmid to add the V5 tag onto the gene sequence. The sequences were then sub-cloned from the pcDNA3.1 vector into the pcDNA5/FRT vector. The template DNA used for the cloning was the pet28a(+) vector containing the CIRP gene as described in Chapter 2 (section 2.1.1). CIRP constructs created in the pcDNA5/FRT plasmid are listed below (figure 5.2).

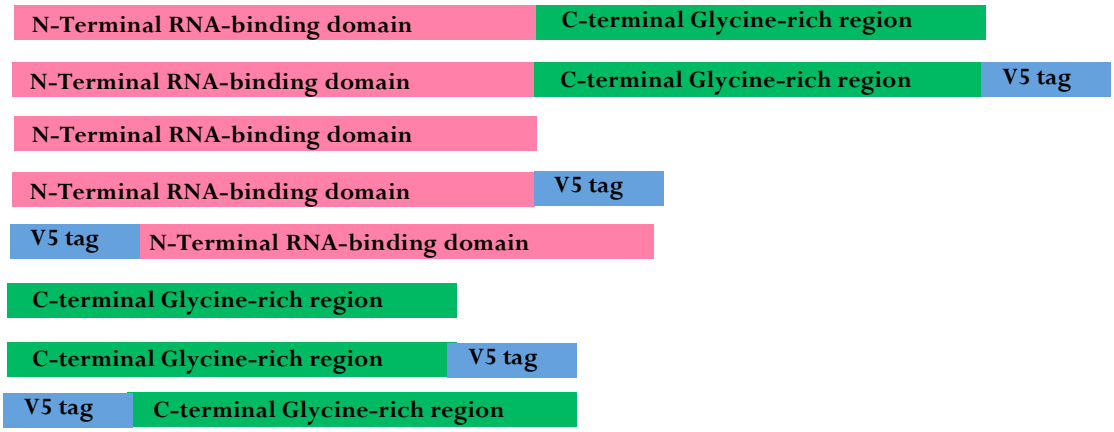

Figure 5.2 Schematic diagram of CIRP constructs generated with and without N/C-terminal V5 tag. N-terminal domain is described as amino acids 1-86 and the C-terminal domain 87-172. These are shown in figure 2.3 in Chapter 2.

To create the CIRP constructs with an N-terminal V5 tag, the region of the vector containing the V5 tag followed by a His_6 tag was amplified by polymerase chain reaction using the V5HIS Forward and Reverse primers (Table 5.1) and pcDNA3.1 plasmid DNA as template. The amplification region was from the Xba1 site to the end of the His tag but did not include the stop codon. An Nhe1 site and start codon was added to the 5' and a BamH1 site to the 3' end to aid with ligation into the FRT vector (figure 5.3). The V5+His tag fragment was then cloned into the pcDNA5/FRT at the BamH1 and Nhe1 sites making a construct with an N-terminal V5 tag which the CIRP genes could be cloned into.

For the FRT plasmid containing the CIRP truncation with an N-terminal V5 tag, pet28a(+) CIRP vector was used as a template to amplify CIRP, N-CIRP and C-CIRP. The primers were designed to add a BamH1 site at 5' and Xho1 site and a stop codon at the 3' end. The restriction sites added were used to ligate the DNA fragments amplified by PCR into the pcDNA5/FRT plasmid with the N-terminal V5 tag described above following restriction digest. To make the FRT plasmid construct expressing CIRP truncations with no tag, the DNA fragments were ligated into the original pcDNA5/FRT plasmid as the primers were designed to add a start codon to each fragment. To amplify full length CIRP, CIRP Forward and CIRP Reverse+STOP primers were used. For the amplification of N-CIRP, CIRP Forward and N-CIRP Reverse+STOP primers were used. C-CIRP was amplified using C-CIRP Forward and CIRP Reverse+STOP primers (Table 5.1).

To make the plasmid expressing CIRP truncations with a C–terminal V5 tag, the DNA fragments were initially sub cloned into the pcDNA3.1 plasmid. The primers used to amplify the CIRP fragments were the same as described above except the reverse primers without STOP were used to allow expression of the C-terminal V5 tag. The forward primers were also designed to include a Pme1 site at the 5' end as well as the start codon and the BamH1 site. The PCR amplified CIRP, N-CIRP and C-CIRP fragments were digested and ligated into the pcDNA3.1 plasmid using the BamH1 and Xho1 sites. The ligated fragments were then digested out of the plasmid using the Pme1 enzyme removing the V5 and His tag with it. Finally, the digested fragments were ligated into the FRT plasmid using the two Pme1 sites. The orientation of the insert was confirmed/checked by sequencing.

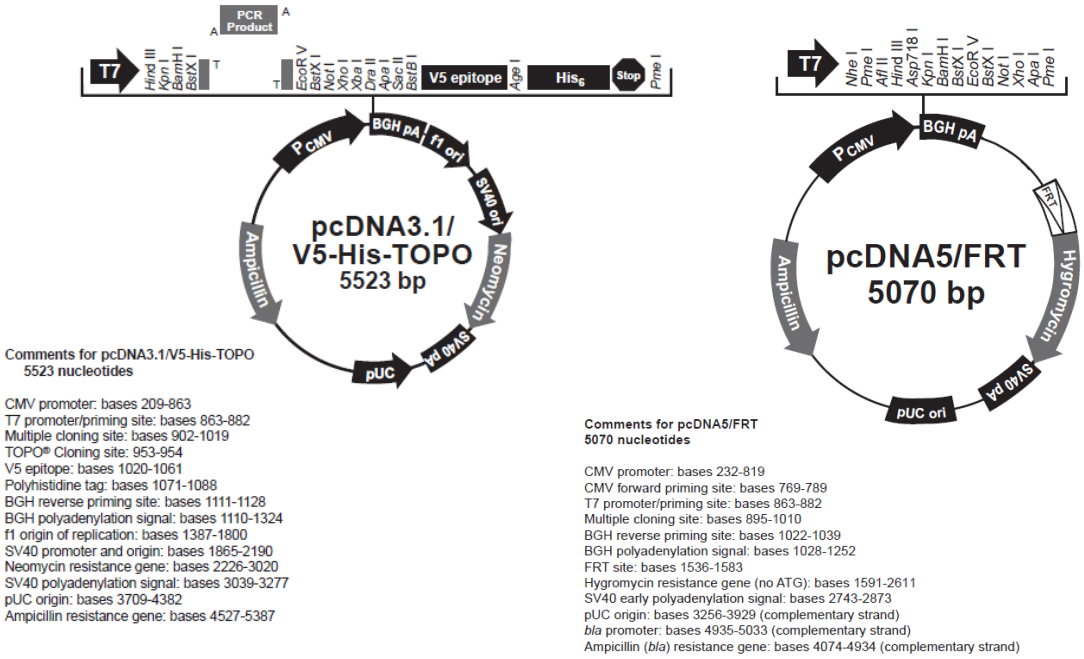

Figure 5.3 Schematic representations of the pcDNA3.1 and pcDNa5/FRT vector maps and multiple cloning sites. (Invitrogen)

Primer Name	Sequence (5'>3')	T_m (°C)
CIRP Forward	TATGGATCCGTTTAAACATGGCTAGCGATGAAGCAAGCTT	71.4
CIRP Reverse	ATACTCGAGCTCTCGTTGTGTGTAGCATAACT	66.9
CIRP Reverse+STOP	ATACTCGAGTTACTCGTTGTGTGTAGCATAACTGT	67.1
C-CIRP Forward	TATGGATCCGTTTAAACATGGACAACCGGTCCCGA	70.6
N-CIRP Reverse	ATACTCGAGCTAGAAGACTTGCCAGCCTGGT	69.5
N-CIRP Reverse+STOP	ATACTCGAGTTAAGAAGACTTGCCAGCCTGGT	68.2
V5HIS Forward	TATGCTAGCAGTCTAGAGGGCCCGCGGTTCGAAATGGGTAAGCCTAT	75.0
V5HIS Reverse	ATAGGATCCGATGGTGATGGTGATGATGACCGGTACGCGTAGAATCGA	75.0

Table 5.1 Primer sequences used for the amplification of CIRP and V5 epitope DNA fragments for the generation of constructs in the pcDNA5/FRT plasmid.

5.2.2 Polymerase Chain Reaction

The pet28a(+) CIRP template DNA to be used was generated in large quantities by transforming into DH5α *E.coli* cells and following the QIAGEN midiprep kit protocols. For the PCR reactions, Pfu DNA polymerase (Promega, 3U/µl) was used as described in Chapter 2 (section 2.2.2). The primer pairs and DNA template used are described above (5.2.1). PCR products were analysed on 2% agarose gels and visualised with ethidium bromide to confirm that the correct sized product was

present. PCR products were purified using a PCR clean up (QIAGEN, QIAquick PCR Purification Kit) or gel extraction kit (QIAGEN, QIAquick Gel Extraction Kit) following the manufacturers protocols.

5.2.3 DNA Gel Electrophoresis

This was undertaken as described in Chapter 2, section 2.2.3.

5.2.4 Restriction Enzyme Digestion

To isolate or to give the correct sticky ends to the CIRP, N-CIRP and C-CIRP DNA fragments, restriction enzyme digestion was used. PCR products or pcDNA3.1 plasmid with the correct inserts, were taken forward and the DNA was digested using the enzymes described above using the optimal buffer for the two enzymes (section 5.2.1). The digest procedure is described in Chapter 2, section 2.2.6.

5.2.5 Ligation of PCR products

The PCR products and plasmids were digested with the appropriate enzymes to give the correct sticky ends as described in section 5.2.1. DNA was quantified using a Nanodrop instrument. Various ratios of insert:vector were tested; 5:1, 2:1, 1:1 with 50 ng vector for each ligation in 10 µl reaction volume. The amount of insert DNA to be added was calculated using the equation 5.1 below.

Equation 5.1:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert: vector molar ratio} = \text{ng of insert}$$

1 µl of T4 DNA ligase (Promega, 3 U/µl) was added per reaction and 1µl of 10x ligase buffer. The samples were incubated for 1.5-2 h at room temperature. After incubation, 3-4 µl of the ligation mix was transformed into competent DH5α cells. Negative (no insert) and positive (1 µl of plasmid DNA) controls were also included. 100 µl of transformed cells were spread on LB agar plates containing 100 µg/ml

ampicillin and incubated overnight at 37°C. Colonies were checked by PCR screening or restriction enzyme digest before being sent for sequencing.

5.2.6 PCR Screening of Colonies

Was undertaken as described in Chapter 2, section 2.2.5.

5.2.7 Competent Cells

Were prepared as described in Chapter 2, section 2.2.8.

5.2.8 Transformation

For DNA purification, competent DH5α cells were transformed with pcDNA3.1 or pcDNA5/FRT vectors with or without the CIRP genes cloned in. For the transformation, 30 µl aliquots of competent cells in microfuge tubes were thawed on ice and the transformation procedure is described in Chapter, section 2.2.9 followed with the exception that transformed cells were spread on to LB agar plates containing 100 µg/ml ampicillin (Sigma).

5.2.9 Creating Stable CHOK1 Cell Lines Expressing Recombinant Mouse CIRP, N-CIRP and C-CIRP With the Invitrogen Flp-In System

Adherent Flp-In CHOK1 cells were maintained in Ham's F12 media (Invitrogen) containing 10% FBS and 1/1000 Zeocin selection agent (Invitrogen) at 37°C. 24 hours before transfection, $5x10^5$ cells were seeded into T25 flasks in 6.25 ml of medium without Zeocin. To transfect the cells, 9 µg of pOG44 Flp-Recombinase expression vector (Invitrogen) plus 1 µg of pcDNA5/FRT vector containing the gene of interest were diluted in 625 µl Opti-mem media (Invitrogen). 30 µl of Lipofectamine 2000 reagent (Invitrogen) was also diluted in 625 µl Opti-mem. The two mixes were incubated for 5 min and then mixed together. After a 20 min incubation of the mix, this was added to the cells in the T25 flask seeded with cells the previous day. The cells and transfection reagent were then incubated for five and a half hours at 37°C before the cells were washed with PBS. The media was then replaced with fresh medium without Zeocin for 24 hours. 24 hours after transfection the cells were washed again with PBS and the media changed. 48 hours after transfection, cells were washed with PBS and trypsinised with 200 µl trypsin-EDTA

(Invitrogen, 0.05%). The detached cells were taken up in 20 ml Ham's F12 (10% FBS) medium containing 500 µg/ml hygromycin B (Invitrogen). Cells were then plated out in 6 well plates with 2 ml of cell suspension per-well. Media containing hygromycin was then used for the remainder of the process. Medium in the wells was changed every third day.

After 7 to 10 days, colonies of cells were visible and clonal selection was carried out once clones were visible by eye (10-12 days). Clonal selection was achieved by removing the medium and PBS washing the cells. A cloning tube was then firmly pressed down around the clones to be removed and 50 µl of trypsin-EDTA added. After a 3 min incubation at 37°C, the detached cells were removed by aspirating up and down with a Gilson pipette and transferred into a 6 well plate containing 2 ml of medium. 3 clones were selected for each cell line alongside generation of a polyclonal population. The media in the wells containing clones was changed every third day until there were sufficient numbers of cells to be transferred to a T25 flask. Once 70% confluent, the cells were harvested and frozen stocks and lysates were made for further analysis.

5.2.10 Growth Curve Profiles of Stably Expressing CIRP FRT CHO Cells

Growth curves were established for the stable FRT CHO cell lines generated as described above (section 5.2.1). To record growth over time an xCELLigence system (RTCA DP, Acea) connected to a PC with the RTCA control software was used which allows real time cell based assays. The instrument was placed in an incubator, and the automated system recorded cell index readings at given intervals. To record growth, 5000 viable cells were seeded per well in a total volume of 200 µl in 16 well plates (E-Plate 16, Acea). The instrument was programmed to take readings every 30 min. Cell index readings were collected until the decline during the death phase plateaued. Growth data for each cell line was collected at both 37°C and 32°C.

5.2.11 Transient Transfection and Expression of CIRP in CHO Cells

For transient transfection of cells in 24 well plates, $6x10^4$ viable cells were seeded per well in 500 µl of media and incubated at 37°C for 24 hours before transfecting. The commercially available transfection reagent Lipofectamine 2000 was used for all

transfections. For 1 transfection (i.e. 1 well), in a 24 well plate, 500 ng of DNA was diluted in 50 µl of Opti-mem and 2 µl of Lipofectamine reagent was diluted in 50 µl of Opti-mem. The two tubes were incubated for 5 min before mixing the two together and incubating for 20 min at room temperature. During the incubation time, the cells to be transfected were PBS washed and fresh media was added to each well. The 100 µl transfection mix was then added to each well drop-wise. For co-transfections with more than one plasmid the same method was followed except 400 ng of each plasmid DNA was used. For transfection in 6 well plates, cells were seeded at $4x10^5$ viable cells per well in 2 ml of media and incubated at 37°C for 24 hours before transfecting. For each well 2.5 µg of DNA was diluted in 150 µl and 5 µl of Lipofectamine 2000 reagent was diluted in 150 µl of Opti-mem. The 300 µl transfection mix was added to each well drop-wise. For co-transfections the same method was followed except 2 µg of each DNA was added to the transfection mix. Transfection of HEK cells was as for CHO cells. The plasmid expressing 3'ATR-UTR-luciferase and all eIF4G constructs were kindly provided by Jo Roobol.

5.2.12 Collection and Lysis of CHO and HEK Cells

Cells were harvested by removing the media and PBS washing the cells briefly. Once the PBS was removed, ice-cold lysis buffer (20 mM HEPES-NaOH pH 7.2, 10 mM NaCl, 10 mM Na β glycerophosphate, 0.5 % TX100) was added. Immediately before using the lysis buffer 50 µl/ml of 1 M NaF and 5 µl/ml of 200 mM activated Na_3VO_4 was added to the lysis buffer to inhibit protein phosphatases. At the same time, 2 µl/ml of 5 mg/ml leupeptin, 2 µl/ml of 1 mg/ml pepstatin and 2 µl/ml of 100 mM PMSF were added to the lysis buffer before use as protease inhibitors. The volume of lysis buffer used to harvest cells was experiment dependent as specified in sections 5.2.15 and 5.2.18. After the addition of the lysis buffer with inhibitors, cells were removed by scraping on ice.

5.2.13 Determination of Protein Concentration in Cell Lysates

The concentration of protein in cell lysates was determined using the method of Bradford (Hammond & Kruger, 1988). Samples were prepared by diluting 5 µl of cell lysate in 50 µl of water. To this, 1 ml of Bradford reagent (0.01% Coomasie Blue G250, 15% ethanol, 10% phosphoric acid) was added, mixed and incubated at room

temperature for 10 min. The absorbance of samples was then read at 595 nm using water as blank and a standard curve was generated using known concentrations of BSA. Unknown protein concentrations were then calculated from the standard curve.

5.2.14 Immunostaining of CHO Cells

For immunostaining experiments, cells were seeded into 24 well plates containing 13 mm glass coverslips. Cells were incubated at 37°C for 48 hours or at 32°C for 48 hours or at 37°C for 24 hours followed by 32°C for 24 hours after transfection. CHO cells stably and transiently expressing CIRP were subjected to immunostaining for CIRP expression. For transient expressing cells, $4x10^4$ viable cells were seeded 24 hours before transfection. Transient transfection was then undertaken as described in section 5.2.11. For stable expression, $3x10^4$ viable cells were seeded for incubation at 37°C for 48 hours whilst $5x10^4$ viable cells were seeded for incubation at 37°C for 24 hours after seeding followed by a shift to 32°C for 24 hours. At the end of the incubation period, the media was aspirated off and cells were rinsed with 2 ml/well of PBS at 37°C. Cells were then fixed and permeabilised with 1 ml/well of methanol at -20°C for 5 min. The permeabilising agent was then aspirated off and cells rinsed for 5 min in PBS (2 ml/well). A 3% BSA in PBS blocking solution was then added for 30 min at room temperature.

Droplets of primary antibody (20 µl) were prepared by diluting the antibody in 3% BSA in PBS on Nescofilm in a moist box. The primary antibodies used were a monoclonal anti-CIRP antibody produced in rabbit (Protein-Tech Group) (1 in 100 dilution) and a monoclonal anti-V5 antibody produced in mouse (1 in 500 dilution) (Sigma V8012). The coverslips were removed from the wells using forceps and placed on the antibody droplets cell side down. The coverslips were then incubated overnight at 4°C and then rinsed on 5 successive drops of 0.1% Tween PBS. The coverslips were then placed on 30 µl of secondary antibody droplets diluted 1 in 100 in 3% BSA in PBS and incubated in the dark at room temperature for 2 hours. The secondary antibodies used were anti-rabbit IgG, TRITC (Sigma, T2526) and anti-mouse IgG, FITC (Sigma, F2012). The coverslips were then rinsed on 5 successive drops of 0.1% Tween PBS. To stain nuclei, coverslips were incubated on 30 µl droplets of DAPI for 1 min and then rinsed on 5 successive drops of 0.1% Tween

PBS. Finally, coverslips were mounted onto microscope slides by placing cell side down onto 10 µl Mowiol+antifade reagent (1,4-phenylenediamine dihydrochloride, Sigma) and left to cure overnight at 4°C in the dark. Once set, the sides of the coverslips were sealed with clear nail varnish and tops of the coverslips were cleaned with a Q-tip dipped in water before visualizing.

5.2.15 Co-Immunoprecipitation Experiments

For co-immunoprecipitation experiments from CHO/HEK cells grown in 6-well plates, cells were harvested by adding 100 µl of lysis buffer (section 5.2.12) into each well and scraped on ice. Lysates were then centrifuged at 13000 rpm for 10 min in a bench top centrifuge at 4°C. The supernatants were then used for IP experiments. To ensure that the binding capacity of the beads were not exceeded, Bradford assays were carried out on each sample to ensure the protein concentration was in the 0.4-10 mg/ml range.

For immunoprecipitation, 2-3 µg of the appropriate antibody (specified in the results section 5.3.3) was added into the lysates and incubated for 3-4 hours at 4°C on a rocking platform. 40 µl of a 1 in 4 suspension of Protein A Sepharose beads (Sigma) (i.e. 10 µl packed beads diluted in 40 µl) was added to each sample and incubated overnight at 4°C on a rocking platform. The beads were then pelleted by centrifuging for 1 min at 13000 rpm on a bench top microfuge at room temperature and the supernatant removed. The beads were then washed 3 times by adding 100 µl lysis buffer and centrifuging at 13000 rpm for 1 min before removing the supernatant. After the last wash, 20 µl of 2xSDS PAGE sample buffer (section 2.2.19) was added to the beads and boiled for 3 min to disassociate the protein from the beads. The beads were then sedimented by centrifuging for 1 min at 13000 rpm and the sample loaded on to, and run on, a 12.5% SDS-PAGE gel as described in section 2.2.19. Gels were then transferred onto nitrocellulose membranes for western blot analysis (section 5.2.16).

5.2.16 Western Blot Analysis

Cell lysate samples generated from harvest samples or after immunoprecipation were run on SDS-PAGE gels as described in section 2.2.19. Proteins were then transferred

onto nitrocellulose membranes for detection using antibodies (specified in section 5.3.2 and 5.3.3). All components for transfer were pre-soaked in transfer buffer (100 mM Tris, 12.5 mM glycine, 0.1% SDS) prior to assembly. The gels were then transferred for 1 hour at 750 milliamps at 4°C. Blots were removed from the transfer tank and placed in a staining box with the protein side up. The blots were washed for 5 min in TBS buffer (10 mM Tris, 140 mM sodium chloride, pH 7.5) and then placed in block solution (25 ml 5% marvel in TBS per blot) for 30 min at room temperature on a rocking platform. The block solution was then discarded, blots briefly washed in TBS and placed in 8 ml of the appropriate 1° antibody dilution (specified in result sections 5.3.2 and 5.3.3) in 3% BSA (bovine serum albumin) in TBS. Blots were then incubated on a rocking platform at 4°C overnight. The 1° antibody was then removed and all subsequent washes carried out at room temperature. Blots were washed under stringent wash conditions of high salt (500 mM NaCl in 0.1% Tween-TBS) for 10 min, followed by three 10 min washes in 0.1% Tween-TBS. The blots were then placed in 8 ml of horseradish peroxidase conjugated secondary antibody against the species which the 1° antibody was raised in, diluted 1/1000 in block solution. Blots were incubated in secondary antibody solution for 1 hour on a rocking platform. Finally, blots were washed in stringent wash for 10 min, followed by 4x5 min washes in 0.1% Tween-TBS. Peroxidase activity of the bound secondary antibody was then detected by ECL (enhanced chemiluminescense). ECL reagent (1 ml) was added to each blot and incubated for 5 min. ECL signal was detected by exposing the blot to film for up to 10 min in an x-ray film cassette in the dark and developing the film in an automated film developer.

5.2.17 Luciferase Assays

For detecting firefly luciferase activity, cells were initially washed with PBS after removing the media. For cell lysis, two 20x buffers were used (A- 348 mM Na_2HPO_4, 70 mM NaH_2PO_4, pH 7.68; B- 18 mM $CaCl_2$, 70 mM KCl, 18 mM $MgCl_2$, 2740 mM NaCl), each containing protease inhibitors as described in section 5.2.11 whilst 0.1% triton was added to the lysis buffer immediately before harvest. Lysis buffer (100 µl) was added to each well and the cells scraped on ice. Luciferase activity was determined by mixing 75 µl of lysate and 75 µl of the steadylite plus reporter gene assay system (Perkin Elmer) reagent in the wells of a 96 well white opaque microplate

(OptiPlate). Each sample was mixed by pipetting up and down and incubated for 10 min before analysis on a luminometer. For the gaussia luciferase assay, the media was removed from the 24 well plates and centrifuged at 13000 rpm to remove any cell debris. In a 96 well white opaque microplate, 25 µl of the cell media was mixed with 25 µl of the Biolux gaussia luciferase assay reagent (New England BioLabs) and mixed by pipetting. Analysis of light was then undertaken on a luminometer immediately. Bradford assays were performed on all samples and used to normalise the luciferase readings to total protein content from the cells.

5.3 Results

5.3.1 Generation of CIRP, N-CIRP and C-CIRP (± V5 tag) constructs

PCR reactions were carried out with the appropriate primers and templates as described in sections 5.2.1 and 5.2.2 to generate CIRP/ N-CIRP/ C-CIRP with and without a V5 tag at the 5' and 3' end of the gene sequence. The analysis of the resulting PCR products is shown in Figure 5.4 below.

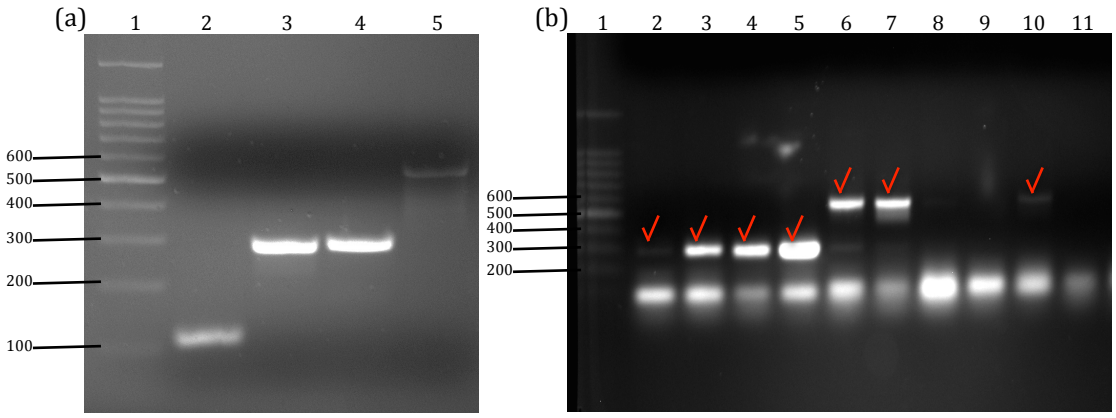

Figure 5.4 2% agarose gel analysis of PCR products (a) PCR products amplified from pcDNA 3.1 plasmid and pet28a(+) plasmid containing the full length CIRP sequence for cloning into the pcDNA FRT plasmid. Lane 1: 100 bp ladder, Lane 2: DNA fragment containing the V5 and His Tag sequences, Lane 3: N-CIRP DNA fragment, Lane 4: C-CIRP DNA fragment, Lane 5: Full-length CIRP DNA fragment. (b) PCR colony screen of DNA from ligation of PCR products in to pcDNA 3.1 plasmid, Lane 1: 100 bp ladder, Lanes 2-5: C-CIRP, Lanes 6-11: Full-length CIRP. Red tick marks indicate correct inserts.

The V5 epitope sequence is a short sequence consisting of only 42 bases. Such a short stretch would be difficult to clone and therefore a longer stretch of DNA upstream from the V5 sequence including the 3' His tag sequence downstream was amplified to generate a fragment with an expected size of 109 bases (figure 5.4, lane 2). Following digestion with the appropriate restriction enzymes, this was successfully ligated into the FRT plasmid between the Nhe1 and BamH1 restriction sites. This plasmid was then used to ligate CIRP fragments downstream of the V5 tag to generate an N-terminal V5 tag. The expected fragment size of the N-CIRP, C-CIRP and full-length CIRP PCR products were 286, 283 and 544 respectively. PCR products of the correct sizes were observed on the DNA gels (figure 5.4, (a)). The CIRP fragments with STOP codons were successfully ligated into the empty FRT vector or into the FRT vector with N-terminal V5 tag. The CIRP fragments without the STOP codon were sub-cloned into the pcDNA 3.1 vector to generate a C-terminal V5 tag and this

fragment was then subsequently digested from this donor vector and ligated into the FRT plasmid. Ligation colonies containing the correct inserts throughout the cloning process were identified by PCR colony screen (figure 5.4, (b)). DNA of colonies with the correct insert were sent for sequencing and all sequences confirmed as correct.

5.3.2 Transient Expression of CIRP, N-CIRP and C-CIRP (± V5 tag) Constructs in the FLP-In CHO Host Cell Line

The CIRP constructs generated with a V5 tag were transiently tranfected into CHO FlpIn cells to test for the expression of each protein whilst an empty FRT vector was transfected as a control. Cells lysates were generated 24 h post-transfection and the expression of the target protein determined by western blot analysis of transfected cell lysates and immuno staining of the transfected cells using the anti-V5 antibody.

Initial Western blot analysis for the expression of the CIRP constructs with a V5 tag showed that only bands for full-length CIRP and CIRP with an N-terminal V5 tag were present 24 h post-transfection (figure 5.5, (a)). No bands were observed for N-CIRP or C-CIRP with an N- or C-terminal V5 tag. An increased amount of protein was observed for the full- length CIRP sample that was incubated at 32°C for 18 hours showing the up-regulation of the protein at mild hypothermic temperatures. This is interesting as the vector construct does not contain the CHO 3' or 5'-UTRs and therefore the normal control of CIRP expression via both transcriptional activation and via the 5'-UTR of CIRP (Al-Fageeh & Smales, 2009) are not present. This may reflect the greater stability of proteins at 32°C compared to 37°C. Although it would be expected that endogenous CIRP would be up-regulated at 32°C, this would not be detected by the anti-V5 antibody.

The initial cell lysates were collected from a 24 well plate where 500 ng of DNA was transfected per well and 20 μg of total protein was loaded on to the subsequent SDS-PAGE gel (figure 5.5 (a)). A further blot where the lysate was taken from a 6 well plate where 2.5 μg of DNA was transfected and 30 μg of total protein loaded on to the gel revealed a band for N-CIRP with an N-terminal V5 tag at the expected size (figure 5.5 (b)). These analyses allowed transient expression of CIRP and N-CIRP with an

N-terminal V5 tag to be confirmed but expression of the other constructs, including none of the C-CIRP variants, was not observed.

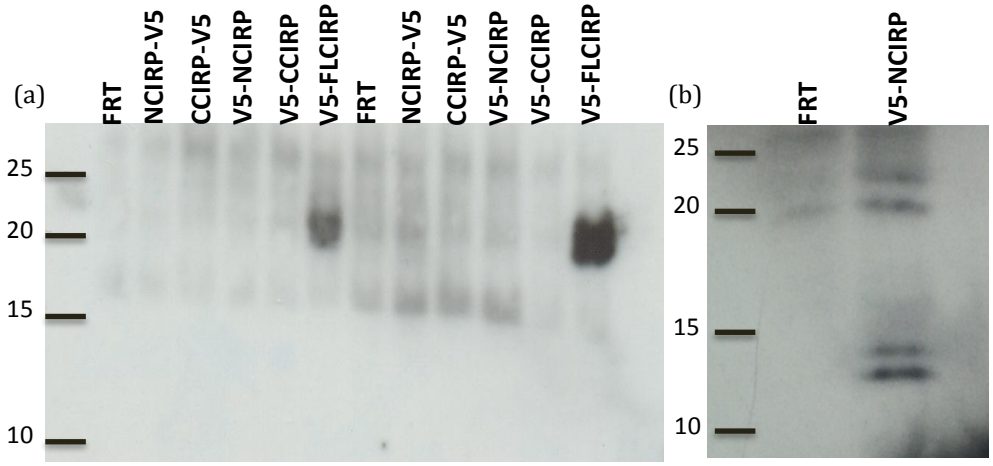

Figure 5.5 Western blot analysis of CHO FlpIn cells transiently transfected with CIRP/N-CIRP/C-CIRP constructs containing a V5 tag on either the N- or C-terminal. Blots were probed with a V5 antibody. (a) Lanes 1-6: cells incubated at 37°C for 24 hours post transfection, lanes 7-12: cells incubated at 37°C for 6 hours, followed by an 18 hour incubation at 32°C post transfection. (b) Cells incubated at 37°C.

To further characterise expression, the CHO FlpIn cells transiently expressing full-length CIRP with an N-terminal V5 tag were subjected to immunostaining. For controls, cells were transfected with blank FRT plasmid. Cells were probed with 1° CIRP or V5 antibodies. For the V5 antibody, fluorescent conjugated FITC (red) and for CIRP, TRITC (green) secondary antibodies were used (figure 5.6).

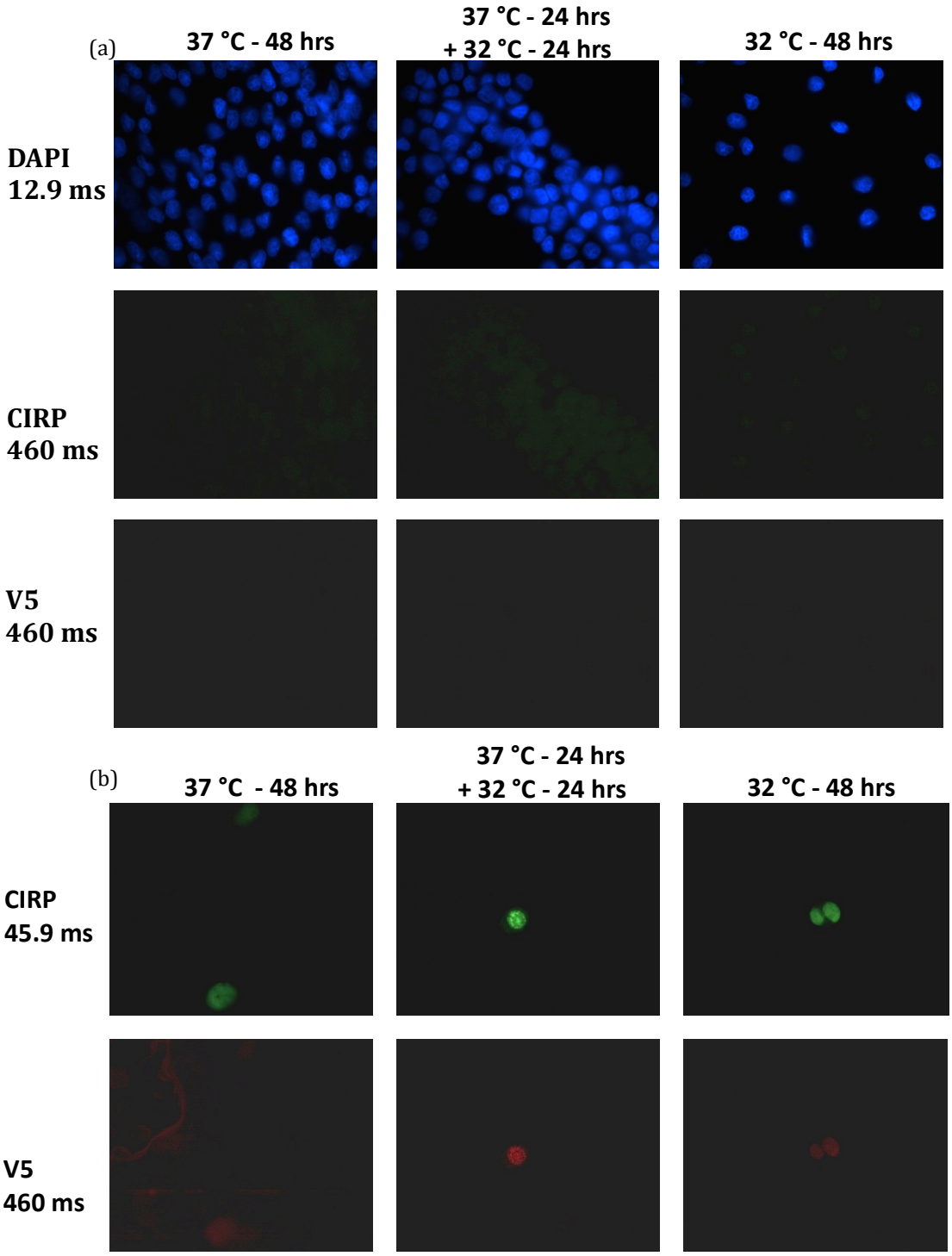

Figure 5.6 Immunofluorescense images collected following immunostaining of CHO FlpIn cells transiently expressing control FRT vector or FRT vector containing full length CIRP with an N-terminal V5 tag incubated at 37 or 32°C following transfection as shown above. (a) Cells expressing control vector (b) Cells expressing full length CIRP with V5 tag.

The control cells transfected with empty FRT plasmid did not show any FITC signal but a small amount of TRITC staining was seen on the cells due to the presence of endogenous CIRP protein in the cells, however the signal for the TRITC fluorescence

was very low (figure 5.6 (a)). The DAPI staining confirmed the presence of many cells, which had been fixed on to the slide. Cells which had been transfected with CIRP showed a signal for both CIRP and V5 staining. Only 1-2 cells per slide showed signal for both CIRP and V5 (figure 5.6, (b)). As the presence of many cells on the slide was confirmed by the DAPI staining, this suggests that the transfection efficiency was low or there was low protein expression in most cells. The reason why expression is observed in so few cells is most likely to be a combination of both these. Cells which had been at 32°C showed higher protein levels compared to the cells incubated at 37°C in the transfected cells.

Another interesting observation made was that if the cells had been incubated only at 32°C following transfection they expressed less protein than if they were first incubated at 37°C followed by temperature shift to 32°C. The temperature shift causes the most up-regulation of CIRP rather than culturing only at colder temperature, this observation was made in multiple fields and can also be seen in figure 5.7 below.

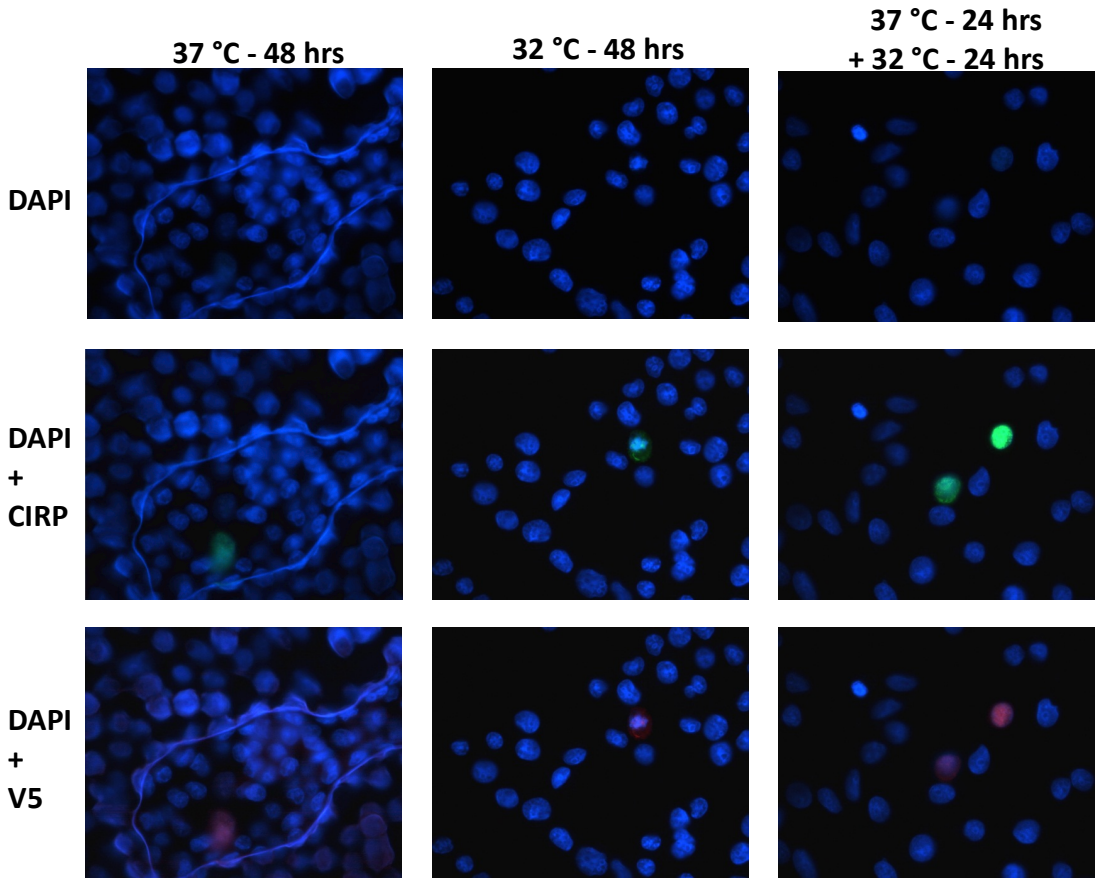

Figure 5.7 Analysis of CIRP localisation in CHO FlpIn cell transiently expressing FRT vector containing full length CIRP with an N-terminal V5 tag incubated at 37 or 32°C following transfection as shown above (figure 5.6) by overlay of immunofluorescence images.

To determine where CIRP was localised in the cell, the DAPI stained images showing the cell nucleus were overlaid with images of the cells showing the CIRP and V5 signal (figure 5.7). These showed that in the cells incubated at 37°C for 48 hours and 32°C for 48 hours the CIRP appeared to be localised to the nucleus as the images overlaid with the nucleus boundaries on the DAPI images. However, some signal for CIRP and V5 was present outside the nucleus in the cytoplasm in the cells that had been subjected to temperature shift from 37°C to 32°C. Therefore, it appears that the temperature shift results in the translocation of CIRP to the cytoplasm from the nucleus as well as leading to an up-regulation in expression, this agrees with the findings of Leeuw et al (2007) who presented that CIRP migrates from nucleus to the cytoplasm upon cell stress by a methylation dependent mechanism (De Leeuw et al., 2007).

5.3.3 Investigating the Potential Interaction between CIRP and eIF4G

Potential interactions between CIRP and the eIF4G protein was investigated by co-immunoprecipation experiments using a V5 primary antibody. eIF4G protein with an N-terminal V5 tag was pulled down with a V5 antibody to determine if CIRP was bound to it (figure 5.8, (a)) and vice versa full-length CIRP protein with an N-terminal V5 tag was pulled down with a V5 antibody to determine if eIF4G was bound (figure 5.8, (b)). Two truncations (kindly provided by Miss Joanne Roobol, University of Kent) of the eIF4G protein termed 642 and 1078 were also investigated (figure 5.18), and these also possessed an N–terminal V5 tag.

The immunoprecipitation experiments to investigate protein-protein interactions between eIF4G and full length CIRP confirmed the presence of an interaction between the two proteins (Figure 5.8). CIRP bands were observed on the blot where eIF4G protein was pulled down in both HEK and CHO cells (figure 5.8, (a)). A strong band was also observed for the eIF4G 642 and 1078 truncations confirming that large amounts of these proteins were pulled down with the anti-V5 antibody and protein A beads. A band for full-length eIF4G could not be observed on this blot due to the large size of this protein. CIRP was therefore shown to be present in immunoprecipitation samples of full-length eIF4G and both eIF4G truncations. The CIRP band intensity was greater when immunoprecipitating eIF4G in HEK cells

compared to CHO cells (see Figure 5.8). The intensity of the CIRP band was greatest when immunoprecipitating the eIF4G 642 truncation in HEK cells suggesting that the strongest interaction is between CIRP and this truncated version of eIF4G. When the interaction between eIF4G and CIRP was investigated by immunoprecipitation of recombinant CIRP with a V5 tag, large amounts of CIRP were observed on the blots and eIF4G protein was present in samples for both CHO and HEK cells (figure 5.8 (b)).

A much stronger eIF4G band was present in the sample from HEK cells compared to CHO cells agreeing with the eIF4G immunoprecipitation results again suggesting that a stronger interaction takes place between mouse CIRP and eIF4G in HEK cells than in CHO cells. Interestingly, the CIRP bands seen on the blots appeared as a double band, (figures 5.5 (b), 5.8 (b), 5.9), this may be due to the methylation event that CIRP undergoes before being translocated to the cytoplasm. Therefore the two bands observed here are potenitially one for the up-regulated methylated CIRP located in the cytoplasm and the second form that is not methylated located in the nucleus.

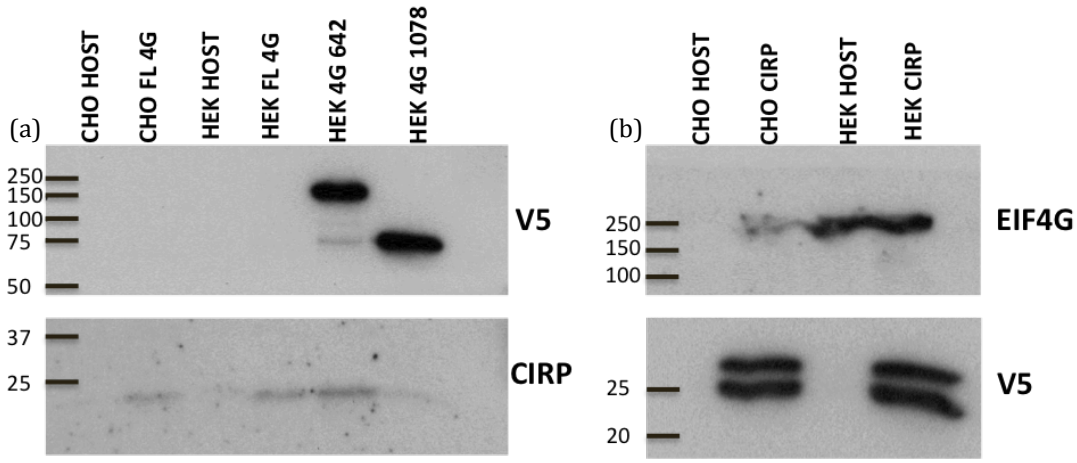

Figure 5.8 Western blot analysis of immunoprecipitation experiments to study the interaction of recombinant CIRP with eIF4G. (a) Immunoprecipitation of the NV5 full-length eIF4G and truncations 642 and 1078 for the interaction with CIRP in CHO and HEK cells. (b) Immunoprecipitation of the NV5 full-length CIRP for the interaction with eIF4G in CHO and HEK cells. All samples were incubated at 37°C for 48 h post transfection before collection and lysis.

The interaction between N-CIRP and eIF4G was also confirmed by pull down of N-CIRP with an N-terminal V5 tag whereby bands for N-CIRP were observed in the resulting blot (figure 5.9). The band intensity of N-CIRP was stronger in the HEK cells. eIF4G was present in both CHO and HEK cell samples containing recombinant

V5 N-CIRP. Once again, a stronger eIF4G band was observed for HEK cell samples. These data collectively suggest that expression of mouse CIRP in HEK cells is higher than in CHO cells, both full-length CIRP and N-CIRP interact with eIF4G and a stronger interaction is present between mouse CIRP and eIF4G in HEK cells although this may simply reflect the relative expression of the recombinant CIRP between the HEK and CHO cells.

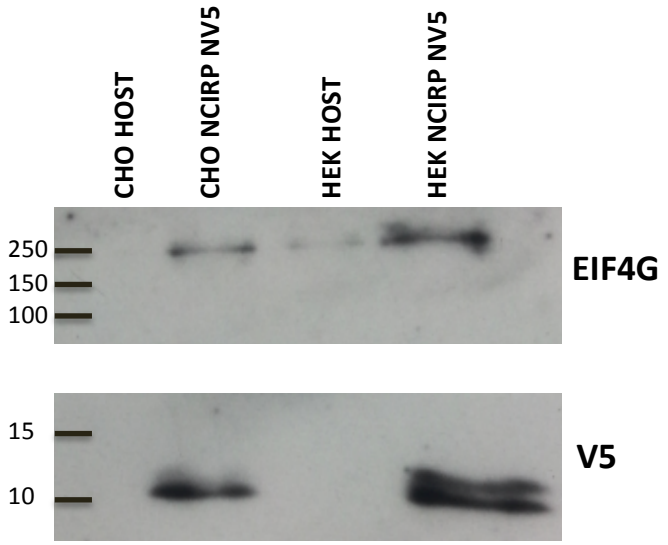

Figure 5.9 Western blot analyses of immunoprecipitation experiments to study the interaction of recombinant N-CIRP with eIF4G. Immunoprecipitation of the NV5 N-CIRP for the interaction with full-length eIF4G in CHO and HEK cells. All samples were incubated at 37°C for 48 h post transfection before collection and lysis.

5.3.4 Investigating The Effect of the Presence of the ATR 3'UTR on Luciferase mRNA when Over-Expressing CIRP and eIF4G on Transient Firefly Luciferase Expression

The data in section 5.3.3 confirmed that CIRP interacts with eIF4G and others have previously shown that CIRP also binds to the 3'UTR of the ATR mRNA. To determine if CIRP acts to increase specific mRNA translation upon mild cold shock by binding to both mRNA and eIF4G, CHO cells were transfected with either all 3 or a combination of ATR-3'UTR-luciferase, luciferase alone, NV5 CIRP and NV5 eIF4G and the subsequent effect on luciferase expression investigated. As well as full-length eIF4G and CIRP, the eIF4G 642 truncation and N-CIRP was also transfected into cells. Following transfection in 24 well plates, cells were incubated at 37°C for 48 hours or for 6 hours at 37°C, followed by 42 hours at 32°C. Expression of CIRP

or eIF4G was then confirmed by western blot analysis and the luciferase activity determined as described in section 5.2.17. The resulting luciferase expression results are shown in Figure 5.10 below.

Full-length CIRP was transiently expressed in CHO cells with eIF4G and luciferase containing the 3'UTR-ATR, the highest luciferase activity was observed when both CIRP and full length eIF4G and the ATR 3'UTR containing luciferase were all present (figure 5.10). Removing the ATR 3'UTR from luciferase or the eIF4G plasmid from transfection experiments resulted in a reduction in the luciferase activity observed. The eIF4G 642 truncation did not have the same effect as the full-length eIF4G suggesting that full-length eIF4G is required to elicit an effect as the presence of the truncation showed reduced luciferase activity (see Figure 5.10). The sample transfected with luciferase without the ATR 3'UTR and the 642 eIF4G truncation had relatively low luciferase activity. The same findings were observed at both 37 and 32°C. The statistical ANOVA test was performed on each data set. A p value of 0.0669 was calculated for the assays carried out at 37°C with full-length CIRP, this suggests that the means in each group tested at 37°C do not significantly differ from each other. However at 32°C the p value was 0.008, which suggest that there may be a significant difference between the groups tested. To further identify which groups in the data set differ from each other, Tukey's test was performed. With this test it was identified that there was a significant difference between the sample transfected with ATR3'UTR-luciferase, eIF4G and CIRP and the sample transfected with luciferase mRNA, 642 eIF4G and CIRP as well as the sample transfected with ATR 3'UTR-luciferase and 642 eIF4G.

When N-CIRP was transiently expressed in CHO cells with the different luciferase and eIF4G combinations, the highest luciferase activity was observed with N-CIRP and full-length eIF4G and luciferase without the ATR 3'-UTR at both 37°C and 32°C (Figure 5.10). It is also noted that in general, luciferase readings were lower for the N-CIRP samples. The ANOVA calculations gave a p value of 0.0011 at 37°C for N-CIRP data sets suggesting a possible significant difference between the groups. The Tukey's test revealed that the samples; +ATR,+N-CIRP,+eIF4G and –ATR,+N-CIRP,+eIF4G are different; +ATR, +N-CIRP, +642 eIF4G and –ATR, +N-CIRP,

+eIF4G are different; –ATR, +N-CIRP, +eIF4G and –ATR, +N-CIRP, +642 eIF4G are different; –ATR, +N-CIRP, +eIF4G and +ATR, -NCIRP,+642 eIF4G are different. However the ANOVA test for the assays carried out at 32°C had a p value of 0.07, which would suggest that there was no significant difference between the data groups.

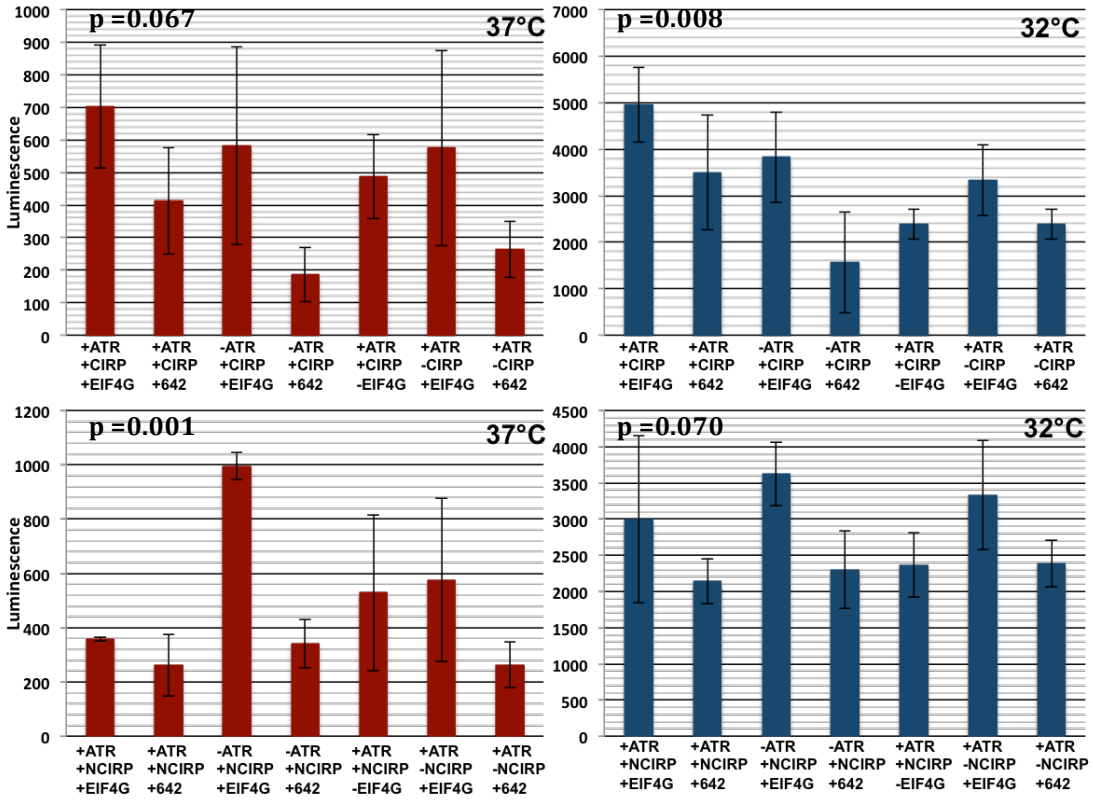

Figure 5.10 Recombinant firefly luciferase protein production in CHO cells transiently expressing CIRP/N-CIRP protein, full-length/642 eIF4G protein and +/-3'ATR-UTR-luciferase at 37 or 32°C. The top graph shows data for full length CIRP and bottom graph shows N-CIRP luciferase readings. The error bars represent standard deviation errors where n=3. The p values shown are from one-way ANOVA test.

The western blot analysis showed that N-CIRP expression was very low compared to full-length CIRP, especially at 37°C where no N-CIRP bands were observed (Figure 5.11). This was also true for the expression of eIF4G i.e. lower expression of the protein was observed on blots when expressed with N-CIRP compared to when expressed with full-length CIRP.

The western blot analysis of the luciferase assay samples showed that of the eIF4G proteins, the 642 truncation showed the highest expression levels (figure 5.11).

Although this protein was expressed at higher levels by the cells, it is the transient over-expression of the full-length eIF4G protein that leads to the higher luciferase levels. Interestingly, in the absence of CIRP, full-length and 642 eIF4G levels were reduced and these samples also show low luciferase activity when compared to samples which contained full-length CIRP. The full-length CIRP western blots also showed that when luciferase was co-transfected without ATR 3'-UTR, the CIRP and 642 eIF4G sample showed the lowest CIRP expression levels, which also showed the lowest luciferase levels. The luciferase activity and western blot data shows that expression of CIRP, eIF4G and 3' ATR UTR-luciferase leads to maximal expression of all three components. Luciferase activity was highest when luciferase contained the ATR 3' UTR, CIRP and eIF4G overexpressed.

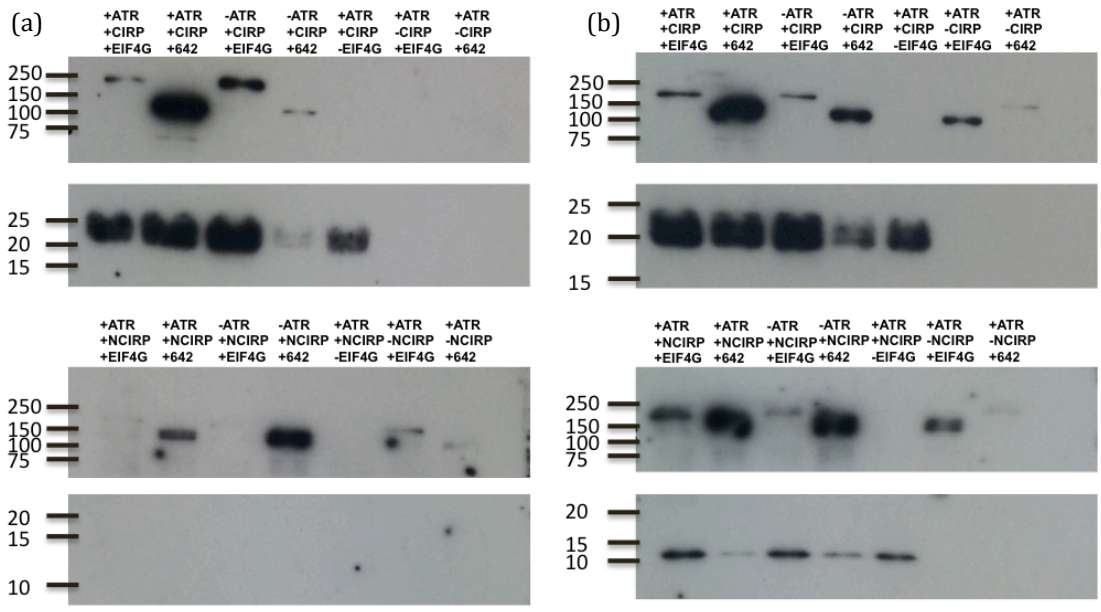

Figure 5.11 Western blot analysis of CHO FlpIn cells transiently transfected with NV5 CIRP/N-CIRP, 3'UTR-ATR-luciferase, NV5 eIF4G/642. Blots were probed with a V5 1° antibody, top blots show samples with full-length CIRP and bottom blots show N-CIRP samples. (a) Cells incubated at 37°C for 48 hours post transfection (b) cells incubated at 37°C for 6 hours, followed by an 42 hour incubation at 32°C post transfection.

Following the initial preliminary experiment described above, a second targeted experiment was undertaken investigating the ATR-3'UTR-luciferase with full-length CIRP and full-length eIF4G as the combinations of these constructs showed the highest recombinant luciferase expression in the preliminary studies. The samples were tested in 6 replicates instead of 3 to give greater confidence to the data and the results are reported in figure 5.12 below.

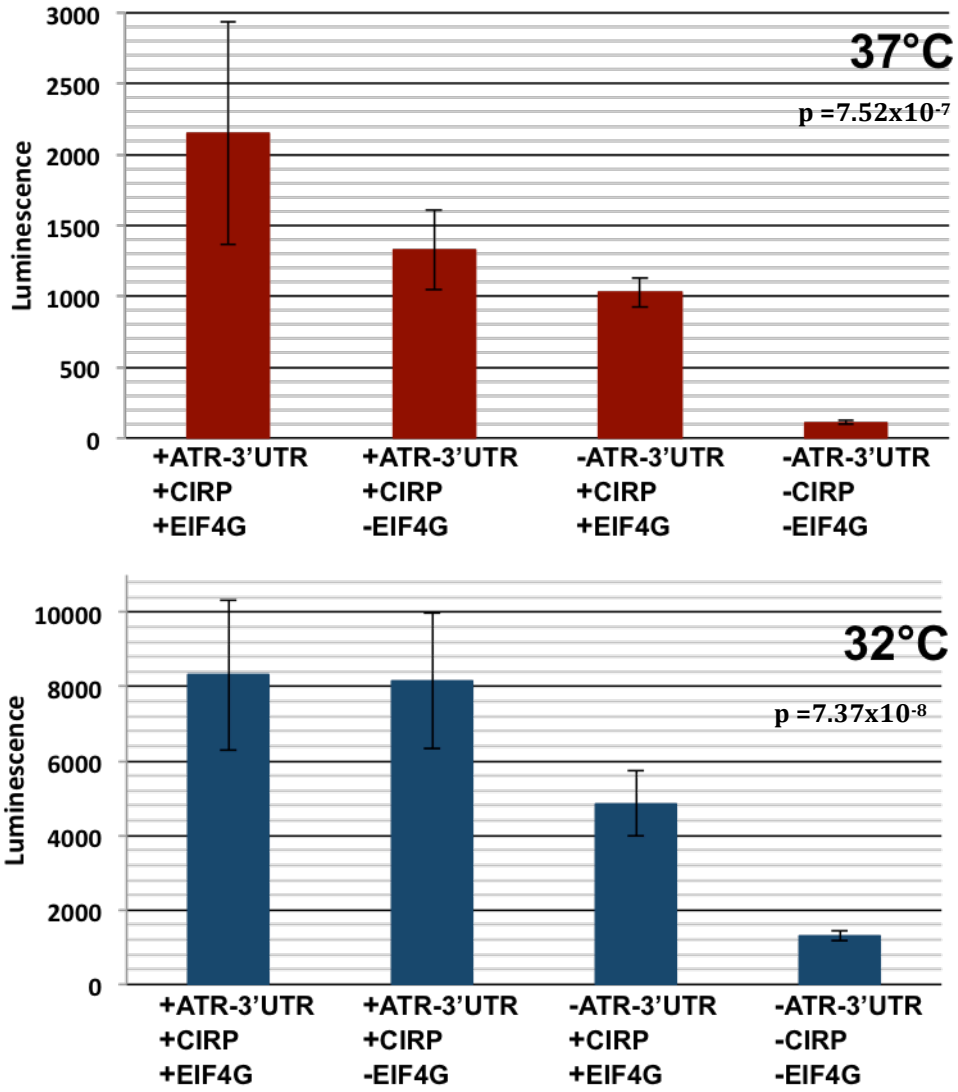

Figure 5.12 Recombinant firefly luciferase protein production in CHO cells transiently expressing CIRP protein, full-length eIF4G protein and 3'ATR-UTR-luciferase at 37 or 32°C. The error bars represent standard deviation errors where n=6. The p values shown are calculated from one way ANOVA test.

In agreement with the preliminary data, the highest luciferase expression was observed when the luciferase containing the ATR 3'UTR and CIRP and eIF4G proteins were present at 37°C (figure 5.12). The luciferase activity was much higher, approximately 21 times, at 37 °C compared to cells transfected with luciferase lacking the ATR 3'-UTR and in the absence of recombinant CIRP and eIF4G. The data also show that removing the ATR 3'UTR from the luciferase gave a greater reduction in luciferase activity compared to when eIF4G was removed, suggesting the presence of the ATR 3'-UTR leads to a greater effect than the presence of additional exogenous eIF4G. Indeed, at 32°C, there was no increase in the observed luciferase activity when

eIF4G was over-expressed alongside CIRP in the presence of the ATR 3'-UTR containing luciferase construct suggesting that the presence of additional eIF4G at 32^{o}C does not result in any additional increase in recombinant protein expression (Figure 5.12). The ANOVA test calculated a very small p value of $7.51x10^{-7}$ for the groups at 37°C and $7.37x10^{-8}$ for the 32°C data set. These value confirm that there is a significant difference between the groups tested here. The Tukey's test revealed that each group differs from one another except the groups: +ATR,+CIRP, -eIF4G and -ATR,+CIRP, +eIF4G at 37°C; +ATR,+CIRP, +eIF4G and +ATR,+CIRP, -eIF4G at 32°C.

5.3.5 The Effect of Over-Expression of Recombinant CIRP and eIF4G on Stable Firefly and Gaussia Luciferase Expression

The results presented in section 5.3.4 showed that transient firefly luciferase expression can be enhanced via the addition of the ATR 3'-UTR to luciferase and the over-expression of CIRP and eIF4G at 37^{o}C and of CIRP at 32^{o}C. To determine if a similar effect could be generated in stably expressing cells, CHO cells stably expressing firefly luciferase and gaussia luciferase were used to study the effect of CIRP protein and a combination of CIRP and eIF4G on the recombinant protein expression levels. Cells were incubated at 37°C for 24 or 48 hours following transfection with CIRP or eIF4G constructs or for 6 hours at 37°C before moving to 32°C for the remainder of the incubation. The resulting luciferase activity is reported in figures 5.13 and 5.14 below.

Cells incubated for 24 hours at 37°C did not show any appreciable differences in luciferase activity compared to control cells transfected with blank FRT vector (figure 5.13) except for when cells were transfected with C-CIRP and eIF4G, where there was a reduction in luciferase activity. At 32°C approximately twice as much luciferase activity was observed and as in the 37°C samples, the majority produced similar amounts of luciferase including the C-CIRP and eIF4G sample. The data collected for cells incubated for 48 hours showed the same trend (figure 5.13). Some samples such as FRT+4G and CIRP+4G showed large differences at 32°C but these had large error bars. The ANOVA test calculations gave p values which were above

0.05 for all data sets suggesting that there is no significant differences between the groups tested here at 37°C and 32°C or 24 h and 48 h.

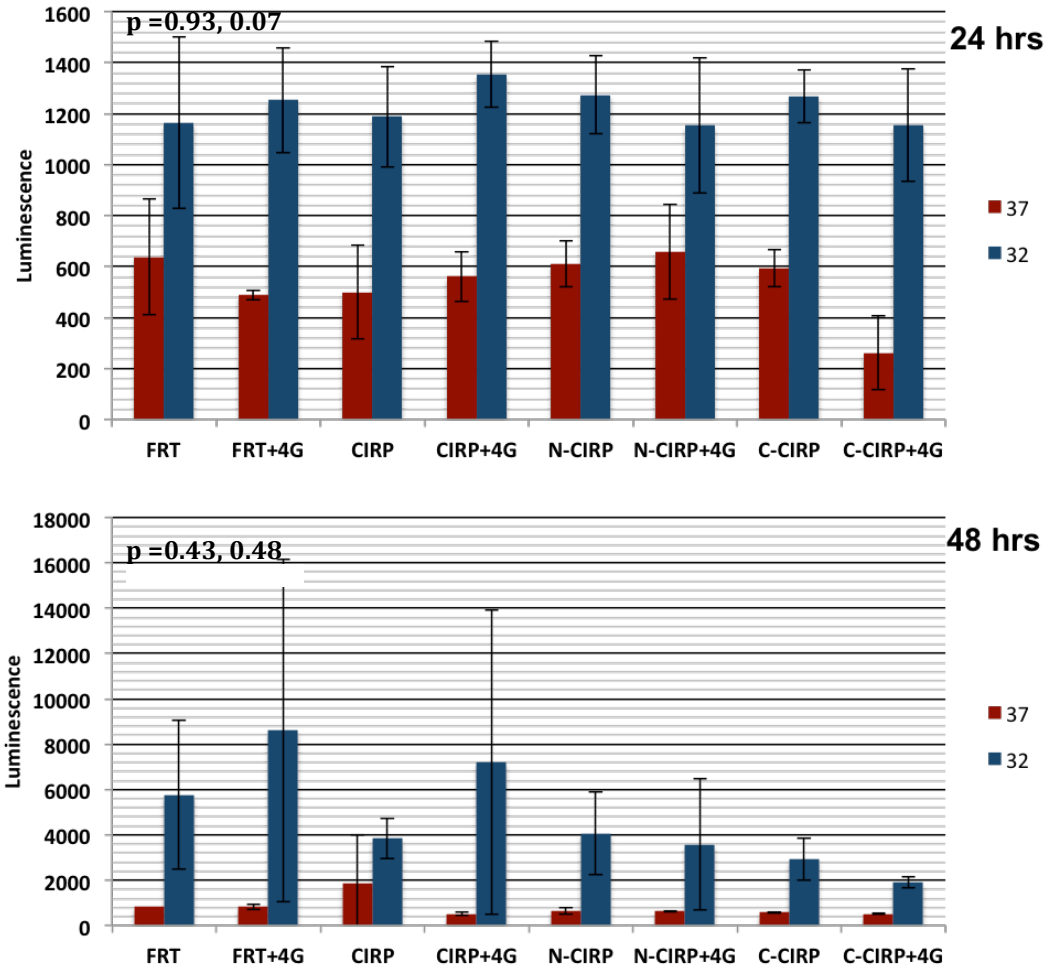

Figure 5.13 Recombinant firefly luciferase protein production in CHO cells stably expressing firefly luciferase and transiently transfected with CIRP/N-CIRP protein, eIF4G protein or both at 37 or 32°C for 24 or 48 hours. As a control in place of CIRP/N-CIRP empty FRT plasmid was transfected. Errors bars represent standard deviation errors where n=3. p values shown were calculated using the one way ANOVA test.

When investigating the effect of over-expression of CIRP or eIF4G on stable, secreted Gaussia luciferase production similar results to the firefly protein expression were observed (figure 5.14). Interestingly, an up-regulation of the amount of secreted protein at 32°C was not observed (figure 5.14) suggesting that the secretory capacity of the cell may be limited at the reduced temperature. As observed with the intracellular firefly data, addition of eIF4G led to an improvement of gaussia luciferase production in most samples, although this was not seen for the CIRP+4G sample at 37°C for both the 24 and 48 hour samples. In general, although a large difference in protein production was not observed between the control and cells

expressing CIRP, the overall trend was that the largest influence on stable recombinant protein expression was observed when both CIRP and eIF4G were both over-expressed at 32°C. This was further confirmed by carrying out the ANOVA test which calculated p vales above 0.05 i.e. no significant differences between the groups tested in each data set.

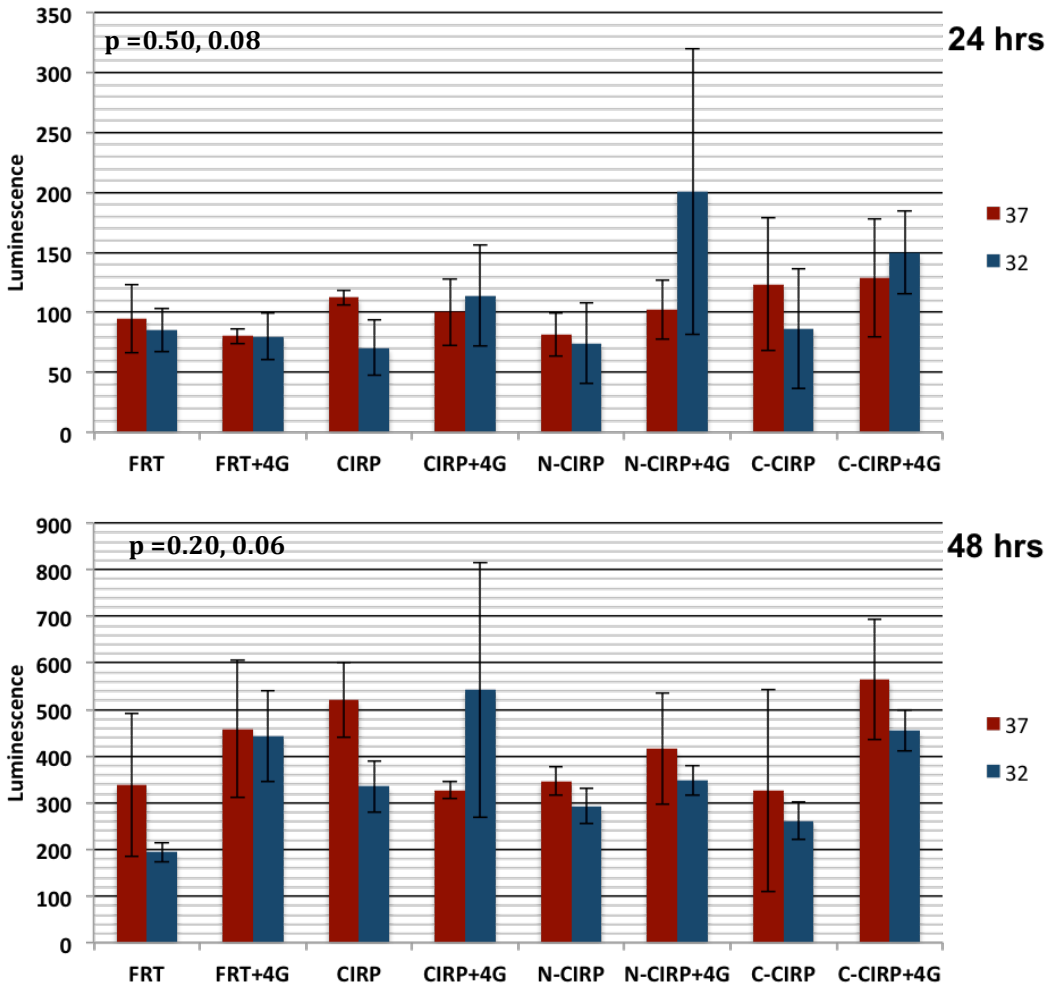

Figure 5.14 Recombinant gaussia luciferase protein production in CHO cells stably expressing gaussia luciferase and transiently expressing CIRP protein, eIF4G protein or both at 37 or 32°C for 24 or 48 hours. Error bars represent standard deviation errors where n=3. p values shown were calculated using the one way ANOVA test.

5.3.6 Generation of CHO FlpIn Cell Lines Stably Expressing CIRP

CHO FlpIn cells stably expressing full-length CIRP, N-CIRP and C-CIRP with an N/C-terminal V5 tag or no tag were generated as described in section 5.2.1 and 5.2.9. For characterisation of these cells lines, growth curves were obtained at 37°C and at 32°C. Immunofluorescence was also undertaken to confirm protein expression and localisation. The resulting data is presented in figures 5.15 below.

The growth profiles at 37°C showed that cell lines expressing CIRP and variations, apart from the C-CIRP cell line, all grew faster compared to the control (figure 5.15 (a)) whilst the C-CIRP cell line growth was similar to that of the FRT control cell line. Specifically, cells stably expressing full-length CIRP without a tag showed the largest difference in growth, being the cell line with fastest growth rate and reaching the highest cell concentration of all cell lines. The differences in the growth curves suggests that the expression of CIRP leads to faster cell growth at 37°C with the exception of C-CIRP which grew at the same rate as the control. This may be due to a lack of expression of the natively disordered C-CIRP in CHO cell as previously observed when transiently attempting to express C-CIRP (see section 5.3.2). The unstructured nature of this domain may lead to its degradation and no or very low expression.

The growth profiles of the stable cell lines at 32°C did not show large difference between growth rates of the control cell line and CIRP cell lines. Cell lines stably expressing full-length CIRP and N-CIRP with N-terminal V5 tag reached the highest cell concentration/density. The growth profile clearly shows the influence of 32°C culturing on growth and cell viability being extended (note the extended time on the x axis on figure 5.15b). These data suggest that whilst CIRP over-expression results in more rapid growth of CHO cells at 37^{o}C, this effect is not observed when culturing such cells at 32^{o}C.

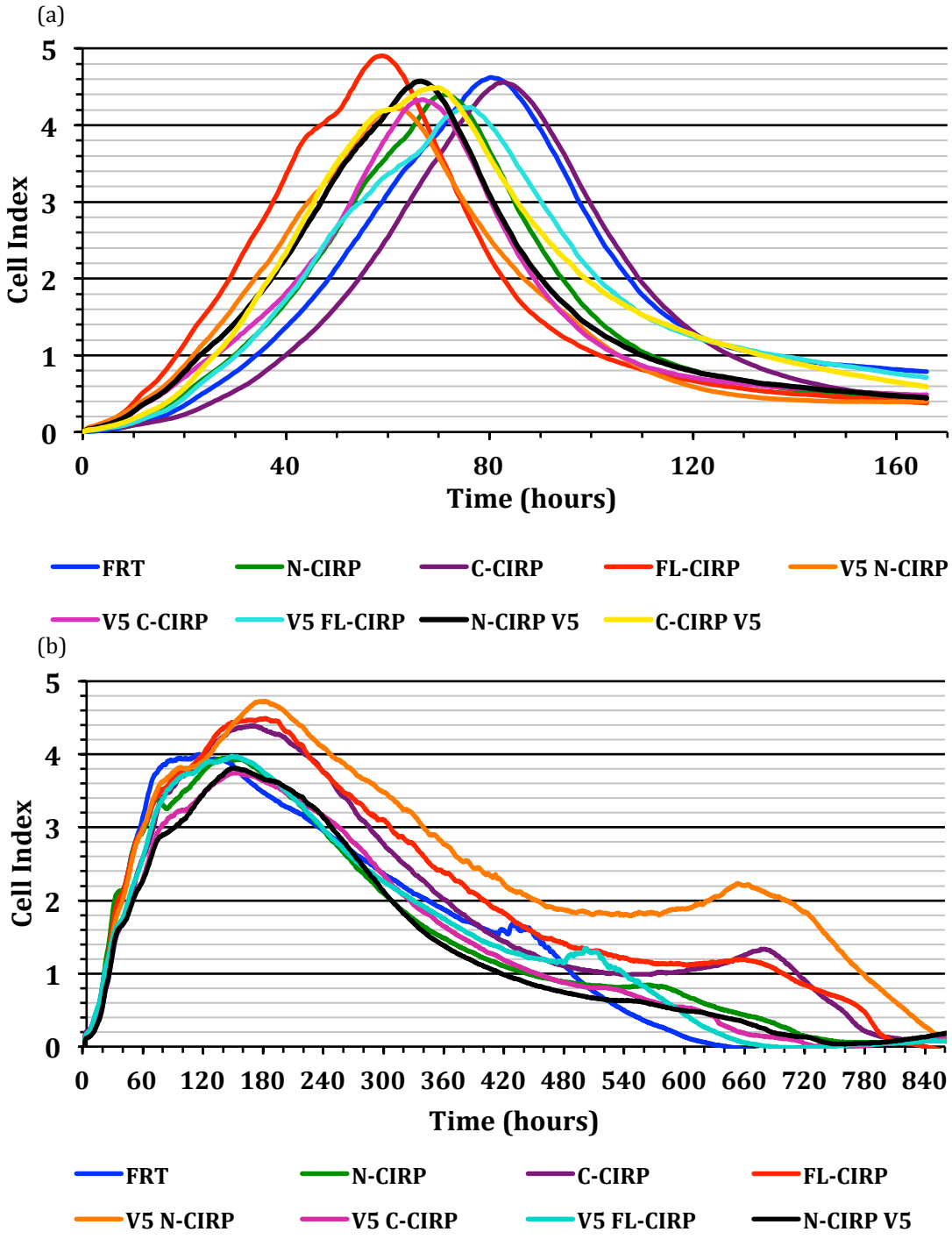

Figure 5.15 Growth curves generated using the xCELLigence instrument for the CHO Flp In cell lines stably expressing full length CIRP and CIRP truncations with and without the V5 tag. (a) Cells incubated at 37°C (b) Cells incubated at 32°C. Note the larger x-axis bar on panel b.

In order to confirm that the cell lines generated were indeed expressing the desired proteins, the stable cell lines expressing full-length CIRP with an N-terminal V5 tag and the control FRT cell line were immunostained using CIRP and V5 antibodies. After seeding, cells were either incubated at 37°C for 48 hours or at 37°C for 6 hours followed by incubation at 32°C for 42 hours before fixing and immunofluorescence as described in section 5.2.14. Resulting representative images are shown in figures 5.16 and 5.17 below.

The immunostaining of the cells expressing full-length CIRP with an N-terminal V5 tag confirmed the expression of CIRP in these cells (Figure 5.16). Some background signal was observed in the control samples for TRITC (CIRP) and FITC (V5) (figure 5.16, (a)) but this was due to the long exposure time and a much higher signal was observed for the CIRP expressing cells confirming that recombinant CIRP is expressed in these cells (figure 5.16, (b)) and it is expressed with a V5 tag as signal for both V5 and CIRP were observed in these samples. Localisation appears cytosolic in these images and all of the cells are expressing as would be expected for stable cell lines.

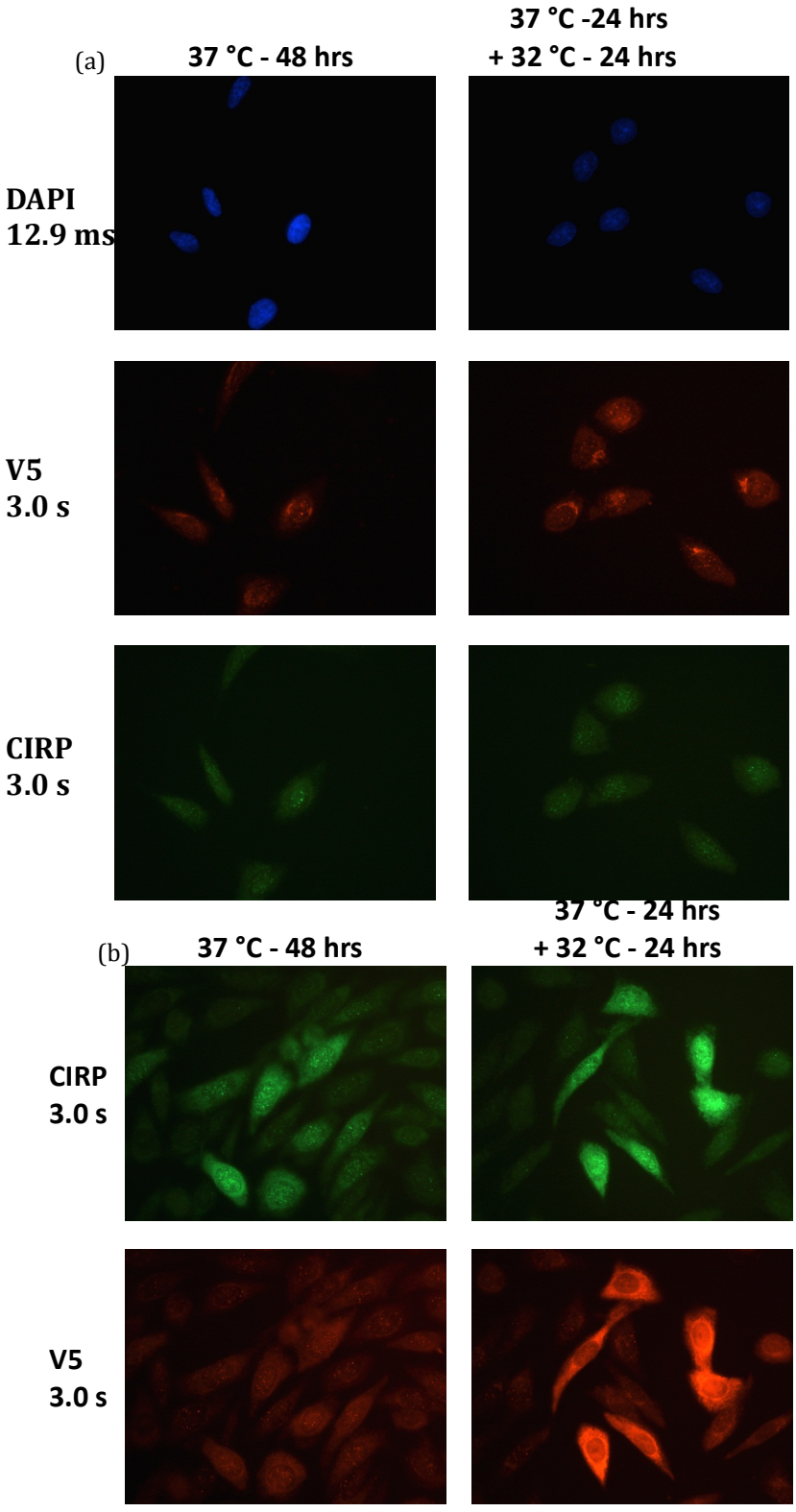

Figure 5.16 Immunofluorescense images collected following immunostaining of CHO FlpIn cells stably expressing control FRT vector or FRT vector containing full length CIRP with an N-terminal V5 tag incubated at 37 or 32°C following transfection as shown above. (a) Cells expressing control vector (b) Cells expressing full length CIRP with V5 tag.

To determine which region of the cell CIRP localises to in these cells, images of cells showing CIRP and V5 staining were overlaid with images of the cells showing the nucleus DAPI staining (figure 5.17). These showed that at 37°C CIRP is present in both the nucleus and cytoplasm but at 32°C the amount of CIRP in the cytoplasm was much higher compared to what was observed at 37°C even though CIRP was still also localised to the nucleus. This result agrees with the immunostained cells transiently expressing CIRP (section 5.3.2), at mild cold temperatures (32°C) CIRP translocates from nucleus to cytoplasm.

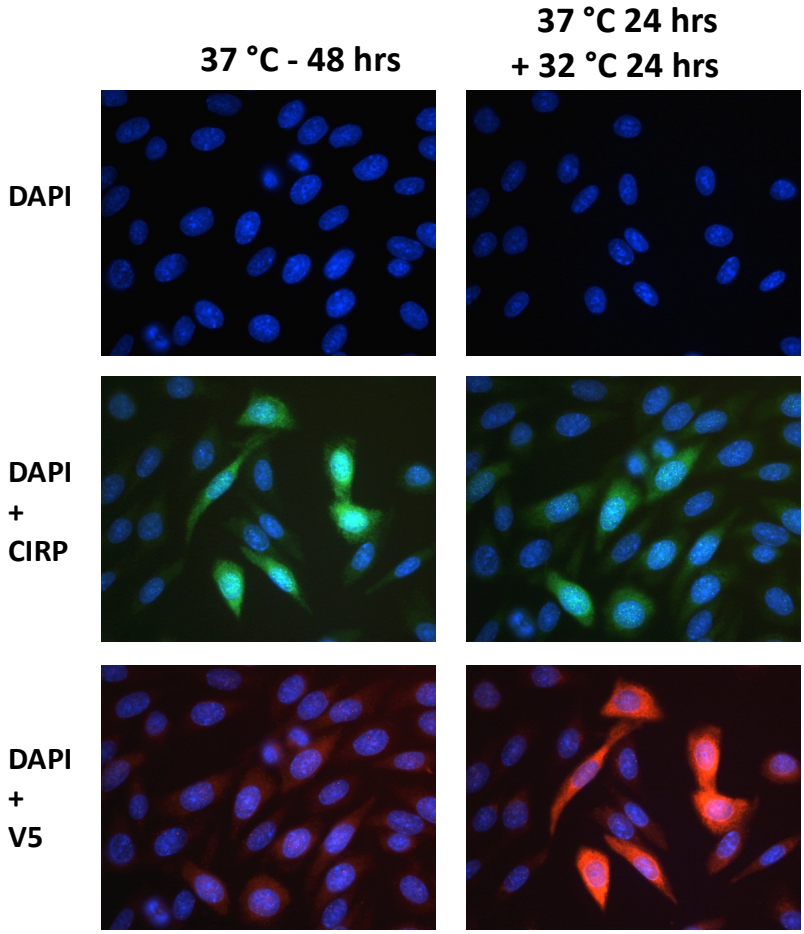

Figure 5.17 Analysis of CIRP localisation in stably expressing cells by overlay of immunofluorescence images.

5.4 Discussion

This chapter describes investigations that were undertaken to further probe the function of CIRP in the cell and how the expression of CIRP relates to recombinant protein expression. To enable the investigation of the roles of the N- and C-terminal domains of CIRP on their own as well as the full-length protein, FRT plasmids expressing N-CIRP, C-CIRP or full-length CIRP were successfully generated by PCR, restriction digest and ligation of the fragments. For the co-immunoprecipitation experiments to study CIRP interactions, each construct was also generated with a V5 tag at either the N-terminal or the C-terminal. All constructs were successfully generated and sequences were confirmed by DNA sequencing.

The plasmids were tested for expression by transiently transfecting into CHO cells and cell lysates collected and analysed by western blotting. Full-length CIRP with an N-terminal V5 tag was successfully expressed transiently in CHO cells. The expression of N-CIRP was more difficult to observe by western blot analysis, a higher concentration of total protein had to be analysed to be able to see the N-CIRP bands on the blots. The protein that had been subjected to storage by freezing was much more difficult to see in samples. This was not found for the full-length protein. N-CIRP is not expressed by the cells as well as the full-length protein and is presumably degraded much more quickly. This might be expected as the cells may recognise that this is not the full-length protein and the missing C-terminal region could make the protein more prone to degradation. In a study by Zhang et al (2007) it was found that C-terminal cleavage of α-crystallin protein which has a chaperone function and supressing protein aggregation or promoting protein folding, significantly reduced its stability and increased its susceptibility to degradation by the ubiquitin proteasome pathway (Zhang et al., 2007). It is also noted that the full-length endogenous protein is still made in the N-CIRP cells and it has been previously suggested that both the N-terminal and C- terminal are required for maximum function of the CIRP (Yang et al., 2006).

The western blot analysis suggested that the C-terminal domain on its own was not expressed at all or that it was degraded too quickly to be observed. The western blot analysis also confirmed that CIRP is up-regulated under mildly hypothermic conditions as also shown by previous studies (Nishiyama, Itoh, et al., 1997). As this

domain is natively unstructured, it is not surprising that it would be a target for degradation or not expressed at all. The C-terminal region of CIRP was also found to be prone to degradation in experiments described in Chapter 2 (section 2.3.3). Studies have shown that unstructured regions of protein are highly susceptible to degradation especially *in vitro* and in the absence of ligands (Mittag, Kay, & Forman-Kay, 2010; Tompa, Prilusky, Silman, & Sussman, 2008).

The successful expression of full-length CIRP with an N-terminal V5 tag was further confirmed by immunostaining of cells transiently transfected with the protein. Only one or two cells on each slide showed expression of CIRP suggesting low expression levels or poor transfection efficiency. As expected, higher CIRP expression was observed at the lower temperature of 32°C. Exposing the cells to hypothermia without first incubating at 37°C did not increase protein expression but shifting the temperature of the cells from being incubated at 37°C to 32°C showed an increase in CIRP levels in the cells. This suggests that it is the temperature shift that activates mechanisms associated with cold shock but exposing them to cold temperatures without the shift most likely activates other stress mechanisms in the cell. The up-regulation of CIRP due to the temperature shift was also confirmed as it was the temperature shift condition that lead to the translocation of CIRP to the cytoplasm and CIRP is predicted to translocate to the cytoplasm from the nucleus under mild hypothermia to be functional (De Leeuw et al., 2007). This is not the case for cells that do not experience the recovery at 37°C before the temperature shift but are only incubated at 32°C, as well as the up regulation of CIRP being absence in these cells, it was not present in the cytoplasm suggesting that mechanisms activated by cold shock and involving CIRP were not activated.

The interaction of CIRP with the eIF4G protein was confirmed through co-immunoprecipitation experiments. eIF4G interacted with both full-length CIRP and the N-terminal domain of CIRP alone. This suggests that the C-terminal of CIRP is not required for the interaction between CIRP and eIF4G. It has been suggested that the N- and C-terminal regions of CIRP interact individually with RNAs but that both are required for maximal binding (Yang et al., 2006). As C-CIRP expression could not be achieved, co-immunoprecipitation experiments could not be carried out to test

the interaction of this domain with eIF4G. Truncations of eIF4G were kindly provided by Jo Roobol, these truncations of eIF4G were made of the C-terminal, removing parts of the sequence from the N-terminus (figure 5.18).

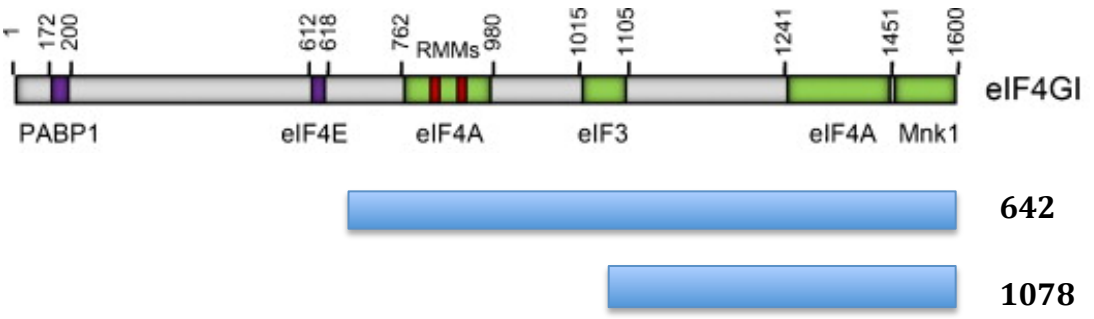

Figure 5.18 Schematic figure of the eIF4G initiation factor, with the binding domains indicated. The truncations tested for binding to CIRP are shown as blue rectangles. Figure adapted from Yángüez & Nieto, 2011.

As described above, the data reported here shows that eIF4G interacts with the N-terminal domain of CIRP. To help determine the binding region for CIRP on eIF4G, the truncations in figure 5.18 were used. The binding was significantly reduced or non-existent when testing the 1078 truncation of eIF4G (figure 5.8), therefore the binding site for CIRP must be between residues 642 and 1078 on eIF4G between the eIF4E and eIF3 binding sites. It was also noticed that higher expression of eIF4G and the truncations was observed in HEK cells compared to CHO, previous studies have shown that in comparison to HEK, CHO is more challenging for transient transfection producing lower yields (Ye et al., 2009).

With the constructs available here, we were to then investigate the hypothesis that through the interaction of CIRP and eIF4G and the interaction of CIRP with the 3'UTR of specific mRNA such as ATR, CIRP acts as a bridge between the 3' and 5'UTR of mRNA through its interaction with eIF4G and circularise the RNA, enhancing the translation of such an mRNA. Luciferase reporter assays were therefore undertaken to investigate the effect of CIRP on recombinant protein production as well as to further investigate the interaction between CIRP, eIF4G and a specific 3'-UTR sequence, that from the ATR 3'UTR mRNA that has previously been shown to bind to CIRP (Yang et al., 2010). In order to investigate this, a luciferase construct containing the ATR 3'-UTR was used.

The luciferase assays revealed that the highest expression of recombinant protein (luciferase) was observed when CHO cells were transfected with CIRP, eIF4G and luciferase containing the ATR 3'UTR both at 37°C and 32°C, although generally the luciferase activity was much higher at 32°C (figure 5.12). This may be due to the up-regulation of CIRP at 32°C and is likely to have been enhanced by the up-regulation of endogenous CIRP as well as the recombinant transfected CIRP. The highest luciferase expression was observed for the CIRP, eIF4G and luciferase containing ATR 3'UTR mRNA combination when full-length eIF4G protein was used and not with the 642 truncation. The significance of the differences between the samples was further confirmed by ANOVA and Tukey's tests. This further supports the theory that CIRP interacts with RNAs and the translational machinery to improve protein production (Yang et al., 2006). For translation to take place the full eIF4G protein would be necessary as it is a scaffold protein with many binding sites for other initiation factors (Hinton, Coldwell, Carpenter, Morley, & Pain, 2007). The removal of eIF4G over-expression resulted in a reduction of the luciferase activity. This is presumably because at reduced temperature protein synthesis rates drop so there may be more 'available' eIF4G (e.g. less translation so more free eIF4G available and hence over-expression of this has no effect as not limiting) so this is not a limitation but the presence of CIRP and ATR 3'-UTR allows more efficient translation at this temperature.

Removal of the ATR 3'-UTR from luciferase also led to a reduction of luciferase expression. Presumably when this 3'UTR is present, CIRP can bind to the ATR 3'-UTR on luciferase and to eIF4G and enhance the translation of the luciferase mRNA to give increased protein production. It has also been shown previously that the interaction of CIRP with eIF4G can lead to increased translation (Yang et al., 2006). The same results were not observed when CIRP was replaced by N-CIRP, in fact lower luciferase expression was observed across all N-CIRP samples. This is in agreement with previous findings that both the N- and C-terminal domains are required for maximum function of CIRP (Yang et al., 2006) although N-CIRP alone binds to RNAs (as described in chapter 3,section 3.3.3) and to eIF4G. The C-terminal domain must contribute to the function of the protein; this maybe by a change in conformation upon ligand binding which could be further studied by in cell NMR in

future experiments (Sakakibara et al., 2009; Tochio, 2012). The binding affinity of the N-terminal to RNA was found to be weak in chapter 3. Disordered regions of proteins can enable recognition of a spectrum of ligands and conformational rearrangement for these (Mittag et al., 2010), this may be true for CIRP as it can bind several RNA ligands and the C-terminal may strengthen the binding which is weak for the N-terminal domain alone. Another point the consider is that it has been suggested that the translocation of CIRP from cytoplasm into the nucleus is dependent on the methylation of the arginine residues found on the C-terminal domain (Aoki, Ishii, Matsumoto, & Tsujimoto, 2002; De Leeuw et al., 2007). Without the methylation of these residues at the C-terminal, N-CIRP may be remaining in the nucleus and unable interact with the translational machinery and mRNAs, leaving only endogenous protein available in the cell to do this.

Increased eIF4G levels were observed when CIRP was overexpressed compared to when CIRP was not, as shown by western blots (figure 5.11). This suggests that over expression of CIRP leads to the general up-regulation of protein expression in the cells which has also been reported in other studies (Tan et al., 2008). These findings all provide evidence that eIF4G, and CIRP act together to enhance protein production in cells by binding to the 3'UTR of mRNAs specifically the 3'UTR of ATR in this case. It was also noticed that removing the ATR 3'UTR from luciferase lead to lower luciferase levels than if exogenous eIF4G was removed. This was probably because when recombinant eIF4G was removed, endogenous eIF4G levels were still sufficient enough to carry out translation of the ATR 3'UTR luciferase mRNA but removing the ATR 3'UTR means that the translation of the luciferase mRNA may not be CIRP mediated and hence not up-regulated as CIRP only binds specific mRNAs for example 3'UTR ATR. The significance of these results were confirmed using the statistical ANOVA and Tukey's tests.

With transient transfection, there can be variables such as the transfection efficiency, which means not every cell is expressing the gene of interest and also the transfected plasmid dilutes through generations of the cells and every cell will have not have the same number of copies of the gene, hence will not make the same amount of protein, which can lead to inconsistency between samples. Recombinant protein production levels were therefore also studied using cells that stably express secreted gaussia or

intracellular firefly luciferase (section 5.3.5). As the desired gene is integrated into genome of the host cell in stably expressing cell lines, every cell expresses the gene of interest; therefore there is less variability in expression. However, there can still be variability in terms of the cells expressing the exogenous CIRP or eIF4G. Neither of the luciferase mRNAs in these stable cell lines contain the ATR 3'UTR either. Most samples showed that the addition of eIF4G increased the luciferase levels observed marginally at best. The ANOVA tests showed that there was no significant differences between the groups in each data set with p values higher than 0.05. The reason that differences were small may have been due to the fact that the cells had a different production load on them, meaning for example the control cell lines had only the load of producing the luciferase protein where as the cells transfected with CIRP and eIF4G were generating two exogenous proteins as well as luciferase. This means that control cells with the empty plasmid may be able to make much higher levels of luciferase as it is the only protein being overexpressed by the cell whereas the cell transfected with CIRP or both CIRP and eIF4G have to make all 3 proteins, which may compromise luciferase expression. This may simply be a competition effect whereby the resources of the cell have to be shared across multiple exogenous genes or just one in the case of luciferase on its own. This may also be the reason why little increase in luciferase readings was observed when CIRP and EIF4G were both exogenously expressed compared to when only CIRP was being expressed. As such one could question whether the luciferase comparisons of these samples are a realistic comparison to each other.

Growth profiles of all cell lines stably expressing CIRP, N-CIRP and C-CIRP at 37°C had clear differences compared to the control showing that over expression of CIRP affect cell growth (figure 5.15). The exception was the C-CIRP construct without the V5 tag, this cell line had the same growth profile as the FRT control cell line. This suggests that this domain was not being expressed by the cell at all, in agreement with the western blot analysis of the transiently expressed cells. The cell line expressing full-length CIRP without the V5 tag showed the largest difference being the cell line that grew the fastest compared to tagged full length CIRP, N-CIRP and C-CIRP. The same result was not observed at 32°C, all cell lines had a similar growth profile except full-length CIRP and N-CIRP with N-terminal V5 tag reached higher cell

densities but all cell lines had similar growth rates. This is most likely due to natural up regulation of endogenous CIRP at 32°C. The profiles suggest that overexpression of full-length CIRP increases the growth rates of the cells. This finding is contradictory to the reports in the literature which suggest that up-regulation of CIRP leads to activation of mechanisms which lead to cell cycle arrest hence supressing cell growth (Tan, Lee, Yap, & Wang, 2008). This effect may have been observed here as the growth profiles were effected at 37°C and mechanisms which lead to cell cycle arrest may not be activated at this temperature, but CIRP could still be having its mRNA stabilising and up-regulating effect leading to increased protein synthesis and hence an increased metabolism.

Expression of CIRP and V5 tag in the stable cell lines and their up-regulation at 32°C was also confirmed through immunostaning of the cells (figure 5.16). Analysis of the localisation of CIRP in these cells showed that it localises to both the nucleus and cytoplasm at 37°C, the ratio of nucleus to cytoplasm being skewed toward the nucleus but at 32°C even though protein is present in the nucleus the ratio was now skewed to more CIRP being present in the cytoplasm. This agrees with the results seen for the cells transiently expressing CIRP and supports the theory that CIRP interacts with the translational machinery under mild hypothermia as translation takes place in the cytoplasm. It has been previously reported that CIRP translocate from nucleus to cytoplasm under stress conditions (De Leeuw et al., 2007). This translocation would be necessary for CIRP to be able to interact with eIF4G and act as a bridge to aid mRNA circularisation under cold stress conditions.

The data reported in this chapter supports the idea that CIRP is up-regulated under mild hypothermic conditions and is involved in cold shock mechanisms which lead to increased protein expression in the cell. It does this by translocation to the cytoplasm under cold conditions and direct involvement in the translation process, specifically through interactions with the eIF4G protein and mRNA 3'UTRs. These conclusions are also supported by the cold shock response model reported suggested by Al-Fageeh & Smales (2006). It has also previously been suggested that RNA constructs containing the 3'UTR CIRP-binding region displayed enhanced resistance to degradation by RNases and that CIRP enhances translation of its target RNA species

through stabilization of the mRNA (Sonna, Fujita, Gaffin, & Lilly, 2002) which is supported by the findings here. From the findings reported in this chapter we can hypothesise a mechanistic method by which CIRP interacts with specific mRNAs and the translational machinery via an interaction with eIF4G and the 3'-UTRs of specific mRNAs to circularise these and up regulate their translation. A schematic of this proposed mechanism is shown below in figure 5.19.

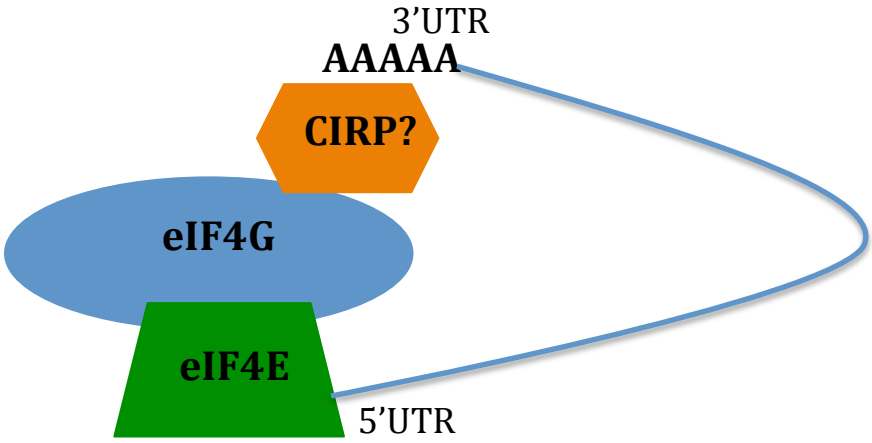

Figure 5.19 Schematic showing the proposed mechanism of how CIRP interacts with translational machinery and specific mRNA to enhance the translation of these.

Chapter 6

General Discussion

6.1 Overview

The objectives of this project were to: (1) to create RNA binding motif mutants of the N- terminal domain of the cold-shock induced protein CIRP to identify the essential residues for RNA binding and use this information to generate inactive version of the CIRP RNA binding domain to help further elucidate the function of the arginine/glycine rich C-terminal domain, (2) express and purify full-length CIRP to produce a soluble protein sample for RNA binding studies, (3) study the RNA binding properties of full-length CIRP, N-CIRP and N-CIRP mutants by EMSA and NMR, (4) solve the 3-dimensional structure and investigate the dynamic properties of N-CIRP in solution by NMR, and (5) undertake studies aimed at furthering our understanding of the molecular mechanisms by which CIRP can influence the rate of recombinant protein production in mammalian cells under normal (37°C) and reduced (32°C) culture temperatures. These objectives were largely achieved and the findings are discussed in detail in the previous chapters of this thesis.

N-CIRP RNA binding mutants were successfully generated, expressed and purified in chapter 2 and in the following chapter (chapter 3) it was shown that these mutated residues (F9 and F49) are essential for CIRP binding to RNA and therefore an inactive RNA binding form of the domain was achieved. Further importance of these residues was confirmed by the NMR dynamics studies where the region (β2-β3 loop) just before the RNP1 sequence including F49 was found to have increased motion compared to the rest of the protein (chapter 4). The three dimensional structure determination of N-CIRP was achieved by first completing the heteronuclear assignments of the backbone and side chain resonances for each amino acid as described in Chapter 4. Even though the structure requires further refinement, the identification of the backbone topology of N-CIRP was accomplished. Finally in Chapter 5, data was presented to support the hypothesis that CIRP interacts with specific mRNAs and the translational machinery via an interaction with eIF4G and the 3'UTR of specific mRNAs to circularise the mRNA and up-regulate translation.

6.2 Conclusions

RNA chaperones have been shown to stabilise mRNA molecules, prevent the formation of secondary structures and prevent or reverse aggregation (Herschlag, 1995; Russell, 2008; Tompa & Csermely, 2004). This study set out to further elucidate the structure and function of CIRP, a cold inducible RNA chaperone and the mechanism(s) by which it is able to enhance mRNA translation and increase protein expression of specific transcripts. This project further confirmed the identity of residues in CIRP essential for the binding of RNA ligands and the importance of these to the function and conformational rigidity of CIRP. The phenylalanine residues 9 and 49 in the amino acid sequence in positions 2 and 3 of RNP2 and RNP1 respectively were confirmed as important for binding RNA as their mutagenesis to Ala prevented RNA binding to N-CIRP(figure 3.7). Although the mutation of these essential residues did not have an effect on the expression of the protein in *E.coli*, their substitution to Ala did result in some conformational change in the structure of the domain as observed by HSQC-NMR, particularly for the F49 residue in RNP1 (figure 3.3). The mutation of one of the phenylalanine residues affected the chemical shift of the other confirming their proximity in space. Even though there is little information on the sequence specificity of the RNAs that CIRP binds, this study has increased the knowledge of how CIRP binds to its RNA ligands.

Proteins which function as RNA chaperones, together with cell signaling and cancer associated proteins, have been reported to often have disordered domains (Iakoucheva, Brown, Lawson, Obradović, & Dunker, 2002; Tompa & Csermely, 2004). Various hnRNP proteins are thought to have disordered regions, for example hnRNPA1, hnRNPC1/C2 and hnRNPU. The protein Hsp12 is induced upon a reduction in temperature (Murata et al., 2006) and it is a natively unfolded protein, which adopts a helical structure in the presence of its ligands (Welker et al., 2010). A study by Tantos et al (2009) has found that disordered proteins have a higher resistance to cold treatment in terms of retaining activity under cold conditions where as the globular proteins showed a drop in activity due to cold denaturation. As a result of this study, Tantos et al suggested that protein disorder is one of the key factors in cold stress (Tantos, Friedrich, & Tompa, 2009). The difficulty with producing a soluble purified full-length CIRP sample during this project limited the experiments that could be carried out to study the function of the C-terminal domain. The data that

was collected suggests that the C-terminal domain contributes to the function of CIRP. It has also been previously shown that both the N- and the C-terminal domain of CIRP are required for maximal RNA binding (Yang, Weber, & Carrier, 2006). The insolubility of the protein when expressed with the C-terminal region suggests that it requires a cellular environment and/or to be in the presence of a ligand to prevent aggregation. The EMSA results in chapter 3 of this work confirmed that both domains are required for maximal RNA binding as at the concentrations investigated only the full-length protein showed binding activity (section 3.3.4). As suggested by Tantos et al (2009), the disordered region may have a function towards resistance to cold stress helping the protein to retain their activity under low temperature conditions. As well as the N-terminal binding to RNA ligands in the absence of the C-terminal, this study has shown that the C-terminal domain is not required for the interaction of CIRP with eIF4G (figure 5.9). Further, the full length-protein lead to the highest increase in luciferase levels compared to N-CIRP when expressed in CHO cells and also cell lines stably expressing full-length CIRP showed an increased growth rate confirming that the C-terminal is necessary for maximal function of CIRP (section 5.3.4 and 5.3.6). When all data is considered together, this suggests that the arginine/glycine rich C-terminal domain plays an important role towards ligand specificity that may provide more stringent recognition or tighter binding after the N-terminal domain has initially recognized RNA binding partners. The low affinity of the N-terminal domain for RNA ligands suggest this may act as an initial binding partner with fast on/off rates whilst the C-terminal region may act to facilitate tighter binding to 'correct' binding partners. Whether this hypothesis is correct or not could be examined in future studies.

Previous studies have shown that CIRP migrates from the nucleus to the cytoplasm under stress conditions such as UV exposure, hypoxia and mild hypothermia (De Leeuw et al., 2007). This translocation has been shown to be regulated by the methylation of C-terminal RGG motifs that CIRP possess in its disordered C-terminal domain. In this project, immunofluorescence data has also shown an increase in the amount of CIRP located in the cytoplasm of cells exposed to cold stress at 32°C and higher concentrations of CIRP located in the nucleus in control cells cultured at 37°C (figures 5.7 and 5.17). This finding supports the function of CIRP acting as a bridge between mRNAs and the translational machinery to enhance the stability and

translation of specific mRNAs under hypothermic conditions, as a cytoplasmic location would be required for this function. This is an interesting point, because as mentioned above, over-expression of the full-length protein lead to a greater increase in luciferase levels compared to expression of N-CIRP (although N-CIRP expression itself was lower than full-length CIRP) and cell lines stably expressing full-length CIRP showed an increased growth rate. Even though the C-terminal is not required for RNA ligand binding of the protein, it is required for CIRP to translocate into the cytoplasm to be able to perform the suggested role its plays in the translation of mRNAs.

Luciferase experiments showed that the overexpression of CIRP in cells transfected with luciferase containing the ATR 3'UTR in its mRNA lead to an increased expression of luciferase protein i.e. the enhanced translation of this mRNA and also the interaction of CIRP with eIF4G (section 5.3.4). This is evidence that CIRP can bind specific mRNAs and enhances their expression through its interaction with the translation machinery leading to increased protein levels. It was possible to up-regulate the expression of luciferase protein by the addition of the ATR 3'UTR sequence, which has been identified as a RNA ligand for CIRP (Yang et al., 2010). These findings provide an opportunity to develop strategies to enhance the level of recombinant protein production in mammalian cells. The data also shows that CIRP binds both proteins via the N-terminal domain and RNA, although exactly where protein binding occurs on the N-terminal domain is not currently known. Presumably this cannot be on the face that interacts with RNA as mapped in the NMR studies if this domain binds RNA and protein simultaneously. For such a small domain to bind a large RNA molecule and a large translation factor at the same time must involve some steric constraints and further work to investigate the binding mode to eIF4G should be undertaken.

6.3 Main Findings and Future Work

As described in chapter 1, there have been many reports on the cellular response and biotechnological application of sub physiological temperatures for improved protein production. Many have investigated how the induction of CIRP at these temperatures contributes to these responses through its role as an RNA chaperone. Several reports have suggested a number of functions for CIRP, for example, in the reduction of testicular damage induced by cryptorchidism (Zhou, Zheng, Yang, Zhang, & Chen, 2009) and in the contribution to human HuR and cyclin E1 deregulation in breast cancer (Guo, Wu, & Hartley, 2010). In this study, we have focused on the effect of CIRP on cells which leads to improved recombinant protein production through investigating the binding of the N-terminal domain of CIRP to RNA ligands and the function of the C-terminal domain as well as gaining a structural understanding. Using this information as well as the investigations described into the function of CIRP in mammalian cells such as its interaction with the translational machinery and its effect on levels of recombinant protein production through these interaction, the work reported here has shown how CIRP can be used to develop new strategies for enhanced recombinant protein expression from mammalian cells grown at mildly hypothermic temperatures.

In conclusion, the work carried out in this project has provided us with data showing that the RNA binding of CIRP is mainly through the phenylalanine residues in position 3 of RNP1 and position 2 of RNP2. These residues are close in space and structurally significant, specifically the phenylalanine in the RNP1 motif. Mouse N-CIRP possesses a typical RNA binding protein structure, which is also highly similar to the human protein and displays little motion in the absence of a ligand. The N-terminal domain of CIRP as well the full-length protein binds to the translation initiation factor eIF4G and does so by translocating from nucleus to the cytoplasm upon cold stress. Full-length CIRP binds to RNA ligands and increases recombinant protein production levels more than N-CIRP when overexpressed, suggesting that the C-terminal may have a role towards ligand specificity rather than binding as N-CIRP binds the ligands tested in this project alone. The final and key finding of this project was that the recombinant protein production from CHO cells can be improved by overexpression of CIRP in cells through it interaction with eIF4G and ATR 3'UTR mRNA.

To further support the conclusions made from this project, further binding experiments should be carried out to identify RNA or protein ligands for CIRP, which will provide more knowledge on the mechanisms CIRP is involved in. As an inactive form of the RNA binding domain was generated in chapter 2, this could be used to build a full length CIRP construct where the N-terminal domain is unable to bind RNA, which would help to investigate the function of the C-terminal domain. As soluble well-behaved recombinant full-length protein could not be produced, it could be studied instead by in-cell NMR (Sakakibara et al., 2009; Tochio, 2012), which would potentially enable the study of the structure and dynamics of the full-length protein and possibly identify an interaction with a ligand. It would also be useful to develop a method to achieve a soluble purified form of the full-length protein to be able to investigate CIRP function and binding as whole in more detail. The dynamics studies of N-CIRP showed little motion in the absence of a ligand and it would be interesting to study the dynamics with a bound ligand present to see which regions, if any, are affected. Studying the dynamic properties of the mutants forms of N-CIRP generated in this project will give further insight as to how these mutations affect the protein structurally and to understand how these affect the stability of the protein some denaturation assays should be carried out. Using further NMR experiments, dihedral angle and hydrogen bonding data should be collected to refine the structure of mouse N-CIRP to a point where it can be deposited in the protein data bank. Finally, this project has uncovered some key points contributing to how CIRP functions in the cells to improve recombinant protein synthesis through interactions with the translational machinery and mRNA. To further understand the mechanism, the interactions between eIF4G and CIRP can be interrogated further by using NMR binding experiments similar to the ones shown in this thesis to identify the binding regions for CIRP on eIF4G and vice versa. A summary of the proposed mechanism that CIRP follows based on the findings of this project is shown below in figure 6.1

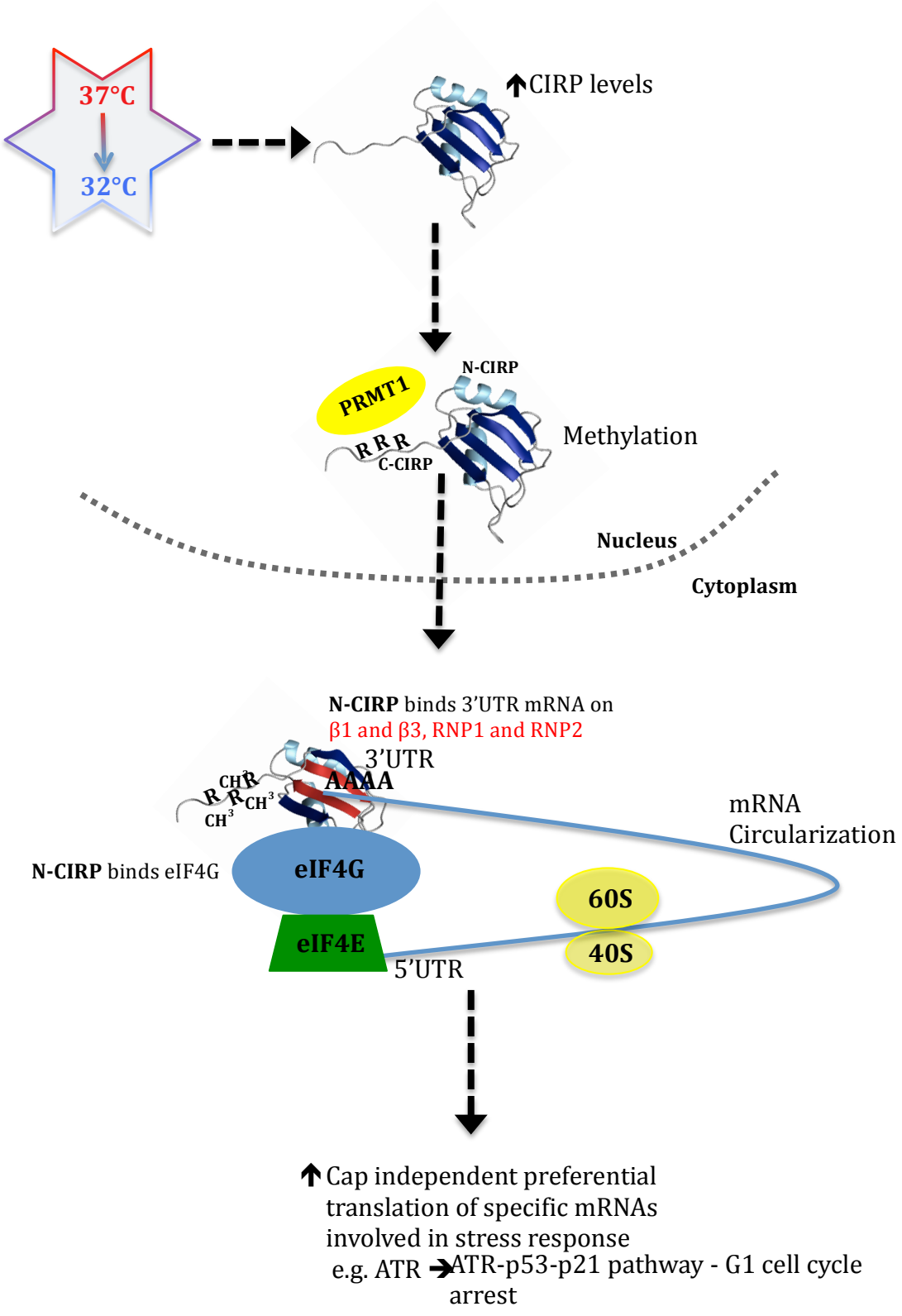

Figure 6.1 Summary of CIRP mechanism in mammalian cells' response to cold-shock based on the findings described in this thesis.

References

Abragam, A. (1961). *The principles of nuclear magnetism.* Oxford:Clarendon.

Abraham, R. T. (2001). Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes & Development*, *15*(17), 2177–96.

Ahmed, H., (2004). Principles and Reactions of Protein Extraction, Purification, and Characterization - CRC Press Book.

Aiyar, A., Xiang, Y. & Leis, J., (1996). Site-directed mutagenesis using overlap extension PCR. *Methods in molecular biology*, 57, 177–91.

Alekhina, O. M., & Vassilenko, K. S. (2012). Translation initiation in eukaryotes: versatility of the scanning model. *Biochemistry. Biokhimiia*, *77*(13), 1465–77.

Al-Fageeh, M. B., Marchant, R. J., Carden, M. J., & Smales, C. M. (2005). The Cold-Shock Response in Cultured Mammalian Cells : Harnessing the Response for the Improvement of Recombinant Protein Production. *Biotechnology and Bioengineering,* 93 (5), 829-35.

Al-Fageeh, M. B., & Smales, C. M. (2006). Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *The Biochemical Journal*, *397*(2), 247–59.

Al-Fageeh, M. B., & Smales, C. M. (2009). Cold-inducible RNA binding protein (CIRP) expression is modulated by alternative mRNAs. *RNA*, 15(6), 1164–1176.

Al-Fageeh, M. B., & Smales, C. M. (2013). Alternative promoters regulate cold inducible RNA-binding (CIRP) gene expression and enhance transgene expression in mammalian cells. *Molecular Biotechnology*, *54*(2), 238–49.

Andersen, D. C., & Krummen, L. (2002). Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology*, *13*(2), 117–123.

Andersson, S., Davis, D. L., Dahlbäck, H., Jörnvall, H., & Russell, D. W. (1989). Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *The Journal of Biological Chemistry*, *264*(14), 8222–9.

Andrec, M., Montelione, G. T., & Levy, R. M. (2000). Lipari-Szabo mapping: A graphical approach to Lipari-Szabo analysis of NMR relaxation data using reduced spectral density mapping. *Journal of Biomolecular NMR*, *18*(2), 83–100.

Aoki, K., Ishii, Y., Matsumoto, K., & Tsujimoto, M. (2002). Methylation of Xenopus CIRP2 regulates its arginine- and glycine-rich region-mediated nucleocytoplasmic distribution. *Nucleic Acids Research*, *30*(23), 5182–92.

Arakawa, T. et al., (2007). Biotechnology applications of amino acids in protein purification and formulations. *Amino acids*, 33(4), 587–605.

Baldi, L., Hacker, D. L., Adam, M., & Wurm, F. M. (2007). Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnology Letters*, *29*(5), 677–84.

Bae, W., Jones, P. G., & Inouye, M. (1997). CspA, the major cold shock protein of Escherichia coli, negatively regulates its own gene expression. *Journal of Bacteriology*, *179*(22), 7081–8.

Baghdoyan, S., Dubreuil, P., Eberlé, F., & Gomez, S. (2000). Capture of cytokine-responsive genes (NACA and RBM3) using a gene trap approach. *Blood*, *95*(12), 3750–7.

Baldi, A., Battista, T., De Luca, A., Santini, D., Rossiello, L., Baldi, F., Paggi, M. G. (2003). Identification of genes down-regulated during melanoma progression: a cDNA array study. *Experimental Dermatology*, *12*(2), 213–8.

Baneyx, F., (1999). Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology*, 10(5), 411–21.

Baranick, B. T., Lemp, N. A., Nagashima, J., Hiraoka, K., Kasahara, N., & Logg, C. R. (2008). Splicing mediates the activity of four putative cellular internal ribosome entry sites. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4733–8.

Barraud, P., & Allain, F. H.-T. (2013). Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. *Journal of Biomolecular NMR*, *55*(1), 119–38.

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P.,Clynes, M. (2011). Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. *Journal of Biotechnology*, *151*(2), 204–11.

Barsukov, I. L., & Lian, L. (1993). Structure Determination from NMR data I: Analysis of NMR data in Roberts, G. C. K. (Ed.) NMR of Macromolecules : A practial approach. *Oxford, OUP*.

Basu, A., Li, X. & Leong, S.S.J., (2011). Refolding of proteins from inclusion bodies: rational design and recipes. *Applied microbiology and biotechnology*, 92(2), 241–51.

Bax, A., Kontaxis, G., & Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. *Methods in Enzymology*, *339*, 127–74.

Bax, & Clore, G. M. (1990). ^{1}H- ^{1}H Correlation via Isotropic Mixing of 13C a New Three-Dimensional Approach for Assigning lH and 13C Spectra of 13C-Enriched Proteins. *Journal of Magnetic Resonance*, 88, 425–431

Becerra, S., Berrios, J., Osses, N., & Altamirano, C. (2012). Exploring the effect of mild hypothermia on CHO cell productivity. *Biochemical Engineering Journal*, *60*, 1–8.

Bedford, M. T. (2007). Arginine methylation at a glance. *Journal of Cell Science*, *120*(Pt 24), 4243–6.

Beer, C., Buhr, P., Hahn, H., Laubner, D., & Wirth, M. (2003). Gene expression analysis of murine cells producing amphotropic mouse leukaemia virus at a cultivation temperature of 32 and 37 degrees C. *The Journal of General Virology*, *84*(Pt 7), 1677–86.

Berjanskii, M. V, Neal, S., & Wishart, D. S. (2006). PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Research*, *34*, 63–9.

Bloch, F., Hansen, W., & Packard, M. (1946). The Nuclear Induction Experiment. *Physical Review*, *70*(7-8), 474–485.

Bloembergen, N., Purcell, E. M., & Pound, R. V. (1947). Nuclear magnetic relaxation. *Nature*, *160*(4066), 475.

Boehr, D. D., Nussinov, R., & Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, *5*, 789–796.

Borys, M. C., Dalal, N. G., Abu-Absi, N. R., Khattak, S. F., Jing, Y., Xing, Z., & Li, Z. J. (2010). Effects of culture conditions on N-glycolylneuraminic acid (Neu5Gc) content of a recombinant fusion protein produced in CHO cells. *Biotechnology and Bioengineering*, *105*(6), 1048–57.

Boshart, M., Weber, F., Jahn, G., Dorsch-Häsler, K., Fleckenstein, B., & Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell*, *41*(2), 521–30.

Broadhurst, R. W., Hardman, C. H., Thomas, J. O., & Laue, E. D. (1995). Backbone dynamics of the A-domain of HMG1 as studied by 15N NMR spectroscopy. *Biochemistry*, *34*, 16608–16617.

Brüschweiler, R. (2003). New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Current Opinion in Structural Biology*, *13*(2), 175–83.

Burgess, R.R., (2009). Refolding solubilized inclusion body proteins. *Methods in enzymology*, 463, 259–82.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J., & Rance, M. (2007). Protein NMR Spectroscopy: Principles and Practice. *Academic Press*.

Chaikam, V., & Karlson, D. T. (2010). Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. *BMB Reports*, *43*(1), 1–8.

Chappell, S. A., Owens, G. C., & Mauro, V. P. (2001). A 5' Leader of Rbm3, a Cold Stress-induced mRNA, Mediates Internal Initiation of Translation with Increased

Efficiency under Conditions of Mild Hypothermia. *The Journal of Biological Chemistry*, *276*(40), 36917–22.

Chen, K., & Tjandra, N. (2008). Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling. *Journal of the American Chemical Society*, *130*(38), 12745–51.

Cheung, Maguire, M. L., Stevens, T. J., & Broadhurst, R. W. (2010). DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance*, *202*(2), 223–33.

Chu, L., & Robinson, D. K. (2001). Industrial choices for protein production by large-scale cell culture. *Current Opinion in Biotechnology*, *12*(2), 180–187.

Clarkson, J., & Campbell, I. D. (2003). Studies of protein-ligand interactions by NMR. *Biochemical Society Transactions*, *31*, 1006–1009.

Coles, M., Heller, M., & Kessler, H. (2003). NMR-based screening technologies. *Drug Discovery Today*, *8*(17), 803–10.

Cok, S. J., Acton, S. J., Sexton, A. E., & Morrison, A. R. (2004). Identification of RNA-binding proteins in RAW 264.7 cells that recognize a lipopolysaccharide-responsive element in the 3-untranslated region of the murine cyclooxygenase-2 mRNA. *The Journal of Biological Chemistry*, *279*(9), 8196–205.

Corish, P., & Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering Design and Selection*, *12*(12), 1035–1040.

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*(3), 289–302.

Craig, N. L. (1988). The mechanism of conservative site-specific recombination. *Annual Review of Genetics*, *22*, 77–105.

Danno, S., Nishiyama, H., Higashitsuji, H., Yokoi, H., Xue, J. H., Itoh, K., Fujita, J. (1997). Increased transcript level of RBM3, a member of the glycine-rich RNA-binding protein family, in human cells in response to cold stress. *Biochemical and Biophysical Research Communications*, *236*(3), 804–7.

Danno, S., Itoh, K., Matsuda, T., & Fujita, J. (2000). Decreased expression of mouse Rbm3, a cold-shock protein, in Sertoli cells of cryptorchid testis. *The American Journal of Pathology*, *156*(5), 1685–92.

Daragan, V. A., Ilyina, E. E., Fields, C. G., Fields, G. B., & Mayo, K. H. (1997). Backbone and side-chain dynamics of residues in a partially folded beta-sheet peptide from platelet factor-4. *Protein Science*, *6*(2), 355–63.

De Leeuw, F., Zhang, T., Wauquier, C., Huez, G., Kruys, V., & Gueydan, C. (2007). The cold-inducible RNA-binding protein migrates from the nucleus to cytoplasmic stress granules by a methylation-dependent mechanism and acts as a translational repressor. *Experimental Cell Research*, *313*(20), 4130–44.

Derry, J. M., Kerns, J. A., & Francke, U. (1995). RBM3, a novel human gene in Xp11.23 with a putative RNA-binding domain. *Human Molecular Genetics*, *4*(12), 2307–11.

Doray, B., Chen, C.D. & Kemper, B., (2001). N-terminal deletions and His-tag fusions dramatically affect expression of cytochrome p450 2C2 in bacteria. *Archives of biochemistry and biophysics*, 393(1), 143–53.

Dresios, J., Aschrafi, A., Owens, G. C., Vanderklish, P. W., Edelman, G. M., & Mauro, V. P. (2005). Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 1865–70.

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., & Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annual Review of Biochemistry*, *62*, 289–321.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics & Modelling*, *19*(1), 26–59.

Eliseeva, I. A., Kim, E. R., Guryanov, S. G., Ovchinnikov, L. P., & Lyabin, D. N. (2011). Y-box-binding protein 1 (YB-1) and its functions. *Biochemistry, 76*(13), 1402–33.

Ermolenko, D. N., & Makhatadze, G. I. (2002). Bacterial cold-shock proteins. *Cellular and Molecular Life Sciences, 59*(11), 1902–13.

Evdokimova, V. M., Kovrigina, E. A., Nashchekin, D. V, Davydova, E. K., Hershey, J. W., & Ovchinnikov, L. P. (1998). The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro. *The Journal of Biological Chemistry*, *273*(6), 3574–81.

Evdokimova, V. M., & Ovchinnikov, L. P. (1999). Translational regulation by Y-box transcription factor: involvement of the major mRNA-associated protein, p50. *The International Journal of Biochemistry & Cell Biology*, *31*(1), 139–49.

Fischer, M. W. F., Majumdar, A., & Zuiderweg, E. R. P. (1998). Protein NMR relaxation: theory, applications and outlook. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *33*(3), 207–272.

Fogolín, M. B., Wagner, R., Etcheverrigaray, M., & Kratje, R. (2004). Impact of temperature reduction and expression of yeast pyruvate carboxylase on hGM-CSF-producing CHO cells. *Journal of Biotechnology*, *109*(1-2), 179–91.

Follis, A. V., Galea, C. A., & Kriwacki, R. W. (2012). Intrinsic protein flexibility in regulation of cell proliferation: advantages for signaling and opportunities for novel therapeutics. *Advances in Experimental Medicine and Biology*, *725*, 27–49.

Fornace, A. J., Alamo, I., & Hollander, M. C. (1988). DNA damage-inducible transcripts in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(23), 8800–4.

Fox, S. R., Patel, U. A., Yap, M. G. S., & Wang, D. I. C. (2004). Maximizing interferon-gamma production by Chinese hamster ovary cells through temperature shift optimization: experimental and modeling. *Biotechnology and Bioengineering*, *85*(2), 177–84.

Fox, S. R., Tan, H. K., Tan, M. C., Wong, S. C. N. C., Yap, M. G. S., & Wang, D. I. C. (2005). A detailed understanding of the enhanced hypothermic productivity of interferon-gamma by Chinese-hamster ovary cells. *Biotechnology and Applied Biochemistry*, *41*(Pt 3), 255–64.

Fujita, J. (1999). Cold shock response in mammalian cells. *Journal of Molecular Microbiology and Biotechnology*, *1*(2), 243–55.

Fussenegger, M. (2001). The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies. *Biotechnology Progress*, *17*(1), 1–51.

Galea, C. A., Nourse, A., Wang, Y., Sivakolundu, S. G., Heller, W. T., & Kriwacki, R. W. (2008). Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *Journal of Molecular Biology*, *376*(3), 827–38.

Geoghegan, K.F. et al., (1999). Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Analytical biochemistry*, 267(1), 169–84.

Gammell, P., Barron, N., Kumar, N., & Clynes, M. (2007). Initial identification of low temperature and culture stage induction of miRNA expression in suspension CHO-K1 cells. *Journal of Biotechnology*, *130*(3), 213–8.

Gaudreault, I., Guay, D., & Lebel, M. (2004). YB-1 promotes strand separation in vitro of duplex DNA containing either mispaired bases or cisplatin modifications, exhibits endonucleolytic activities and binds several DNA repair proteins. *Nucleic Acids Research*, *32*(1), 316–27.

Ginzinger, S. W., & Coles, M. (2009). SimShiftDB; local conformational restraints derived from chemical shift similarity searches on a large synthetic database. *Journal of Biomolecular NMR*, *43*(3), 179–85.

Giuliodori, A. M., Brandi, A., Gualerzi, C. O., & Pon, C. L. (2004). Preferential translation of cold-shock mRNAs during cold adaptation. *RNA*, *10*(2), 265–76.

Givalos, N., Gakiopoulou, H., Skliri, M., Bousboukea, K., Konstantinidou, A. E., Korkolopoulou, P., Karatzas, G. (2007). Replication protein A is an independent prognostic indicator with potential therapeutic implications in colon cancer. *Modern Pathology*, *20*(2), 159–66.

Golovanov, A.P. et al., (2004). A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society*, 126(29), 8933–9.

Goswami, J., Sinskey, A. J., Steller, H., Stephanopoulos, G. N., & Wang, D. I. (1999). Apoptosis in batch cultures of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *62*(6), 632–40.

Gräslund, S. et al., (2008). Protein production and purification. *Nature methods*, 5(2), 135–46.

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

Grishaev, A., & Llinás, M. (2005). Protein structure elucidation from minimal NMR data: the CLOUDS approach. *Methods in Enzymology*, *394*, 261–95.

Gritz, L., & Davies, J. (1983). Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in Escherichia coli and Saccharomyces cerevisiae. *Gene*, *25*(2-3), 179–88.

Gronwald, W., Brunner, K., Kirchhöfer, R., Nasser, A., Trenner, J., Ganslmeier, B.,Kalbitzer, H. (2004). AUREMOL, a New Program for the Automated Structure Elucidation of Biological Macromolecules. *Bruker Rep.*, *154/155*, 11 – 14.

Grzesiekt, S., & Bax, A. (1992a). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, *99*(1), 201–207.

Grzesiekt, S., & Bax, A. (1992b). Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *Journal of the American Chemical Society, 114*(16), 6291–6293.

Gualerzi, C. O., Maria Giuliodori, A., & Pon, C. L. (2003). Transcriptional and Post-transcriptional Control of Cold-shock Genes. *Journal of Molecular Biology*, *331*(3), 527–539.

Guntert, P. (1997). Calculating Protein Structures from NMR data in Reid, D. G. (Ed.) Protein NMR techniques. *Totowa, NJ., Humana Press*.

Guntert, P. (2003). Automated NMR protein structure calculation. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *43*(3-4), 105–125.

Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods in Molecular Biology* , *278*, 353–78.

Guo, X., Wu, Y., & Hartley, R. S. (2010). Cold-inducible RNA-binding protein contributes to human antigen R and cyclin E1 deregulation in breast cancer. *Molecular Carcinogenesis*, *49*(2), 130–40.

Hammond, J. B., & Kruger, N. J. (1988). The bradford method for protein quantitation. *Methods in Molecular Biology*, *3*, 25–32.

Harris, A. L. (2002). Hypoxia--a key regulatory factor in tumour growth. *Nature Reviews. Cancer*, *2*(1), 38–47.

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*, 964–972.

Herschlag, D. (1995). RNA chaperones and the RNA folding problem. *The Journal of Biological Chemistry*, *270*(36), 20871–4.

Hinton, T. M., Coldwell, M. J., Carpenter, G. a, Morley, S. J., & Pain, V. M. (2007). Functional analysis of individual binding activities of the scaffold protein eIF4G. *The Journal of Biological Chemistry*, *282*(3), 1695–708.

Ho, S.N. et al., (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77(1), 51–59.

Hofmann, S., Cherkasova, V., Bankhead, P., Bukau, B., & Stoecklin, G. (2012). Translation suppression promotes stress granule formation and cell survival in response to cold shock. *Molecular Biology of the Cell*, *23*(19), 3786–800.

Holcik, M., & Sonenberg, N. (2005). Translational control in stress and apoptosis. *Nature Reviews. Molecular Cell Biology*, *6*(4), 318–27.

Holcik, M., & Pestova, T. V. (2007). Translation mechanism and regulation: old players, new concepts. Meeting on translational control and non-coding RNA. *EMBO Reports*, *8*(7), 639–43.

Holm, L., & Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics*, *16*(6), 566–567.

Horn, G., Hofweber, R., Kremer, W., & Kalbitzer, H. R. (2007). Structure and function of bacterial cold shock proteins. *Cellular and Molecular Life Sciences*, *64*(12), 1457–70.

Hsieh, S., (2010). Structural and functional characterization of the mammalian cold shock protein CIRP. PhD Thesis.

Huang, Y.-M., Hu, W., Rustandi, E., Chang, K., Yusuf-Makagiansar, H., & Ryll, T. (2010). Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology Progress*, *26*(5), 1400–10.

Hughes, T. A. (2006). Regulation of gene expression by alternative untranslated regions. *Trends in Genetics*, *22*(3), 119–22.

Hwang, S.-J., Yoon, S. K., Koh, G. Y., & Lee, G. M. (2011). Effects of culture temperature and pH on flag-tagged COMP angiopoietin-1 (FCA1) production from recombinant CHO cells: FCA1 aggregation. *Applied Microbiology and Biotechnology*, *91*(2), 305–15.

Idicula-Thomas, S. & Balaji, P. V, (2005). Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein science*, 14(3), 582–92.

Igumenova, T. I., Frederick, K. K., & Wand, A. J. (2006). Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chemical Reviews*, *106*(5), 1672–99.

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*, *323*(3), 573–84.

Ise, T., Nagatani, G., Imamura, T., Kato, K., Takano, H., Nomoto, M., … Kohno, K. (1999). Transcription factor Y-box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen. *Cancer Research*, *59*(2), 342–6.

Ishima, R., & Torchia, D. a. (2000). Protein dynamics from NMR. *Nature Structural Biology*, *7*(9), 740–3.

Jacob, E., & Unger, R. (2007). A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics*, *23*(2), e225–30.

Jacobsen, N. E. (2007). *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. *Quadrature* (p. 643).

Jana, S. & Deb, J.K., (2005). Strategies for efficient production of heterologous proteins in Escherichia coli. *Applied microbiology and biotechnology*, 67(3), 289–98.

Janson & Ryden, (1998). Protein purification: Principles, high-resolution methods, and applications.

Jarymowycz, V. A., & Stone, M. J. (2006). Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chemical Reviews*, *106*(5), 1624–71.

Jayapal, K. P., Wlaschin, K. F., Hu, W.-S. H., & Yap, M. G. S. (2007). Recombinant Protein Therapeutics from CHO Cells — 20 Years and Counting. *CEP Magazine*, 40–47.

Jenkins, N., & Hovey, A. (1993). Temperature control of growth and productivity in mutant Chinese hamster ovary cells synthesizing a recombinant protein. *Biotechnology and Bioengineering*, *42*(9), 1029–36.

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., & Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(17), 7028–33.

Jiang, W., Hou, Y., & Inouye, M. (1997). CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *The Journal of Biological Chemistry*, *272*(1), 196–202.

Joshi, M., Kulkarni, A., & Pal, J. K. (2013). Small molecule modulators of eukaryotic initiation factor 2α kinases, the key regulators of protein synthesis. *Biochimie*, *95*(11), 1980–90.

Jungbauer, A. & Kaar, W., (2007). Current status of technical protein refolding. *Journal of biotechnology*, 128(3), 587–96.

Kay, L. E. (2005). NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance*, *173*(2), 193–207.

Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., & Forman-Kay, J. D. (1993). A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H2O Samples of Proteins. *Journal of Magnetic Resonance, Series B*, *101*(3), 333–337.

Keeler, J. (2005). Understanding NMR Spectroscopy. *Chichester, John Wiley*.

Keiper, B. D., Gan, W., & Rhoads, R. E. (1999). Protein synthesis initiation factor 4G. *The International Journal of Biochemistry & Cell Biology*, *31*(1), 37–41.

Kempf, J. G., & Loria, J. P. (2003). Protein dynamics from solution NMR: theory and applications. *Cell Biochemistry and Biophysics*, *37*(3), 187–211.

Kim, J. Y., Kim, Y.-G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology and Biotechnology*, *93*(3), 917–30.

Kim, S., Merrill, B. M., Rajpurohit, R., Kumar, A., Stone, K. L., Papov, V. V, Williams, K. R. (1997). Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif. *Biochemistry*, *36*(17), 5185–92.

Kita, H. (2002). Modulation of polyglutamine-induced cell death by genes identified by expression profiling. *Human Molecular Genetics*, *11*(19), 2279–2287.

Klann, E., & Dever, T. E. (2004). Biochemical mechanisms for translational regulation in synaptic plasticity. *Nature Reviews. Neuroscience*, *5*(12), 931–42.

Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta*, *1814*(8), 942–68.

Koradi, R., Billeter, M., & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *Journal of Molecular Graphics*, *14*(1), 51–5, 29–32.

Korneeva, N. L., Lamphear, B. J., Hennigan, F. L., & Rhoads, R. E. (2000). Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1. *The Journal of Biological Chemistry*, *275*(52), 41369–76.

Kou, T.-C., Fan, L., Zhou, Y., Ye, Z.-Y., Liu, X.-P., Zhao, L., & Tan, W.-S. (2011). Detailed understanding of enhanced specific productivity in Chinese hamster ovary cells at low culture temperature. *Journal of Bioscience and Bioengineering*, *111*(3), 365–9.

Krieg, P. A., & Melton, D. A. (1984). Functional messenger RNAs are produced by SP6 in vitro transcription of cloned cDNAs. *Nucleic Acids Research*, *12*, 7057–7070.

Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., & Wright, P. E. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(21), 11504–9.

Kumar, N., Gammell, P., & Clynes, M. (2007). Proliferation control strategies to improve productivity and survival during CHO based production culture : A summary of recent methods employed and the effects of proliferation control in product secreting CHO cell lines. *Cytotechnology*, *53*(1-3), 33–46.

Kumar, N., Gammell, P., Meleady, P., Henry, M., & Clynes, M. (2008). Differential protein expression following low temperature culture of suspension CHO-K1 cells. *BMC Biotechnology*, *8*, 42.

Lamphear, B. J., Kirchweger, R., Skern, T., & Rhoads, R. E. (1995). Mapping of Functional Domains in Eukaryotic Protein Synthesis Initiation Factor 4G (eIF4G) with Picornaviral Proteases: Implications for cap-dependent and cap-independent translational initiation. *Journal of Biological Chemistry*, *270*(37), 21975–21983.

Lee, A. L., Volkman, B. F., Robertson, S. A., Rudner, D. Z., Barbash, D. A., Cline, T. W.,Wemmer, D. E. (1997). Chemical shift mapping of the RNA-binding interface of the multiple-RBD protein sex-lethal. *Biochemistry*, *36*, 14306–14317.

Li, Z., & Srivastava, P. (2004). Heat-shock proteins. *Current Protocols in Immunology*.

Lim, Y., Wong, N. S. C., Lee, Y. Y., Ku, S. C. Y., Wong, D. C. F., & Yap, M. G. S. (2010). Engineering mammalian cells in bioprocessing - current achievements and future perspectives. *Biotechnology and Applied Biochemistry*, *55*(4), 175–89.

Ling, M.M. & Robinson, B.H., (1997). Approaches to DNA mutagenesis: an overview. *Analytical biochemistry*, 254(2), 157–78.

Linge, J. P., O'Donoghue, S. I., & Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods in Enzymology*, *339*, 71–90.

Lipari, G., & Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. *Journal of the American Chemical Society*, *104*, 4546–4559.

Liu, Q., & Dreyfuss, G. (1995). In vivo and in vitro arginine methylation of RNA-binding proteins. *Molecular and Cellular Biology*, *15*(5), 2800–8.

Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., & Yan, J. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Scientific Reports*, *3*, 2054.

Lleonart, M. E. (2010). A new generation of proto-oncogenes: cold-inducible RNA binding proteins. *Biochimica et Biophysica Acta*, *1805*(1), 43–52.

López-Méndez, B., & Güntert, P. (2006). Automated protein structure determination from NMR spectra. *Journal of the American Chemical Society*, *128*, 13112–13122.

Los, D. A., & Murata, N. (2004). Membrane fluidity and its roles in the perception of environmental signals. *Biochimica et Biophysica Acta*, *1666*(1-2), 142–57.

Luan, C.-H. et al., (2004). High-throughput expression of C. elegans proteins. *Genome research*, 14(10B), 2102–10.

Manival, X., Ghisolfi-Nieto, L., Joseph, G., Bouvet, P., & Erard, M. (2001). RNA-binding strategies common to cold-shock domain- and RNA recognition motif-containing proteins. *Nucleic Acids Research*, *29*(11), 2223–33.

Mantovani, R. (1998). A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Research*, *26*(5), 1135–43.

Maquat, L. E., Tarn, W.-Y., & Isken, O. (2010). The pioneer round of translation: features and functions. *Cell*, *142*(3), 368–74.

Marchant, R. J., Al-Fageeh, M. B., Underhill, M. F., Racher, A. J., & Smales, C. M. (2008). Metabolic rates, growth phase, and mRNA levels influence cell-specific antibody production levels from in vitro-cultured mammalian cells at sub-physiological temperatures. *Molecular Biotechnology*, *39*(1), 69–77.

Maris, C., Dominguez, C. & Allain, F.H.-T., (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS journal*, 272(9), 2118–31.

Marston, F.A.O. & Hartley, D.L., (1990). Solubilisation of Protein Aggregates. *Methods in Enzymology,* 182, 264-276.

Masterton, R. J., Roobol, A., Al-Fageeh, M. B., Carden, M. J., & Smales, C. M. (2010). Post-translational events of a model reporter protein proceed with higher fidelity and accuracy upon mild hypothermic culturing of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *105*(1), 215–20.

Masterton, R., & Smales, C. M. (2014). The impact of process temperature on mammalian cell lines and the implications for the production of recombinant proteins in CHO cells, *2*, 49–61.

Masuda, T., Itoh, K., Higashitsuji, H., Higashitsuji, H., Nakazawa, N., Sakurai, T., Fujita, J. (2012). Cold-inducible RNA-binding protein (Cirp) interacts with Dyrk1b/Mirk and promotes proliferation of immature male germ cells in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(27), 10885–90.

Maurer, T. (2005). NMR studies of protein-ligand interactions. *Methods in Molecular Biology*, *305*, 197–214.

Maxwell, K.L. et al., (2003). Refolding out of guanidine hydrochloride is an effective approach for high-throughput structural studies of small proteins. *Protein science : a publication of the Protein Society*, 12(9), 2073–80.

Mayr, C., & Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, *138*(4), 673–84.

McBride, A. E., & Silver, P. A. (2001). State of the arg: protein methylation at arginine comes of age. *Cell*, *106*(1), 5–8.

Mcconell, J. (1987). The theory of nuclear magnetic resonance in liquids. *Cambridge,Cambridge University Press.*

McIntosh, L. P., & Dahlquist, F. W. (1990). Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Quarterly Reviews of Biophysics*, *23*(1), 1–38.

Middelberg, A.P., (2002). Preparative protein refolding. *Trends in Biotechnology*, 20(10), 437–443.

Mielke, S. P., & Krishnan, V. V. (2009). Characterization of protein secondary structure from NMR chemical shifts. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *54*(3-4), 141–165.

Miernyk, J. a, & Thelen, J. J. (2008). Biochemical approaches for discovering protein-protein interactions. *The Plant Journal : For Cell and Molecular Biology*, *53*(4), 597–609.

Mihailovich, M., Militti, C., Gabaldón, T., & Gebauer, F. (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(2), 109–18.

Millner, P., (1999). High Resolution Chromatography. *Oxford University Press*.

Mittag, T., Kay, L.E. & Forman-Kay, J.D., (2010). Protein dynamics and conformational disorder in molecular recognition. *Journal of molecular recognition*, 23(2), 105–16.

Mittermaier, A. K., & Kay, L. E. (2009). Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences*, *34*(12), 601–11.

Monti, M., Orrù, S., Pagnozzi, D., & Pucci, P. (2005). Interaction proteomics. *Bioscience Reports*, *25*(1-2), 45–56.

Murata, Y., Homma, T., Kitagawa, E., Momose, Y., Sato, M. S., Odani, M., Iwahashi, H. (2006). Genome-wide expression analysis of yeast response during exposure to 4 degrees C. *Extremophiles : Life under Extreme Conditions*, *10*(2), 117–28.

Müller, F., Demény, M. A., & Tora, L. (2007). New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *The Journal of Biological Chemistry*, *282*(20), 14685–9.

Myers, J.K. & Pace, C.N., (1996). Hydrogen bonding stabilizes globular proteins. *Biophysical journal*, 71(4), 2033–9.

Nekrasov, M. P., Ivshina, M. P., Chernov, K. G., Kovrigina, E. A., Evdokimova, V. M., Thomas, A. A. M., Ovchinnikov, L. P. (2003). The mRNA-binding protein YB-1 (p50) prevents association of the eukaryotic initiation factor eIF4G with mRNA and inhibits protein synthesis at the initiation stage. *The Journal of Biological Chemistry*, *278*(16), 13936–43.

Nelson, J. A., Reynolds-Kohler, C., & Smith, B. A. (1987). Negative and positive regulation by a short segment in the 5'-flanking region of the human cytomegalovirus major immediate-early gene. *Molecular and Cellular Biology*, *7*(11), 4125–9.

Neudecker, P., Lundstrom, P., & Kay, L. E. (2009). Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding. *Biophysical Journal*, *96*, 2045–2054.

Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *Journal of Molecular Biology*, *245*(5), 645–60.

Nishiyama, H., Higashitsuji, H., Yokoi, H., Itoh, K., Danno, S., Matsuda, T., & Fujita, J. (1997). Cloning and characterization of human CIRP (cold-inducible RNA-binding protein) cDNA and chromosomal assignment of the gene. *Gene*, *204*(1-2), 115–20.

Nishiyama, H., Itoh, K., Kaneko, Y., Kishishita, M., Yoshida, O., & Fujita, J. (1997). A glycine-rich RNA-binding protein mediating cold-inducible suppression of mammalian cell growth. *The Journal of Cell Biology*, *137*(4), 899–908.

Nishiyama, H. et al., (1998). Decreased expression of cold-inducible RNA-binding protein (CIRP) in male germ cells at elevated temperature. *The American journal of pathology*, 152(1), 289–96.

Noggle, & Schirmer, R. E. (1971). The Nuclear Overhauser Effect. *Academic Press:New York*.

Noike, T., Miwa, S., Soeda, J., Kobayashi, A., & Miyagawa, S. (2008). Increased expression of thioredoxin-1, vascular endothelial growth factor, and redox factor-1 is associated with poor prognosis in patients with liver metastasis from colorectal cancer. *Human Pathology*, *39*(2), 201–8.

Oezguen, N., Adamian, L., Xu, Y., Rajarathnam, K., & Braun, W. (2002). Automated assignment and 3D structure calculations using combinations of 2D homonuclear and 3D heteronuclear NOESY spectra. *Journal of Biomolecular NMR*, *22*(3), 249–63.

O'Gorman, S., Fox, D. T., & Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science*, *251*(4999), 1351–5.

Ohsaka, Y., Ohgiya, S., Hoshino, T., & Ishizaki, K. (2002). Phosphorylation of c-Jun N-terminal kinase in human hepatoblastoma cells is transiently increased by cold exposure and further enhanced by subsequent warm incubation of the cells. *Cellular Physiology and Biochemistry , 12*(2-3), 111–8.

Oldfield, C. J., & Dunker, A. K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry*, *83*, 553–84.

Oswald, T. et al., (1994). Comparison of N-terminal affinity fusion domains: effect on expression level and product heterogeneity of recombinant restriction endonuclease EcoRV. *Applied microbiology and biotechnology*, 42(1), 73–7.

Pace, C.N. et al., (1996). Forces contributing to the conformational stability of proteins. *FASEB journal*, 10(1), 75–83.

Palmer, A. G. (1997). Probing molecular motion by NMR. *Current Opinion in Structural Biology*, *7*(5), 732–7.

Palmer, A. G. (2001). Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annual Review of Biophysics and Biomolecular Structure*, *30*, 129–55.

Panadero, J., Pallotti, C., Rodríguez-Vargas, S., Randez-Gil, F., & Prieto, J. A. (2006). A downshift in temperature activates the high osmolarity glycerol (HOG) pathway, which determines freeze tolerance in Saccharomyces cerevisiae. *The Journal of Biological Chemistry*, *281*(8), 4638–45.

Pancsa, R. & Fuxreiter, M., (2012). Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB life*, 64(6), 513–20.

Peng, & Wagner, G. (1992a). Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *Journal of Magnetic Resonance (1969)*, *332*, 308–332.

Peng, & Wagner, G. (1992b). Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. *Biochemistry*, *31*(36), 8571–86.

Phadtare, S., & Severinov, K. (2005). Nucleic acid melting by Escherichia coli CspE. *Nucleic Acids Research*, *33*(17), 5583–90.

Pham, P. L., Kamen, A., & Durocher, Y. (2006). Large-scale transfection of mammalian cells for the fast production of recombinant protein. *Molecular Biotechnology*, *34*(2), 225–37.

Pilotte, J., Cunningham, B. A., Edelman, G. M., & Vanderklish, P. W. (2009). Developmentally regulated expression of the cold-inducible RNA-binding motif protein 3 in euthermic rat brain. *Brain Research*, *1258*, 12–24.

Piotto, M., Saudek, V., & Sklen, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR*, *2*, 661–665.

Qiang, X., Yang, W.-L., Wu, R., Zhou, M., Jacob, A., Dong, W., Wang, P. (2013). Cold-inducible RNA-binding protein (CIRP) triggers inflammatory responses in hemorrhagic shock and sepsis. *Nature Medicine*, *19*(11), 1489–95.

Rajagopal, P., Waygood, E. B., Reizer, J., Saier, M. H., & Klevit, R. E. (1997). Demonstration of protein-protein interaction specificity by NMR chemical shift mapping. *Protein Science*, *6*, 2624–2627.

Ramón, A., Señorale-Pose, M. & Marín, M., (2014). Inclusion bodies: not that bad…. *Frontiers in microbiology*, 5, 56.

Rao, P. N., & Engelberg, J. (1965). HELA Cells: Effects of Temperature on the Life Cycle. *Science*, *148*(3673), 1092–4.

Ren, L., Emery, D., Kaboord, B., Chang, E., & Qoronfleh, M. W. (2003). Improved immunomatrix methods to detect protein:protein interactions. *Journal of Biochemical and Biophysical Methods*, *57*(2), 143–157.

Rezaei, M., Zarkesh-Esfahani, S. H., & Gharagozloo, M. (2013). The effect of different media composition and temperatures on the production of recombinant human growth hormone by CHO cells. *Research in Pharmaceutical Sciences*, *8*(3), 211–7.

Rio, D. C. (2014). Electrophoretic mobility shift assays for RNA-protein complexes. *Cold Spring Harbor Protocols*, *2014*(4), 435–40.

Rollin, B. V, & Hatton, J. (1947). Nuclear magnetic resonance at low temperatures. *Nature*, *159*(4032), 201.

Roobol, A., Carden, M. J., Newsam, R. J., & Smales, C. M. (2009). Biochemical insights into the mechanisms central to the response of mammalian cells to cold stress and subsequent rewarming. *The FEBS Journal*, *276*(1), 286–302.

Roobol, A., Roobol, J., Carden, M. J., Bastide, A., Willis, A. E., Dunn, W. B.,Smales, C. M. (2011). ATR (ataxia telangiectasia mutated- and Rad3-related kinase) is activated by mild hypothermia in mammalian cells and subsequently activates p53. *The Biochemical Journal*, *435*(2), 499–508.

Rosenberg, A.H. et al., (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, 56(1), 125–35.

Russell, R. (2008). RNA misfolding and the action of chaperones. *Frontiers in Bioscience : A Journal and Virtual Library*, *13*, 1–20.

Ryan, B.J., (2011). Avoiding proteolysis during protein chromatography. *Methods in molecular biology*, 681, 61–71.

Sahara, T., Goda, T., & Ohgiya, S. (2002). Comprehensive expression analysis of time-dependent genetic responses in yeast cells to low temperature. *The Journal of Biological Chemistry*, *277*(51), 50015–21.

Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Ito, Y. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature*, *458*(7234), 102–5.

Sakurai, T., Itoh, K., Higashitsuji, H., Nonoguchi, K., Liu, Y., Watanabe, H., Fujita, J. (2006). Cirp protects against tumor necrosis factor-alpha-induced apoptosis via activation of extracellular signal-regulated kinase. *Biochimica et Biophysica Acta*, *1763*(3), 290–5.

Sati, S.P. et al., (2002). Extra terminal residues have a profound effect on the folding and solubility of a Plasmodium falciparum sexual stage-specific protein over-expressed in Escherichia coli. *European Journal of Biochemistry*, 269(21), 5259–5263.

Sauer, B. (1994). Site-specific recombination: developments and applications. *Current Opinion in Biotechnology*, *5*(5), 521–7.

Sawicka, K., Bushell, M., Spriggs, K. A., & Willis, A. E. (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochemical Society Transactions*, *36*, 641–7.

Schatz, S. M., Kerschbaumer, R. J., Gerstenbauer, G., Kral, M., Dorner, F., & Scheiflinger, F. (2003). Higher expression of Fab antibody fragments in a CHO cell line at reduced temperature. *Biotechnology and Bioengineering*, *84*(4), 433–8.

Schenborn, E. T., & Mierendorf, R. C. (1985). A novel transcription property of SP6 and T7 RNA polymerases: Dependence on template structure. *Nucleic Acids Research*, *13*, 6223–6236.

Schlitz, H., (1984). Protein Purification: Principles and Practice. *Biochemical Education*, 12(3), 143.

Shammas, S. L., Travis, A. J., & Clarke, J. (2013). Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX. *The Journal of Physical Chemistry. 117*(42), 13346–56.

Sheikh, M.S., (1997). Identification of Several Human Homologs of Hamster DNA Damage-inducible Transcripts. Cloning and Characterization of a Novel UV-inducible cDNA that codes for a putative RNA-Binding Protein. *Journal of Biological Chemistry*, 272(42), 26720–26726.

Shen, Y., Delaglio, F., Cornilescu, G., & Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, *44*(4), 213–23.

Shukla, D. & Trout, B.L., (2011). Understanding the synergistic effect of arginine and glutamic acid mixtures on protein solubility. *The journal of physical chemistry,* 115(41), 11831–9.

Simeonov, A., Jadhav, A., Thomas, C. J., Wang, Y., Huang, R., Southall, N. T., Inglese, J. (2008). Fluorescence Spectroscopic Profiling of Compound Libraries. *Journal of Medicinal Chemistry,* 51(8), 2363-71.

Simpson, R.J., (2009). Basic Methods in Protein Purification and Analysis: A Laboratory Manual. *Cold Spring Harbor Labaratory Press.*

Singh, S.M. & Panda, A.K., (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of bioscience and bioengineering*, 99(4), 303–10.

Smart, F., Aschrafi, A., Atkins, A., Owens, G. C., Pilotte, J., Cunningham, B. A., & Vanderklish, P. W. (2007). Two isoforms of the cold-inducible mRNA-binding protein RBM3 localize to dendrites and promote translation. *Journal of Neurochemistry*, *101*(5), 1367–79.

Smith, H.E., (2007). The transcriptional response of Escherichia coli to recombinant protein insolubility. *Journal of structural and functional genomics*, 8(1), 27–35.

Smock, R. G., & Gierasch, L. M. (2009). Sending signals dynamically. *Science*, *324*, 198–203.

Sonna, L. a, Fujita, J., Gaffin, S. L., & Lilly, C. M. (2002). Invited review: Effects of heat and cold stress on mammalian gene expression. *Journal of Applied Physiology*, *92*(4), 1725–42.

Sorensen, H.P. & Mortensen, K.K., (2005a). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of biotechnology*, 115(2), 113–28.

Sorensen, H.P. & Mortensen, K.K., (2005b). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial cell factories*, 4(1), 1.

Spronk, C. a. E. M., Nabuurs, S. B., Krieger, E., Vriend, G., & Vuister, G. W. (2004). Validation of protein structures derived by NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *45*(3-4), 315–337.

Studier, F.W. et al., (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in enzymology*, 185, 60–89.

Studier, F.W. & Moffatt, B.A., (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*, 189(1), 113–30.

Sumitomo, Y., Higashitsuji, H., Higashitsuji, H., Liu, Y., Fujita, T., Sakurai, T., Fujita, J. (2012). Identification of a novel enhancer that binds Sp1 and contributes to induction of cold-inducible RNA-binding protein (cirp) expression in mammalian cells. *BMC Biotechnology*, *12*, 72.

Sunley, K., Tharmalingam, T., & Butler, M. (2008). CHO cells adapted to hypothermic growth produce high yields of recombinant beta-interferon. *Biotechnology Progress*, *24*(4), 898–906.

Sureban, S. M., Ramalingam, S., Natarajan, G., May, R., Subramaniam, D., Bishnupuri, K. S., Anant, S. (2008). Translation regulatory factor RBM3 is a proto-oncogene that prevents mitotic catastrophe. *Oncogene*, *27*(33), 4544–56.

Sutcliffe, M. (1993). Structure determination from NMR data II: computational approaches. in Roberts, G. C. K. (Ed.) NMR of Macromolecules: A Practical Approach. *Oxford, Oxford University Press*.

Takaya, M., Nomura, Y., & Kato, I. (2002). Cold-inducable expression vector USA: Takara Shuzo Co. Ltd.

Takeuchi, K., & Wagner, G. (2006). NMR studies of protein interactions. *Current Opinion in Structural Biology*, *16*(1), 109–17.

Tan, J. S., Mohandas, N., & Conboy, J. G. (2006). High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function. *Blood*, *107*(6), 2557–61.

Tan, H. K., Lee, M. M., Yap, M. G. S., & Wang, D. I. C. (2008). Overexpression of cold-inducible RNA-binding protein increases interferon-gamma production in Chinese-hamster ovary cells. *Biotechnology and Applied Biochemistry*, *49*(Pt 4), 247–57.

Tang, J., Frankel, A., Cook, R. J., Kim, S., Paik, W. K., Williams, K. R., Herschman, H. R. (2000). PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *The Journal of Biological Chemistry*, *275*(11), 7723–30.

Tantos, A., Friedrich, P., & Tompa, P. (2009). Cold stability of intrinsically disordered proteins. *FEBS Letters*, *583*(2), 465–9.

Thorne, N., Inglese, J., & Auld, D. S. (2010). Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chemistry & Biology*, *17*(6), 646–57.

Tochio, H. (2012). Watching protein structure at work in living cells using NMR spectroscopy. *Current Opinion in Chemical Biology*, *16*(5-6), 609–13.

Tompa, P., Prilusky, J., Silman, I., & Sussman, J. L. (2008). Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins*, *71*(2), 903–9.

Tompa, P., & Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. *FASEB Journal*, *18*(11), 1169–75.

Tsumoto, K. et al., (2010). Non-denaturing solubilization of inclusion bodies. *Current pharmaceutical biotechnology*, 11(3), 309–12.

Vasina, J.A. & Baneyx, F., (1996). Recombinant protein expression at low temperatures under the transcriptional control of the major Escherichia coli cold shock promoter cspA. *Applied and environmental microbiology*, 62, 1444–1447.

Vergara, M., Becerra, S., Berrios, J., Osses, N., Reyes, J., Rodríguez-Moyá, M., Altamirano, C. (2014). Differential effect of culture temperature and specific growth rate on CHO cell behavior in chemostat culture. *PloS One*, *9*(4), 93865.

Wagner, G. (1993). Prospects for NMR of large proteins. *Journal of Biomolecular NMR*, *3*(4), 375-85.

Wagstaff, J. L., Masterton, R. J., Povey, J. F., Smales, C. M., & Howard, M. J. (2013). ^{1}H NMR spectroscopy profiling of metabolic reprogramming of Chinese hamster ovary cells upon a temperature shift during culture. *PloS One*, *8*(10), 77195.

Walsh, G. (2010). Biopharmaceutical benchmarks 2010. *Nature Biotechnology*, *28*(9), 917–24.

Welker, S., Rudolph, B., Frenzel, E., Hagn, F., Liebisch, G., Schmitz, G., Buchner, J. (2010). Hsp12 is an intrinsically unstructured stress protein that folds upon membrane association and modulates membrane function. *Molecular Cell*, *39*(4), 507–20.

Wellmann, S., Bührer, C., Moderegger, E., Zelmer, A., Kirschner, R., Koehne, P., Seeger, K. (2004). Oxygen-regulated expression of the RNA-binding proteins RBM3 and CIRP by a HIF-1-independent mechanism. *Journal of Cell Science*, *117* (9), 1785–94.

White-Grindley, E., & Si, K. (2006). RISC-y Memories. *Cell*, *124*(1), 23–6.

Whitehead, B., Craven, C. J., & Waltho, J. P. (1997). Double and triple resonance NMR methods for protein assignment. *Methods in Molecular Biology* , *60*, 29–52.

Wishart, Bigam, C., Holm, A., Hodges, R., & Sykes, B. (1995). 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR*, *5*(5), 67–81.

Wishart, D. (2005). NMR spectroscopy and protein structure determination: applications to drug discovery and development. *Current Pharmaceutical Biotechnology*, *6*(2), 105–20.

Wishart, D. S., & Nip, A. M. (1998). Protein chemical shift analysis: a practical guide. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, *76*(2-3), 153–63.

Wishart, D. S., & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods in Enzymology*, *239*, 363–92.

Wishart, Sykes, B. D., & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, *31*(6), 1647–51.

Wolffe, A. P. (1994). Structural and functional properties of the evolutionarily ancient Y-box family of nucleic acid binding proteins. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *16*(4), 245–51.

Wright, C. F., Oswald, B. W., & Dellis, S. (2001). Vaccinia virus late transcription is activated in vitro by cellular heterogeneous nuclear ribonucleoproteins. *The Journal of Biological Chemistry*, *276*(44), 40680–6.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. *Wiley*.

Yang, C. & Carrier, F., (2001). The UV-inducible RNA-binding protein A18 (A18 hnRNP) plays a protective role in the genotoxic stress response. *The Journal of biological chemistry*, 276(50), 47277–84.

Yang, R., Weber, D.J. & Carrier, F., (2006). Post-transcriptional regulation of thioredoxin by the stress inducible heterogenous ribonucleoprotein A18. *Nucleic acids research*, 34(4), 1224–36.

Yang, R., Zhan, M., Nalabothula, N. R., Yang, Q., Indig, F. E., & Carrier, F. (2010). Functional significance for a heterogenous ribonucleoprotein A18 signature RNA motif in the 3'-untranslated region of ataxia telangiectasia mutated and Rad3-related (ATR) transcript. *The Journal of Biological Chemistry*, *285*(12), 8887–93.

Yángüez, E., & Nieto, A. (2011). So similar, yet so different: selective translation of capped and polyadenylated viral mRNAs in the influenza virus infected cell. *Virus Research*, *156*(1-2), 1–12.

Ye, J., Kober, V., Tellers, M., Naji, Z., Salmon, P., & Markusen, J. F. (2009). High-level protein expression in scalable CHO transient transfection. *Biotechnology and Bioengineering*, *103*(3), 542–51.

Yeon-Gu, K. (2012). Omics-Based CHO Cell Engineering – Entrance into Post-Genomic Era. *Advances in Genetic Engineering & Biotechnology*, 1,1.

Yoon, S. K., Kim, S. H., & Lee, G. M. (2003). Effect of low culture temperature on specific productivity and transcription level of anti-4-1BB antibody in recombinant Chinese hamster ovary cells. *Biotechnology Progress*, *19*(4), 1383–6.

Yoon, S. K., Song, J. Y., & Lee, G. M. (2003). Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *82*(3), 289–98.

Yoon, S. K., Kim, S. H., Song, J. Y., & Lee, G. M. (2006). Biphasic culture strategy for enhancing volumetric erythropoietin productivity of Chinese hamster ovary cells. *Enzyme and Microbial Technology*, *39*(3), 362–365.

Yuan, X., Davydova, N., Conte, M. R., Curry, S., & Matthews, S. (2002). Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. *Nucleic Acids Research*, *30*(2), 456–62.

Zeng, Y., Kulkarni, P., Inoue, T., & Getzenberg, R. H. (2009). Down-regulating cold shock protein genes impairs cancer cell survival and enhances chemosensitivity. *Journal of Cellular Biochemistry*, *107*(1), 179–88.

Zhang, Y. et al., (1998). Expression of eukaryotic proteins in soluble form in Escherichia coli. *Protein expression and purification*, 12(2), 159–65.

Zhang, X., Dudek, E. J., Liu, B., Ding, L., Fernandes, A. F., Liang, J. J., Shang, F. (2007). Degradation of C-terminal truncated alpha A-crystallins by the ubiquitin-proteasome pathway. *Investigative Ophthalmology & Visual Science*, *48*(9), 4200–8.

Zhang, X. H., Zhao, C., & Ma, Z. A. (2007). The increase of cell-membranous phosphatidylcholines containing polyunsaturated fatty acid residues induces phosphorylation of p53 through activation of ATR. *Journal of Cell Science*, *120*(23), 4134–43.

Zhou, K.-W., Zheng, X.-M., Yang, Z.-W., Zhang, L., & Chen, H.-D. (2009). Overexpression of CIRP may reduce testicular damage induced by cryptorchidism. *Clinical & Investigative Medicine*, 32(2), 103-11.

Appendices

Appendix A: Minimal Media

Compound	Concentration
Na_2HPO_4	6.8 g/l
KH_2PO_4	3 g/l
Sodium Chloride	0.5 g/l
Na_2SO_4	0.3 mM
EDTA	0.05 mM
$MnCl_2$	160 mM
$FeCl_3$	5 μg/ml
$ZnCl_2$	0.5 μg/ml
$CuCl_2$	0.1 μg/ml
$CoCl_2$	0.1 μg/ml
H_3BO_3	0.1 μg/ml
$MgSO_4$	1 mM
$CaCl_2$	0.3 mM
d-Biotin	1 mg/ml
Thiamine	1 mg/ml
$(NH_4)_2SO_4$	1 g/l
Glucose	4 g/l (^{12}C) or 2 g/l (^{13}C)

Appendix B: Resonance Assignments of CIRP

Residue Number	Residue	N	HN	CA	CB	CG	CD	CE	HA	HB	HG	HD	HE
1	Met	122.05	8.35	55.32	32.94	31.85	-	-	4.43	1.93,2.05	2.49	-	-
2	Ala	125.75	8.38	52.51	19.21	-	-	-	4.33	1.39	2.45	-	-
3	Ser	114.86	8.29	-	-	-	-	-	-	-	-	-	-
4	Asp	122.25	8.42	54.64	40.99	-	-	-	4.66	2.75,2.62	-	-	-
5	Glu	120.44	8.34	58.25	-	36.94	-	-	4.18	2.07,1.99	2.3	-	-
6	Gly	108.32	8.77	45.18	-	-	-	-	3.8,4.12	-	-	-	-
7	Lys	121.86	7.67	55.94	34.33	24.65	29.47	-	4.95	1.93,1.59	1.44,1.14	1.35	-
8	Leu	127.65	9.58	53.09	43.5	26.62	22.83,26.41	-	5.14	1.34,1.78	1.89	0.98,0.81	-
9	Phe	123.93	9.24	57.09	41.32	-	-	-	4.67	2.74,3.10	-	-	-
10	Val	126.61	8.04	-	33.15	20.92,22.89	-	-	4.76	1.49	0.82,0.57	-	-
11	Gly	111.48	9.07	43.33	-	-	-	-	4.71,3.64	-	-	-	-
12	Gly	106.89	8.36	45.62	-	-	-	-	4.22,3.75	-	-	-	-
13	Leu	114.87	7.53	53.77	43.43	26.91	23.19,25.67	-	4.12	1.12,1.38	1.3	0.48,0.71	-
14	Ser	115.07	9.25	-	-	-	-	-	-	-	-	-	-
15	Phe	126.72	8.77	59.5	37.45	-	-	-	4.33	3.35,3.01	-	-	-
16	Asp	115.62	8.04	54.41	41.53	-	-	-	4.71	2.77,2.43	-	-	-
17	Thr	118.79	7.31	64.48	68.33	22.54	-	-	4.05	3.94	1.16	-	-
18	Asn	126.69	7.87	50.51	40.14	-	-	-	4.86	2.95,3.33	-	-	-
19	Glu	118.41	9.89	61.61	28.4	38.23	-	-	3.55	2.05,1.83	2.68,2.12	-	-
20	Gln	119.12	8.12	59.37	27.84	34.4	-	-	4.08	2.13,2.03	2.42	-	-
21	Ala	122.42	7.91	55.02	19.08	-	-	-	4.18	1.59	-	-	-
22	Leu	116.82	7.82	57.37	42.43	26.48	24.79,23.62	-	4.13	1.37,1.68	1.55	0.66	-
23	Glu	120.27	8.7	60.38	29.44	36.23	-	-	3.69	2.15,2.20	2.18	-	-

24	Gln	117.49	8.08	59.14	28.29	33.99	-	-	4.03	2.19,2.25	2.43,2.55	-	-
25	Val	115.4	7.58	64.83	32.34	21.81,21.45	-	-	3.99	2.19	1.10,0.78	-	-
26	Phe	115.85	8.25	62.29	39.63	-	-	-	4.63	2.96,3.40	-	-	-
27	Ser	115.62	8.51	61.35	63.08	-	-	-	4.57	4.18,4.08	-	-	-
28	Lys	119.61	7.14	57.59	31.46	23.51	28.47	41.54	3.98	1.48,1.28	0.74,0.95	1.44,1.49	2.85
29	Tyr	115.66	7.68	59.47	38.93	-	-	-	4.25	3.27,2.91	-	-	-
30	Gly	104.98	7.37	44.24	-	-	-	-	3.83,4.89	-	-	-	-
31	Gln	116.07	8.27	56.81	28.99	33.67	-	-	4.22	2.00,2.09	2.28,2.37	-	-
32	Ile	129.3	8.78	60.4	39.82	18.06,27.84	14	-	4.03	1.59	0.03,1.47,0.51	0.39	-
33	Ser	122.46	9.08	58.81	64.2	-	-	-	4.53	3.62	-	-	-
34	Glu	120.11	7.27	56.58	33.44	37.13	-	-	4.47	1.79,1.98	1.97,2.14	-	-
35	Val	124.36	8.43	61.37	34.74	22.4,22.27	-	-	4.73	1.9	0.89,0.87	-	-
36	Val	124.68	8.78	60.76	36.86	21.24	-	-	4.55	1.93	0.91	-	-
37	Val	126.12	8.46	62.52	33.18	21.88,20.87	-	-	3.98	1.71	0.56,0.76	-	-
38	Val	130.09	8.37	64.6	30.33	21.69,22.79	-	-	3.59	0.84	0.91,0.79	-	-
39	Lys	124.72	8.18	-	35.51	24.66	28.78	41.95	4.78	1.40,1.28	1.06,1.40	1.59,1.44	2.98
40	Asp	123.16	8.63	54.22	42.8	-	-	-	4.47	2.50,2.93	-	-	-
41	Arg	127.14	8.91	58.91	30.01	27.12	43.09	-	3.96	1.9	1.69	3.24	-
42	Glu	117.29	8.59	58.63	30.22	36.7	-	-	4.27	2.16,2.12	2.24,2.30	-	-
43	Thr	107.44	8.17	61.83	70.74	21.03	-	-	4.33	4.31	1.2	-	-
44	Gln	114.34	8.14	57.41	25.95	33.91	-	-	3.87	2.35,2.28	2.19,2.30	-	-
45	Arg	118.31	7.65	55.31	31.55	27.23	43.2	-	4.36	1.59,1.78	1.64	3.1	-
46	Ser	116.73	8.75	-	-	-	-	-	-	-	-	-	-
47	Arg	122.9	9.05	55.99	30.9	27.83	43.42	-	4.39	2.28,1.05	1.61	3.05,3.23	-
48	Gly	108.41	9.61	45.63	-	-	-	-	4.18,3.26	-	-	-	-
49	Phe	112.34	7.11	54.82	41.58	-	-	-	5.21	2.98,3.27	-	-	-

50	Gly	106.2	8.84	-	-	-	-	-	-	-	-	-	-
51	Phe	114.11	9	55.9	44.12	-	-	-	5.89	2.59,2.86	-	-	-
52	Val	122.12	8.78	60.97	35.07	20.68,21.65	-	-	4.42	1.34	0.27,0.27	-	-
53	Thr	123.23	8.98	61.44	69.34	22.11	-	-	5.05	3.9	1.02	-	-
54	Phe	127.45	8.83	58.11	40.42	-	-	-	4.68	2.81,3.69	-	-	-
55	Glu	120.11	7.92	58.84	30.25	36.22	-	-	4.03	1.90,2.06	2.17	-	-
56	Asn	116.37	9.29	51.41	40.2	-	-	-	5.12	2.98,2.79	-	-	-
57	Ile	124.58	8.73	64.89	38.16	17.71,28.26	14.05	-	3.68	1.79	0.94,1.13,1,53	0.9	-
58	Asp	121.29	8.38	57.62	39.91	-	-	-	4.35	2.75,2.57	-	-	-
59	Asp	121.25	7.31	56.94	40.06	-	-	-	4.22	2.65,2.89	-	-	-
60	Ala	120.13	6.93	54.48	18	-	-	-	3.27	1.52	-	-	-
61	Lys	117.45	7.75	59.91	31.66	25.33	29.15	41.93	3.86	1.93	1.40,1.63	1.7	2.95
62	Asp	120.1	7.67	57.16	40.54	-	-	-	4.27	2.93,2.66	-	-	-
63	Ala	122.46	8.11	54.44	20.07	-	-	-	2.58	1.19	-	-	-
64	Met	117.66	8.35	59.64	32.93	31.08	-	-	3.74	2.02,2.29	2.40,2.37	-	-
65	Met	115.74	7.77	57.39	31.93	32.01	-	-	4.22	2.16,2.11	2.68,2.76	-	-
66	Ala	119.53	7.71	54.03	20.21	-	-	-	4.18	1.26	-	-	-
67	Met	111.2	8.01	-	31.63	-	-	-	4.76	1.77,2.03	-	-	-
68	Asn	117.19	7.67	56.77	38.52	-	-	-	4.37	2.95,3.16	-	-	-
69	Gly	116.44	8.66	45.86	-	-	-	-	3.76,4.26	-	-	-	-
70	Lys	120.93	7.81	54.57	32.29	25.49	27.85	42.5	4.43	1.87,1.96	1.76,1.31	1.69	3.06
71	Ser	113.54	8.15	-	-	-	-	-	-	-	-	-	-
72	Val	125.7	8.97	61.92	34.55	21.84,20.79	-	-	4.08	1.87	0.80,0.79	-	-
73	Asp	129.6	9.5	55.47	39.07	-	-	-	4.28	2.42,2.96	-	-	-
74	Gly	101.43	8.41	45.38	-	-	-	-	3.55,4.09	-	-	-	-
75	Arg	120	7.56	53.02	32.07	26.43	42.24	-	4.61	1.71,1.97	1.57,1.51	3.13	-

76	Gln	121.74	8.45	55.8	28.82	34.25	-	-	4.37	1.83,1.91	2.19,2.12	-	-
77	Ile	122.22	8.48	-	-	-	-	-	-	-	-	-	-
78	Arg	123.03	8.43	53.71	32.69	27.25	43.31	-	5.01	1.72,1.69	1.66,1.55	3.30,3.23	-
79	Val	123.81	8.71	60.02	34.42	24.71,22.63	-	-	5.41	1.93	1.08,1.06	-	-
80	Asp	125.2	9.15	52.12	45.26	-	-	-	5.2	2.52,2.75	-	-	-
81	Gln	121.61	9.18	57.14	29.77	34.71	-	-	4.43	2.22,2.00	2.47	-	-
82	Ala	126.79	8.63	53.27	19.33	-	-	-	4.41	1.46	-	-	-
83	Gly	107.68	8.66	-	-	-	-	-	-	-	-	-	-
84	Lys	120.6	8.02	55.96	33.1	24.62	28.64	41.88	4.37	1.73,1.84	1.37,1.41	1.65	2.96
85	Ser	117.47	8.31	-	-	-	-	-	-	-		-	-
86	Ser	122.94	7.88	-	-	-	-	-	-	-		-	-

Appendix C: CIRP Relaxation Data

Residue	Residue Number	T1/s	T1 Error /s	T2/s	T2 Error /s	hnNOE on /arb units	hnNOE off /arb units	hnNOE as (I-0)/I0)	Order Parameter S2	S2 Error	Order Parameter Te	Te Error
Met	1	0.764	0.001	0.238	0.000	28101.500	79849.800	-0.648	0.544	0.047	52.103	11.248
Ala	2	0.583	0.001	0.181	0.000	29733.400	160024.000	-0.814	0.697	0.016	212.340	50.035
Ser	3	0.643	0.001	0.246	0.000	90076.900	183749.000	-0.510	0.493	0.018	30.869	5.087
Asp	4	0.644	0.001	0.189	0.000	167704.000	273680.000	-0.387	0.609	0.011	44.226	8.900
Glu	5	0.591	0.001	0.151	0.000	287771.000	471476.000	-0.390	0.731	0.013	60.523	15.861
Gly	6	0.602	0.001	0.142	0.000	276860.000	374206.000	-0.260	0.777	0.012	36.114	19.587
Lys	7	0.587	0.001	0.136	0.000	309519.000	394705.000	-0.216	0.805	0.016	13.888	25.890
Leu	8	0.619	0.001	0.137	0.000	260088.000	306973.000	-0.153	0.794	0.017	0.000	26.152
Phe	9	0.624	0.001	0.135	0.000	276156.000	315898.000	-0.126	0.803	0.013	0.000	28.702
Val	10	0.580	0.001	0.130	0.000	303239.000	323408.000	-0.062	0.844	0.016	0.000	44.716
Gly	11	0.626	0.001	0.136	0.000	301014.000	304286.000	-0.011	0.796	0.018	0.000	37.007
Gly	12	0.667	0.001	0.131	0.000	309488.000	343122.000	-0.098	0.808	0.015	0.000	39.707
Leu	13	0.597	0.001	0.123	0.000	418271.000	494064.000	-0.153	0.848	0.022	0.000	52.388
Ser	14	0.621	0.001	0.133	0.000	312230.000	342581.000	-0.089	0.809	0.019	0.000	35.323
Phe	15	0.591	0.001	0.139	0.000	107749.000	91356.900	0.179	0.809	0.000	0.000	0.000
Asp	16	0.601	0.001	0.145	0.000	340498.000	385501.000	-0.117	0.760	0.009	13.039	20.591
Thr	17	0.630	0.001	0.155	0.000	469340.000	534256.000	-0.122	0.738	0.015	8.990	18.789
Asn	18	0.615	0.001	0.142	0.000	388128.000	407190.000	-0.047	0.784	0.014	0.000	25.746
Glu	19	0.652	0.001	0.146	0.000	301775.000	351865.000	-0.142	0.750	0.018	0.000	21.046
Gln	20	0.665	0.001	0.132	0.000	479742.000	576860.000	-0.168	0.789	0.015	0.000	32.050
Ala	21	0.587	0.001	0.124	0.000	541600.000	664480.000	-0.185	0.858	0.015	0.000	45.833
Leu	22	0.617	0.001	0.121	0.000	493964.000	570852.000	-0.135	0.843	0.000	0.000	0.000
Glu	23	0.655	0.001	0.120	0.000	492125.000	598051.000	-0.177	0.797	0.026	0.000	50.037
Gln	24	0.672	0.001	0.137	0.000	538244.000	590957.000	-0.089	0.771	0.016	0.000	31.851

Val	25	0.631	0.001	0.128	0.000	447595.000	520284.000	-0.140	0.798	0.024	0.000	51.483
Phe	26	0.618	0.001	0.130	0.000	358196.000	429707.000	-0.166	0.808	0.017	0.000	36.874
Ser	27	0.621	0.001	0.117	0.000	517460.000	559578.000	-0.075	0.852	0.068	0.000	260.210
Lys	28	0.648	0.001	0.134	0.000	607602.000	677183.000	-0.103	0.796	0.015	0.000	32.624
Tyr	29	0.635	0.001	0.127	0.000	348730.000	413902.000	-0.157	0.816	0.020	0.000	46.524
Gly	30	0.680	0.001	0.132	0.000	395228.000	447984.000	-0.118	0.787	0.016	0.000	40.123
Gln	31	0.696	0.001	0.152	0.000	379966.000	471920.000	-0.195	0.712	0.012	0.000	15.058
Ile	32	0.601	0.001	0.150	0.000	235877.000	322848.000	-0.269	0.780	0.018	34.853	20.841
Ser	33	0.593	0.001	0.145	0.000	244057.000	310645.000	-0.214	0.757	0.009	45.679	18.900
Glu	34	0.628	0.001	0.139	0.000	403206.000	494537.000	-0.185	0.784	0.009	0.000	19.057
Val	35	0.640	0.001	0.140	0.000	312692.000	389944.000	-0.198	0.774	0.015	0.000	21.204
Val	36	0.614	0.001	0.134	0.000	317962.000	350770.000	-0.094	0.810	0.010	0.000	30.220
Val	37	0.660	0.001	0.143	0.000	304563.000	392547.000	-0.224	0.755	0.011	0.000	17.830
Val	38	0.639	0.001	0.135	0.000	334082.000	395334.000	-0.155	0.791	0.016	0.000	26.639
Lys	39	0.667	0.001	0.135	0.000	319867.000	358939.000	-0.109	0.790	0.012	0.000	31.387
Asp	40	0.631	0.001	0.135	0.000	399821.000	439862.000	-0.091	0.803	0.011	0.000	29.813
Arg	41	0.598	0.001	0.132	0.000	298243.000	351118.000	-0.151	0.824	0.022	0.000	33.402
Glu	42	0.661	0.001	0.164	0.000	317467.000	488479.000	-0.350	0.670	0.007	58.997	10.034
Thr	43	0.659	0.001	0.142	0.000	270076.000	362979.000	-0.256	0.754	0.013	3.358	17.106
Gln	44	0.616	0.001	0.130	0.000	375294.000	427324.000	-0.122	0.819	0.018	0.000	41.521
Arg	45	0.620	0.001	0.145	0.000	352496.000	470843.000	-0.251	0.754	0.008	36.090	16.038
Ser	46	0.695	0.001	0.151	0.000	272046.000	321808.000	-0.155	0.711	0.023	0.000	17.726
Arg	47	0.634	0.001	0.138	0.000	280627.000	327883.000	-0.144	0.786	0.016	0.000	25.715
Gly	48	0.582	0.001	0.135	0.000	309652.000	339905.000	-0.089	0.814	0.009	0.000	29.860
Phe	49	0.584	0.001	0.117	0.000	550004.000	582270.000	-0.055	0.899	0.000	0.000	0.000
Gly	50	0.617	0.001	0.126	0.000	406090.000	450994.000	-0.100	0.838	0.020	0.000	78.506

Phe	51	0.582	0.001	0.147	0.000	307838.000	399385.000	-0.229	0.765	0.011	86.358	19.777
Val	52	0.605	0.001	0.150	0.000	317629.000	370103.000	-0.142	0.758	0.013	31.874	19.598
Thr	53	0.608	0.001	0.139	0.000	300823.000	332911.000	-0.096	0.790	0.009	0.000	24.773
Phe	54	0.617	0.001	0.138	0.000	273759.000	305546.000	-0.104	0.794	0.018	0.000	28.914
Glu	55	0.641	0.001	0.139	0.000	379541.000	431782.000	-0.121	0.778	0.020	0.000	26.638
Asn	56	0.619	0.001	0.134	0.000	266516.000	327801.000	-0.187	0.801	0.022	0.000	26.744
Ile	57	0.635	0.001	0.132	0.000	444051.000	547421.000	-0.189	0.800	0.015	0.000	27.110
Asp	58	0.568	0.001	0.131	0.000	405444.000	505733.000	-0.198	0.841	0.014	13.639	32.531
Asp	59	0.590	0.001	0.151	0.000	316438.000	411755.000	-0.231	0.773	0.020	26.863	21.282
Ala	60	0.609	0.001	0.123	0.000	536983.000	575890.000	-0.068	0.824	0.000	0.000	0.000
Lys	61	0.633	0.001	0.121	0.000	432673.000	596604.000	-0.275	0.821	0.017	0.000	32.547
Asp	62	0.626	0.001	0.127	0.000	481318.000	585417.000	-0.178	0.804	0.020	0.000	40.920
Ala	63	0.630	0.001	0.147	0.000	427697.000	547321.000	-0.219	0.752	0.019	9.669	19.197
Met	64	0.654	0.001	0.121	0.000	476038.000	578190.000	-0.177	0.822	0.060	0.000	203.150
Met	65	0.627	0.001	0.123	0.000	467010.000	569079.000	-0.179	0.841	0.020	0.000	43.391
Ala	66	0.600	0.001	0.128	0.000	454660.000	584254.000	-0.222	0.829	0.016	0.000	29.824
Met	67	0.655	0.001	0.126	0.000	363442.000	383661.000	-0.053	0.828	0.000	0.000	0.000
Asn	68	0.649	0.001	0.112	0.000	531127.000	629352.000	-0.156	0.799	0.000	0.000	0.000
Gly	69	0.599	0.001	0.109	0.000	340963.000	374240.000	-0.089	0.899	0.000	0.000	0.000
Lys	70	0.630	0.001	0.124	0.000	427688.000	482847.000	-0.114	0.848	0.018	0.000	78.370
Ser	71	0.685	0.001	0.150	0.000	371135.000	473773.000	-0.217	0.728	0.006	0.000	14.979
Val	72	0.615	0.001	0.143	0.000	284482.000	349874.000	-0.187	0.779	0.015	12.824	21.194
Asp	73	0.651	0.001	0.119	0.000	197360.000	205182.000	-0.038	0.810	0.055	0.000	57.615
Gly	74	0.642	0.001	0.118	0.000	366500.000	410483.000	-0.107	0.841	0.000	0.000	0.000
Arg	75	0.611	0.001	0.145	0.000	350315.000	390647.000	-0.103	0.789	0.019	0.000	27.143
Gln	76	0.690	0.001	0.140	0.000	411763.000	489819.000	-0.159	0.772	0.008	0.000	23.676

Ile	77	0.645	0.001	0.142	0.000	304141.000	417243.000	-0.271	0.755	0.016	15.334	17.605
Arg	78	0.622	0.001	0.145	0.000	290348.000	291755.000	-0.005	0.765	0.013	0.000	23.357
Val	79	0.610	0.001	0.148	0.000	270604.000	334971.000	-0.192	0.761	0.016	18.124	19.986
Asp	80	0.635	0.001	0.142	0.000	329772.000	388058.000	-0.150	0.767	0.010	0.000	20.999
Gln	81	0.638	0.001	0.157	0.000	436037.000	513983.000	-0.152	0.721	0.013	17.245	16.518
Ala	82	0.622	0.001	0.154	0.000	401918.000	525264.000	-0.235	0.735	0.015	25.777	16.933
Gly	83	0.616	0.001	0.211	0.000	197636.000	320119.000	-0.383	0.637	0.011	90.735	11.186
Lys	84	0.637	0.001	0.233	0.000	129481.000	299171.000	-0.567	0.546	0.010	59.876	5.787
Ser	85	0.895	0.001	0.306	0.000	-58080.000	45619.100	-2.273	0.274	0.019	184.480	42.271
Ser	86	1.012	0.001	0.421	0.000	-303127.000	294526.000	-2.029	0.225	0.008	134.020	6.075