UNIVERSITY OF
BIRMINGHAM

Structural Studies of the DNA Partitioning Protein IncC from the Plasmid RK2

By

MUHAMMAD FAYYAZ UR REHMAN

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

School of Biosciences
The University of Birmingham
January, 2018

Dedicated to Abu and Ammi !

Acknowledgements

I would like to acknowledge my extreme gratitude to Allah, the Lord, the Almighty, the Maintainer and the Sustainer of this universe. It was only through His unlimited mercy and bounties, that it was possible for me to complete this thesis. In every stressful and difficult situation during my PhD studies, I was able to turn towards Him and seek His help. I would like to convey my special gratitude to my supervisor Dr Eva Hyde for her invaluable scientific guidance at each step of this difficult project. It is difficult to overstate my gratitude to Dr Mark Jeeves for his enthusiasm, inspiration and great efforts to make NMR analysis possible. I also wish to extend my deepest thanks to my supervisor Dr Scott White for valuable scientific discussions and guidance throughout the project.

I am thankful to Islamic Development Bank (IDB) for financial support, without that, it was not possible to carry out this project. I am also thankful to Prof Grainger and Prof Thomas for scientific discussions and guidance. I would like to thank to all seventh-floor people who helped me a lot to learn and go forward with my project. Rosemary and Nick deserves big thanks from my side. I am also thankful to Dr. Tim Knowles, Dr Muhammad Jamshad, Dr. Ian Cadby, Dr. Klaus Futterer, Dr. Rich Logan, Dr. Rich Meek, Dr. Maria, and Pooja Sridhar for helping me in different experiments. I am thankful to Dr Melissa Grant, Dr. Norman and Dr. Jinglei Yu for helping me out for crosslinking MS experiments. I also appreciate help and support from Dr Sara Whittaker and Dr Sue in NMR facility for their help during the NMR experiments and say thanks to Allen Bowden, in School of Chemistry, for heling me in HPLC experiments. My special thanks to Dr Anmol Gautam for making lab an enjoyable place to work, to Dr. Prateek Sharma for his constant support and guidance, to Dr Yasir, Dr Ghoswami and Dr Zeeshan for encouragement. I would like to thanks Dr Shabbir and Dr Safwan for guidance. I would like to express my thanks to Dr Aslam for continuous encouragement and guidance throughout my PhD. I would like to thank to my friends back in my country, Arif, Zubair, Waqar, Asim, Sarfraz, Shazi, Shahbaz, Nasir, Tahir and Dr Junaid for encouragement and support. I would also like to say thanks to my colleagues at University of Sargodha, Prof Akram Chaudhary, Prof Yasin, Prof Tariq, Prof Anwar and Dr Ijaz for encouraging me to pursue my degree.

I would like to convey my special gratitude to my parents, especially to my beloved father, whose inspiration and motivation were the driving forces for me to pursue and complete my PhD. Without support from Ammi and Abu, I wouldn't get anywhere. I would like to thanks my sisters, Saima, Naima, Dr Aima and brothers Ali, Tahir and Dr Aamer for all their support during my PhD.

M Fayyaz ur rehman

Abstract

Plasmid DNA partitioning is a crucial process for the transfer of at least a single copy of plasmid to the daughter cells during cell division in bacteria. With many low-copy number plasmids the partitioning process involves a DNA-binding protein, called ParB; a centromere-like DNA site, called *parS*; and a ParA-family protein ATPase; similar to those of bacterial chromosomes. Most plasmid ParA proteins are longer than those of bacterial chromosomes, however, interestingly, the RK2 plasmid encodes two ParA proteins, IncC1 and IncC2 proteins. The two proteins are expressed in two different lengths from different start codons within same gene. The longer protein is IncC1 (364 a.a), while IncC2 lacks a N- terminal domain of 105 amino acids (IncC NTD). The secondary structure of IncC NTD was determined, using NMR spectroscopy and other biophysical methods, and shown to be predominantly random coil. Carbon-detected NMR experiments were performed to get the chemical shifts and hence information about the backbone conformation for IncC NTD. It appears to bind DNA weakly and non-specifically, from Electrophoretic Mobility Shift Assays (EMSAs) and NMR. The expression and purification of IncC1 and IncC2 proteins was optimized. The two proteins and IncC NTD were characterized using various biophysical methods including Circular Dichroism, Analytical Ultracentrifugation, Small Angle X-ray Scattering, Size Exclusion Chromatography-Multi Angle Light Scattering, and EMSAs. IncC NTD was disordered even within full-length IncC1 protein, as shown using NMR. Bacterial two hybrid assays (BACTH assay) and chemical crosslinking showed the two IncC proteins form homo- and hetero-dimers and interact with KorB protein (the ParB homologue in RK2). IncC1 and IncC2 proteins bind to DNA, non-specifically. IncC1 binds DNA weakly in the absence of nucleotides but IncC2 protein was found to bind DNA only in the presence of nucleotides (ADP, ATP, ATPγS), hence suggesting a role of the IncC NTD in aiding DNA binding in the absence of nucleotides.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 6

Chapter 7

List of Tables

List of Abbreviations

°C	Degree Celsius
µg	Microgram
µL	Microliter
A, C, G, T	Nucleotides: adenosine, cytosine, guanine, thymine
aa	Amino acid
Abs	Absorbance
ADP	Adenosine diphosphate
APS	Ammonium persulphate
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
BACTH	Bacterial adenylate cyclase two hybrid
bp	Base pair(s)
BS3	Bis-Sulfo Succinimidyl Suberate
BSA	Bovine Serum Albumin
CBP	Centromere Binding Protein
CD	Circular Dichroism
CL-MS	Crosslinking combined with Mass Spectrometry
CS Rosetta	Chemical shift rosetta
CTD	C-terminal domain
CV	Column Volume
DBD	DNA binding domain
DNA	Deoxyribonucleic Acid
DSS	Di-Succinimidyl Suberate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
hr	Hour(s)
HRP	Horseradish peroxidase
HSQC	Heteronuclear Single Quantum Correlation
HTH	Helix-turn-helix motif
IDP	Intrinsically Disordered protein
IDR	Intrinsically disordered region
IMAC	Immobilised Metal-ion Affinity Chromatography
Inc	Incompatibility (group)
IPTG	Isopropyl-!-thiogalactopyranoside
JCSG	Joint Center for Structural Genomics
kb	Kilobases
kDa	Kilo Daltons
LB	Luria Bertani
MD	Molecukar Dynamics
mdeg	millidegree
MST	MicroScale Thermophoresis
MWCO	Molecular Weight Cut–Off
NaCl	Sodium Chloride
NEB	New England Biolabs
ng	nanogram
nm	nanometer
NMR	Nuclear Magnetic Resonance
NN	Neural Network
ns	nanosecond
NTD	N–Terminal Domain
NTP	Nucleotide triphosphate
o/n	Overnight
ONPG	O-nitrophenol-β-galactoside
PAGE	Polyacrylamide Gel Electrophoresis

par	Partitioning
PCR	Polymerase Chain Reaction
psi	pound per square inch
PTM	Post translational modifications
R_g	Radius of Gyration
RHH	Ribbon Helix Helix motif
RMSD	Root mean square deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room Temperature
S	Seconds
SAXS	Small angle light scattering
SDS	Sodium Dodecyl sulphate
SEC	Size exclusion chromatography
SEC-MALLS	Size exclusion chromatography with multi angle laser light scattering
SIM XL	Spectrum Identification machine
TAE	Tris-acetate buffer
TBS	Tris-buffered saline
TEMED	N,N,N',N'-tetramethylene diamine
Tris	Tris (hydroxymethyl) amino methane
TSA	Thermal Shift Assay
WB	Western Blot
X-gal	5-Bromo-4-Chloro-3-indolyl-!-D-galactopyranoside

CHAPTER 1

INTRODUCTION

INTRODUCTION

All organisms, including bacteria, need to reproduce by cell division to continue their existence. To ensure inheritance in bacteria, plasmid and chromosomal DNA segregation is crucial before cell division. After replication, plasmids and chromosomal DNA are partitioned to bacterial daughter cells, to transfer genetic information. Unlike eukaryotes, bacteria are thought not to possess any dedicated mitotic machinery. From the replicon model (Jacob and Brenner, 1963) to the diffusion ratchet partitioning model (Zhang and Schumacher, 2017, Hwang et al., 2013, Vecchiarelli et al., 2010, Vecchiarelli et al., 2013), there have been many studies on active DNA partitioning systems in bacteria. The replicon theory proposes that partitioning involves anchoring of daughter chromosomes to the cell envelope at mid-cell. During cell division, as a cell divides, the chromosomes are pulled towards the opposite poles of the cell. In recent decades, new studies have shown the presence of partitioning proteins in the bacteriophage P1 and in the low copy number F plasmid. These proteins are encoded by *par* genes (Gordon and Wright, 2000) and *sop* genes, respectively (Draper and Gober, 2002). These studies also show the presence of mitotic-like machinery in bacteria (Ptacin et al., 2010).

1.1 Plasmid partitioning

Bacteria adapt and acclimatize themselves to stringent environmental conditions with the help of self-replicating, extra-chromosomal DNA segments, called plasmids (Leplae et al., 2006, Heuer and Smalla, 2012). Plasmids do not contain genes that are crucial for cellular functions but may express genes involved in antibiotic resistance, toxins, or detoxification/degradation of harmful compounds (Zielenkiewicz and Ceglowski, 2001).

Plasmids are widely distributed, range from a few kb to 100 kb in size, and exist independently of chromosomal DNA. Sometimes, plasmids make up to 50% of the total DNA present in bacterial cells (Fraser et al., 1997). Bacterial plasmids can be divided into more than 30 Inc (incompatibility) groups, based on their ability to co-exist in the same bacterial cell (Jain and Srivastava, 2013).

Bacteria can carry large plasmids but these can only be in low copy number to avoid the metabolic burden involved in their maintenance and replication. The random distribution of such low copy number plasmids cannot guarantee faithful plasmid inheritance to daughter cells (Sengupta and Austin, 2011). Thus, active plasmid partitioning systems are required for the stable

maintenance and transfer of plasmids to the daughter cells. The first discovery of active partitioning systems (*par*) was in the P1 and F plasmids in *E. coli* (Austin and Abeles, 1983, Gerdes et al., 1985, Ogura and Hiraga, 1983).

Many bacteria, including *Pseudomonas aeruginosa*, *Pseudomonas putida, Bacillus subtilis, Caulobacter crescentus* and *Borrelia burgdorferi,* encode partitioning systems for chromosomal DNA that are homologous to those for plasmids (Glaser et al., 1997, Ireton et al., 1994, Tilly et al., 2012, Bartosik et al., 2009, Wake and Errington, 1995, Bignell and Thomas, 2001). In most bacteria, partitioning requires two proteins (ParA, ParB or their homologues) and a DNA binding site: *parS*. Chromosomal segregation in *Bacillus subtilis* involves Soj and Spo0J proteins which are equivalent to ParA and ParB, respectively (Leonard et al., 2005). Spo0J and ParB are involved in binding to DNA and its localization of to the cell. ParA and Soj are ATPases that interact with ParB and DNA (Draper and Gober, 2002). An illustration of *par* operons in different partitioning systems is given in Figure 1.1.

1.2 Partitioning systems

Based on functional and sequence diversity, partitioning proteins can be divided into different groups. ParA proteins have either one of three typical NTPase motifs; a Walker Type ATPase motif, an Actin-like motif or a tubulin-like GTPase motif. On the basis of structural similarity among the ParA proteins and their interactions with ParB, several plasmid partitioning models have been proposed (Schumacher, 2008).

1.2.1 Type I partitioning system

In the Type I partitioning system, it is proposed that DNA partitioning occurs in a mitotic-like fashion, where ParA filaments are involved in the movement of plasmids or chromosomes (Gerdes et al., 2010). The chromosomes or plasmids are positioned in the middle of the dividing cell, while ParA separates the chromosomes/plasmids by forming filaments that polymerise and depolymerise (Gordon et al., 2004). The ParA proteins (ParA-ATP) extend from one cell extreme (new cell pole) to the ParB-DNA complex by making filaments. In turn, these filaments exert a pulling force on the ParB-DNA complex. The pulling force is generated by the change in the ParA structure upon ATP hydrolysis, that retracts the ParB-DNA complex to new pole (Fogel and Waldor, 2006).

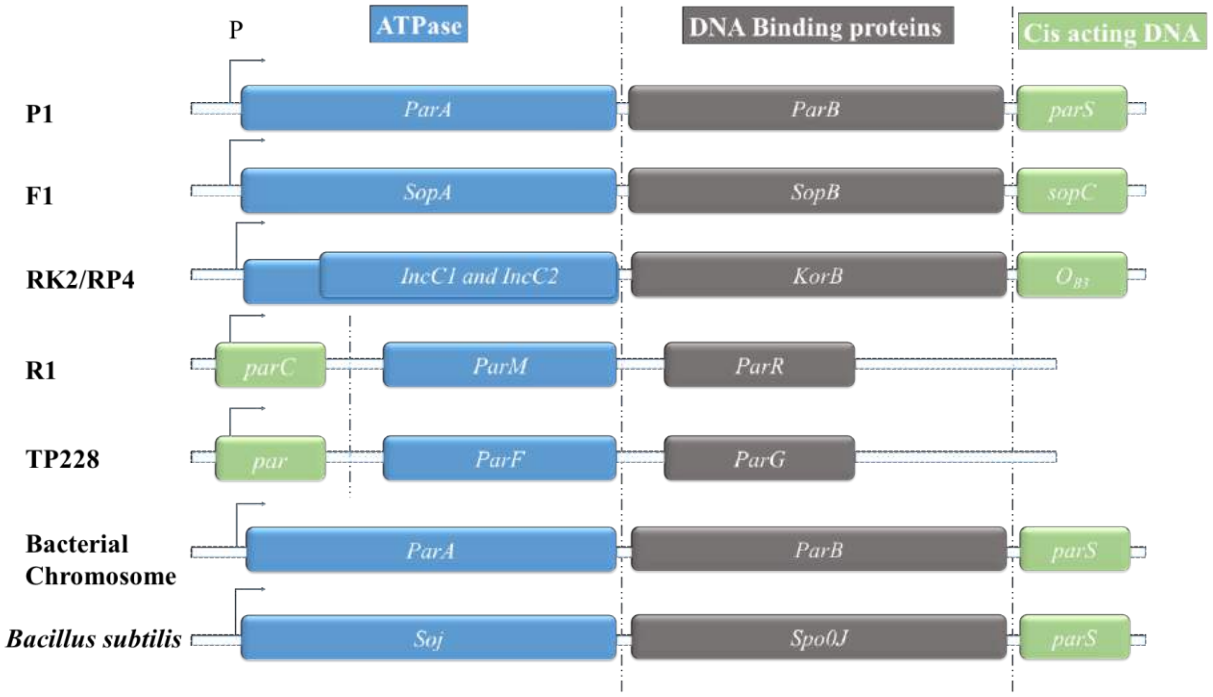


Figure 1.1 An illustration of *par* operons in different partitioning systems. P1 phage and F1 plasmids encode ParA/SopA proteins which are ATPases, ParB and SopB are the DN- binding proteins, while *parS* and *sopC* are the Cis acting DNA-binding sites (Bouet and Funnell, 1999, Libante et al., 2001). RK2 and RP4 plasmids encode two ParA proteins, IncC1 and IncC2, from the same gene but using different start codons (Thomas and Smith, 1986), and KorB protein that is equivalent to ParB protein (Thomas and Hussain, 1984). RK2 plasmid contains 12 O_B or KorB DNA-binding sites; O_{B3} has been shown in the figure (Rosche et al., 2000, Lukaszewicz et al., 2002) and is essential for partitioning. The R1 plasmid encodes ParF and ParG proteins equivalent in function to ParA and ParB proteins, while the DNA-binding site, *parC,* is present upstream of *ParM* (Jensen et al., 1994). The TP228 plasmid has a DNA- binding site upstream of the ParF (ParA) and ParG (ParB) proteins (Hayes, 2000, Schumacher, 2008). Bacterial chromosomal segregation systems in many hosts consist of ParA, ParB proteins and a *parS* DNA binding site. The *Bacillus subtilis* chromosome encodes Soj and Spo0J as the ParA/ParB equivalents (Ireton et al., 1994).

ParB is attached to chromosomal or plasmid *parS* sequence site in the bacterial cell. ParA from the Type I partitioning system contains a deviant Walker-Type motif containing a phosphate-binding-loop/P-Loop, KGGXXK[T/S]. This highly conserved motif is involved in ATP -binding and is usually present in various DNA-binding ATPase proteins (Ramakrishnan et al., 2002, Leipe et al., 2002, Lutkenhaus, 2012, Koonin, 1993). The classical Walker-type motif (GXXGXGK[T/S]) has also been found in many proteins that interact with phosphorylated substrates (Leipe et al., 2002). ParA proteins bind to nucleotides (ATP, ADP) using this Walker-type motif. ATP binding to ParA promotes its dimerization/polymerization that is required for its function (Davey and Funnell, 1994).

In this model, ParA (ParA-ATP and ParA-ADP) are randomly distributed in a cell and may interact with the nucleoid, while ParB dimeric proteins interact with the plasmid DNA. When ParA-ATP proteins interact with ParB proteins bound to plasmid DNA, the ParA-ATP proteins start to polymerize and make filaments (Figure 1.2 a-c). The ParB proteins stimulates ATP hydrolysis by ParA that causes retraction of the ParA filaments and release of the ParA-ADP protein. ParB-plasmid complex is then pulled forward, as ParB interacts with the next ParA-ATP in the filament, while ParA-ATP keep on polymerizing, and hydrolyzing/depolymerizing in the presence of ParB. This leads ParB-plasmid complex towards cell poles. The ParB-DNA complex dissociates at the cell poles (Figure 1.2 d-e) (Guynet and de la Cruz, 2011).

The Type I partitioning system can be subdivided into two categories, Ia and Ib, based on ParA size and the genetic organization of the *parABS* operon (Gerdes et al., 2000) (Figure 1.3). The Type Ia system encodes longer ParA proteins, comprising of 321-420 amino acids, in comparison to ParA from the type Ib partitioning system which contain 192-308 amino acids. The Type II ParA proteins are also Walker-like ATPases (Gerdes et al., 2000, Roberts et al., 2012). The ParA proteins involved in chromosomal partitioning are usually short, while plasmid partitioning systems may encode either longer or shorter ParA proteins (Figure 1.3).

The position of '*parS'* varies in these two systems; it is present downstream of the ParAB operon in type Ia system, and upstream of the operon in Type Ib partitioning system. ParB or DNA-binding proteins share little sequence homology to each other. These proteins contain either a ribbon-helix-helix (RHH) or a helix-turn-helix (HTH) fold (Figure 1.3) (Funnell, 2016, Sanchez et al., 2013). These folds can easily rotate around a flexible linker and bind to DNA at the partition site. ParB proteins usually exist as dimers or multimeric forms in solution.

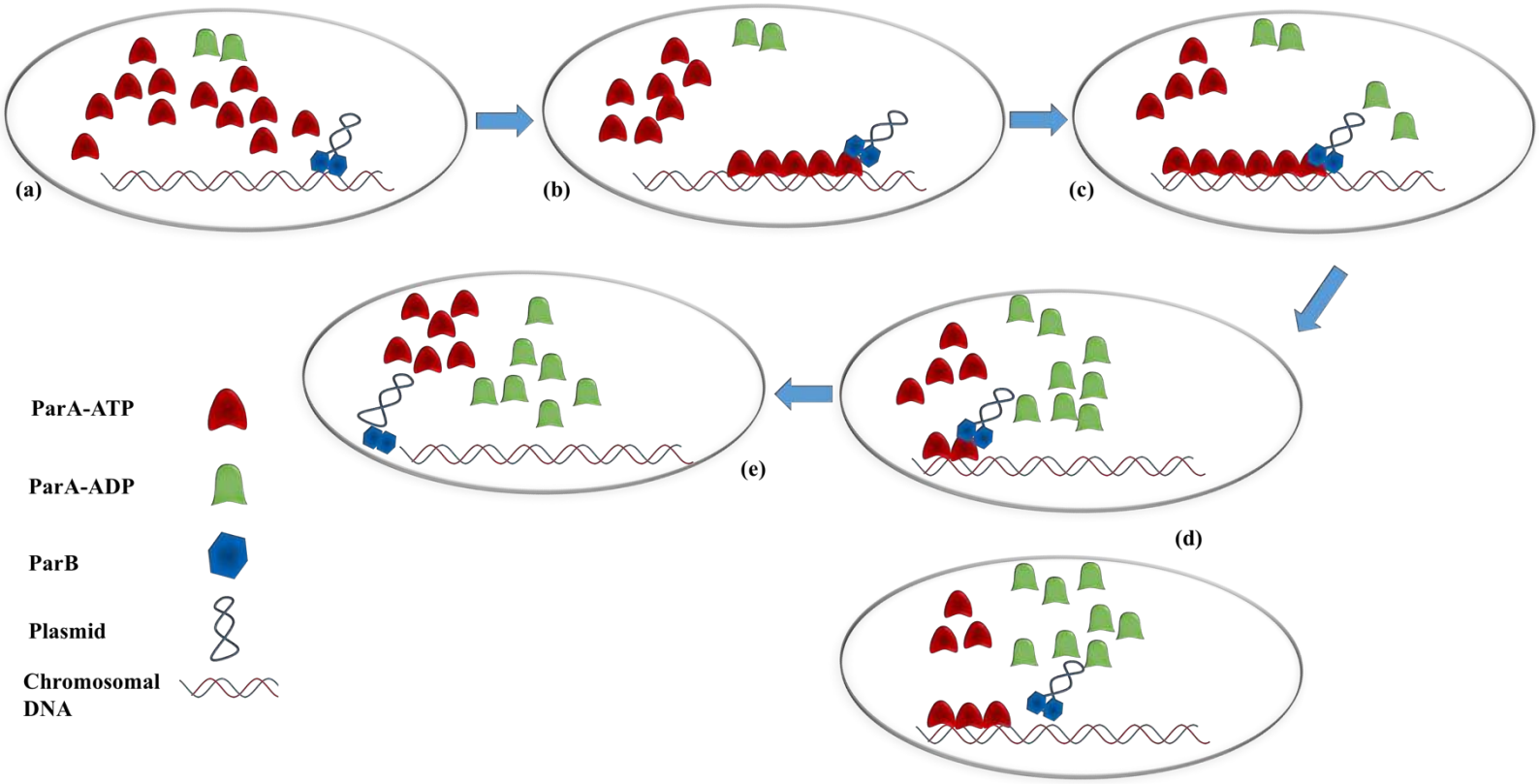


Figure 1.2 Type I partitioning model. **(a)** The ParA-ATP proteins (red) are distributed in a cell and may interact with the nucleoid, (double helix, red/blue). The ParB dimeric proteins (blue hexagons) interact with the plasmid DNA (blue, coil) **(b)** When ParA-ATP proteins interact with ParB proteins bound to plasmid DNA, the ParA-ATP proteins start to polymerize, giving filaments. **(c)** ParB stimulates the ATPase activity of the ParA-ATP proteins. ATP hydrolysis causes retraction of the ParA filaments and release of the ParA-ADP protein (green). **(d)** Plasmid bound to the ParB proteins is pulled forward, as ParB interacts with the next ParA-ATP in the filament. The ParA-ATP proteins keep on polymerizing, and hydrolyzing/depolymerizing in the presence of ParB, and lead ParB-plasmid complex towards cell poles. Sometimes the ParB-plasmid complex is released from the ParA filament but some of the ParA proteins again find the complex and lead it to cell poles **(d)** The ParB-plasmid complex reaches the cell pole, while interacting with the ParA filament. **(e)** The ParB-DNA complex dissociates when it reaches the cell poles (adapted from (Guynet and de la Cruz, 2011))

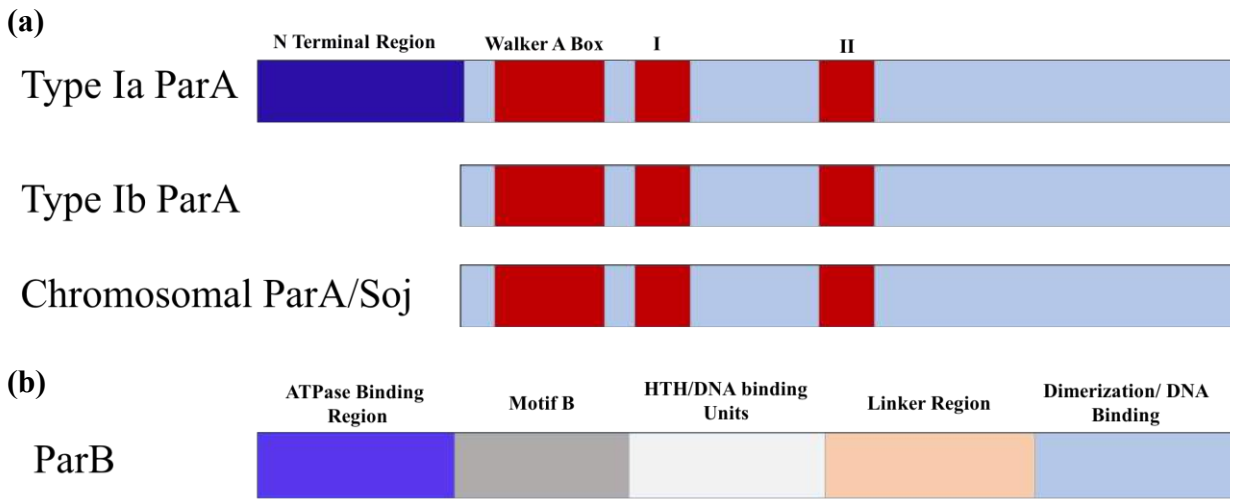


Figure 1.3 ParA proteins in the Type Ia and Ib partitioning systems. (a) Type Ia ParA proteins are longer than Type Ib ParA proteins. An N-terminal region (~100 amino acids, Blue) is absent in the 1b ParA proteins and in the chromosomal ParA proteins. A Walker-like motif (shown in red) and two conserved regions (I and II, red) are present in all Type 1 ParA proteins. **(b)** The ParB protein of Type I partitioning system consist of an N-terminal ATPase (ParA) -binding region (Dark blue), a conserved motif B (Grey), a central HTH or DNA binding region (White), a flexible linker region (Orange) and a dimerization/DNA binding domain (Light blue) (adapted from (Schumacher, 2008)).

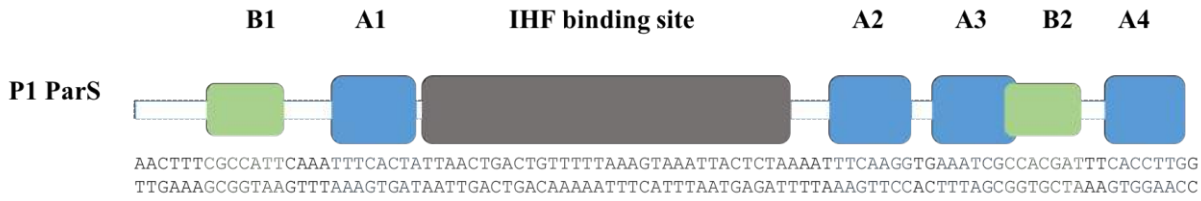
1.2.1.1 P1 bacteriophage – a Type Ia partition system

P1 is a low copy number bacteriophage that replicates in *E. coli* (Austin and Abeles, 1983). One of the widely studied type Ia partitioning system is the ParABS partitioning system from the P1 plasmid (Austin and Abeles, 1983, Ogura and Hiraga, 1983), where the *parS* centromeric site is present downstream of the *parA* and *parB* genes (Figure 1.1) (Abeles et al., 1985).

The 85 bp *parS* site consists of 4 A boxes, 2 B boxes and an IHF (integration host factor) binding site (Figure 1.4a). Mutational analyses showed that the presence and relative spacing of boxes A2, A3, B1 and B2 are vital for the function of the wild-type *parS* site *in vivo,* while boxes A1 and A4 are dispensable (Funnell and Gagnier, 1994). Both ParB and IHF augment each other's binding affinity towards the *parS* DNA (Funnell, 1991), IHF binding creates a large bend in *parS* and a ParB dimer binds to the DNA by a Helix-Turn-Helix motif present in the DNA-binding domain. The C-terminal dimerization domain of ParB interacts with Boxes B1 and B2, present in different positions in *parS* (Figure 1.4b). The C-terminal domain of ParB needs to dimerize in order to interact with DNA (Schumacher and Funnell, 2005). As a result, a ParB-*parS* complex is formed in which ParB and *parS* site are wrapped around each other (Funnell, 1991). This DNA-

protein complex allows bridging of the *parS* arms and subsequently leads to DNA condensation (Surtees and Funnell, 2001). IHF is not directly involved in partitioning but it bends DNA to bring ParB and its interactions sites in close proximity (Funnell, 1991). After formation of the initial ParB-DNA complex, ParB spreads along the DNA (500 bp) (Rodionov et al., 1999).

(a)



(b)

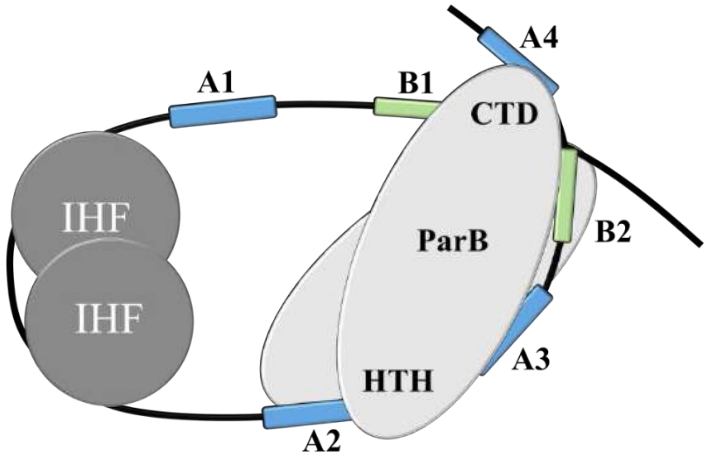


Figure 1.4 The P1 *parS* site and partitioning complex. (a) An illustration of the *parS* site of P1. The site consists of four conserved A boxes (A1-A4, blue), 2 B Boxes (B1, B2, green) and an IHF (integration host factor) binding site (Grey). The *parS* sequence is also shown. **(b)** The P1 partitioning complex. The Helix turn helix (HTH) regions of the ParB dimer (white) bind to the A2 and A3 boxes (blue), while the C-terminal domain (CTD) of the ParB dimer interacts with the B1 and B2 boxes (green). The IHF dimer (Grey) binds to the IHF binding site and causes bending of the *parS* site (adapted from (Surtees and Funnell, 2001)).

For the P1 plasmid partitioning, ATP hydrolysis was considered to be needed to provide energy for DNA transport, while this hydrolysis is not necessary for the ParA-ParB interactions, only ParA-ATP binds to ParB (Bouet and Funnell, 1999). The ATP-ADP switch allows the ParA proteins to have two functions, the ATP bound form interacts with ParB-DNA complex, and ATP hydrolysis leads to plasmid partitioning; while ParA-ADPs act as the repressor of par operon (Figure 1.5) (Bouet and Funnell, 1999).

The *parS* site of P1 is involved in *par* operon regulation (Hao and Yarmolinsky, 2002). The ParA-ParB ratio is very important in regulation of the *par* operon as well as in stabilizing the nucleoprotein complex (Pratto et al., 2009, Friedman and Austin, 1988, Funnell, 1988, Abeles et al., 1985). The weak ATPase activity of ParA is stimulated by interaction with non-specific DNA, ParB or both (Davis et al., 1992). ATP- and ADP-bound ParA proteins show different properties in terms of repression, conformation, dimerization and in their interaction with other partners (Davey and Funnell, 1994, Davey and Funnell, 1994, Bouet and Funnell, 1999).

1.2.1.2 F plasmid – another type Ia partition system

F plasmid is a low copy number plasmid. Its partitioning system consists of SopA (a ParA protein), SopB (a ParB protein) and *sopC* (the DNA binding site) (Figure 1.1) (Ogura and Hiraga, 1983). The gene organization of the *par* operon of F plasmids is shown in Figure 1.6. The SopA protein polymerizes as filaments radiating from the SopB-plasmid complex, in the presence of SopB and s*opC*, pushing the plasmids apart (Lim et al., 2005, Bouet et al., 2007). SopB dimerizes to interact with *sopC* (43 bp) and forms the initial partitioning complex. In contrast to the P1 plasmid partitioning system, the SopA-DNA interaction inhibits SopA (ParA) polymerization and SopB is required to mask the DNA and form DNA- depleted zone (Bouet et al., 2007). SopA also acts as a weak auto-repressor when binds to *sopAB* promoter region (Figure 1.6).

The ATPase activity of SopA is critical for the partitioning process (Gerdes et al., 2000) as SopA polymerizes into long filaments when bound to ATP. When SopA-ATP interacts with the SopB-DNA complex, it forms long filaments radiating from the complex, like eukaryotic mitotic spindles (Figure 1.7) (Lim et al., 2005). Recent studies on P1 and F1 plasmids suggest diffusion of SopB-*sopC* complex along a SopA-ATP gradient (Hatano and Niki, 2010, Vecchiarelli et al., 2013) (detailed in section 1.3.1).

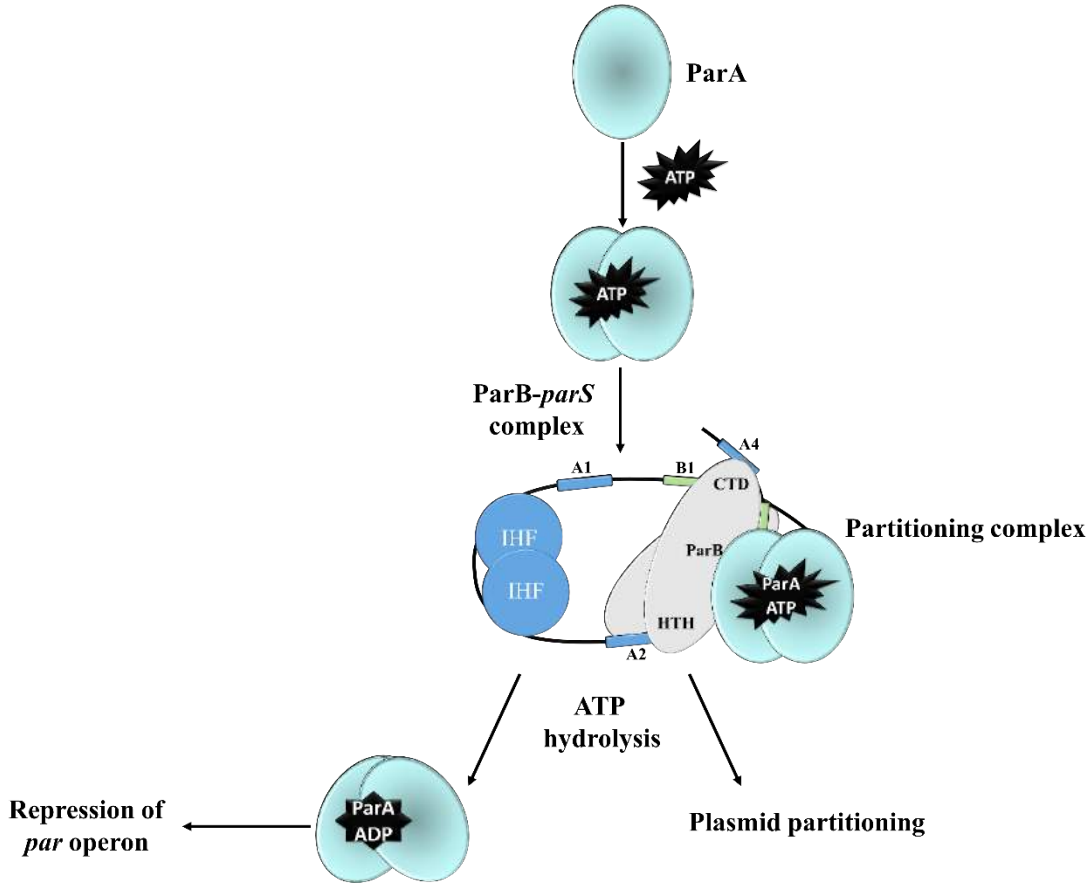


Figure 1.5 Action of ParA-ATP and ParA-ADP in P1. ATP (shown in black) binding to the ParA protein leads to its dimer formation. ParA-ATP interacts with ParB and ParB-*parS* (detailed in Figure 1.4) to give a partitioning complex. IHF is integrating host factor These interactions lead to plasmid partitioning involving ATP hydrolysis and formation of ParA-ADP. ParA-ADP dissociates from the partitioning complex to make room for the next ParA-ATP to bind to the complex (as shown in Fig 1.2). ParA-ADP is also involved in the repression of *par* operon (adapted from (Bouet and Funnell, 1999))

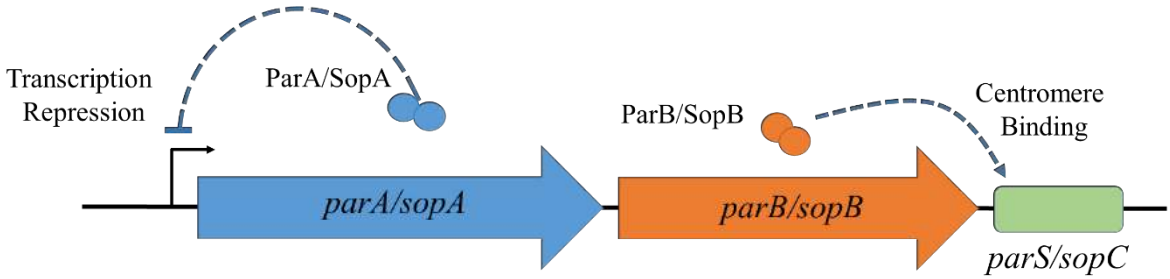


Figure 1.6 Organization of the *par* operon of the P1/F plasmids. The *parS*/*sopC* site (green box) is present downstream of the P1/F plasmid *par* operon. The ParB/SopB proteins (orange circles, showing dimeric protein) bind to *parS*/*sopC* (blue arrow line shows binding of ParB to *parS*), while the ParA/SopA proteins (blue circles showing dimeric protein) are involved in repression of the *par* operon (adapted from (Hayes and Barilla, 2006)).

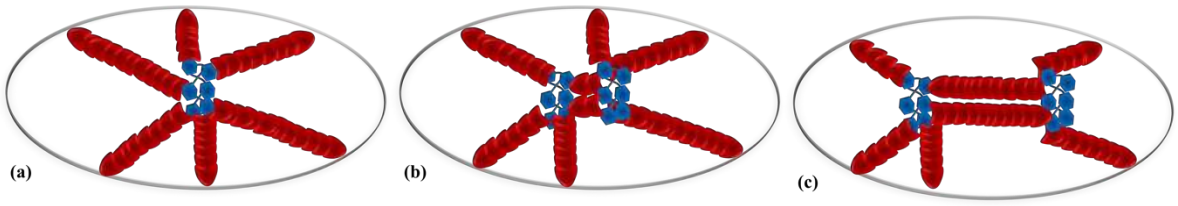


Figure 1.7 An illustration of partitioning of F plasmid. (a) SopA proteins (red) radiate from the nucleoprotein complex where SopB (blue diamonds) interacts with *sopC* (blue lines). **(b, c)** A pulling force generated by ATP hydrolysis drives plasmids to the cell poles resulting in plasmid partitioning (adapted from (Lim et al., 2005))

1.2.1.3 TP228 Plasmid - a type Ib partition system

The Type Ib partition system is similar to Type Ia but instead of ParA, ParB performs auto-regulatory activity. The ParB protein homologs, like in the Type Ia system, do not share sequence similarity with each other, except for the helix- turn-helix region. The partitioning system from the multi-antibiotic resistant, low copy number plasmid, TP228 from *E. coli* and *Salmonella newport* is the best-studied Type Ib partition system (Hayes, 2000). In this system, the NTPase activity of ParF (a ParA equivalent) interacts with the ParG protein (a ParB equivalent) (Zampini et al., 2009, Barilla et al., 2007). ParG activates the ATPase activity of the ParF protein that is crucial for its polymerization and de-polymerization. ParG act as a transcription repressor at the *par*H (centromere-like) site (Figure 1.8) (Schumacher, 2012, Schumacher, 2008).

ParF polymerization has been observed in the presence of ATP and ATPγS, while the presence of ADP ParF polymerization becomes unstable (Barilla et al., 2005). The ratio of ParF to ParG also affects polymerization, as the weak ATPase activity of ParF is stimulated by ParG, which leads to polymer destabilization.

ParF and ParG homologs have been found in various plasmids including pTAR plasmid from *Agrobacterium tumifaciens* (Kalnin et al., 2000), a pRA2 plasmid from *Pseudomonas* (Hayes, 2000), pB171 from *E. coli* and pVS1 plasmid from *Pseudomonas aeruginosa* (Hayes, 2000). The pB171 plasmid encodes ParA2 protein which is found to form filaments on the nucleoid (Ebersbach and Gerdes, 2005).

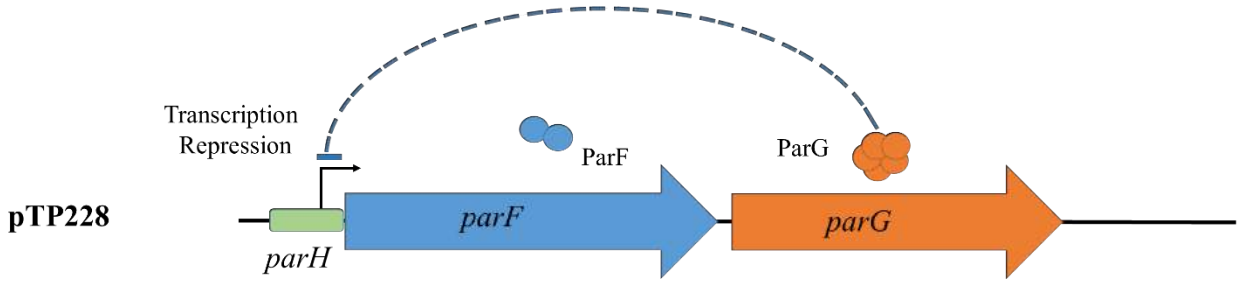


Figure 1.8 Genetic organization of the operon *parFGH* of pTP228. The ParF protein (blue circles) is equivalent to a ParA protein, ParG (orange circles) is a DNA-binding protein, while *parH* (green box) is a DNA-binding site and is present upstream of the *par* operon. The ParG protein is involved in the repression of the *par* operon (adapted from (Hayes and Barilla, 2006, Schumacher, 2008).

1.2.2 Chromosomal partitioning systems

1.2.2.1 *Bacillus Subtilis*

The ParA/ParB homologs Soj and Spo0J are involved in chromosomal DNA partitioning (Lee and Grossman, 2006) as well as in regulation of sporulation, a process of asymmetric cell division under stringent conditions (Mori et al., 1989). Proper chromosomal partitioning is required during the sporulation process (Thomaides et al., 2001). Spo0J is involved in sporulation initiation, while Soj inhibits this process if Spo0J is not present by regulation of sporulation genes including *spoIIA*, *spoIIE* and *spoIIG* (Quisel et al., 1999).

In the partitioning process, Spo0J interact with one of the 10 *parS* sites clustered on the bacterial chromosome near the origin of replication. The absence of Spo0J results in 1-2 % anucleate cells, elongated cells and cells with chromosomal de-condensation (Autret et al., 2001). The crystal structure of the dimeric form of Spo0J from *Thermus thermophilus* has been determined (Leonard et al., 2004). There are two primary C-terminal dimerization domains as well as two helix-turn-helix (HTH) DNA binding domains in the Spo0J dimer (Figure 1.16). Soj is very important for partitioning as well as for sporulation regulation. It localizes to cell poles in the presence of the Spo0J and is responsible for DNA condensation (Lee and Grossman, 2006). The non-specific Soj-DNA interactions require the presence of nucleotides (ATP/ADP). ATP is necessary for Soj dimerisation as well as for its further polymerisation (Leonard et al., 2005).

The Soj protein has been reported to show a dynamic location pattern in the cell and can assemble/disassemble itself in large, nucleoid-associated structures (Marston and Errington, 1999). The change in Soj oligomeric states suggests that Soj may mimic filamentous cytoskeletal structures *e.g.* actin filaments, the activity of which is also controlled by nucleotide binding, ATP hydrolysis and by regulatory proteins (Schmidt and Hall, 1998). The crystal structure of Soj has been determined as a dimeric protein in the presence of ATP (Figure 1.18). The dimerization interface is located next to the nucleotide-binding surfaces. Soj interacts with N-terminal region of Spo0J that, in turn, stimulates its ATPase activity (Leonard et al., 2005).

1.2.2.2. *Vibrio Cholera*

This rod-shaped, gram-negative, bacterium utilizes a ParAB partitioning system for segregation of its two chromosomes (ChrI and ChrII). Chromosomal segregation in ChrI is similar to that of *B. subtilis*, involving replication at one cell pole and then transfer of copies towards the other cell pole, whereas in ChrII, this process mimics P1and F plasmid partitioning (Gordon et al., 1997, Niki and Hiraga, 1997). In Chromosome I, the *par* genes (*parAI* and *parBI*) and their 3 DNA-binding sites are present near to the origin of replication. During segregation, the ParAI protein first polymerizes from mid-cell to the poles, while interacting with the replicated origin of replication and ParBI. Then ParAI shrinks towards the new poles by either pulling or pushing the origin of replication towards the new poles (Fogel and Waldor, 2006). In case of ChrI, deletion of the *par* genes has no effect on segregation but *parA*/*parB* gene deletion in ChrII (*parAII, parBII*) results in loss of ChrII (Heidelberg et al., 2000).

1.2.2.3 *Caulobacter crescentus*

The ParA and ParB protein homologs are involved in the partitioning of the *C. crescentus* chromosome. The stalked cells undergo asymmetric cell division after chromosomal segregation (Ryan and Shapiro, 2003). During segregation, the ParB dimer can binds to DNA at one of the 6 *parS* sites on the chromosome (Figge et al., 2003, Mohl et al., 2001). The *par* operon is regulated by ParB repression, which can be reversed by its interactions with ParA (Easter and Gober, 2002). In this segregation system, ParA-ADP can interact with single-stranded DNA, while ParA-ATP does not interact with DNA, but relieves *par* operon repression by ParB (Easter and Gober, 2002). The normal protein expression of both ParA and ParB is crucial to chromosomal segregation.

Overexpression or depletion of either of these proteins results in severe defects in cell division, chromosomal organization and segregation (Mohl et al., 2001).

1.2.2.4 *Pseudomonas putida*

ParA, ParB homologs and 6 *parS* sites are involved in the segregation of the *P. putida* chromosome. Unlike the segregation systems in *B. subtilis* and *C. crescentus,* the DNA binding sites (*parS*) of *P. putida* are widely distributed in the chromosome. Mutations in the ParA protein do not affect cell viability (Mohl and Gober, 1997) and chromosome segregation may not require a complete ParAB partitioning system (Bignell et al., 1999). Homologs of the Par proteins are not found in *E. coli* and *Haemophilus influenza* (Gordon and Wright, 2000).

1.2.3 Archaeal partitioning system

Chromosomal segregation in archaea, like *Sulfolobus solfataricus,* involves two proteins SegA and SegB (Kalliomaa-Sanford et al., 2012). The SegA protein is an ATPase and is an ortholog of Walker-type ParA in bacteria. The SegB protein is involved in specific DNA interactions. The DNA-binding protein SegB interacts with SegA only in the presence of nucleotides and polymerisation of SegA is stimulated in the presence of SegB.

1.2.4 Type II partitioning systems

In type II systems, the ParA proteins are structurally homologous to the Actin like/Hsp70 superfamily of proteins (Bork et al., 1992). ParA does not directly interact with DNA at the partition sites but is instead recruited into the partition nucleoprotein complex through interactions with ParB (Barilla and Hayes, 2003, Bouet and Funnell, 1999). Partition is facilitated by the formation of ParA filaments, resulting in movement of plasmids to cell poles (Moller-Jensen et al., 2003, Moller-Jensen et al., 2002).

1.2.4.1 R1 Plasmid

Par partitioning system in the R1 plasmid is the best example of Type II partitioning system. Like other ParA partitioning systems, it consists of a DNA -binding protein (ParR), a protein with ATPase/GTPase activity (ParM) and a centromere-like binding site (*parC*) (Figure 1.1). In contrast to the Type I partitioning system , the DNA binding site, *parC,* is an 160 bp site located upstream of *parR* and *parM* (Figure 1.1) (Dam and Gerdes, 1994, Jensen and Gerdes,

1997, Jensen et al., 1998, Jensen et al., 1994). The ParR protein, in complex with the *parC* site, interacts with growing filaments of the actin-like ParM proteins (Bharat et al., 2015). The plasmid copies are positioned in the middle of the cell after replication and bind to the ParR protein that moves the plasmids to the opposite poles of the cell (Perez et al., 1999).

Unlike the Type I partition system, the ParM (ParA) proteins do not interact with the nucleoid to polymerise, instead, the polymer ends get extended by ParM-ATP subunits, while pushing the plasmids to the cell poles (Figure 1.9). The ParR protein interacts with *parC* and ParM-ATP. The ParA-ATP when interact with the ParR-DNA nucleoprotein complex, it starts polymerizing (Garner et al., 2004) as elongated polymers/bidirectional-filaments that push the ParR-DNA complex to cell poles. ParM Polymers dissemble after ATP hydrolysis (Moller-Jensen et al., 2003). The ParR-DNA-ParM complex does not dissociate until ParM-ATP protein is present at polymer cap (Garner et al., 2004) (Figure 1.9). Both *in vivo* and *in vitro* studies show a pushing plasmid model with the arrangement of bead-like ParM polymers (Schumacher, 2012, Schumacher, 2008). The DNA binding protein, ParR, binds to *parC* autoregulating its own operon (Figure 1.10) (Jensen et al., 1994).

ParM is an ATPase and the motor protein of the partitioning system. *In vivo* and *in vitro* studies show that ParM makes actin-like, left-handed, double-stranded filaments when it binds to the plasmid (Gayathri et al., 2012, Gayathri et al., 2013). ParM-DNA binding requires ParM to be in an active form bound to ATP, ParM-ATP. Plasmid pairing and polymerisation do not occur with ParM-ADP or ParM-ATPγS; that shows that ATP hydrolysis destabilizes the ParM filament (Garner et al., 2004, Jensen et al., 1998).

Cryo-EM studies show differences in the helical symmetry of ParM-ATP filaments and destabilizing ParM-ADP filaments after ATP hydrolysis (Bharat et al., 2015). Filament formation also depends on the ParR protein bound to plasmid at *parC*, which stimulates polymerization when it interacts with ParM (Moller-Jensen et al., 2002). Microscopy studies show that ParM-ATP is added to the positive end (the end attached to the ParR-*parC* complex) of the bidirectional filament (Garner et al., 2007), and photobleaching experiments are in agreement with this observation (Moller-Jensen et al., 2003).

1.2.4.2 *Bacillus thuringiensis* pBMB67 plasmid

The *Bt*ParM protein from the *Bacillus thuringiensis* pBMB67 plasmid forms antiparallel, 2-4 stranded, helical supercoiled filaments when interacting with ParR (Jiang et al., 2016). It forms filaments only in the presence of ATP, while ATP hydrolysis acts as a switch for *Bt*ParM polymerization and depolymerization. The filament formation stimulates ATP hydrolysis and after hydrolysis *Bt*ParM-ADP filaments are destabilized (Jiang et al., 2016).

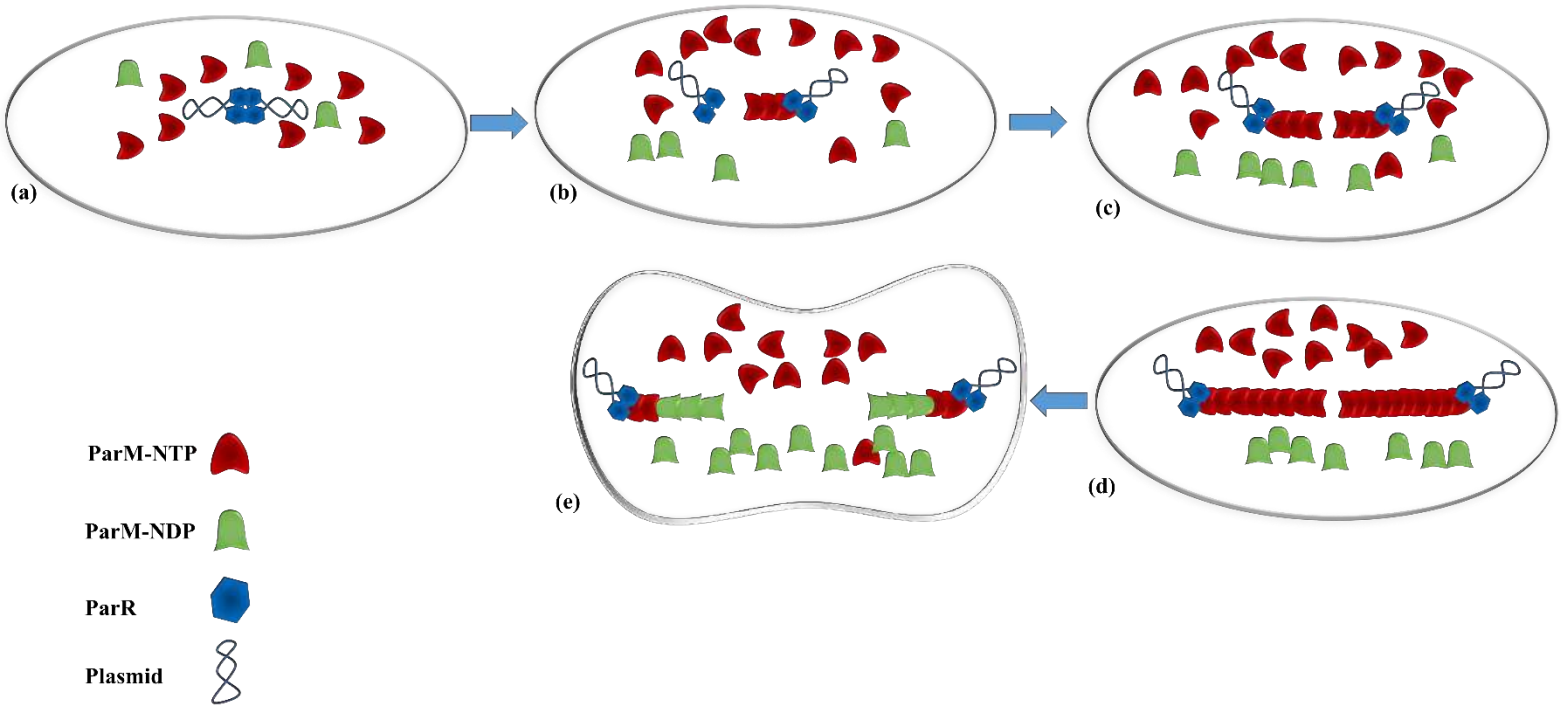


Figure 1.9 Pushing partitioning model (a) The ParR protein (Blue diamonds) binds to the *parC* site on the plasmid DNA (blue), while plasmids are located in mid-cell after replication. **(b)** The ParM-ATP protein complex (Red) explores the entire cell to interact with the ParR-DNA nucleoprotein complex and starts forming polymers (Garner et al., 2004) **(c)** The ParM-ATP proteins form elongated polymers/bidirectional-filaments that push the ParR-DNA complex to cell poles **(d)** Plasmids are moved to cell poles while pushed by the ParR/ParM complex. **(e)** ParM Polymers dissemble after ATP hydrolysis (Moller-Jensen et al., 2003). The ParR-DNA-ParM complex does not dissociate until ParM-ATP protein is present at polymer cap (Garner et al., 2004). (adapted from (Garner et al., 2004, Garner et al., 2007, Million-Weaver and Camps, 2014))

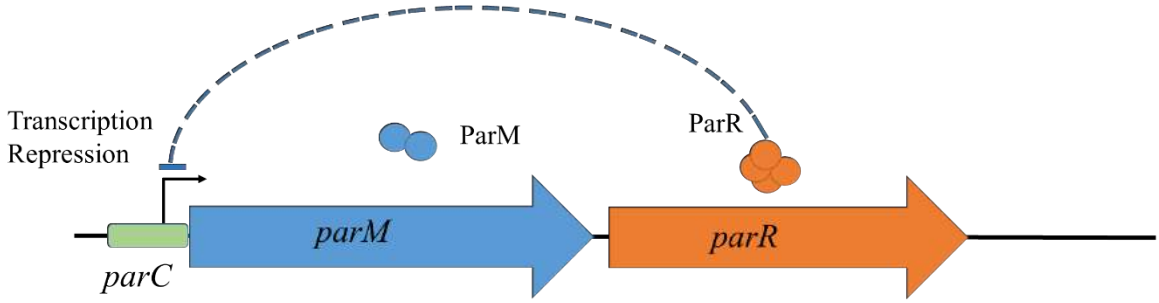


Figure 1.10 Genetic organization of the *par*MRC operon. ParM (blue circles) is equivalent to a ParA ATPase. ParR (orange circles) is a DNA binding protein and *parC* (green box) is a DNA binding site present upstream to the operon. ParR is involved in regulation of its own operon by binding to the *parC* site (adapted from (Hayes and Barilla, 2006, Schumacher, 2008)).

Another low copy number, plasmid encoding multiple antibiotic resistance genes from *Staphylococcus aureus*, pSK41plamsid, shows the presence of a Type II partitioning system. The *par* operon encodes ParM and ParR proteins (Popp et al., 2010). ParR forms dimers that bind to *parC* located upstream of *parM* (Schumacher et al., 2007). Like ParR from R1 plasmid, ParR from pSK41 autoregulates its own operon. The details of the partitioning process for pSK41plamsid are still to be elucidated (Liu et al., 2013).

1.2.5 Type III partitioning systems

The type III partitioning system is different from the previously described partitioning systems and has been well studied in the pBtoxis plasmid from *B. thuringiensis* and in the pX01 plasmid of *Bacillus anthracis* (Tinsley and Khan, 2006, Anand et al., 2008, Fink and Lowe, 2015). In the *tubZRC* partitioning system of pBtoxis, the DNA-binding protein, TubR, interacts with the TubZ protein (a GTPase). The TubR protein forms dimers and its N-terminal domain binds to the DNA major groove. TubR binds to the plasmid and oligomerizes at four, 12 bp, repeated sequences (*tubC*) (Ni et al., 2010). TubR is also involved in transcription autoregulation (Figure 1.11). TubZ is a tubulin/FtsZ-like protein that forms parallel, right-handed, double-stranded, stable filaments in the presence of GTP (Aylett and Lowe, 2012).

Many TubR dimers bind to one pBtoxis plasmid. The C-terminal part of TubZ interacts with the TubR-plasmid complex. That stimulates treadmilling of TubZ, leading to plasmid tramming. TubZ subunit is added at one end (+) of the TubZ filament and disassembles at the other

end (-). The TubR-plasmid complex is moved to the continuously growing + end of the TubZ filament, until it reaches the cell poles. After the plasmid reaches the cell pole, the filaments of TubZ bends and moves towards opposite, pole, leaving the plasmid at the cell pole (Figure 1.12). The bending of TubZ filaments stimulates TubR-plasmid detachment (Ni et al., 2010, Schumacher, 2012), but TubZ movement does not depend on TubR interactions (Larsen et al., 2007).

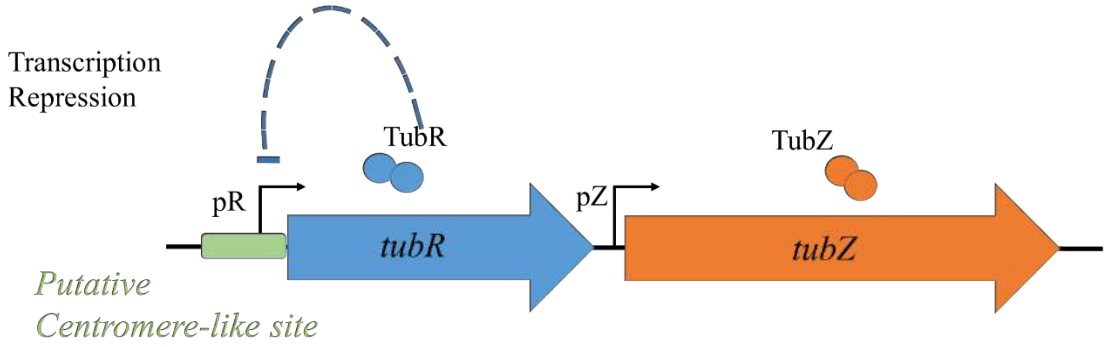


Figure 1.11 Genetic organization of *TubR and TubZ genes*. The TubR protein (blue circles) is involved in regulation of its own promoter (pR, green box shows the centromere-like site where TuR binds) and binds to plasmid DNA. TubZ (orange circles) is a GTPase and is expressed from the pZ promoter (adapted from (Guynet and de la Cruz, 2011).

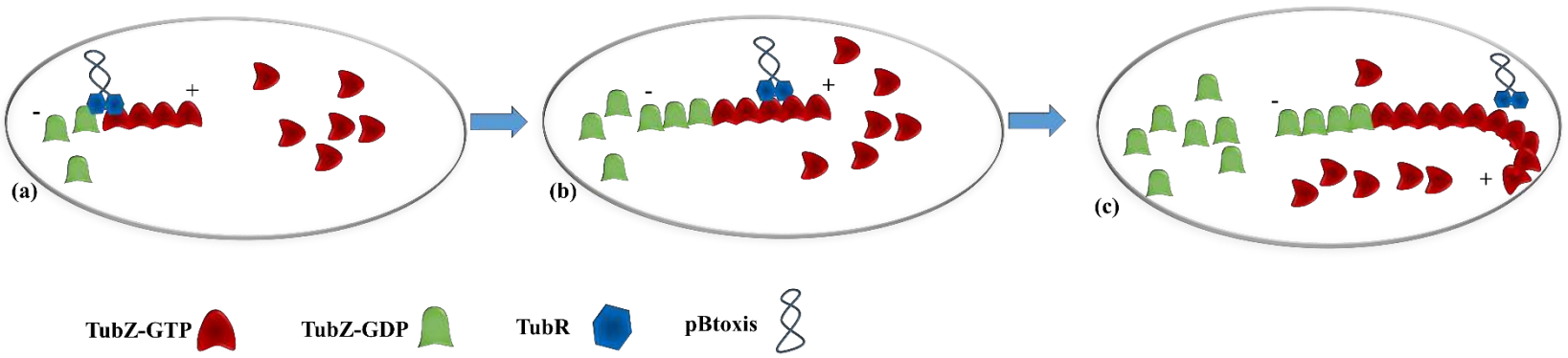


Figure 1.12 Tramming/Treadmilling partitioning model (a) The TubR protein (Blue diamonds) binds to plasmid (Blue) and TubZ (Red and green). The TubR interactions with TubZ-ATP (Red) stimulate formation of microtubule-like filaments **(b)** The interaction of the TubR-Plasmid complex with TubZ starts filament formation at the + end and the ParR-plasmid complex is moved by tramming to the cell poles **(c)** At the cell pole, the plasmid is released, while TubZ filaments bend and move back towards the other cell pole (Larsen et al., 2007). (adapted from (Baxter and Funnell, 2014)

1.2.6 Type IV partitioning system

Some low-copy-number plasmids do not possess proper *par* partitioning systems. This suggests there may be some other partitioning systems that remain to be explored (Guynet and de la Cruz, 2011). The pSK1 plasmid from *Staphylococcus aureus* and plasmid R388 from *E. coli* show significant differences from usual *par* systems (Simpson et al., 2003). A proposed partitioning model for these plasmids is the 'Pilot-fish model' which is independent of shape, width or length of the host cell. That makes it a good system for broad host range plasmid partitioning. The Pilot Fish partitioning system does not require a motor protein or ATPase activity.

Plasmid R388 is a low copy number, broad host range plasmid (Suzuki et al., 2010) that does not contain typical partitioning system but maintains at least four plasmid copies in *E. coli* (Fernandez-Lopez et al., 2006). The *stbABC* operon of this plasmid (Figure 1.13) contain genes for StbA (a DNA binding protein), StbB (a NTPase) and StbC but StbA alone is sufficient for segregation (Guynet and Cruz, 2011). StbB (an ATPase) is not involved in partitioning, and segregation is completed with the help of the host nucleoid by some mechanism that has still to be explored (Figure 1.14). StbA binds specifically to a *stbDRs* site on R388 plasmid via its N-terminal and subsequently, this nucleoprotein complex interacts with host chromosome that acts as "Pilot Fish" to carry out plasmid segregation (Guynet and de la Cruz, 2011).

In pSK1 partitioning, only one Par protein is required. The Par protein is involved in DNA binding and no ATPase activity is needed. The structure of this Par protein contains both a putative HTH and a coiled-coil region, which may provide it with DNA binding and polymer-forming properties, respectively (Simpson et al., 2003, Guynet and de la Cruz, 2011).

Plasmid pRC4 is a low copy (2-4 copies per genome) number plasmid from *Rhodococcus erythropolis* (Hashimoto et al., 1992) and does not contain any known partitioning protein. There are also some small size plasmids that do not contain partitioning systems including pSC101(from *Salmonella panama*) (Meacock and Cohen, 1980), pLS11 (from *Bacillus subtilis*) (Chang et al., 1987) and pYAN-1(from *Sphingobium yanoikuyae*) plasmids (Hayashi and Kurusu, 2014). Such plasmids are proposed to use cellular partitioning components *i.e.* the cell membrane and the host nucleoid for stable maintenance and partitioning (Hayashi and Kurusu, 2014).

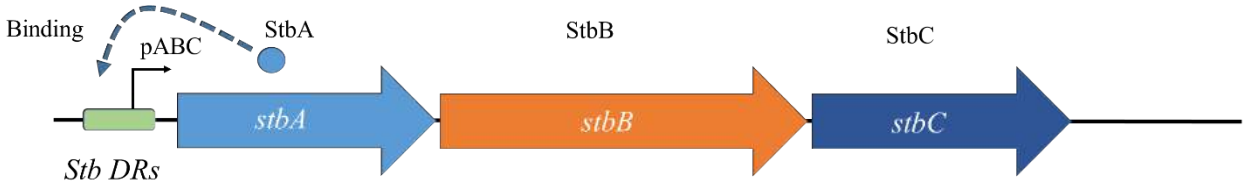


Figure 1.13 Genetic organization of the *stbABC* operon. Three proteins are expressed from the operon *i.e.* StbA, StbB, StbC but only StbA (blue circle) is involved in binding to plasmid DNA (*stbDRs,* green) and alone is sufficient for the partitioning process (adapted from (Guynet and de la Cruz, 2011)

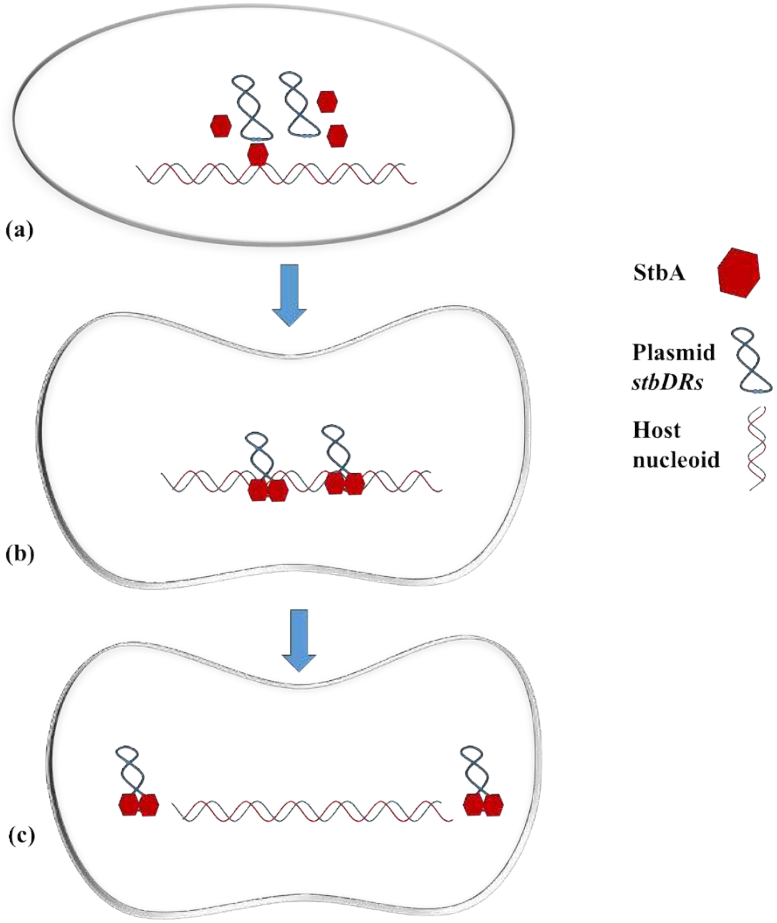


Figure 1.14 Pilot fish partitioning model for plasmid R388 segregation (a, b) The StbA protein (Red diamonds) binds to the *stbDRs* site on the plasmid (blue). The complex further interacts with host nucleoid (red/blue wave) and **(c)** without any motor protein, the plasmids are moved to cell poles by a still unknown mechanism. (adapted from (Guynet and de la Cruz, 2011)

1.3 Recent plasmid partitioning models

Although Type I and Type II partitioning systems have been widely studied for plasmid, as well as chromosomal, segregation, recent advances show many differences from conventional partitioning models. Recently reported, plasmid diffusion-based models consider that a dynamic ParA-ATP gradient allows plasmid transport, instead of ParA filaments that act by pushing or pulling plasmids (Brooks and Hwang, 2017). *In vivo* and *in vitro* studies suggest diffusion/Brownian ratchet models that are the most acceptable partitioning model to date (Hwang et al., 2013, Vecchiarelli et al., 2010, Vecchiarelli et al., 2012, Vecchiarelli et al., 2014, Hu et al., 2017).

1.3.1 Diffusion/Brownian ratchet model

This model has been proposed from experiments using the *par* system of the P1 plasmid (Hatano and Niki, 2010) and the F1 plasmid (Vecchiarelli et al., 2013). In this partitioning model, ParA proteins do not form filaments. Several ParB protein molecules bind non-specifically to the *parS* site on plasmids and cluster them to form a partitioning complex (PC). ParB binds to the parS site on the plasmid and then spreads along the plasmid, binding non-specifically. Active ParA (ParA-ATP) forms diffuse clouds on the nucleoid by non-specific DNA interactions. The partitioning complex (PC) interacts with ParA-ATP to stimulate the ATPase activity of ParA, while ATP hydrolysis causes ParA to come off the nucleoid as ParA-ADP (Baxter and Funnell, 2014). The ParB-plasmid partitioning complex moves towards a new ParA-ATP, while leaving ParA-ATP depletion zones behind. The depletion zones are refilled by diffusing ParA-ATP from other areas. The transient depletion zone allows directed movement of the partitioning complex. In the case of single partition complex, there is oscillatory movement of the plasmid from pole to pole, but, when more than one partitioning complex is present, directed motion allows the complex to make its way to the cell poles (Figure 1.15) (Hu et al., 2017, Brooks and Hwang, 2017). *In vivo* and *in vitro* studies favour diffusion ratchet model instead of pulling or pushing filament partitioning models. A ‘DNA relay’ model has been suggested on the basis of computer modelling, describing ParA-DNA complex randomly tethers and move along the chromosomal DNA (Lim et al., 2014). The complex leaves a ParA-ATP depletion zone behind that allows the directed movement. In the Brownian ratchet model, collective ParA-ParB bond interactions provide a driving force to transport plasmid via directed diffusion.

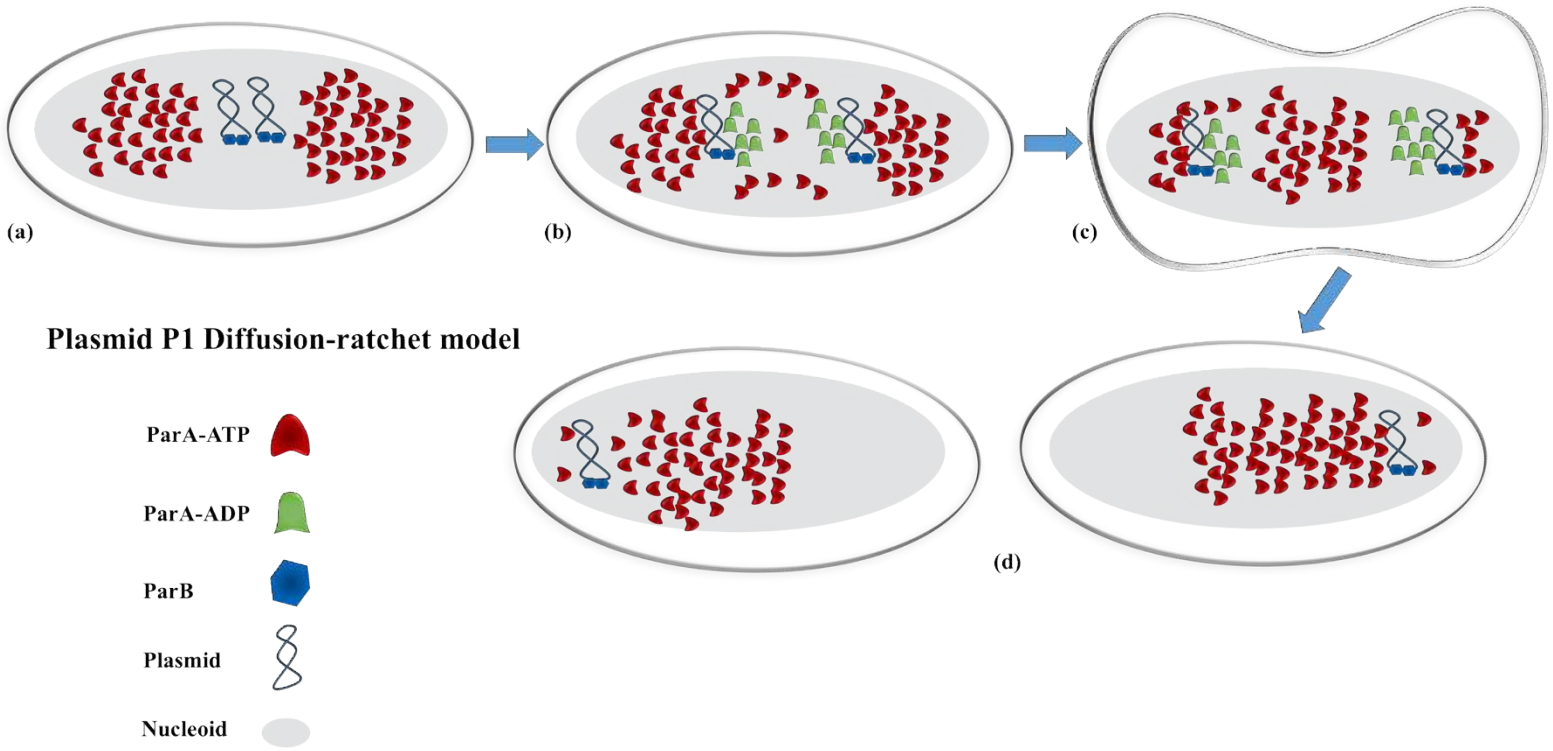


Figure 1.15 An illustration of the **Diffusion Ratchet model: (a)** DNA binding protein (ParB, shown as blue diamonds) interacts with plasmid *parS* site on plasmid (plasmid shown as blue lines) to form a partitioning complex **(b)** The partitioning complex interacts with ParA-ATP protein (red) bound to nucleoid (light blue circle) and stimulates ATP hydrolysis. After ATP hydrolysis, ParA-ADP (green) is released from the partitioning complex and ParB-DNA complex moves to the next available ParA-ATP by Brownian movement, leaving behind an ParA-ATP depletion zone. The ParA-ATP depletion zone allows directional movement of the partitioning complex that diffuses towards the cell poles. The ATP concentration gradient provides a driving force to the partitioning complex. **(c)** The adjacent activated ParA-ATP proteins diffuse towards the depletion zone to refill it. The partitioning complex may oscillate between the poles if only a single plasmid is present **(d)** By directional movement and diffusion, the daughter plasmids reach the cell poles

1.4 Structural biology of partitioning proteins

A few structures of plasmid and chromosomal segregation proteins have been reported in Protein Data Bank (PDB). These structures may contribute to the understanding of the mechanism and models of partitioning and provide information on protein-protein and protein-DNA interactions. The structure-function relationship of these molecules shows the different structural motifs that occur and how these are involved in partitioning. Binding of ATP and its hydrolysis effects the activation/deactivation of the ParA proteins in the presence of ParB and DNA. The ParA homologs may form dimers and filaments in an ATP-dependent fashion in some partitioning models (Leonard et al., 2005).

1.4.1 DNA binding proteins

DNA- or centromere- binding proteins (CBPs) have many different sizes and structures in plasmid- and chromosomal- partitioning systems (Baxter and Funnell 2014). These proteins are usually smaller than their ATPases. Most of the CBPs use either a Ribbon-Helix-Helix (RHH) (Moller-Jensen et al., 2007) or a Helix-Turn-Helix (HTH) (Fisher et al., 2017, Schumacher, 2008) motif fold for DNA binding.

1.4.1.1 CBPs – with RHH motif

The RHH fold plays an important role in protein dimerization. The beta strands (ribbons) of the dimer interact in an antiparallel manner and the helices interlace tightly to form a dimeric protein structure (Schumacher et al., 2007). RHH-containing proteins may also form a dimer of dimers when binding to DNA. In a dimeric structure, the antiparallel β-sheet from the $(RHH)_2$ motif interacts with the major groove of DNA (Schumacher, 2008).

The ParG protein from the TP228 plasmid (Figure 1.16a) is a partitioning protein. NMR studies show that ParG forms a dimer but, according to the proposed model, ParG binds to DNA as a tetramer (dimer of dimers) (Golovanov et al., 2003). The interaction between the C-terminal helices results in protein dimerization, while the extreme N-terminal of the protein consists of flexible arms (not shown in the figure). ParG contains an RHH motif where the β-sheet is involved in DNA interactions.

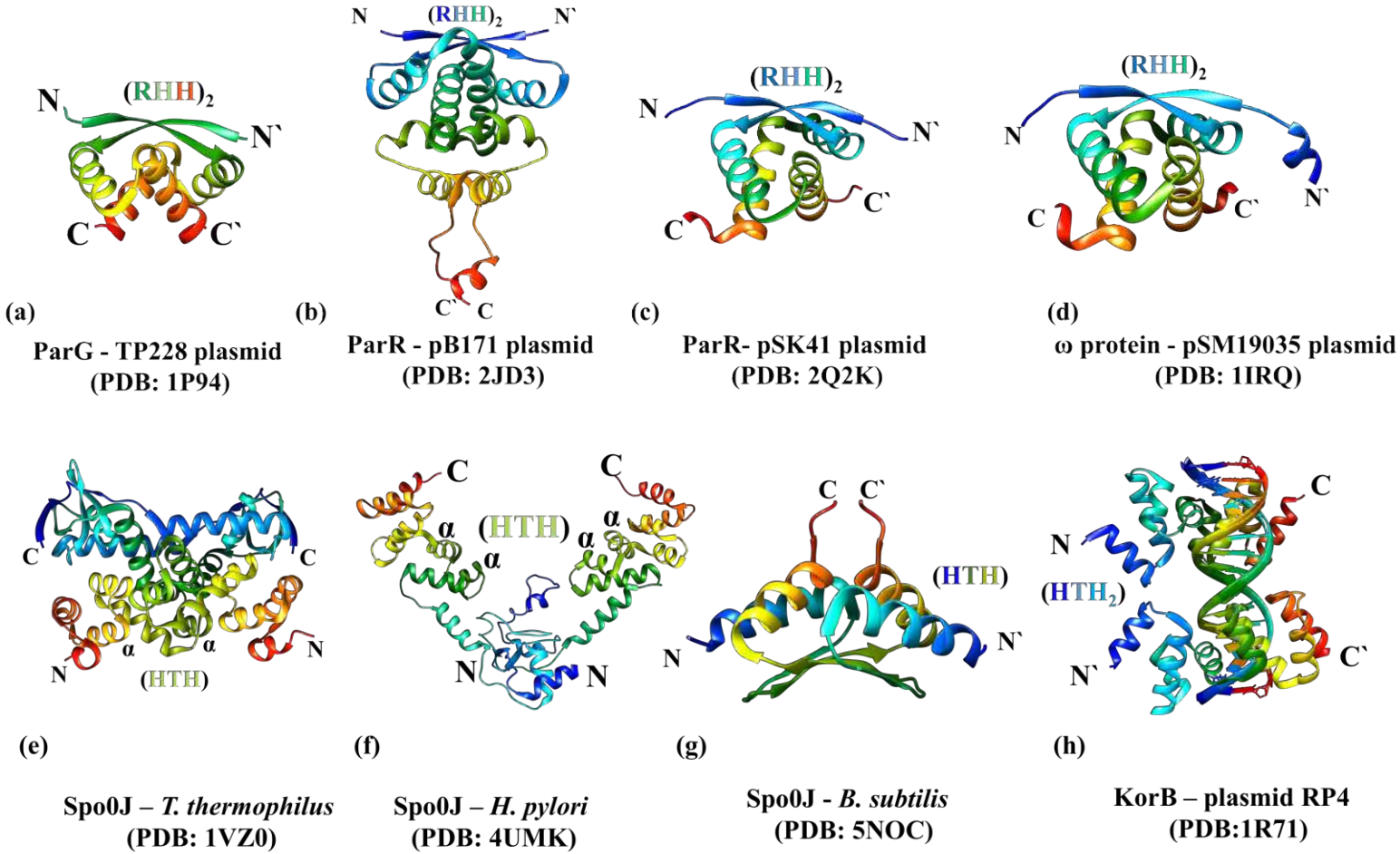


Figure 1.16 Structures of DNA/centromere binding proteins (CBP). (a-d) CBPs with RHH; (**a)** an RHH fold in ParG structure (PDB ID 1P94) from the TP228 plasmid (the N-terminal flexible arms have been removed), (**b)** RHH fold in ParR structure (PDB ID 2JD3) from the pB171 plasmid, **(c)** The RHH in the structure of ParR (PDB ID 2Q2K) from pSK41 plasmid, **(d)** RHH from the ω repressor protein structure (PDB ID 1IRQ) from the pSM19035 plasmid. **(e-f) CBPs with HTH; e)** Structure of the N-terminal domain of Spo0J from *T. thermophilus* (PDB ID 1VZ0) showing the HTH (shown in green), **(f)** The Spo0J structure (PDB ID 4UMK) from *H. pylori* showing HTH (green), **(g)** Structure of the dimeric C-terminal Domain of Spo0J (PDB ID 5NOC) from *B. subtilis* and the HTH fold is shown in blue and green colours, **(h)** The structure of the dimeric DNA binding domain of KorB from the RP4 plasmid in the presence of DNA (PDB ID 1R71) (All structures are shown in secondary structure succession rainbows colours) .

Studies of the ATP-dependent polymerization and depolymerization of the ParA-like (ParF) proteins show that the flexible N-terminal region of ParG interacts with ParF and stimulates its ATPase activity (Barilla et al., 2007). In many partitioning systems, ATP hydrolysis is stimulated by Arginine finger motif (Park et al., 2012) that is present in the flexible N-termini of ParB-like proteins. The N-terminal region of the ParG protein is also crucial to interact with DNA and to form a complex, its deletion leads to weak DNA binding (Carmelo et al., 2005).

An RHH motif -containing CBP from the pB171 plasmid, ParR, dimerizes to interact with DNA (Figure 1.16b). The N-terminal $(RHH)_2$ fold of the dimeric protein specifically interacts with the *parC* DNA, while its C-terminal is involved in interaction with the ParM filament (Moller-Jensen et al., 2007).

The ParR protein, from the pSK41 plasmid, binds to the centromeric site via RHH motif and interacts with the ParM (partner ATPase) by its C-terminal (Schumacher et al., 2007) (Figure 1.16c). The crystal structure of another CBP, Omega (ω) repressor protein from pSM19035, shows that the $(RHH)_2$ motif is involved in DNA binding (Figure 1.16d). The β-sheet from $(RHH)_2$ inserts into the major groove of DNA (Murayama et al., 2001). The N-terminal of the protein, like ParG, consists of a flexible region and an N-terminal deletion mutant of ω protein has shown higher affinity to DNA (Soberón et al., 2011).

Electron microscopy studies of the ParR protein from plasmid R1 and pB171 suggest that RHH containing proteins form ring-like complexes when interacting with DNA (Moller-Jensen et al., 2007).

1.4.1.2 CBPs – with HTH motif

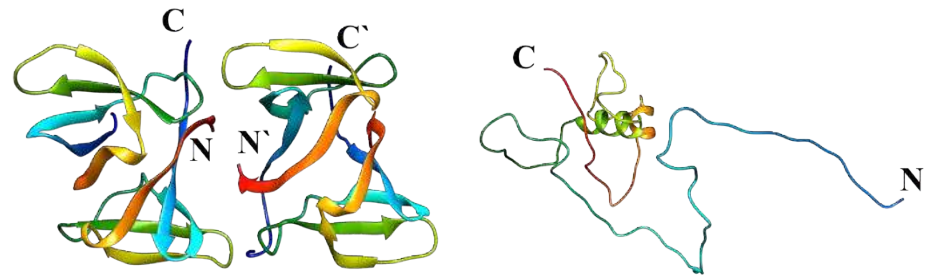
A CBP involved in chromosomal segregation, Spo0J, belongs to the Type I partition system. Structures for different domains of Spo0J protein have been determined using X-ray crystallography and NMR for Spo0J from *Thermus thermophilus* (Leonard et al., 2004), *Helicobacter pylori* (Chen et al., 2015) and *Bacillus subtilis* (Fisher et al., 2017). The N-terminal domain, consisting of 222 residues, from *Thermus thermophilus* Spo0J (T*t*Spo0J) protein was crystallized showing a dimeric, predominantly helical, structure containing a HTH motif (Figure 1.16e) (Leonard et al., 2004), although it is monomeric in solution. There is also a B motif which may be involved in the formation of oligomers when binds to DNA. The sedimentation studies of

the C-terminal domain show it forms dimers in solution (Leonard et al., 2004). The sequence of T*t*Spo0J is homologous to that of other ParB proteins.

A structure of a C-terminal deleted Spo0J protein from *Helicobacter pylori* (*Hp*Spo0J) has been reported in DNA-bound form (ct-*Hp*Spo0J) (Chen et al., 2015). In the presence of *parS* DNA, the monomeric protein was found to form a dimer and the dimeric form of the protein shows a higher affinity towards DNA. The N-terminal domain of the ct-*Hp*Spo0J has been proposed to be involved in spreading and in DNA condensation during chromosome segregation (Chen et al., 2015).

The structure of the C-terminal domain (CTD) of *B. subtilis* Spo0J (BtSpo0J) has been solved using NMR (Fisher et al., 2017) (Figure 1.16f). The CTD has a dimeric structure with 2 α-helices and 2 β- sheets. The NMR results show that the CTD interacts with DNA non-specifically and suggest that the CTD-DNA interactions play a major role in DNA condensation (Fisher et al., 2017).

Another extensively studied DNA -binding protein is KorB, which is present in the low copy number, broad host range plasmids RP4 and RK2 (Adamczyk and Jagura-Burdzy, 2003, Lukaszewicz et al., 2002). KorB is a 358-amino acid containing protein with a HTH motif present in the central domain (aa 137-252); a C-terminal dimerization domain (aa 293-353), along with two intrinsically disordered regions (Rajasekar et al., 2010). The crystal structures of the C-terminal and DNA binding domains of KorB from plasmid RP4 have been determined (Delbruck et al., 2002, Khare et al., 2004). The C-terminal of KorB is essential for protein dimerization (Lukaszewicz et al., 2002) and does not directly interact with DNA. This C-terminal domain facilitates nucleoprotein complex formation. The DNA -binding domain dimerizes only in the presence of DNA and contains a HTH motif (Figure 1.16h). The two domains (C- and N- terminal) are linked together by the flexible linker (Rajasekar et al., 2010). The structure of the N-terminal domain of KorB (30-150 amino acid) has been recently determined by Dr Hyde's group using NMR and has been found to contain two small stretches of α-helix while the rest of the domain is disordered (unpublished results) (Figure 1.17).



(a) C-terminal domain KorB – RP4 plasmid (PDB: 1IGQ)

(b) N-terminal domain KorB – RK2 plasmid (Dr Hyde group)

Figure 1.17 C- and N-terminal domains of KorB protein. (a) Structure of C-terminal dimerization domain of KorB (PDB ID 1IGQ) from plasmid RP4. **(b)** NMR structure of the N-terminal domain of KorB (aa 30-150). (All structures are shown in secondary structure succession rainbows colours).

1.5 ParA and ParA-like proteins/Motor Proteins

ParA and ParA-like proteins are ATPases with either Walker-type or Actin-like NTPases motifs. Proteins structures for some ParA have been determined using X-ray crystallography and cryo-EM. These structures provide insights about conformational changes due to ATP binding and hydrolysis, as well as due to binding with other partners including ParB and DNA (Leonard et al., 2005, Orlova et al., 2007, Zhang and Schumacher, 2017). ParA proteins from Type Ia partitioning system contain an N-terminal (~100 amino acids) extension that is involved in DNA binding. In contrast, the ParA proteins from Type Ib partitioning system and most of the chromosomal ParAs do not contain an N-terminal extension (Hester and Lutkenhaus, 2007).

The best-studied ATPases include ParM and Soj protein. ParM mimics actin-like proteins structurally and biochemically and forms filaments in presence of either ATP or GTP (Popp et al., 2008, Garner et al., 2004) (Figure 1.18a-c). These polymers, during the partitioning process, move across the cell, pushing the plasmids to the cell poles.

In vivo studies of the R1 plasmid show that both ends of the ParM filament attach to plasmids, pushing them apart (Moller-Jensen et al., 2003, Salje et al., 2010, Gayathri et al., 2012). ParM filaments also form bundles consisting of double helical filaments. The structure of filaments of a ParM protein from *E. coli* (Figure 1.18a) has been studied by Cryo-EM. The structure of ParM from R1 plasmid has been determined, using X-ray crystallography, in the presence and absence of ADP (Figure 1.18b). Despite biochemical similarities to Actin, ParM filaments

assume a different subunit interface from that in F-actin (Campbell and Mullins, 2007, Orlova et al., 2007). In the presence of ADP, ParM shows a closed structure (van den Ent et al., 2002). The Cryo-EM modelling of the ParM protein from plasmid R1 shows that the ParM-ADP closed structure is due to disruption of salt bridges within the protofilaments (Figure 1.18c) (Bharat et al., 2015).

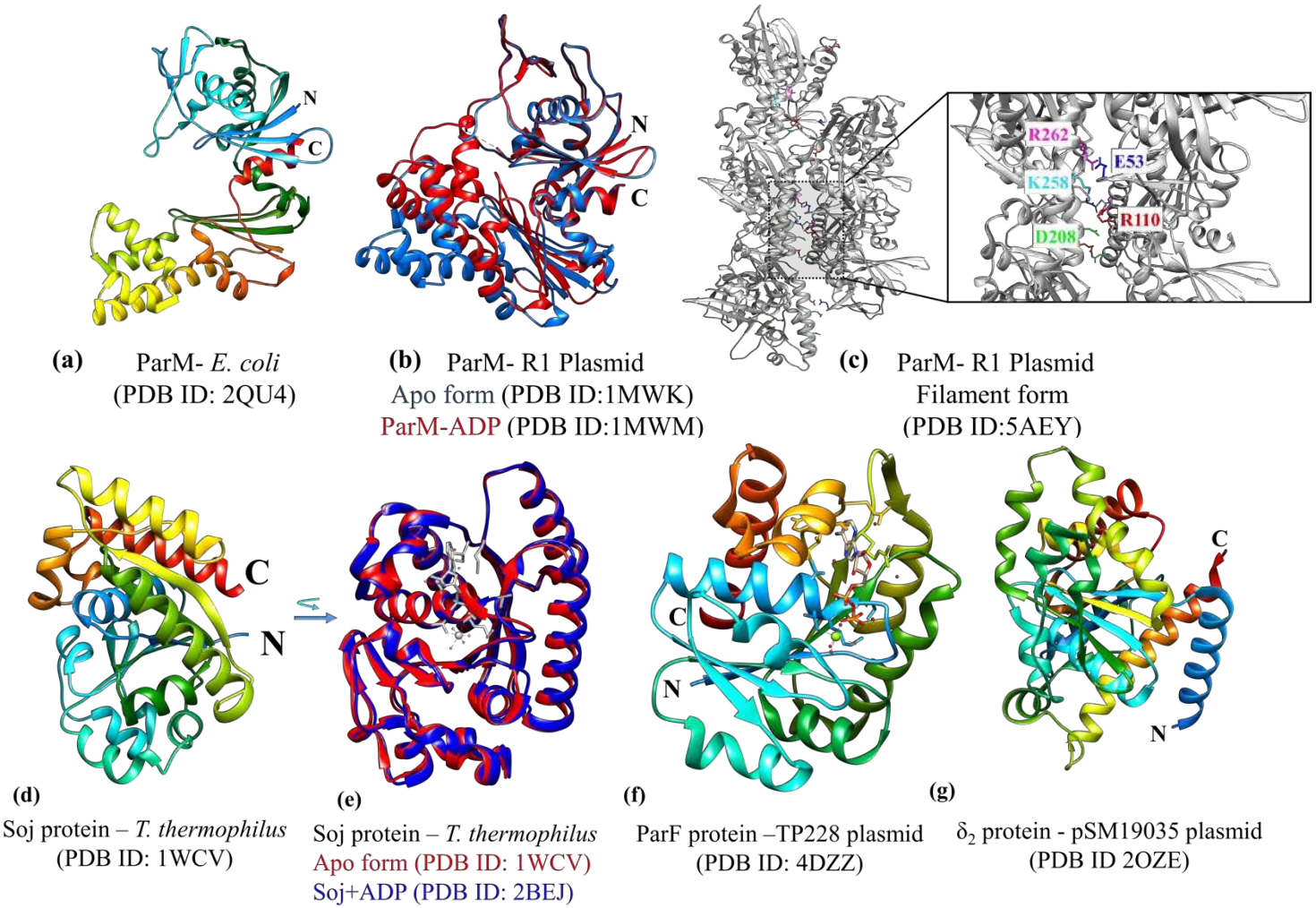


Figure 1.18 Structures of the ParA proteins from various partitioning systems. (a) Structure of ParM (shown as a monomer) from *E. coli* (PDB ID 2QU4). **(b)** Structural comparison of Apo (PDB ID 1MWK, blue) and ADP bound (PDB ID 1MWM, red) form of ParM. In the presence of ADP, ParM shows a closed structure. **(c)** The interface residues hold ParM filaments together by salt bridges (shown between R262, K258, and D208 of one subunit and E53 and R100of the other). ADP (white) disrupts the interactions to give a closed structure (PDB ID 5AEY). **(d)** Soj protein Apo structure (monomer) from *T. thermophilus* (PDB ID 1WCV). **(e)** The P-Loop in Soj protein undergoes conformational changes upon binding to ADP (PDB ID 2BEJ). **(f)** Structure of ParF protein (shown as a monomer) from the TP228 plasmid (PDB ID 4DZZ) **(g)** Structure of the δ_2 protein (monomer) from the pSM19035 plasmid (PDB ID 2OZE). (All structures are shown in secondary structure succession rainbows colours)

The structure of the ParM protein from the R1 plasmid has also been determined in the presence of AMP-PNP (Galkin et al., 2009) and GMP-PNP (Popp et al., 2008), by cryoEM. similar ParM filament structures were observed in both cases. The structure of a ParM-ADP complex from a plasmid of *Bacillus thuringiensis* has also been determined (PDB ID 4XHP).

A ParA protein, Soj, is involved in chromosomal segregation in *T. thermophilus* and interacts with Spo0J bound to *parS* (Lin and Grossman, 1998). Soj interacts with DNA non-specifically (with the help of surface arginine residues), forms dimers and then forms filaments in ATP -dependent manner (Leonard et al., 2005, Hester and Lutkenhaus, 2007). Soj proteins have been found to oscillate between cell poles while associating with the nucleoid (Leonard et al., 2005). The crystal structure of Soj (Figure 1.18d), in the presence and absence of ADP/ATP, shows that the P-loop undergoes some conformational changes upon Soj binding to these nucleotides (Figure 1.18e). The Soj structure shows similarities to MinD, a membrane-associated ATPase of MinCDE cell division system in *E. coli* (Rowlett and Margolin, 2013). The ATPase activity of Soj is augmented by its interaction with the N-terminal region of its DNA-binding partner Spo0J but, surprisingly, Soj binding to DNA alone does not increase the ATP hydrolysis rate (Leonard et al., 2005).

ParF is another filament-forming ATPase from the ParFGH partitioning system of the TP228 plasmid (Hayes, 2000). In the presence of ADP, ParF remains in monomeric form, but it forms dimers and extensive filaments when bound to ATP (Schumacher, 2012). ParF oscillates between the cell poles while transporting the nucleoprotein complex (ParG-*parC*. The ParF-ADP crystal structure showed a monomer (Figure 1.18f). Structural and biochemical studies show that ParF filaments are constructed as units of a dimer of dimers, and that the loss of polymerization causes plasmid instability (Schumacher, 2012).

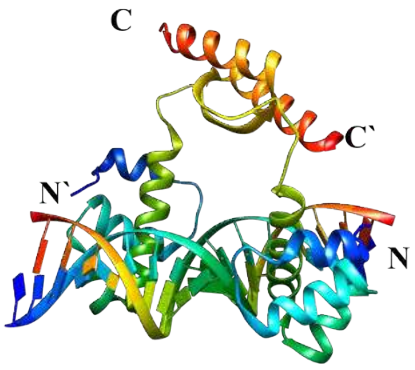
Recently, the structure of a ParA protein from the pNOB8 plasmid, present in *Sulfolobus* NOB8H2, (Schleper et al., 1995) has been determined (Zhang and Schumacher, 2017) in the presence of β, γ-imidoadenosine 5′-triphosphate (AMP-PNP) and DNA. The protein was found to form a dimer in the presence of NTPs. The pNOB8 ParA-AMP-PNP dimer interacts with ParB using its dimer interface. The structure of the δ_2 ParA protein from the pSM19035 plasmid also has been determined (Figure 1.18g). δ_2 forms spiral-like structures, while oscillating between cell poles using ATP as the driving force (Zhang and Schumacher, 2017). The negligible ATPase activity of the δ_2 was stimulated in the presence of its partner protein ω_2 and *parS* DNA.

1.6 Partitioning in RK2 plasmid

The RK2 plasmid (60 kb) is one of the best studied, low copy number (5-7 copies per *E. coli* chromosome), broad host range, *Inc*P α plasmids (Figurski and Helinski, 1979). The RK2 plasmid partitioning system consists of the ParA and ParB homologs, IncC and KorB, encoded by the central control region (CCR) of the plasmid (Figure 1.20) (Meyer and Hinds, 1982, Gerdes et al., 2000). The RK2 *par* operon is regulated by KorB and a repressor protein, KorA (101 a.a), expressed from an alternative ORF within the *incC* gene (Rosche et al., 2000).

1.6.1 KorA protein

KorA is a DNA -binding protein expressed from the central control region (CCR) of the RK2 plasmid. It is involved in transcription repression at seven operator (O_A) sites. The O_A site is a 12 bp palindromic DNA sequence, with the consensus 5´GTTTAGCTAAAC3´. The KorA protein also shows cooperative binding to KorB protein at five of these seven O_A sites (Bechhofer and Figurski, 1983; Kornacki *et al*., 1987; Theophilus & Thomas, 1987). The crystal structure of the KorA dimer in the presence of O_A DNA has been determined (Rajasekar et al., 2016) and the region between amino acids 37-56 shows homology to helix-turn-helix sequences and is involved in binding to DNA (Thomas and Smith, 1986, Kostelidou et al., 1998).



KorA – RK2 plasmid
(PDB: 5CLV)

Figure 1.19 Structure of the KorA protein bound to O_A DNA. The KorA protein forms dimer in the solution (Jagura-Burdzy and Thomas, 1995) and its dimeric form interact with O_A DNA (shown in rainbow colours). The structure of KorA shows the dimeric KorA is involved in interactions with O_A DNA (PDB ID 5CLV).

1.6.2 KorB and IncC Proteins

KorB is a DNA/centromere binding protein. As described in section 1.4.1, it employs a HTH motif present in the DNA-binding domain (DBD) to interact with DNA. The crystal structures of the CTD and DBD of KorB have been elucidated (Figure 1.16h and Figure 1.17). The RK2 plasmid contains 12 DNA binding sites (O_{B1-12}) that interact specifically with KorB protein (Williams et al., 1993) (Figure 1.20), although deletion of the OB_3 site alone contributes to plasmid instability (Williams et al., 1998). The presence of multiple KorB DNA binding sites suggests that KorB might be involved in different functions apart from partitioning *i.e.* as repressor.

The IncC proteins of RK2 belong to the ParA family of Walker-type ATPase and potentiate KorB binding to the O_B sites. The RK2 plasmid partitioning system is unique as it involves two ParA proteins, IncC1 and IncC2. The *incC* gene in *par* operon encodes two polypeptides of different lengths starting from two different start codons present in the same gene; the full-length IncC1 protein (364 a.a, 38.1 kDa) and a shorter form, IncC2 protein (259 a.a, 27.5 kDa) (Figure 1.20 inset). The amino acid sequence of IncC2 is identical to the last 259 aa of IncC1, and shows significant relatedness to the ParA and SopA partitioning proteins of P1 and F plasmids, respectively. The two IncC proteins differ by the presence of the N-terminal 105 amino acids in IncC1 that are absent in IncC2. The IncC proteins have been found to bind DNA non-specifically and their ATPase activity is augmented by KorB and DNA (Batt et al., 2009). IncC proteins directly interact with the KorB and stabilizes KorB–DNA complex (Rosche et al., 2000). IncC1 is similar in size to the longer ParA proteins present in most plasmids, while the size of IncC2 mimics the chromosomal ParA proteins. The function of N-terminal region (105 a.a) in IncC1 was unknown previously and is described in this thesis, while the IncC1 and IncC2 proteins have been characterized in this study using various biophysical methods, chemical crosslinking, Mass spectrometry and homology modelling.

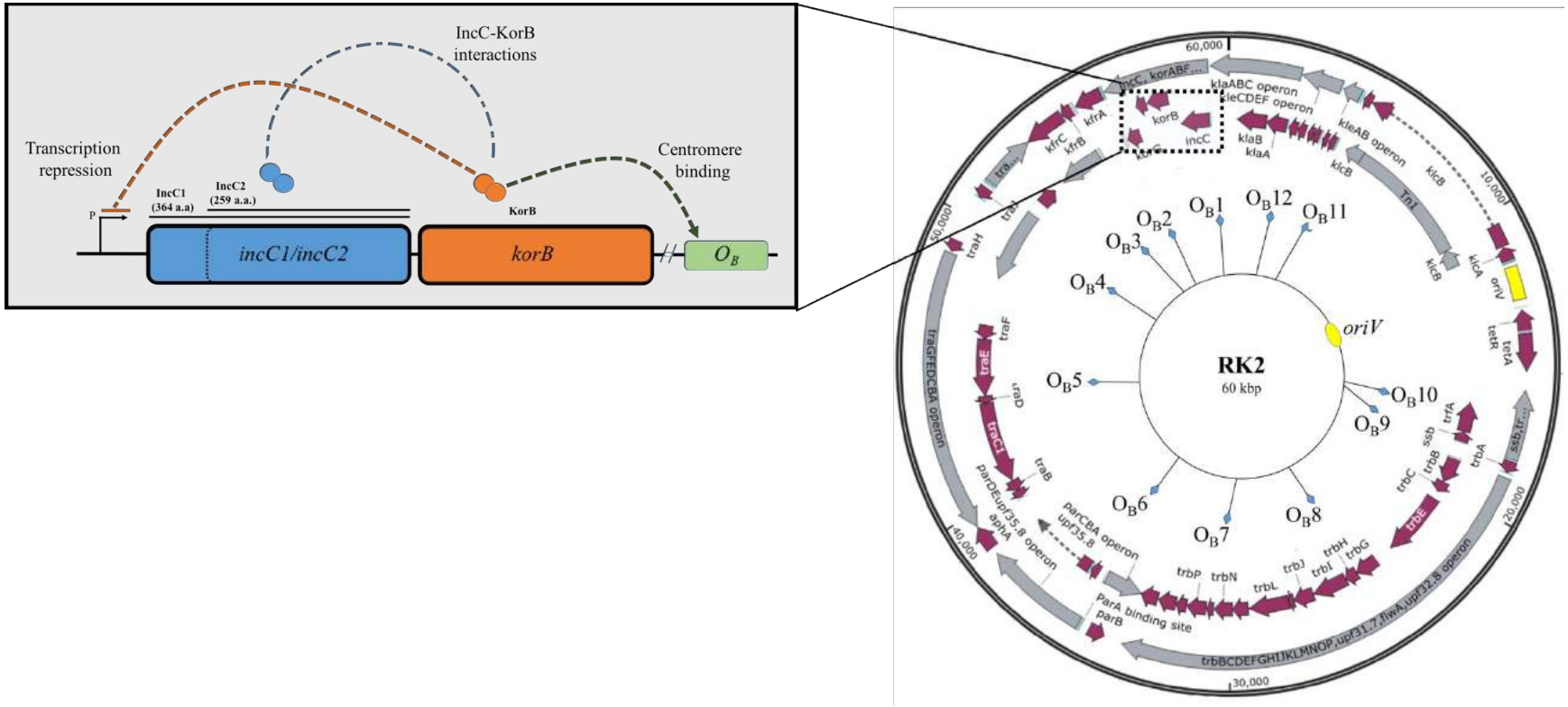


Figure 1.20 RK2 plasmid map. RK2 is a ~60 kb plasmid. The map shows the presence of different operons (Grey) and the proteins expressed from the operons (Magenta). The vegetative origin of replication, *OriV* is shown in yellow. 12 O_B sites (blue diamonds) are distributed throughout the RK2 plasmid (Williams et al., 1993). The inset shows that the RK2 plasmid encodes two ParA proteins from the same gene but with different start codons, the IncC proteins; IncC1 and IncC2 (Blue), and one DNA binding protein KorB (Orange). IncC1 consists of 364 amino acids, while IncC2 lacks the N-terminal 105 amino acids of IncC1 and consists of 259 amino acids. The IncC and KorB proteins interact with each other (blue dotted line) (Thomas et al., 1980). KorB interacts with all 12 O_B sites, one of which, OB3, (green) is the centromere-like site, (distribution is shown in RK2 map, in inset a green box depicts KorB-O_B binding) specifically and acts as a repressor and autoregulates its own operon.

1.7 Aims and objectives of the research

The purpose of this research is to study the structure and function of RK2 plasmid's two ParA proteins, IncC1 and IncC2. The presence of the IncC NTD (the N-terminal domain of IncC1 consisting of 105 a.a which is missing in IncC2) leads to questions about its structure and function, as one of these ParA proteins is sufficient for plasmid partitioning. It is also of interest to characterize IncC1 and IncC2 using different biophysical and biochemical methods to study how two proteins are affected by the presence of IncC NTD.

Aim 1

One major aim of the research is to determine the backbone structure of the IncC NTD using NMR and other biophysical tools. The specific objectives are:

i) Purification of IncC NTD

ii) Its characterization using CD, AUC, SAXS and EMSAs

iii) Use of carbon-detected NMR experiments for IncC NTD backbone assignment.

iv) Determination of IncC NTD – DNA interaction by NMR.

The IncC NTD is characterized in Chapter 3 and elucidation of its secondary structure by NMR is discussed in Chapter 4. The DNA binding of IncC NTD is reported in Chapter 3 and Chapter 4.

Aim 2

Another major aim of this research is to compare the properties of IncC1 and IncC2 proteins. Different biochemical and biophysical methods are required to characterize the proteins. The objectives include:

i) The optimization of the overexpression and purification of the two proteins.

ii) Characterization of the IncC1 and IncC2 proteins and comparison of their DNA-binding and nucleotide-binding properties using CD, AUC, SAXS, SEC-MALLS and EMSAs.

The purification and characterization of these proteins are described in Chapter 5.

Aim 3

The final aims of this research are to analyze the 3D structure of IncC and the interactions of IncC with KorB protein *in vivo* and *in vitro*. The objectives include:

i) Cloning of *incC*, *korB* and *korA* genes into BACTH vectors to express fusion proteins with T18 and T25 cyaA fragments in different orientations.

ii) Determination of protein-protein interactions, indirectly using β-galactosidase assay.

iii) Chemical crosslinking to evaluate the presence of homo-dimers and higher order multimers species for IncC, KorB and KorA proteins.

iv) Analysis of the structure of the IncC2 protein. This was done using homology modelling and the model was verified by the use of chemical cross-linking. The cross-linked peptides were, identified using Mass spectrometry and mapped on the protein model.

The BACTH assay and chemical crosslinking experiments for IncC, KorB and KorA protein are described in Chapter 6.

CHAPTER 2

Materials and Methods

Chapter 2 Materials and Methods

2.1. Materials

All chemicals (unless mentioned) were purchased from SigmaAldrich (Dorset, UK), Fisher Scientific (Leicestershire, UK) and Bioline (London, UK).

2.1.1 Bacterial strains

Different *Escherichia coli* strains were used (Table 2.1). The *E. coli* DH5α, JCB387 and XLI Blue strains are used to obtain plasmid DNA, while the *E. coli* BL21(λDE3), BL21 (DE3) pLysS and T7 Express strains were used to overexpress different proteins. The *E. coli* strain BTH101 was used in the bacterial two hybrid experiment.

Table 2.1 Bacterial strains used in this study and their genotypes

	Bacteria (*E. coli*)	Genotype	Antibiotic Resistance	Reference
1	BL21(λDE3)	F⁻ *ompT gal dcm lon hsdSB* (r_B^- m_B^-) λ(DE3)	None	(Studier and Moffatt, 1986)
2	BL21 (DE3) pLysS	F⁻ *ompT gal dcm lon hsdSB*(r_B^- m_B^-) λ(DE3) pLysS	Chloramphenicol	(Moffatt and Studier, 1987)
3	BL21 Star (DE3)	F⁻ *ompT hsdSB* (r_B^-, m_B^-) *gal dcm rne*131 λ (DE3)	None	(Studier and Moffatt, 1986)
4	DH5α	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15* Δ(*lacZYA-argF*) *U169*, *hsdR17*, λ⁻	None	(Grant et al., 1990, Sambrook and Russell, 2001)
5	T7 Express	*fhuA2 lacZ*::*T7 gene1* [*lon*] *ompT gal sulA11 R*(*mcr-73*::*miniTn10--Tet*^{S})2 [*dcm*] R(*zgb-210*::*Tn10--Tet*^{S}) *endA1* Δ(*mcrC-mrr*)*114*::*IS10*	None	Invitrogen
6	JCB387	Δ*nir* Δ*lac*	None	(Typas and Hengge, 2006)
7	XLI-Blue	*endA1 gyrA96*(*nal*^{R}) *thi-1 recA1 relA1 lac glnV44 F' hsdR17*(r_K^- m_K^+)	Tetracycline	Stratagene
8	BTH101	F⁻, *cya*$^{-99}$, *araD139*, *galE15*, *galK16*, *rpsL1* (*Str* ^{r}), *hsdR2*, *mcrA1*, *mcrB1*.	Streptomycin	Euromedex BACTH system kit

2.1.2 Bacterial growth media preparation

Following bacterial growth media have been used in this thesis. Media were prepared as described elsewhere (Sambrook and Russell, 2001).

a) LB media

All bacterial stains were grown at 37°C (unless mentioned) using the autoclaved Luria Bertani (LB) broth or the LB agar along with appropriate antibiotic.

The LB medium (1 L) consists of 10 g Tryptone, 5 g yeast extract, and 10 g sodium chloride. pH of medium set to ~7.4 using 1 M sodium hydroxide (NaOH). For the LB agar, 15 g of agar added to 1 L LB and autoclaved.

b) MacConkey agar

The MacConkey agar is a selective medium for gram negative and bacilli bacteria. The commercial MacConkey agar medium (Oxoid) (20 g Peptone, 10 g lactose, 5 g Bile salts, 5 g sodium chloride, 75 mg Neutral red, 12 g agar, pH ~7.4) autoclaved and used to make plates with required antibiotics.

2.1.3 Antibiotics (Stock solutions)

Different antibiotics are used as a selection marker. Stock solution concentrations and working concentrations of antibiotics have been given in Table 2.2.

Table 2.2 Antibiotic concentrations for stock and working solutions

	Antibiotics	Stock Concentrations	Working Concentrations
1	Ampicillin (Amp)	100 mg/mL	100 µg/mL
2	Kanamycin Sulphate (Kan)	50 mg/mL	50 µg/mL
3	Chloramphenicol (Cam)	25 mg/mL	25 µg/mL

2.1.4 Plasmids

Plasmids used in this thesis have been given in Table 2.3 (plasmid maps have been shown in Appendix 1). The *incC1*, *incC2* and *incC-NTD* genes were previously cloned in a pET21a vector and plasmid DNA was obtained for protein overexpression and purification.

Table 2.3 Plasmid used in this thesis and their properties

	Plasmid (respective protein)	Selectable Marker	Size (kb)	Replicon	Properties	Reference
1	pGBT342 (IncC1)	Kan	6.4	pMB1	pET28a with T7 tag removed, and *incC1* inserted in *EcoRI-SalI* restriction site. KorA Start codon was ATG has been changed to ACG without altering *incC1*	(Jagura-Burdzy et al., 1999)
2	pSMB315 (IncC NTD)	Kan	5.7	pMB1	pET28a with T7 tag removed, and *incC1-ntd* (N-terminal domain of 315 bp) inserted in *EcoRI-SalI* restriction site.	(Batt et al., 2009)
3	pGBT343 (IncC2)	Kan	6.2	pMB1	pET28a with T7 tag removed, and *incC2* inserted in *EcoRI-SalI* restriction site.	(Jagura-Burdzy et al., 1999)
4	pSMB330 (KorA)	Kan	5.7	pMB1	pET28a with T7 tag removed, and *korA* inserted in *EcoRI-SalI* restriction site.	(Batt, 2008)
5	pSMB322 (KorB)	Kan	6.5	pMB1	pET28a with T7 tag removed, and *korb* inserted in *EcoRI-SalI* restriction site.	(Batt, 2008)
6	pKK113	Amp	2.7	pMB1	pUC19 with O_B site cloned in it	(Lukaszewicz et al., 2002)
7	RK2	Kan, Tet	60	IncP α	Wild type broad-host range plasmid, can transfer between all Gram-negative bacteria	(Ingram et al., 1973)

2.1.5 Restriction enzymes and PCR enzymes

Most enzymes, including restriction enzymes (EcoRI, SalI, KpnI, PstI, PvuII), T4 DNA ligase, Calf Intestinal Alkaline Phosphatase (CIAP) and Antarctic Phosphatases purchased from NEB (New England Biolabs) (Unless mentioned) and stored at -20°C. Enzymes were used according to the manufacturer's instructions and reaction requirement. The Velocity DNA polymerase (Bioline) was used for high fidelity PCR. MyRedTaq Mix (Invitrogen) was used for the colony PCR.

2.1.6 Kits used

Isolate-II Mini Plasmid Kit (Bioline), Pure Yield Plasmid Maxiprep kit (Promega), QIAquick Gel Extraction Kit (Qiagen), Slice pH kit (Hampton Research).

2.1.7 Computational tools

Different servers and software were used in the thesis to analyze and visualize data. Software include Spectra Manager™ Suite (JASCO), REDATE, SEDNTREP, SEDFIT, GUSSI, ATAS, PRIMUS, CLC Genomics, SnapGene, Genesnape, ImageJ, Affinity Analysis, GraphPad

Ver. 4 (Prism), CCPN-Analysis NMR, YASARA, Modeller, Swissprot PDB Viewer, ClustalX, JalView, Amicon Software for AKTA (Amersham, GE), Chromas 2.4.3. NanoDrop® software by NanoDrop (Wilmington, USA). The servers include BLAST (Basic Local Alignment Search Tool), ExPASy server (www.expasy.ch), and DichroWeb server.

2.2 Competent cells preparation

To transform plasmid DNA, *E. coli* first should be made competent. Different *E. coli* strains were grown to mid-log and made competent by chemical treatment ($RbCl_2$-$CaCl_2$ and $CaCl_2$ methods).

2.2.1 Preparation of competent cells by $RbCl_2$- $CaCl_2$ Method

Competent cells of the *E. coli* strains, DH5α, BL21 (DE3), BL21 (DE3) pLysS, T7 Express, were prepared using chemical treatment of $RbCl_2$ and $CaCl_2$.

Bacterial strains from glycerol stocks (stored at -80°C) were streaked onto LB-agar plates with respective antibiotic. Plates were incubated overnight at 37°C and a bacterial colony was used to inoculate 5 mL LB media followed by an overnight incubation at 37°C. 1 mL of overnight culture was inoculated to 100 mL LB. Bacteria were grown until OD_{600} ~ 0.4 - 0.6 in about 1.5 to 2 hr. Bacterial cells were centrifuged, at 4,000x *g* for 10 min at 4°C in a Jouan bench top centrifuge. Pellet was resuspended in 25 mL of the ice-cold TF buffer 1 (30 mM KOAc, 100 mM $CaCl_2$, 10 mM $RbCl_2$, 50 mM $MnCl_2$, 15 % Glycerol (v/v), pH 5.8). Resuspended cells were incubated on ice for 30 min followed by centrifugation (4,000x *g*) for 10 min at 4°C. After that, pellet was resuspended in 4 mL of the TF buffer 2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15 % Glycerol (v/v), pH 6.5) and incubated on ice for 1 hr. 50 µL of cells were aliquoted in autoclaved eppendorf tubes and stored at -80°C until further use.

2.2.2 Preparation by $CaCl_2$ Method

Competent cells of the *E. coli* strains, JCB387, XL1-Blue, and BTH101 were prepared using $CaCl_2$ chemical treatment. One mL of an overnight culture of a *E. coli* strain was used to inoculate 100 mL of fresh LB medium and incubated at 37°C in shaking incubator to obtain bacterial cell in mid-log phase (OD_{600} ~ 0.4-0.6). Then, cells were centrifuged at 4,000x *g* and 4°C for 10 min in a Jouan bench top centrifuge. Cell pellet was resuspended into 25 mL of ice-cold 100 mM $CaCl_2$ and kept on ice for 30 min. Cells were again centrifuged using same

conditions. Pelleted cells were resuspended 2.5 mL of 100 mM $CaCl_2$ (containing 15 % v/v Glycerol) and incubated on ice for 60 min. Competent cells were stored at -80°C in 1.5 mL eppendorf as aliquots of 50 µL till further use.

2.3 Transformation of bacterial cells

Competent cells can be transformed with foreign DNA. Competent bacterial cells (50-100 µL) were transformed with 20-50 ng foreign DNA or 10 µL ligation mixture in 1.5 mL microfuge tube and mixed gently. Cells were placed on ice for 30 min followed by heat shock for 90 s in a water bath at 42°C. After heat shock treatment, 1 mL of LB (containing no antibiotic) was added to each tube and mixed gently. Microfuge tubes were incubated for 1 hr at 37°C. After incubation, culture (50-100 µL) was plated onto agar plates and incubated for 18 hr at 37°C. Control plates (with respective antibiotic) were also prepared in each transformation to check the competent cell growth with or without selective antibiotic.

2.4 Plasmid DNA isolation and purification

2.4.1 Plasmid DNA prep by alkaline lysis/chloroform/phenol extraction

This method is adapted from Birnboim and Doly (1979). The alkaline lysis step involves selective denaturation of chromosomal DNA to an insoluble clot, while circular plasmid DNA, rapidly re-anneals after neutralization and remains double-stranded.

The transformed bacteria were grown in 3-5 mL LB medium (containing required antibiotic) for 18 hr at 37°C. Overnight grown culture was centrifuged at 4,000x *g* for 10 min at 4°C in a Jouan bench top centrifuge. Cell pellet was resuspended in 200 µL of lysis buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0) containing freshly prepared lysozyme solution (50 µg/mL). The cell lysate was transferred to 1.5 mL microfuge tubes and placed on ice for 30 min. Then 200 µL of a freshly prepared alkaline solution (0.4 M NaOH, 2 % w/v SDS) was added to each tube, mixed and tubes were placed on ice for 10 min. After incubation 350 µL of neutralizing solution was added that contained 3 M sodium acetate pH 5.0 (3 M sodium acetate, 11.5 % v/v acetic acid, pH 5.0) and tubes were mixed gently by inversion. All centrifugation steps were performed at 14,000x *g* unless mentioned. The tubes were centrifuged for 10 min at room temperature using a bench top micro centrifuge. Supernatant was transferred to a new microfuge tube while discarding pellet. After adding isopropanol (750 µL) to

supernatant, tubes were incubated at -20°C for an hour. After 1 hr, pellet was obtained by centrifugation for 15 min. The pellet was resuspended in 100 µL of TEN buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM EDTA). To get rid of DNA, DNase free RNase A (final conc. 100 µg/ml) was added and tubes were incubated for 1 hr at 37°C. After incubation, the supernatant was treated with 100 µL of Phenol:Chloroform (1:1) solution to remove proteins and lipids. Solution was mixed by vortexing for 10 s and centrifuged for 2 min. Aqueous layer was separated to fresh tube and an equal volume of chloroform was added to remove any phenol contamination. Centrifugation step was repeated to get aqueous layer followed by addition of 2.5 volume of ice cold 95% ethanol to precipitate plasmid DNA. Then, the tubes were centrifuged for 10 min and top supernatant layer was removed to obtain pellet. The pellet was then washed with 500 µL of ice cold 70 % ethanol. Centrifugation was repeated and the plasmid DNA pellet was dried in a vacuum evaporator after removing supernatant. 50 µL of TE buffer (10 mM Tris.HCl, 1 mM EDTA, pH 8) was added to the pellet and stored at -20°C.

2.4.2 Plasmid DNA isolation by Isolate II, Bioline mini prep plasmid isolation kit

Alternatively, the Isolate-II Miniprep DNA purification kit (Bioline) was used to purify high-quality plasmid DNA sample for DNA sequencing and cloning experiments. All solutions/buffers were used according to instruction by manufacturer.

2.4.3 Plasmid Maxi prep by PureYield™ Plasmid Maxiprep System

For purification of low copy number plasmids, DNA was purified using the Promega PureYield™ Plasmid Maxiprep System. This process is time efficient and completes in 1 hr as it uses vacuum pump and manifold (*e.g.* Vac-Man® Laboratory Vacuum Manifold). An overnight LB culture (250 mL), inoculated with desired bacterial strain, was used for maxiprep according to manufacturer's instructions.

2.5 Plasmid DNA quantitation

2.5.1 DNA quantitation by Hoechst 33258 Dye

The Hoefer DyNA Quant 200 Fluorometer and a fluorescent dye, Hoechst 33258 (Bisbenzimide) was used to quantify DNA. This dye shows excitation in near UV (350 nm) and emission in blue region (450 nm). Sensitivity of the Hoechst 33258 depends on its concentration (approx. 10 ng/mL). Stock solution of Hoechst 33258 dye (1 mg/mL) was diluted to 100 ng/mL using TEN buffer (10 mM Tris base, 1 mM EDTA, 200 mM NaCl) (10 µL dye + 100 mL TEN buffer). The calf thymus DNA (ctDNA) (stock solution, 1 mg/mL in TE buffer) was used to obtain a standard curve. Fluorometer was turned on and blank was set with 2 mL of diluted dye solution. Then, ctDNA (100 µg/mL) was titrated (1 µL at one time) to obtain standard values. Unknown DNA sample was titrated against 2 mL of dye and sample DNA concentration was obtained from the standard curve.

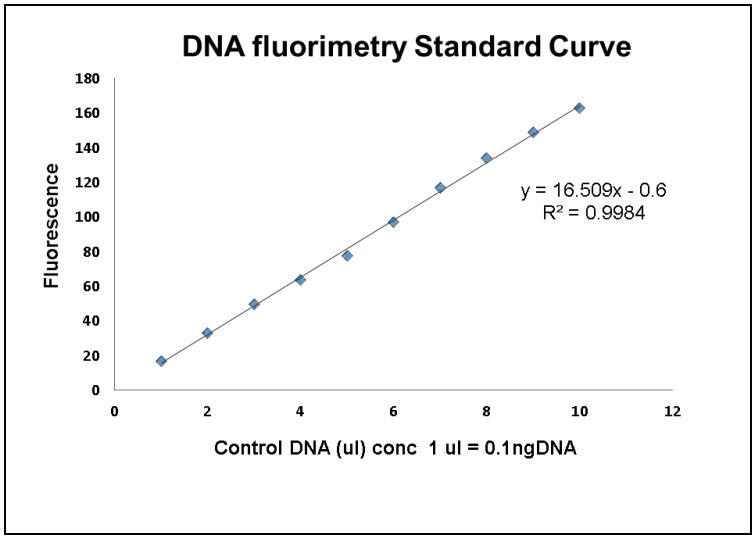


Figure 2.1 Standard curve for DNA fluorimetry using ctDNA (calf thymus DNA)

2.5.2 DNA quantitation by NanoDrop

The NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, USA) was used to measure light absorption of 2 µL aliquots of DNA at 260 nm (A_{260}). Purity of DNA was calculated using 260/280, 230/260 ratios. An average extinction coefficient for double-stranded DNA (1.0 A_{260} = 50 µg/mL), single stranded DNA (1.0 A_{260} = 33 µg/mL), or RNA (1.0 A_{260} = 40 µg/mL) is used to quantify DNA from A_{260}.

2.6 DNA Restriction digestion and Agarose gel electrophoresis

Plasmid DNA digestion was performed using high fidelity (HF) restriction enzymes and compatible buffers (usually CutSmart buffer from NEB). Heat activation was performed where applicable (70-85°C for 10 min). Digested DNA was analysed using agarose gels (usually 1%).

Agarose gel electrophoresis (AGE) was used to analyze plasmid DNA and digested fragments. Typically, 0.6 - 1% w/v agarose gels were run at 10 V/cm, using TAE buffer (40 mM Tris acetate, 1 mM EDTA, pH 8 at 25°C) or TBE buffer (89 mM Tris borate, 2 mM EDTA, pH 8 at 25°C) as running buffer, while DNA loading buffer (6X) contained 30 % (v/v) glycerol, 0.25 % (w/v) Bromophenol Blue, and 0.25 % (w/v) Xylene Cyanol. Agarose gel was stained with ethidium bromide (0.5 µg/mL) and de-stained with water. DNA ladder marker (1 µL, 1 kb or 100 bp ladder, NEB) was also loaded each time. DNA was visualized using a Gel Doc transilluminator and image was edited/saved using the Genesnape Software.

2.7 Polymerase chain reaction (PCR)

To amplify DNA, PCR was performed using various primers (designed and analyze for dimer formation and non-specific binding using SnapeGene software). PCR involved three steps, denaturation, (90-96°C), annealing (50-60°C), and extension (68-72°C). PCR cycles were selected according to requirement of an experiment.

2.8 Colony PCR

Colony PCR was used to confirm the transformants/clones. A transformant colony was picked and resuspended in 22 µL nuclease free water followed by addition of 1 µL of each forward/reverse primer (10 pmol), and 1 µL of DMSO (where mentioned). MyRedTaq Mix (Invitrogen) (2X, 25 µL), containing DNA polymerase and PCR buffer, was added just before start of the PCR reaction. PCR was performed using various primers, conditions, temperatures, and cycles (mentioned in results).

2.9 Purification of PCR products

The QIAquick PCR Purification Kit was used to purify PCR products (>100 bp-10 kb). PB Buffer was added to of the PCR reaction in 5:1 ratio. An alkaline pH was observed by yellow colored mixture or 3 M sodium acetate, pH 5.0 was added to turn mixture orange or violet. PCR

mixture was passed through a QIAquick column and flow through was discarded. After washing column with 750 µL of Buffer PE and discarding flow though, DNA was eluted using 30-50 µL Buffer EB (10 mM Tris-HCl, pH 8.5) and stored at -20°C.

2.10 Purification of DNA from Agarose gels

PCR amplified digested inserts and digested vectors were gel purified for efficient ligation and cloning. The QIAquick Gel Extraction Kit (Qiagen) was used to purify DNA fragments from gels (up to 400 mg slices). Procedure was carried out according to manufacturer's instructions.

2.11 DNA ligation

Double digested, and purified DNA fragments were used for DNA ligation in appropriate vectors. Plasmid vectors (1-10 µg) were digested with same restriction enzymes as used for inserts and gel purified. After digestions, inserts and linear vectors were analyzed by using agarose gel electrophoresis. Ligation reaction contained in different ratios (1:1, 1:3, 1:5) of vector: insert to obtain conditions for optimum ligation. T4 DNA ligase (Invitrogen™) was used to catalyze ligation in the presence ligation buffer containing ATP. Reaction mixture was prepared by adding, vector: insert (200 ng vector: 0.2 pmol insert) 1 µL of 10X ligation buffer (1X buffer contains 10 mM $MgCl_2$, 50 mM Tris.HCl, 1 mM ATP, 10 mM DTT, pH 7.5, 25°C) and 1 µL T4 DNA Ligase (1 U/µL). Vector DNA and inserts were added in respective ratios, reaction volume was made to 10 µL and incubated overnight at room temperature. Ligation products (5 µL) were used to transform DH5α competent cells.

2.12 DNA sequencing

For DNA sequencing Sanger Dideoxy-chain termination method (Sanger et al., 1977), was employed using the ABI Prism" Big Dye' Terminator Cycle Sequencing Kit, Ver. 3 (Perkin-Elmer). All sequencing was performed by commercial genomic services, University of Birmingham. For sequencing, ~ 200-500 ng template DNA and forward or reverse primer (3.2 pmol) were added in reaction mixture of 10 µL (adding DNase free water). To sequence pET21a following primers were used (Table 2.4).

Table 2.4: Primers used for pET21a sequencing

	Primer Name	Sequence
1	T7 Promoter	3` TAATACGACTCACTATAGGG 5`
2	T7 terminator	3` GCTAGTTATTGCTCAGCGG 5`

2.13 SDS polyacrylamide gel electrophoresis

SDS Polyacrylamide Gel Electrophoresis (SDS PAGE) was used to examine purity and separation of various proteins according to their molecular weight. A protein sample (10-30 µL) was prepared using 4X SDS PAGE loading buffer (2 mL 1 M Tris.HCl pH 6.8, 0.8 g SDS, 4 mL 100 % Glycerol, 4 mL β-mercaptoethanol, 10 µL 10% Bromophenol blue) and mixture was boiled to denature the proteins. Sodium dodecyl sulphate (SDS) binds to linear proteins and gives them a negative charge.

In this thesis, most of gels were prepared manually. For the crosslinking Mass spectrometry experiments, commercial gradient gels (Bio-Rad) were used. A Mini-PROTEAN® Electrophoresis System (Bio-Rad) was used to run gels in all cases.

2.13.1 Tris Glycine SDS PAGE

In general, Laemmli SDS PAGE or Tris-Glycine SDS PAGE was used to separate proteins. Solutions used for SDS PAGE are given in Appendix 4. Protein samples were prepared using the 4X SDS loading buffer and boiled for 5 min at 95°C before loading onto gel using sample loading tips. Tris-Glycine buffer was used as running buffer.

2.13.2 Tris tricine SDS PAGE

Low molecular weight protein (10-100 kDa) can easily be separated using Tris Tricine SDS PAGE. Tris tricine SDS PAGE was performed as described by (Schagger, 2006). Different anode and cathode buffers are used in this system along with 15 % PAGE. Solutions for making Tris tricine SDS PAGE gel, sample loading buffer, anode and cathode buffers have been given in Appendix 4. Anode and cathode buffers were used to run gel electrophoresis at 150 V.

2.14 Staining of SDS PAGE gels

Proteins separated on SDS PAGE gels, were visualized by staining with either coomassie blue (CBB R250) or colloidal coomassie (CBB G250). Latter is more sensitive (detects proteins ~ 10 ng) and do not need de-staining to visualize protein bands.

2.14.1 Coomassie blue staining

Coomassie blue stain (250 mg Coomassie Brilliant Blue, CBB R250, 10% v/v glacial acetic acid, 40 % v/v methanol, 50 % v/v distilled water) was used to stain SDS PAGE gel for 1 hr. After that, gel was de-stained for 1 hr to overnight in de-staining solution (10 % v/v glacial acetic acid, 40 % v/v methanol, 50 % v/v distilled water). The gels were then subjected to 7.5 % glycerol for 30 min on a shaker and then dried using the cellophane paper sheets and gel drier at 65°C for 2 hr.

2.14.2 Colloidal coomassie staining

Colloidal coomassie staining is more sensitive than coomassie blue and it does not require de-staining to develop protein bands. Gel was stained 4 hr to overnight to visualize very light protein bands. Staining solution can be made by adding following in a sequence, 5 % (w/v) ammonium sulphate, 10 % (v/v) ethanol (96%), 0.02 % (w/v) CBB G-250. These were dissolved in distilled water and 2 % (v/v) orthophosphoric acid (85 %) was added that turns coomassie molecules into colloidal particles and solution itself was used as unfiltered. For de-staining (if required), solution was made using 10 % (v/v) ethanol (96 %), 2 % (v/v) orthophosphoric acid (85%).

2.15 Electrophoretic mobility shift assays (EMSA)

2.15.1 PAGE for electrophoretic mobility shift assays

EMSA is an efficient and rapid technique to study protein-DNA interaction. This technique compares mobility of protein bound DNA with unbound DNA. Unbound DNA runs fast in EMSA gel. A modified method was used for EMSA. Binding buffer (10 X), used to allow protein-DNA interactions, contained 100 mM Tris (pH 7.5 at 20°C), 10 mM EDTA, 1 M KCl, 1 mM DTT, 50 % v/v glycerol, 0.10 mg/mL BSA (10X buffer). Protein samples were prepared using various ratios of DNA and protein (0.1-10 μM) in a reaction of 20 μL and incubated for 30 min at room temperature. Usually DNA concentration was kept constant and optimal protein

concentration was determined to interact with DNA. Either short oligonucleotides or long linear DNA fragments were used for assay. TAE was used as running buffer for PAGE (10X, Tris acetate EDTA buffer contains 400 mM Tris, 25 mM EDTA pH 7.8). 40% w/v acrylamide-bisacrylamide stock solution (40: 1.2) was used to prepare different percentages of gels; for 10 % PAGE (1 mL (10 x) TAE buffer, 2.5 mL 40% w/v acrylamide-bisacrylamide, 6.5 mL H_2O, 250 µL APS and 6 µL TEMED).

PAGE was performed at 4°C in cold room at 5 V/cm. Low temperature helps to avoid DNA smears in gel. The EMSA gel was stained using ethidium bromide (0.5 µg/mL) and de-stained with water. Gel was visualized using a gel Doc transilluminator (Cambridge Biosciences) and Images were recorded.

Table 2.5: DNA oligo used for EMSAs

Oligonucleotide	Sequence
O_AO_B	5`GAGTTTTAGCGGGTAAAGGTGTGATGCGAGAAATGTT TAGCTAAACTTC 3`

2.15.2 Agarose gel electrophoresis (AGE) for electrophoretic mobility shift assays

Agarose gel electrophoresis (0.1-1.5 %) was also used for EMSA with long DNA fragments. 10X binding buffer (mentioned above) was used to prepare 10 µL sample mixture. Gels were run in cold room at 4°C at 3 V/cm. Gels were stained with ethidium bromide (0.5 µg/mL) and detained with water. Gel was visualized using a Gel Doc transilluminator (Cambridge Biosciences) and images were stored.

2.15.3 EMSA for IncC NTD by radiolabeled PCR fragments

Three DNA fragments from the RK2 plasmid were amplified using the primer sets given in Table 2.6. The amplified fragments were gel purified and radiolabeled by mixing amplified fragments (20 µL, ~200 nM), T4 polynucleotide kinase (1 µL), 2 µL [γ-32P]-ATP (6 µCi/µL) and kinase buffer (3.5 µl, 10 X). Reaction mixture was incubated at 37 °C before removing the free [γ-32P]-ATP using the Sephadex G-50 spin column. Eluted labelled fragments were stored and used for EMSA.

Table 2.6. Primers used to amplify fragments from RK2 plasmid

Primer	Sequence
RK2 Primer 1F	GGCTGCGAATTCTCCGGTAATTGGTAAGAATAATAATAC
RK2 Primer 1R	GCCCGAAGCTTCGACCGTGCGGGGCTTTTTTTGCCTG
RK2 Primer 2F	GCGCCTTGGTTGAACATAGCGG
RK2 Primer 2R	GGCAAGAGACGAAAGCCCGGTTTC
RK2 Primer 3F	AGGTCGCCGTTGCGAACCACC
RK2 Primer 3R	GACGCAAGTGGCAAGGCACAGGTAA

2.15.4 EMSA with PKK113 fragments

To obtain longer linear plasmid DNA, the pKK113 plasmid (pUC18, map given in Appendix 1, plasmid contain an O_B site (TTTAGCGGGCTAAA) inserted within the MCS) was digested using PvuII restriction enzyme and two fragments (291 bp and 2364 bp) were obtained. The smaller fragment obtained from PKK113 digestion consists of following sequence.

5` CTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTT
CCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTTTTAGCCCGCTAAAG
AATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC
CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG
AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAG 3`

The DNA fragments from the digested PKK113 plasmid were gel purified using a QIAquick Gel Extraction Kit (Qiagen) and ESMAs were performed.

2.16 Protein expression and isolation

All *incC* and *korB* genes have been previously cloned in modified pET21a vector (Batt et al., 2009). This vector encodes an N-terminal His_6 tag before protein sequence. A thrombin cleavage site separates the His_6 tag and protein. Proteins were overexpressed in the *E. coli* strains BL21(DE3), BL21(DE3) pLysS and T7 Express using the LB medium. Gene expression in all these strains was induced by IPTG. IncC NTD was expressed using only BL21(DE3).

2.16.1 Pilot scale protein expression

A single colony from transformed cells was inoculated to 5 mL LB in the presence of antibiotic (Kanamycin for BL 21 (λDE3) and T7 Express; Kanamycin and Chloramphenicol for BL21 (DE3) pLysS). Overnight cultures were used to inoculate fresh medium. For small scale protein overexpression, a fresh 50 mL LB medium was used with appropriate antibiotic(s). Inoculated culture was incubated at different temperatures (37°C and 25°C) and with shaking (190 rpm). IPTG (0.1-1 mM) was added to cells grown to mid-log phase (OD_{600} ~0.6) to induce protein expression. 1 mL culture was aliquoted for SDS PAGE analysis before and after induction. Bacterial cells were further grown for 4 hr at 30-37°C, or overnight at 16-25°C, followed by centrifuging at 4°C for 15 min at 10,000x *g* in a Beckman Avanti J-25 Ultracentrifuge. Supernatant was discarded and cell pellet was resuspended in a minimum amount of STE (TEN) buffer (100 mM NaCl, 10 mM Tris.HCl pH 8 and 0.1 mM EDTA). Centrifugation step was then repeated, supernatant was discarded and cell pellet was air dried and kept at -20°C till further use. For the IncC1 and IncC2 proteins, an initial trial expression was obtained using the different *E. coli* strains and various conditions were tried to express proteins in a soluble form. To check proteins in insoluble fraction, cell pellet was resuspended in 6 M urea and SDS PAGE analysis was performed.

2.16.2 Large scale protein overexpression

Optimisation of temperature, IPTG concentration and *E. coli* strain resulted in soluble protein overexpression. Optimised conditions were used for large scale protein overexpression to get sufficient protein for biophysical characterisation. A small culture of 50 mL was inoculated with a single colony of a selected *E. coli* strain (with appropriate antibiotic and 2 % sterilised glucose to avoid leaky overexpression) and grown overnight at 37°C. 2-4 media flasks (each with 500 mL LB medium) were inoculated using 5 mL of starter culture. Flasks were incubated with continuous shaking at 37°C until an OD_{600} ~ 0.6 was reached. IPTG (0.5-1 mM) was added to culture flasks to induce protein expression. Temperature was adjusted to optimum setting and cultures were incubated for 6-16 hr (optimum growth time established by pilot protein production). Cells were centrifuged as described above and cell pellet was stored at -20°C till further use.

2.17 Bacterial cell lysis by sonication

Bacterial cell pellet, stored at -20°C, was thawed and resuspended in a lysis buffer. Different lysis buffers, containing an EDTA-free protease inhibitor tablet (Roche Diagnostics) were used for different proteins at different pH to optimize protein solubility (Tris, Phosphate, HEPES). About 3 mL of a lysis buffer was added every gram of cell pellet. Cell pellet was thoroughly resuspended. To get rid of DNA, DNAse I or Benzonase (Roche) was added to final concentration of 10 µg/mL. Sonication was performed on ice using Sonicator (Sonics) by applying 10 sec on/off pulse periods (20-30 V power) for 4 min for lysis of different bacterial strains. The cell lysate was centrifuged in a Beckman Avanti J-25 Ultracentrifuge at 4°C and 40,000 rpm for 30 min (Beckman, Ti50 rotor). The cell pellet contained insoluble protein fraction. Soluble protein fraction (supernatant) was then loaded on a Ni-NTA column for purification.

2.18 Protein purification

2.18.1 Ni-NTA column chromatography

(a) Using in-lab made nickel agarose column

Protein purification was carried out using a Nickel NTA (Ni-NTA) resin (Qiagen). 15 mL column was prepared for the His_6 tagged protein purification (usually used 2 mL resin for 10-15 mg of protein). The Ni-NTA column was equilibrated with a buffer (usually the same buffer used for lysis unless mentioned) containing 10 mM imidazole, 0.5 mM PMSF, pH 7.5-8.0. Lysate was loaded on column and unbound proteins were allowed to pass. Column was washed with a wash buffer (3 volumes, optimum buffer plus 20 mM imidazole pH 7.5-8.0). After washing, the His_6 tagged proteins remained bound to column and eluted with a linear imidazole gradient (20 mM-250 mM imidazole) or step elution (50 mM, 100 mM, 150 mM, 200 mM and 300 mM imidazole). For step elution, at least 3 column volume buffer was used for each step. Fractions (2.5 mL) were collected and protein concentrations were quantified using Bradford reagent. Samples (20 µL) were analysed on SDS PAGE from every 2^{nd} fraction to check the purity of a protein.

(b) Using G.E. Healthcare HiTrap Ni Sepharose 6 Fast Flow (FF) column

In some cases, where proteins were not very stable, instead of lab-made columns, Hi-Trap chelating column (5mL) from Pharmacia Biotech was used. This column contains high

performance chelating Sepharose resin cross-linked to the agarose beads. This column can be charged either with Nickel or Copper to capture the His_6 tagged proteins and can be used either by gravity method/peristaltic pump or as connected to Äkta chromatography system. Binding capacity of the column is ~ 40 mg/mL and has a maximum flow rate of 20 mL/min. A 5 mL Hi-Trap FF column was used with peristaltic pump (flow rate of 1 mL/min). Column was charged with 10 mL of 100 mM $NiSO_4$ and washed with 5 column volume (CV) of deionized water. The column was equilibrated with selected buffer (5 CV) containing 10 mM imidazole. Clear lysate was loaded onto column after centrifugation and column was washed with 5 column volume of wash buffer. The His_6 tagged protein was eluted by step elution (50, 100, 150, 200, 300 mM) using imidazole containing buffers (at least 3 column volume for each step). All buffers contained PMSF to inhibit any protease activity. Eluted fractions were checked for protein concentration by Bradford assay and for purity by SDS PAGE.

2.19 SP-Sepharose cation exchange chromatography

SP-Sepharose^{TM} Fast Flow (Amersham Biosciences) resin was used for cation exchange chromatography. Desired buffer with pH 7.5-8.0 (mentioned in results) was used to equilibrate the 15 mL SP-Sepharose column. Protein sample was loaded on column and fractions (2.5 mL) were collected. Proteins were separated using the same buffer with sodium chloride gradient (100 mM-1M). Protein fractions were assayed by Bradford reagent and run on SDS PAGE. Purified fractions were stored at 4°C.

2.20 Protein dialysis and buffer exchange

To remove salts from proteins or to exchange protein buffer, dialysis was performed. Semi-permeable membranes of different cut off sizes were used to remove small ions and molecules. Dialysis tubing was boiled in 0.1 mM EDTA for 60 min and rinsed thoroughly with distilled water. The treated membrane was stored at 4°C. Tubing was cut to appropriate size and clamped at both ends after pouring protein. Tubing was placed in beaker with appropriate buffer to exchange. Generally, dialysis was performed with continuous stirring for 18 hr at 4°C with 4-6 buffer changes. Imidazole was removed using dialysis and protein was quantified using the UV spectroscopy and Bradford method while purity was determined using SDS PAGE.

2.20.1 PD-10 Columns

PD-10 columns (Amersham Biosciences) provide an easy and quick way to exchange buffer and to de-salt proteins. These columns contain Sephadex™ G-25 matrix to separate high (Mr >5000) from low molecular weight substances (Mr <1000). The PD-10 columns were first equilibrated with selected buffer and then 2.5 mL protein sample was loaded on column. After loading protein, 3 mL of elution buffer (same as equilibration buffer) was used to elute protein by gravity.

2.21 Storing of proteins

Most of procedures for protein purification/analysis (chromatography, PAGE) were performed at room temperature. PMSF (final conc. 1 mM) was added to all purified protein samples to inhibit any protease activity. Sodium Azide (NaN_3) was added to some protein samples at concentration 0.02 % to inhibit microbial growth in samples. Purified proteins were usually stored at 4°C. Some unstable protein samples were flash frozen in liquid nitrogen and stored at -80°C before further analyses.

2.22 Concentrating protein using ultrafiltration/ centrifugation

The Amicon ultrafiltration cells and the Millipore YM filter membranes (required cut off size) were used to concentrate proteins (using 60 psi pressure and N_2 gas). About 15-30 mL of a protein sample was added to a ultrafiltration cell (different size cells) and N_2 gas was allowed to pass till protein was sufficiently concentrated.

The purified protein fractions were also concentrated to feasible protein concentration by centrifugation using Vivaspin or Centricon concentrator (cut off 5-10 kDa Mol. wt.). A 2-10 mL of protein sample was added to concentrator tube and centrifuged at 4,000x *g* and 4°C for 15 min to an hour or until desired protein concentration was acquired. Proteins were checked for aggregation at different intervals during concentration.

2.23 Protein concentration determination

2.23.1 UV-Visible spectrophotometry

Protein concentration can be measured by UV-Visible Spectrophotometry, at 260 and 280 nm, using either a quartz cuvette in spectrophotometer or using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

Protein concentration is calculated by any of following equations:

$$\text{Protein concentration (mg/mL)} = 1.55\ (A_{280}) - 0.76\ (A_{260})$$

$$A = \varepsilon cl$$

Where 'ε' is extinction coefficient (specific for different protein, calculated theoretically) and 'l' is path length of cuvette and 'A' is absorbance. Some components of protein buffers also absorb in 260-280 nm range. In that case, protein concentration was estimated by Bradford method.

Table 2.7 Extinction coefficient (ε_{280}) for different proteins

Protein	ε_{280}	Mol. weight (kDa)
IncC NTD-His_6	6990	13.1
IncC1-His_6	18500	29.5
IncC2-His_6	27500	39.5

2.23.2 Bradford assay

This is a colorimetric assay for measurement of protein concentration. Assay uses a dye coomassie brilliant blue (CBB) G-250 which bind to proteins. Binding to protein stabilizes blue anionic dye which can be measured at 595 nm. BSA was used to make a standard curve for protein estimation. For a sample assay, 20 µL protein sample was added to 0.5 mL of H_2O and 0.5 mL of Bradford reagent (Bio-Rad). Absorbance was noted at 595 nm and protein concentration was determined using the calibration curve obtained using different concentrations of BSA.

2.24 Size exclusion chromatography (SEC)

SEC is a non-binding method that can be used to purify proteins on the basis of size and shape. It also provides a close assessment of molecular weight for globular proteins. Various SEC resins are available that allow fractionation of proteins with wide range of molecular

weights and sizes. SEC column separates proteins depending not only on their size and shape but also on their frictional coefficients (*f*). Stokes radius (R_s) is usually used instead of frictional coefficient. R_s allows better assessment of protein size and can be defined as "radius of a smooth sphere" that would have an actual *f* of a protein" (Erickson, 2009). R_s is larger in case of elongated/flat proteins like IDPs that have more bound water. Protein elution from a SEC column closely depends upon Stokes radius (R_s) of proteins (Siegel and Monty, 1966, Erickson, 2009). Proteins, with known Stokes radii, can be run to calibrate SEC column and to obtain a standard R_s curve. From standard curve, Stokes radius of an unknown protein can be calculated. For a protein with smooth sphere, Stokes radius (R_s) can be calculated theoretically from the following equation using frictional coefficient.

$$f = 6\pi\eta R_s$$

Where η is viscosity of the sample and *f* is frictional coefficient.

Here, a 120 mL Hiload 16/60 Superdex 75 pg column (Amersham Pharmacia) attached to an ÄKTA Purifier chromatography system (GE Healthcare) was used to run size exclusion chromatography. Proteins with a wide range of molecular weight (3-70 kDa) can be purified using Superdex 75 pg column. A flow rate of 0.4 mL/min was used to collect 2.5 mL sample fractions. A maximum of 2.5 mL of sample was loaded onto column using a 5-mL sample loop. UV absorbance, conductivity, pressure and fractionation were monitored during run and recorded using the Unicon ver. 5.1 software. The peak fractions, monitored by UV absorbance, were pooled, concentrated and analyzed by SDS PAGE to check purity of protein samples. Calibration curve, for molecular weight (mol. wt.) and Stokes radius (R_s), were prepared by running standard proteins and by measuring their elution volumes (Figure 2.2).

2.25 Size Exclusion Chromatography-Multi Angle Laser Light Scattering (SEC-MALLS)

Size exclusion chromatography can be combined with MALLS to determine molecular weight and monodispersity of a protein sample with more confidence. In this technique, a size exclusion column is attached to RI (refractive index) monitor and a multi angle light scattering detector. Protein is fractionated using a SEC column and passed through a MALLS detector. Light from 18 different angles passes through separated fractions and scatters depending upon shape and size of molecules. Refractive index give concentration of the protein, that together with intensity of light scattered, is used to calculate molecular mass.

The IncC proteins (IncC1, IncC2, IncC NTD) were analyzed using the SEC-MALLS instrument at the Harwell Research Complex, UK. A SEC column Superdex 200 10/300 GL (GE Healthcare Life Sciences) was attached to a ÄKTA pure chromatography system (Figure 2.4) was used. Column was washed and equilibrated using Tris Buffer (20 mM Tris.HCl pH 7.5, 150 mM NaCl and 0.1 mM EDTA). An SIL-20A Auto-sampler was used to inject 300 µL protein sample (0.5-2 mg/mL). A Dawn HELEOS-II 18-angle light scattering detector (Wyatt Technology Corp.) and an Optilab T-rEX refractive index monitor (Wyatt Technology Corp.) (Figure 2.3) were used to record light scattering and refractive index. Light scattering was observed at 682 nm while UV absorbance was measured at 280 nm.

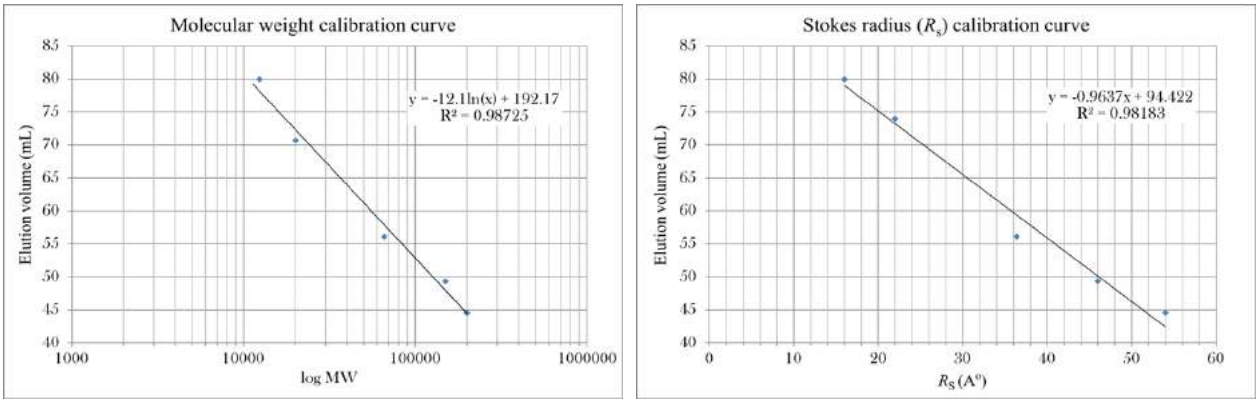


Figure 2.2 Standard curves for molecular weight and Stokes radius (R_S) were obtained using five standard proteins, β-Amylase, Sweet Potato (mol wt. 200 kDa, R_s 5.4 nm), Alchol Dehydrogenase tetramer (mol wt. 150 kDa, R_s 4.6 nm), Bovine Serum Albumin (For monomer, mol wt. 66 kDa, R_s 3.64 nm), Trypsin inhibitor (mol wt. 20.1 kDa, R_s 1.88 nm), Cytochrome C (mol wt. 12.4 kDa, R_s 1.69 nm). Proteins were run individually, as well as a mixture to get best calibration.

Figure 2.3: The SEC-MALLS instrument at the Harwell Research Complex (UK). SEC column is attached to ÄKTA pure chromatography system, HELEOS-II 18-angle light scattering detector and Optilab T-rEX refractive index monitor

SEC-MALLS was performed at 20°C with a flow rate of 0.7 mL/min. Data were analyzed using Astra Ver. 6.0. Peaks were analyzed giving weight-averaged molecular weight, radius of hydration and polydispersity of samples. Basic light scattering equation is used to determine average molecular weight.

$$\frac{K^* c}{R(\theta)} = \frac{1}{M}\left[1 + \frac{16\pi^2}{3\lambda^2}\,\langle r^2{}_g\rangle sin^2\left(\frac{\theta}{2}\right) + \cdots..\right] + 2A_2 c$$

Where

$$K^* = \frac{\left[4\pi^2 n^2 \left(\frac{dn}{dc}\right)\right]}{\lambda_o^4 N_A}$$

(Wen et al., 1996)

K^* is an optical parameter, R(θ) is excess light scattered at angle θ (intensity due to solute), c is solute concentration, λ is wavelength of light in a vacuum, M is weight-averaged molecular weight, $<r^2_g>$ is mean square radius of gyration, A_2 is second viral coefficient, n is refractive index, dn/dc is refractive index increment of solute and N_A is Avogadro's number. If molecular weight does not exceed 5 x 10^7 Da, viral coefficient A_2 and term $[16\pi^2 \sin^2(\theta/2)]/(3\lambda^2)$ is negligible (Wen et al., 1996). So, following equation can be used for average molecular weight (M) determination.

$$M = K'\left(\frac{LS}{RI}\right)$$

Where

$$K' = \frac{K_{RI}}{\left[K_{LS}\left(\frac{dn}{dc}\right)\right]}$$

(Wen et al., 1996)

LS is light scattering and RI is refractive index. Where K_{LS} and K_{RI} are instrument calibration constants.

2.26 Biophysical characterization of proteins

Different biophysical tools were used for structural and functional characterization of proteins.

2.26.1 Analytical ultracentrifugation (AUC)

AUC is a low-resolution technique that allows quantitative analysis of macromolecules on the basis of their sedimentation in response to centrifugal force. Centrifugal force makes macromolecule moves down the cell as a function of its mass and shape. This motion is opposed by frictional force, which depends on the shape of a molecule, and buoyancy, which depends on the solvent. As centrifugation progresses, macromolecules are separated with defined boundaries due to particle movement. AUC can be performed either as sedimentation velocity experiment (svAUC) or sedimentation equilibrium experiment (seAUC). The svAUC provides information for size, shape and molecular interactions while seAUC give additional information including stoichiometry or association constants. In a sedimentation velocity (sv) experiment, rate of sedimentation is measured in real time, while centrifuging at high speed; while in sedimentation equilibrium (se) experiment, force is lower so that a concentration gradient is maintained within cell. Real-time collection of the sedimentation data allows hydrodynamic and thermodynamic characterisation of macromolecules. Sedimentation is optically monitored by recording absorbance at 230 nm or 280 nm, caused by peptide bonds and aromatic residues, relative to the reference cell containing buffer. Molecular complex within a range of 100 kDa to 10 GDa can be probed for shape and sizes as well as for any aggregation.

In this thesis, svAUC has been used to characterise mass of protein samples in native state. The svAUC experiment involved sedimentation using high speed centrifugation (40 K

rpm). Under the influence of centrifugal force, macromolecules migrate from meniscus forming a concentration distribution across cell and an optical detector (absorbance, interference and fluorescence) collects data in real time. Protein can be characterized under various conditions of temperature, pH, ionic strength in their native environment using AUC. Sedimentation/Svedberg coefficients (s) can be calculated from real time data and depend upon molecular weight, conformation and molecular interactions. s can be calculated from the following equation.

$$s = \frac{v}{\omega^2 r} = \frac{M(1 - \bar{V}\rho)}{N_A f}$$

Where *s* is sedimentation coefficient, *v* is terminal velocity, ω is angular velocity of rotor and *r* is radius of particle to the rotor axis. Using this equation as well as Lamm equation (given below), hydrodynamic properties of protein samples can be determined.

$$\frac{\partial c}{\partial t} = D\left[\left(\frac{\partial^2 c}{\partial r^2}\right) + \frac{1}{r}\left(\frac{\partial c}{\partial t}\right)\right] - s\omega^2\left[r\left(\frac{\partial c}{\partial t}\right) + 2c\right]$$

Where *c* is solute concentration, *t* and *r* are time and radius, and *D* is solute diffusion constant.

AUC was performed for the all IncC proteins. For IncC NTD three different protein concentrations were used. Data were acquired using the Beckman Coulter Optima XL-A Analytical Ultracentrifuge (Beckman Coulter, Palo Alto, CA, USA) with an An-50Ti Analytical Rotor at 40,000 rpm and 20°C. Continuous scans (every 5 min) were collected for absorbance of 280 nm.

Concentration of a protein in various positions of cell are monitored as a function of time, and sedimentation coefficient is determined, based on the Lamm equation, using the program SEDFIT (Schuck, 2000, Lebowitz et al., 2002). SEDNTREP 2.0 was used to get $\bar{V}$, η and ρ (Table 2.8) (Laue et al., 1992). Sedimentation in water at 20°C (s20, w) was calculated. Different parameters used to fit Lamm equation, while analysed data have been shown in Table 2.8. The time lapses in recorded AUC data files can restrain the precision of sedimentation coefficient (Zhao et al., 2013). To overcome this problem REDATE (Rectifying Emender of Data Acquisition Time Errors) software was used. The c(S) and residual plots were exported to GUSSI 1.0.8e version and different concentrations were overlaid each other and figures were exported.

Table 2.8: AUC parameters for IncC NTD protein

Protein	Conc. mg/mL	Meniscus	Bottom	Buffer Viscosity (η)	Buffer Density (ρ)	Partial Volume ($\bar{V}$)	Hydration
IncC NTD	0.25	6.10	7.15	0.0101	1.009	0.702	0.424
IncC NTD	0.84	6.12	7.14	0.0101	1.009	0.702	0.424
IncC NTD	1.20	6.15	7.17	0.0101	1.011	0.702	0.424
IncC1	0.25	6.12	7.16	0.0107	1.011	0.734	0.394
IncC1+ATP	0.25	6.13	7.19	0.0107	1.011	0.734	0.394
IncC2	0.20	6.16	7.16	0.0107	1.011	0.745	0.383
IncC2+ATP	0.20	6.16	7.19	0.0107	1.011	0.745	0.383

2.26.2 Circular Dichroism (CD)

CD can be used to get assessment of the proportion of α helix, β sheets, random coils in a protein secondary structure. CD calculates difference in light absorption, as a function of wavelength, when left and right circularly polarized light passes through a sample. In plane-polarized light, waves vibrate in single plane and radiation electric field has fixed orientation, while in circularly polarized light, direction of polarization rotates either clockwise (right circularly polarized light) or anticlockwise (left circularly polarized light) at the frequency of radiation (Figure 2.4).

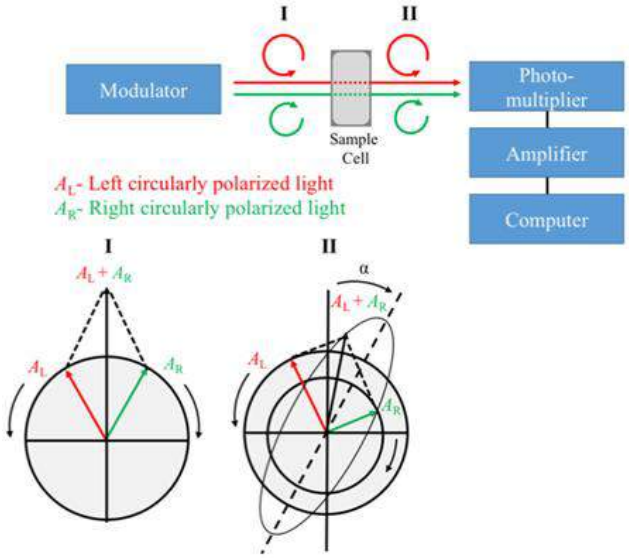


Figure 2.4 Instrumentation and working of two beam CD instrument. The Field vectors of left (Red) and right (Green) circularly polarized light and their superposition (black), before (I) and after (II) passage through the sample. CD Instrumentation is shown above where light from a modulator passes through the protein sample and change in absorbance is detected and signal is passed from photomultiplier and amplifier to computer. (adapted from Greenfield, 2006)

Circularly polarized light is absorbed by chiral molecules in a different fashion. The difference in absorbance (ΔA) gives either a positive or negative value. CD instrumentation is shown in Figure 2.4.

The ΔA is calculated using following equation.

$$\Delta A = \frac{A_L - A_R}{A}$$

where A_L is the absorbance for left circularly polarized light, A_R is the absorbance for right circularly polarized light, and A is the absorbance for non-polarized light. Although the data is measured in form of ΔA, it is displayed in degrees of ellipticity (θ) that originates from magnitudes of the electric field vectors A_L and A_R. θ can be calculated using the following equation and can be converted into molar ellipticity [θ].

$$\theta = \Delta A \left(\frac{ln10}{4}\right)\left(\frac{180}{\pi}\right)$$

Multiplicative relation between circular dichroism Δε and molar ellipticity can be determined by the following equation.

$$[\theta] = \frac{100\theta}{dc}$$

Where *d* is path length [cm] and *c* is concentration of the protein [mg/mL]. For secondary structure, peptide bond absorbs depending on its orientation, so results are usually given in mean residue ellipticity $[\theta]_{mre}$ *i.e.* per residue.

$$[\theta]_{mr} = \frac{[\theta]100M}{cdN}$$

$$[\theta] = 100\Delta\varepsilon\left(\frac{\ln 10}{4}\right)\left(\frac{180}{\pi}\right) = 3289.2\Delta\varepsilon$$

Where M is molar weight [g/mol], c is concentration [mg/mL], d is path length [cm] and N is number of amino acids.

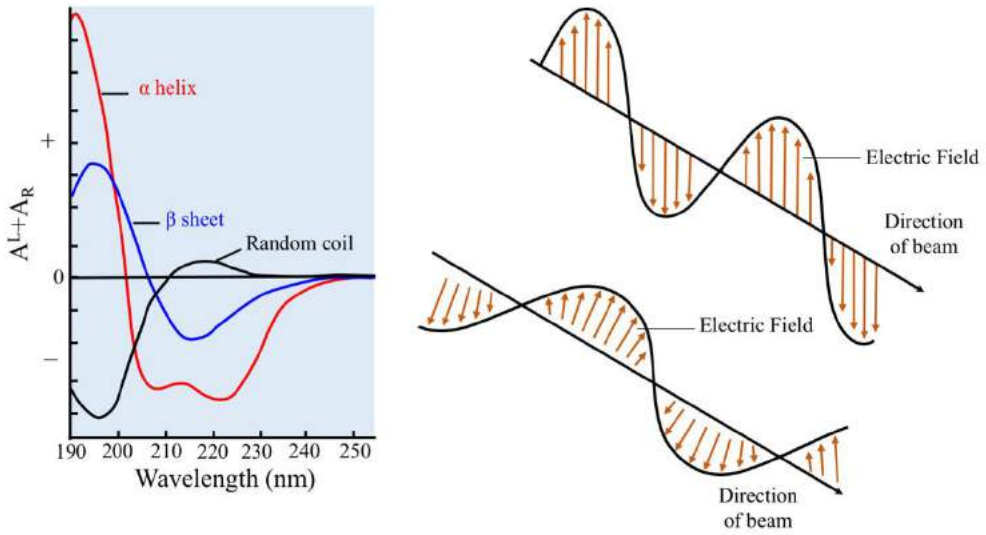
All amino acids with exception of Glycine show enantiomeric properties, therefore, CD can be used for characterization of proteins. Moreover, peptide bond among amino acids is a most ubiquitous chromophore. To characterize protein backbone and secondary structure, CD spectrum in farUV region (190-260 nm) is obtained. The α-helical contents in proteins and polypeptides can be estimated by measuring double minimum at 208 nm and 222 nm and positive CD signals between 190-200 nm. Although a specific region for β sheet in CD spectrum

is not well defined but a minimum between 216-220 nm and a positive peak at ~195 nm can be used to estimate β-sheets in a protein structure (Figure 2.5). β turns show positive peak between 200-205 nm and a negative peak between 220-225 nm and 180-190. Random coils give positive signal in CD spectra around 210-225 (Figure 2.5). Polyprolines are unusual and show a positive peak between 225-230 nm in CD spectra. Composition of a complex protein can be determined as follows.

$$CD_{protein} = AX_{\alpha\ helix} + BX_{\beta\ sheet} + CX_{random\ coil}$$

Where X represents contents for each component of secondary structure and A, B and C are the data points of a corresponding spectra at a certain wavelength.

(a) **(b)**



(c)

Figure: 2.5 (a) CD standard spectra for protein showing α helix, β sheets and random coil conformations. Red line showing a negative CD signal at 222 nm and 208 nm represent a typical CD spectrum for α helix. Blue line represents a CD spectrum for β-sheets and has a negative peak at 216 nm and a positive peak at 195 nm. Black line shows a CD spectrum for a random coil-like protein with a positive peak around 210-225 and a negative peak around 190-200 nm. **(b)** Plane polarized light where electric field is distributed in one direction. **(c)** Circularly polarized light where direction of electric field is change with progression of light wave (adapted from Greenfield, 2006)

The CD spectrum can be compared with standard data sets of known proteins and an assessment of secondary structure is obtained. For that purpose, many online servers and software are available like the DichroWeb server (http://dichroweb.cryst.bbk.ac.uk, (Whitmore and Wallace, 2004), the K2d server (http://www.bork.embl.de/~andrade/k2d/) (Andrade et al.,

1993) CDPro (Sreerama et al., 1999, Sreerama and Woody, 2004), CONTIN-CD (Provencher and Glockner, 1981) and Dicroprot (https://dicroprot-prabi.ibcp.fr/index.php) (Deleage and Geourjon, 1993). CDSSTR is an algorithm that uses a protein database for standard spectra and secondary structures (Sreerama and Woody, 2000). These servers and software have reference datasets that include proteins with crystal or NMR structures.

For CD analysis of the IncC proteins, proteins (~2-10 μM/ 0.1-0.5 mg/mL) in the CD buffer (10 mM Tris, 100 mM $NaClO_4$, 0.1 mM EDTA, pH 7.5) were used. Spectra were collected using the Jasco J-1500 Spectro-polarimeter in an absorbance range 185-260 nm (0.1 mm to 2 mm sample cuvettes (Hellma), bandwidth 1.0 nm, data pitch 0.2 nm and scan speed of 50 nm/min). Data were submitted to the Dichroweb server using the CDSSTR algorithm and the reference data set 3 to assess secondary structure.

To check the thermostability of the IncC proteins, CD melting curve was obtained using the JASCO J-1500 spectro-polarimeter fitted with a Peltier heating block. Experiments were performed using a 1 cm path length quartz cuvette. CD spectra were collected while monitoring 190-260 nm as a function of increasing temperature (5°C steps between 25°C and 95°C).

The nearUV region (260-290 nm) in CD can be analysed for tertiary structure (aromatic R groups) information. Ligand binding can also analyse by CD spectra by looking at various regions (nearUV/farUV) depending upon ligand absorbance properties. Temperature and time-dependent unfolding/refolding of proteins can be analysed at specific wavelength under various desired conditions in presence/absence of a ligand.

2.26.3 Small angle X-ray scattering (SAXS)

SAXS is a low-resolution technique mainly used for determination of radius of gyration (R_g) and distance distribution function $P_{(r)}$. SAXS can complement analytical ultracentrifugation to characterize proteins for size and shape. Macromolecules in a solution can be studied by this technique around resolution of 13 Å.

a. Radius of gyration (R_g)

R_g provides information for overall size of molecule and can be defined as an average root mean square distance of a points within a particle to the centre of density in the molecule weighted by the scattering length density at that point. For a compact protein, R_g is smaller in comparison to rod-shaped and disordered proteins. R_g can be estimated by the Guinier Law using the following equation.

$$I(q) = I(0)\exp(-\frac{q^2 R_g^2}{3})$$

Where I(q) is scattering intensity. Scattering intensity plotted as ln $I_{(q)}$ vs. q^2 should be a linear function for protein of any shape. Linear fit R_g and I (0) can be determined from slope. These are found proportional to protein concentration and molecular weight. Guinier fit accuracy is limited and it is linear over a spectrum region extended to qRg <1.3 for compact protein and for disordered protein this range may decrease to ≤ 0.8 due to multiple conformations. In case of an IDPs, an alternative way to determine R_g is provided by the Debye approximation that can be applied to much wider spectrum region up to qRg <3 following this equation:

$$\frac{I(q)}{I(0)} = \frac{2}{x^2}(x - 1 - e^{-x})$$

Where $x = (qR_g)^2$.

b. Distance distribution function P(r)

Distance distribution function P(r) can be applied to a protein sample in real space. P(r) function is a histogram of distances between all atoms in a protein, from which we can get D_{max}, maximum length of a protein, a maximum value of r, at which P(r) is equal to zero. D_{max} as well as P(r) histogram provides information for shape, anisotropy, and degree of compactness of proteins. P(r) can be obtained by a Fourier transform of the scattering intensity F(I(q)) using the programs GNOM.

$$P_{(r)} = F\big(I(q)\big) = \int I(q)e^{-i\overrightarrow{q.r}}\,dq$$

For IDPs, P(r) functions are asymmetric and display extended tails (protein aggregation also shows an extended tail) in comparison to smooth and symmetric P(r) function for folded proteins. Radius of gyration and forward intensity I(0) can also be obtained using the P(r) function using the following equation.

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 P(r)dr}{2\int_0^{D_{max}} P\,(r)dr} \text{ and } I(0) = 4\pi \int_0^{D_{max}} P(r)dr$$

c. Protein flexibility and Kratky plot

Another measurement for globular nature and flexibility of proteins can be obtained using the Kratky plot that plots scattering pattern as $q^2I(q)$ versus q. In case of a compact protein, scattering intensity decays at high angles approximately as I(q) ~ $1/q^4$ and this gives a bell-

shaped plot with a defined peak, while for extended proteins $1/q^2$ is asymptotic of I(s) showing plateau over the specific range of 's' followed by a monotonic increase (Figure 2.6). The Kratky plot helps to identify extent of flexibility and folding propensity of protein sample by showing featured plots. To compare various protein with different folding states, normalized or dimensionless Kratky plot can be used. To normalize data for concentration and protein size, I(q)/I(0) is used rather than I(q), and qR_g is used rather than q.

A theoretical assessment of R_g for IDPs can be made by Flory's equation. The equation shows R_g for an unfolded protein correlates to its number of residues present in the protein. Flory's equation is as follows.

$$R_g = R_o N^{\nu}$$

Where N is number of amino acids. For proteins, denatured using some chemicals, R_0 is 1.33 ± 0.076 and 'ν' is 0.598 ± 0.028 and for IDPs R_0 is 2.54 ± 0.01 and 'ν' is 0.522 ± 0.01(Bernado and Svergun, 2012).

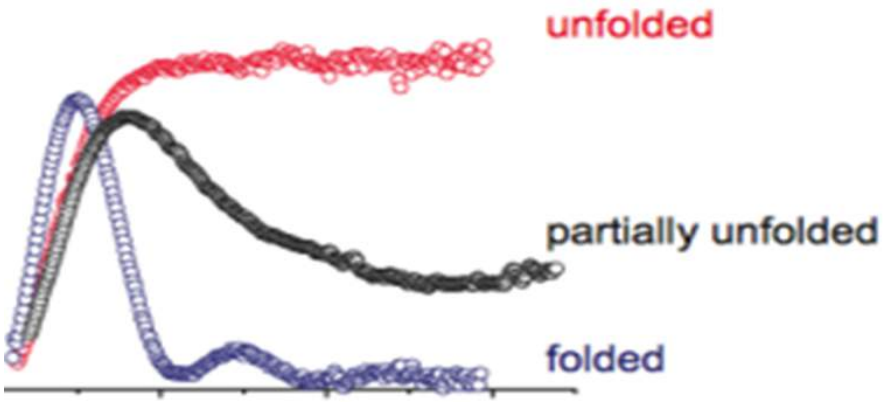


Figure 2.6 A standard Kratky plot showing unfolded, partially unfolded and folded proteins plots. Red colour line shows an unfolded protein curve and does not have any defined peak. Black line shows partially folded, while blue coloured line has a defined peak and shows a folded protein.

The IncC Protein samples were concentrated in respective buffer (mentioned in results). Protein concentrations were determined using the Bradford reagent and a NanoDrop ND-1000 spectrophotometer. Protein samples with varying concentrations were prepared for SAXS analysis. IncC1 and IncC2 (1-5 mg/mL), with/without nucleotides (ATP, ADP), were used to obtain SAXS data sets. The beamline BM29 and the Bio-SAXS instrument at European Synchrotron Radiation Facility (ESRF) in Grenoble, France were used to get datasets. Data were analyzed to one dimension and using the program PRIMUS (ATSAS software package). The Guinier plots, $P_{(r)}$ distributions and Porod-Debye plots were generated and molecular mass was determined from the Guinier plot using the BSA (bovine serum albumin) scattering as a standard. Data from different protein concentrations were compared to examine the protein aggregation.

2.26.4 Thermal shift assay (TSA)

In a thermal shift assay, protein unfolding is observed by increasing temperature gradually. TSA can be used to determine melting temperature of proteins as well as to determine best conditions (temperature, pH, ligand binding) for protein stability. TSA make use of dyes including SYPRO orange, Nile red, and Dapoxyl sulfonic acid. These dyes show high fluorescence when interact with hydrophobic patches on proteins (Niesen et al., 2007). In hydrophilic environments, no or little fluorescence is observed. SYPRO orange is widely used TSA dye due to its low signal to noise ratio. Folded and unfolded proteins show different fluorescence properties, as in unfolded protein hydrophobic patches are more exposed and bind to the dye. In case of folded and structurally ordered proteins, a sharp melting temperature is seen, so increases in dye binding with temperature can be measured. In different buffer conditions (variable salts and pH) protein stability and solubility can be determined as they change melting temperature of a protein. For TSA, a real-time PCR (rtPCR) machine is used and fluorescence is monitored and recorded in a temperature range 25-90°C. With an increase in temperature, protein unfolding increases unrevealing hydrophobic patches in the protein. Fluorescent dye binds to hydrophobic patches and increase in fluorescence is monitored. Fluorescence data can be plotted as a function of temperature and T_m can be obtained from melting curve.

For the IncC proteins, best buffer/pH conditions were determined using TSA. A real-time PCR (Mx3005P qPCR System) was used. Buffers from a Slice pH kit (Hampton Research) as

well as 48 different home-made buffer conditions were used in screening. In each condition, a total volume of 50 µL reaction mixture was prepare containing 30 µM protein and 2.5 X SYPRO Orange (Stock 5000X) (Molecular Probes/Invitrogen). 20 µL of 1X buffers were added to different wells. The PCR machine was programmed to perform 60 steps with a temperature rise of 1°C at every step from 25°C to 85°C, using 1 min for each step. Fluorescence emitted after excitation at 480 nm was recorded at each step. (Either homemade buffers or pH slice kit from Hampton Research was used for pH optimization).

2.27 Crystal trials

The IncC proteins, concentrated to 1-10 mg/mL, were used to obtain crystals. The crystal screens from Molecular Dimensions were used (JCSG-plus™ Screen, 1 and 2) for protein crystallization. Crystal plates were set up using sitting drop method and plates were monitored for crystal growth after every few days using light microscopy.

2.28 Microscale thermophoresis (MST)

MST involves detection of macromolecule interactions and complex formation by their migration along with temperature gradient. Thermophoretic behaviour depends upon nature and properties of macromolecule (size, shape, mass, hydration shell and interactions). Changes in the macromolecule properties along with thermal mobility allow the measurement of any binding event or complex formation. This principle lead to development of a techniques called Microscale thermophoresis (MST, Nanotemper Technologies GmbH).

MST involves use of an IR (infrared) laser to create a thermal gradient in a micro-capillary tube containing fluorescent tagged macromolecule and its unlabelled interaction partner(s). Movement of macromolecules and their complexes along a thermal gradient can be detected. Detector allows the measurement at very low concentration of fluorescently-labelled partner (5-100 nM) to which unlabelled ligand is titrated. Changes in dye fluorescence are instant along with temperature. Infrared laser takes ~1 s to get to final temperature, the temperature jump allows to monitor dynamic changes within timeframe (Figure 2.7).

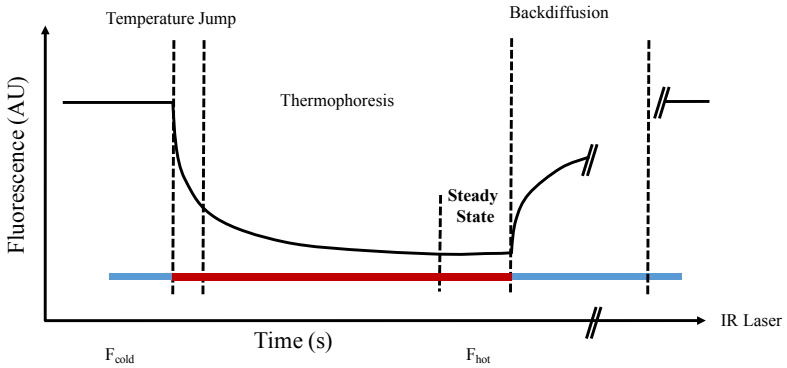


Figure 2.7 Determination of fluorescence of a labelled sample is recorded over time using MST. Black line shows a trace from MST run. Temperature jump is a rapid change in temperature as IR laser falls on the protein sample. Then, in thermophoresis change in fluorescence occur along with moment of protein complexes till a steady state is achieved and IR laser stops leading to back diffusion. F_{cold} and F_{hot} show fluorescence detection when IR is off and on respectively. Adapted from (Scheuermann et al., 2016)

At steady state level, after which laser light stops, there is no net change only due in/out movement of macromolecules in hotspot. Reverse of temperature jump can be observed once laser is switched off and back diffusion ends concentration gradient. For data analysis, initial (F_{cold}) and final (F_{hot}) fluorescence average and F_{hot}/F_{cold} ratio is calculated. From titrations, relative values for macromolecules alone and along with their partners are measured and used for determination of complex formation at respective ligand concentration. If the fluorescence tag is present in vicinity of ligand binding site, it can affect the microenvironment leading to change in dye absorption or emission maxima. Conformation changes in macromolecules upon binding with ligand can also effect dye fluorescence. The competing partners, that share same binding site can be determined using MST.

The IncC1-IncC2 protein interactions were monitored by MST using a Monolith NT-115 Instrument (Red/Blue). The MST buffer contained 20 mM HEPES, pH 7.4, 100 mM KOAc, 10 mM $MgSO_4$, and 2 mM DTT. Standard coated microcapillaries were used to load 10 µl of sample. The MST power (IR intensity) was set to 40 %. The LED intensity was set to get at least 300 fluorescence counts. For IncC2 labelling, the Monolith NT^{TM} Protein Labeling Kit RED-NHS (Nanotemper, Germany) was used and protein was labelled according to kit instruction. The IncC1 protein, in different concentrations, was mixed with the labelled IncC2 and using standard coated capillaries and MST measures were obtained.

2.29 Mass spectrometry

To confirm exact molecular mass of the IncC proteins (IncC1, IncC2 and IncC NTD), Mass spectrometry was used. Protein samples were prepared by diluting the protein samples to 50-100 μg/mL into methanol: acetonitrile (50:50) mixture. The samples were given to Mass spectrometry facility in School of Chemistry, University of Birmingham, for analysis and results were obtained from ESI-MS analysis.

2.30 ATPase Assay

ATPase activity of IncC1 and IncC2 was initially monitored using Malachite Green Assay kit (Sigma-aldrich) and EnzChek® Phosphate Assay Kit (Thermo Fisher Scientific, E-6646) but activity could not be determined in a reliable way, due to low levels. The ATPase activity was then determined by using High-Performance Liquid Chromatography (HPLC). Samples were taken at intervals after reaction and the amount of ATP and ADP was monitored after separation by HPLC, using the absorbance at 254 nm, Isocratic gradient elution (0.7 mL/min) was performed at 4°C using the solvent system using the 50 mM KH_2PO_4/Methanol (95:5) as the mobile phase. The amounts of ADP produced was calculated using standard ADP curve. A C18 column (Shodex) was used and chromatograms were collected by monitoring wavelength 254 nm. The sample injection volume was 20 μL. For some samples, 2-4 folds dilution was performed before loading on the column and corrected accordingly. The ATP and ADP peak areas were quantified using ChemStation (B.04.03) software.

CHAPTER 3

IncC NTD - An N-Terminal Intrinsically Disordered Region (IDR) in IncC1

Chapter 3

IncC NTD - An N-Terminal Intrinsically Disordered Region (IDR) in IncC1

3.1 IDPs and IDRs

Under physiological conditions, the lack of stable tertiary structure results in disorder within a protein. In term of structure, proteins may adopt different conformations and transition from random coil to an ordered form (Figure 3.1) (Staby et al., 2017). Intrinsically disordered proteins (IDPs) do not possess a folded structure in their native environment. IDPs may contain patches of amino acids, forming unfolded regions, called intrinsically disordered regions (IDRs) (Wright and Dyson, 1999, Dunker et al., 2001, Tompa, 2002, Dyson and Wright, 2005, Oldfield and Dunker, 2014, Babu, 2016, Choura and Rebai, 2017). IDRs may consist of short stretches of amino acids to long regions compassing the entire protein (Tompa, 2002). Intrinsically disordered proteins are prevalent in all kingdoms of life: Eukarya, Bacteria, Archaea. IDPs are mostly involved in cell processes as transcription factors and cell signalling proteins (Pazos et al., 2013, Forman-Kay and Mittag, 2013, Kumar et al., 2017). IDPs are more prevalent in viral and higher eukaryal cellular systems in comparison to bacteria and archaea (Oldfield et al., 2005, Chen et al., 2006).

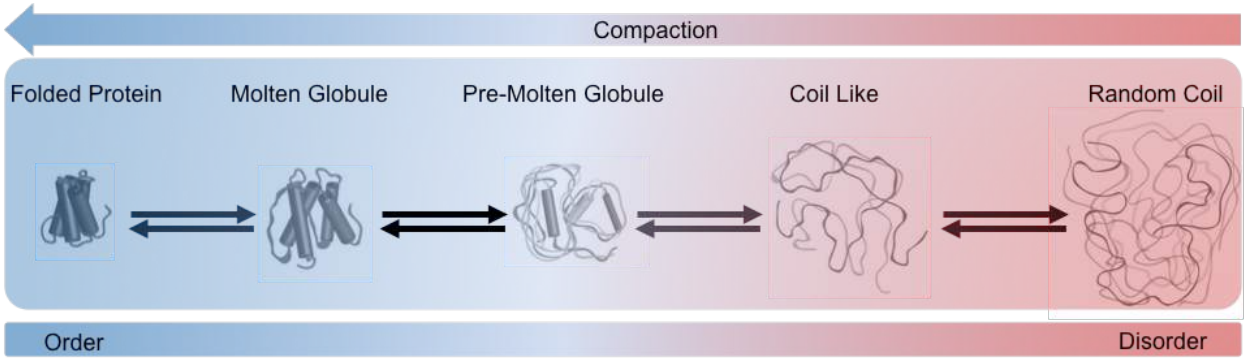


Figure 3.1 Different conformations and transitions of protein structure: Proteins can adapt any native conformation from random coil to folded protein depending on the specific function. Proteins may undergo different transitions from random coil to pre-molten globule, molten globule or fully folded form during interactions with other partners. Adapted from (Uversky, 2002, Staby et al., 2017)

Some IDPs and IDRs may transition from disorder to order when interacting with some partner molecules (Tompa and Fuxreiter, 2008, Dogan et al., 2013, Toto et al., 2016). There are several characteristics of IDPs that distinguish them from ordered proteins including an

abundance of charged, flexible, disorder-promoting amino acids and the paucity of aromatic and hydrophobic residues (Romero et al., 2001, Habchi et al., 2014).

3.2 Functions of IDPs

The extent of disorder makes IDPs moonlighting proteins and enables them to interact with multiple targets asynchronously or in concurrent fashion to assemble large complexes (Tompa, 2005) (Figure 3.2). IDPs are more flexible than folded proteins and thus provide a large surface area for protein-protein interactions (Uversky, 2015). The interactions of IDPs usually occur with a low affinity that allows them to interact, associate, engage in a particular function before dissociating (Dyson and Wright, 2002, Fuxreiter et al., 2004, Oldfield et al., 2005). Another key feature of IDPs is the prevalence of amino acids prone to post-translational modifications (PTMs). These modifications may regulate the interactions of IDPs with various partners including proteins and DNA (Babu, 2016, Babu et al., 2011).

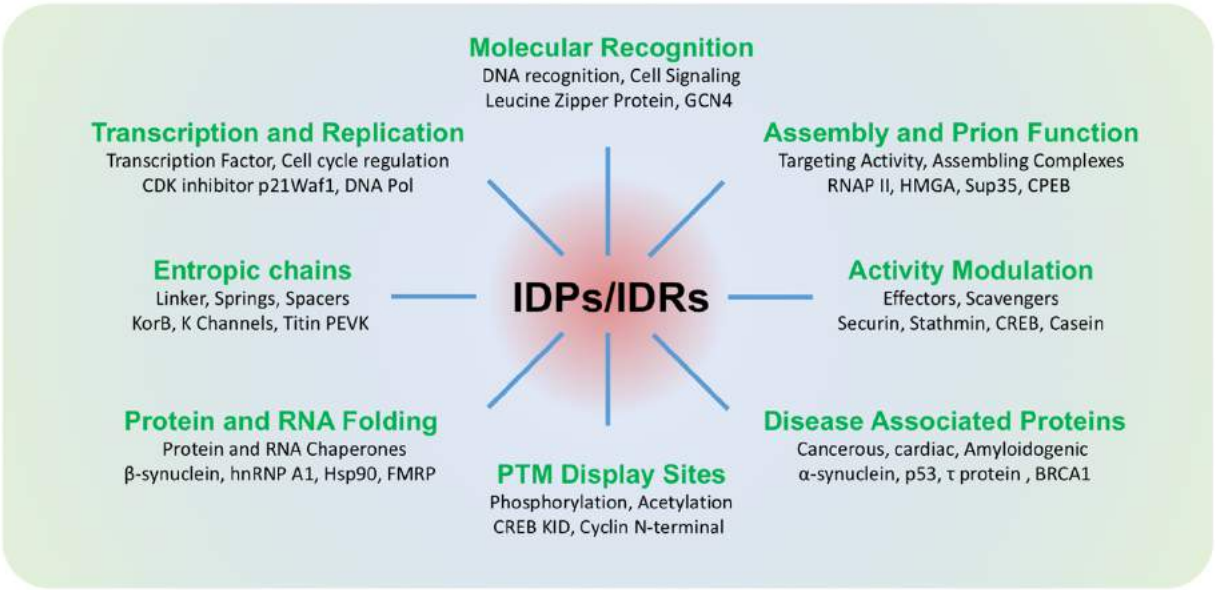


Figure 3.2 Different functions of IDPs and IDRs: The conformational heterogeneity and flexibility enable IDPs/IDRs to perform several functions in the cell. IDPs are involved in recognizing the specific partners and may activate/deactivate them. IDRs are present in many transcription and replication proteins. The IDRs may present as spacers or linkers and provide flexibility to a particular protein. IDPs are also involved in Protein/RNA folding, activity modulation of others proteins and in assembling complexes. IDRs may display the potential PTM regions in the proteins. Adapted from (Tompa, 2005, Sun et al., 2012, Uversky, 2013)

IDPs perform various functions in the cell including molecular recognition, activity modulation, protein folding and entropic chain formation (Tompa, 2005) (Figure 3.2). Some proteins are required in the cell only for a specific time period or for a particular function and the

presence of IDRs make these proteins prone to proteasomal degradation (van der Lee et al., 2014, Prakash et al., 2004), PTMs (Dyson and Wright, 2005, Gao and Xu, 2012), allosteric regulation (Motlagh and Hilser, 2012, Li et al., 2017) and regulated gene expression (Gsponer et al., 2008).

There are some short motifs of hydrophobic amino acids, present in IDPs, called preformed structural elements (PSE), pre-structured motifs (PSM) and molecular recognition features (MoRFs). Such regions interact with partner molecules to give rise continuously fluctuating or stable secondary structure transitions (Fuxreiter and Tompa, 2012). IDPs interactions depend upon the location of disordered regions in the structure. Proteins may adopt "Flanking interaction model" when the disordered region is present in either C- or N-terminal of protein and remains disordered even in the bound state (Tompa and Fuxreiter, 2008). Sometimes, the disordered regions are present between the folded segments and bind to two different regions on interactive partner following "Clamp interaction model" (Tompa and Fuxreiter, 2008). In case of "Random model" disordered regions are randomly present in the protein structure and may form more than one interaction sites (Sharma et al., 2015)

3.3 IDPs/IDRs and DNA interactions

The charge on IDPs has a high impact on the formation of the compact structure due to electrostatic interactions (Mao et al., 2013, Muller-Spath et al., 2010). The intrinsic flexibility of IDPs exists even when the protein is compact. This flexibility remains when a protein complex is formed. IDPs may adopt different mechanisms to interact with DNA. Quick DNA binding, with slow kinetics, shows the "fly-casting mechanism" (Shoemaker et al., 2000). In this mechanism, IDRs interact with DNA weakly, (Fuxreiter et al., 2011) while specificity/affinity can be increased by disorder to order transitions (Figure 3.3) (Spolar and Record, 1994). Another mechanism involves the sliding, hopping, and jumping of IDPs by binding to various non-specific DNA sequences. Such a mechanism is called the Monkey-Bar mechanism (Figure 3.3) (Vuzman et al., 2010).

The net positive charge, on IDPs plays an important role in DNA binding. The positively charged amino acid patches are present predominantly in the DNA-binding regions of IDPs. This also shows evolution might have adopted the positive charged IDRs for non-specific DNA binding (Figure 3.4) (Uversky, 2002). The affinity, specificity and kinetics of IDRs towards DNA are strongly affected by the composition and net charge in the disordered regions. The presence of IDRs in

the DNA binding proteins enhances the protein affinity towards DNA, facilitating linear as well as 3D diffusion, sliding, hopping and searching a particular region on DNA. In the absence of DNA, IDRs may perturb protein folding, or the propensity to form ordered structures, by electrostatic repulsions among positively charged amino acids.

3.4 PTMs in disordered regions

Many functions of IDPs, e.g. signaling and non-specific DNA-protein interactions, are modulated by electrostatic interactions and the charge distribution responsible for this function is greatly influenced by PTMs. Phosphorylation, acetylation and ubiquitination are commonly found PTMs in disordered regions (Iakoucheva et al., 2004). Phosphorylation and acetylation are usually involved in perturbing the non-specific protein-DNA interactions by reducing the protein binding affinity towards DNA (Borg et al., 2007) *e.g.,* IDRs in histones reduce their affinity towards DNA upon acetylation (Wells and McClendon, 2007). Phosphorylation leads to the reduction of the net positive charge of the DNA-binding region while acetylation alters the amino acid chemical properties. Short-lived protein-DNA interactions are greatly affected by Serine/Threonine phosphorylation in a reversible regulatory process (Amoutzias et al., 2012, Collins et al., 2008, Gnad et al., 2009, Gsponer et al., 2008, Iakoucheva et al., 2004, Marchini et al., 2011). In *Mycobacterium tuberculosis,* S/T phosphorylation plays a vital role in pathogenesis. The IDRs were found to be more prone to phosphorylation than other regions in *M. tuberculosis* (Cousin et al., 2013, Pereira et al., 2011). IDRs make approximately 10% of all the proteome in *M. tuberculosis*.

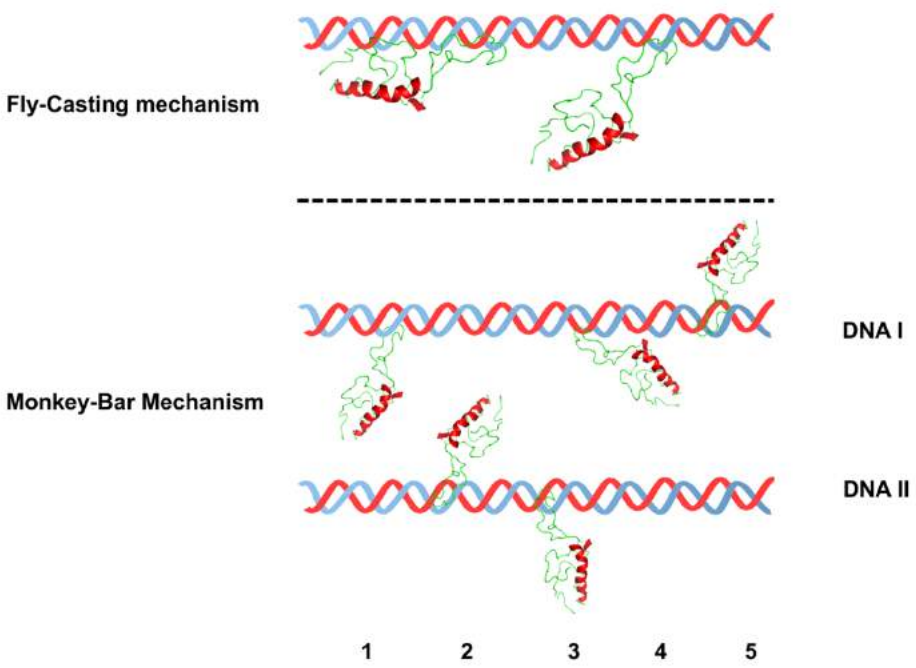


Figure 3.3 Schematic illustration of Fly-Casting and Monkey-Bar mechanisms: In the Fly-Casting mechanism, the positively charged IDRs bind to DNA with more affinity than other regions and may show protein folding upon binding. The binding efficacy determines the extent of sliding on DNA (Vuzman et al., 2010). One or more IDR segments may interact with DNA simultaneously. The monkey-bar mechanism involves hopping or slipping of DNA binding proteins on different DNA segments via disordered tails or linker regions. The 1-5 different protein-DNA binding states show sliding and hopping of the protein on DNAI and DNAII. Adapted from (Staby et al., 2017)

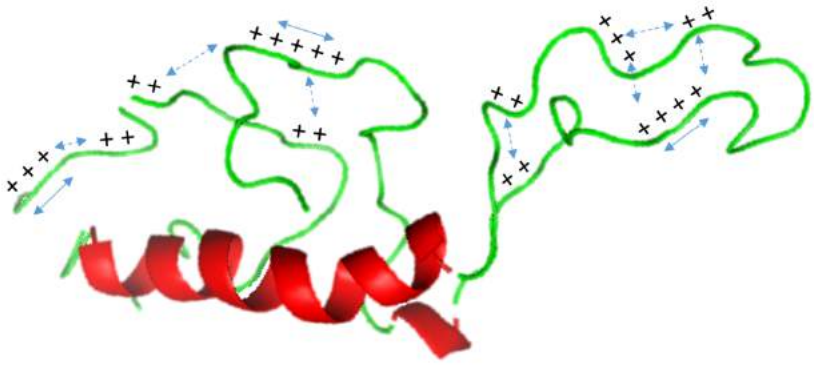


Figure 3.4 A schematic illustration of the distribution of positive charge in IDRs: The disordered regions are greatly influenced by the presence and distribution of predominantly positively charged amino acids. The + sign shows the positively charged segments while arrows show the electrostatic interactions/repulsions that affect protein folding. Adapted from (Staby et al., 2017)

3.5 Prediction of disorder in proteins

Several predictors, based on different methods and parameters, have been developed to predict the extent of disorder in proteins (Table 3.1) (Uversky et al., 2000, Romero et al., 2006, Melamud and Moult, 2003).

3.5.1 Sequence-based predictors

IUpred (Dosztanyi et al., 2005), CH-plot (Huang et al., 2014), GlobPlot (Linding et al., 2003) use physicochemical parameters like charge and hydrophobicity of the protein. IUpred computes the pairwise interactions energy for each residue and assesses the absence of stable associations among amino acids (Dosztanyi et al., 2005). GlobPlot determines globular domains in the proteins. ANCHOR (Dosztanyi et al., 2009) determines disordered regions involved in binding to other partners. ANCHOR is based on inter-chain interaction probabilities of amino acids. It considers the length, state and physicochemical properties of amino acids involved in the potential interactions.

3.5.2 Machine learning predictors

MLAs (Machine learning algorithms) serve as the best disorder predictors and are usually based on either trained "Neural Networks" (NN) or a "Support Vector Machine" (SVM) (Table 3.1). The databases usually used for such MLA training include PDB datasets from X-ray crystal structures or NMR ensembles.

PONDR® (Predictors Of Naturally Disordered Regions, http://www.pondr.com) uses a family of neural network based predictors that have been trained on PDB datasets (Romero et al., 1997, Romero et al., 2001). PONDR®-VSL2 employs neural networks that predict both long (>30 amino acid) and short disordered segments. It also used the PSI-Pred server to verify and compare the disorder predictions with predictions of protein secondary structure. PONDR®-VLXT uses three different neural networks to predict longer disordered regions. The two neural networks have been trained using X-ray characterized Terminal (XT) disordered regions. The third neural network is trained on the disorder regions present randomly in protein structures, with the exception of C- and N-terminals (Romero et al., 2001). PONDR®-VL3 predicts long disordered segments using ten neural network predictions. CaN-XT uses two neural networks, that have been trained using disordered segments from 13 known human calcineurin proteins. RONN (Yang et al., 2005) is another predictor that uses neural networks trained on only known disordered proteins. It works well with longer disorder regions. Spritz (Walsh et al., 2012) is

another neural network -based disorder predictor. It uses Bi-directional Recursive Neural networks (BRNN's) trained on short X-Ray, longer Disprot datasets (from the DISPROT database, www.disprot.org/) and NMR mobile regions (from Mobi Server, http://protein.bio.unipd.it/ mobi/). DISOPRED (Ward et al., 2004) is another neural network-based server and has been trained using PSI-BLAST and support vector machine (SVM) algorithms.

Table 3.1: List of some disorder prediction servers (Deng et al., 2012)

No	Name	Method	Webserver
Physicochemical sequence-based Predictors			
1	IUPred	*Ab initio*	http://iupred.enzim.hu/
2	SLIDER	*Ab initio*	http://biomine.cs.vcu.edu/servers/SLIDER/
Structure-based Predictors			
3	DISOclust	Clustering	http://www.reading.ac.uk/bioinf/DISOclust/
4	IntFOLD-DR	Clustering	http://www.reading.ac.uk/bioinf/IntFOLD/
Machine Learning Predictors			
	Neural Networks (NN)		
5	PONDR (Also provide CH plot)	*Ab initio*	http://www.pondr.com/
6	DISpro	*Ab initio*	http://www.ics.uci.edu/~baldig/dispro.html
7	RONN	*Ab initio*	http://www.strubi.ox.ac.uk/RONN
8	DisEMBL	*Ab initio*	http://dis.embl.de
	Support Vector Machines (SVM)		
9	Spritz	*Ab initio*	http://distill.ucd.ie/spritz/
10	DISOPRED	*Ab initio*	http://bioinf.cs.ucl.ac.uk/disopred
11	PrDOS	PSSM Based	http://prdos.hgc.jp/cgi-bin/top.cgi
Meta Servers			
12	metaPrDOS	Meta method	http://prdos.hgc.jp/cgi-bin/meta/top.cgi
13	MFDp2	Meta method	http://biomine.cs.vcu.edu/servers/MFDp2/
14	Metadisorder	Meta method	http://iimcb.genesilico.pl/metadisorder/predict_protein_disorder_by_metadisorder.html
15	PONDR-FIT (combines Iupred with PONDR)	NN based meta method	http://disorder.compbio.iupui.edu/metapredictor.php

3.5.3 Meta servers

Meta-servers combine predictions from different predictors and filter the consensus disorder probability. Mets disorder (Kozlowski and Bujnicki, 2012) is a meta server that combines predictions from DisEMBL, DISOPRED2, DISpro, Globplot, iPDA, IUPred, Pdisorder, Poodle-s, Poodle-l, PrDOS, Spritz, and RONN. The PONDR-FIT (Xue et al., 2010) is another meta-predictor, that uses predictions from PONDR family predictors and combines them for a consensus prediction.

3.6 Biophysical techniques for characterization of IDPs

Several biophysical techniques can be used for the determination of disorder in the protein structure. These techniques include spectroscopic methods like Circular Dichroism (CD), Raman Spectroscopy, Optical Rotary Dispersion and Fourier-Transform Infrared Spectroscopy (FTIR). The other direct or indirect methods that determine the mode and extent of disorder include Size Exclusion Chromatography (SEC), Analytical Ultracentrifugation (AUC), Small Angle X-Ray Scattering (SAXS), Small Angle Neutron Scattering (SANS), Viscometry, Electron Microscopy (EM), Atomic Force Microscopy (AFM), Single Molecule Fluorescence Resonance Energy Transfer (SM-FRET), Fluorescence Anisotropy, Tryptophan Quenching and Nuclear Magnetic Resonance (NMR).

3.7 IncC NTD – An N-terminal region in the IncC1 protein

The RK2 plasmid, as discussed in Chapter 1, has been isolated from various hosts as a multi antibiotic resistance plasmid (Meyer and Helinski, 1977). RK2 plasmid depends on ParABS partitioning system for its stable transfer to daughter bacterial cells. RK2 expresses two ParA protein homologs, IncC1 and IncC2 proteins. Both proteins are ATPases. The two proteins are expressed simultaneously, with different lengths, from the same operon and reading frame but with different start codons (Batt et al., 2009).

The presence of two ATPases (IncC1, 364 a.a and IncC2 259 a.a) in the RK2 plasmid partitioning system (Figure 3.5) is not unique. The highly related low copy number plasmids like RK2, RP4, RP1, R68 (Burkardt et al., 1979) also possess two ParA proteins. The presence of IncC NTD (N-terminal 105 a.a.) makes IncC1 different from the IncC2 protein. It has been observed that IncC1 and IncC2 proteins are found in different ratios in different hosts and presence of one IncC protein was sufficient for plasmid partitioning (Williams et al., 1998,

Bignell et al., 1999, Siddique and Figurski, 2002). The N-terminal of IncC1 (IncC NTD) might play an important role in plasmid partitioning when IncC1 is dominant ParA protein in some hosts. Therefore, it is important to characterize the N-terminal domain of IncC1 to elucidate its function in plasmid partitioning.

Previously, a helix-turn-helix (HTH) motif was proposed in IncC NTD and the domain itself was anticipated to bind DNA, but IncC NTD was found to show little similarity to the HTH region present in other ParA proteins (Batt et al., 2009). This chapter describes the initial secondary structure for IncC NTD protein, predicted using several web servers. Then, it details how the IncC NTD protein was overexpressed and purified; followed by the characterisation of the purified protein using various techniques including circular dichroism, Small Angle X-ray scattering, thermal shift assay and EMSA. The next chapter describes the determination of the backbone structure of IncC NTD using carbon-detected NMR (Chapter 4).

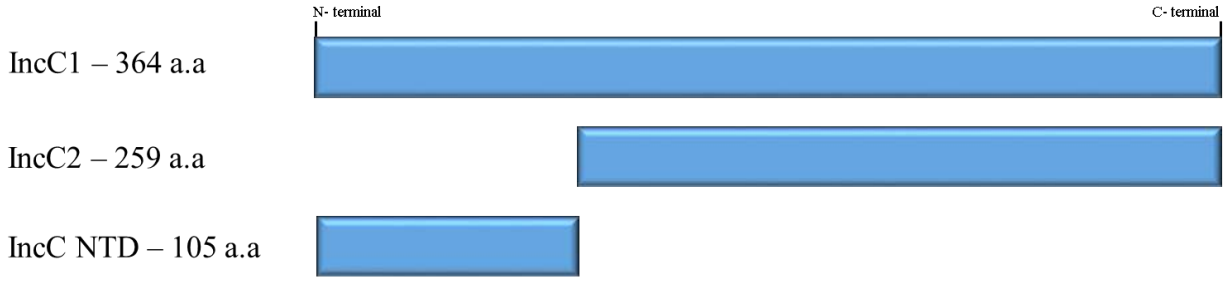


Figure 3.5 ParA proteins of the RK2 plasmid. RK2 plasmid overexpresses two ParA proteins, IncC1 and IncC2. The two proteins are encoded from different start codons in the same operon. IncC1 and IncC2 have different lengths *i.e.* IncC1 (364 a.a) and IncC2 (259 a.a). The N-terminal domain (IncC NTD) consists of 105 amino acids which are not present in the IncC2 protein.

Materials and Methods

IncC NTD has been characterized using computational tools as well as by biophysical methods described here or in Chapter 2.

3.8 Computational studies

The amino acid sequence analysis for IncC NTD (Appendix 3) was performed using ProtParam tool and SEDNTERP ver. 1.09. The ProtParam tool from the Expasy server (http://web.expasy.org/protparam/) provides physiochemical properties of protein sequence and calculates isoelectric point (pI), theoretical molecular mass, extinction coefficient (280 nm), calculated half-life, instability index and hydrophobicity. SEDNTREP (http://www.jphilo.mailway.com/) is a freeware that provides information about charge, extinction coefficients (250-320 nm), molecular mass, solvent density (ρ), viscosity (η) and partial specific volume ($\overline{V}$).

3.8.1 Disorder prediction

Disorder in IncC NTD structure was predicted using various web servers. The web servers used include PONDR predictors (VSL2, VLXT), DisEMBL, RONN, DisPro, IUPRED and PONDR-FIT (Table 3.2). IUPRED predicts on the basis of the amino acid properties, while PONDR, RONN and DisEMBL use machine learning predictors (Table 3.2). PONDR-FIT is a meta-server that combines the prediction from different servers for confident protein disorder prediction.

3.8.2 Secondary structure prediction

The secondary structure of IncC NTD was predicted using Jpred4 (Drozdetskiy et al., 2015), PsiPred (Jones, 1999), SOPMA (Geourjon and Deleage, 1995), YASPIN (Lin et al., 2005), SSpro 4.0 (Magnan and Baldi, 2014), HMMSTR (Bystroff and Shao, 2002), SABLE (Adamczak et al., 2004) and PSSpred (Yan et al., 2013) servers.

JPred4 uses neural networks (NN) and bootstrap networks. It provides secondary structure predictions as well as propensities for solvent accessible and coiled-coil regions. PsiPred is also a neural network-based secondary structure prediction server that used PSSM (position-specific scoring matrices) obtained using PSI-BLAST. SOPMA is a self-optimized prediction method that predicts using the sets of aligned proteins. YASPIN is an HNN (Hidden Neural Network)-based secondary structure prediction server that uses the PSI-BLAST algorithm

and PSSM (position-specific site matrix) for its predictions. The sspro4 prediction was obtained from the SCRATCH software suite that uses neural networks based on PDB datasets. HMMSTR is based on Hidden Markov models (HMMs) for protein structure prediction. SABLE (Solvent AccessiBiLitiEs) and PSSpred (Protein Secondary Structure PREDiction) uses two neural networks to get a protein structure prediction.

3.8.3 Comparison of IncC NTD with N-terminal regions of the ParA proteins

Protein sequence logos were created for the N-terminal regions of different plasmid ParA proteins using the WebLogo3 (http://weblogo.threeplusone.com/). N-terminal sequences (100 a.a), from 16 ParA proteins (Appendix 4), were used to create sequence logos.

3.9 IncC NTD purification and characterization

IncC NTD was overexpressed and purified to homogeneity as described in Chapter 2. Chromatographic techniques, including Ni-Affinity chromatography and size exclusion chromatography, were used to purify IncC NTD. The protein was characterized using SEC-MALLS, AUC, CD, SAXS, MS and NMR. Protein-DNA interactions were studied using EMSA, TSA and NMR.

3.10 Removal of the His tag from IncC NTD by Thrombin digestion

To elucidate any change in the NTD protein structure due to the His_6 tag, Thrombin was used to remove the His_6 tag. The His_6 tag was incorporated in the protein to allow easy purification by Ni-NTA chromatography His tag. It consists of the following 23 amino acids.

MGSSHHHHHH SSGLVPRGSHSEF

Thrombin recognizes the Leu-Val-Pro-Arg-Gly-Ser (LVPRGS) site and cleaves the peptide bond between Arg and Gly. Thrombin (GE Healthcare), with high cleavage efficiency, was used to remove the His_6 tag (1 Unit ~1 µl cleaving 200 µg of protein). IncC NTD (1-5 mg) was subjected to Thrombin cleavage at room temperature and in Thrombin buffer (20 mM Tris-HCl, 100 mM NaCl, pH 7.5). Initially, Thrombin cleavage assays were performed to optimize conditions. IncC NTD (10 µL, 10 mg/mL) was incubated with 1 U of thrombin. Different aliquots of IncC NTD (10 µL, 2 mg/mL) were taken at various time intervals and Thrombin digestion was stopped by adding Benzamidine (1 mg/mL final concentration). IncC NTD samples were then analysed by 15 % Tricine SDS PAGE. After His_6 tag cleavage (1U thrombin for 1 mg of protein, 2 h reaction at room temperature), Thrombin and the cleaved His_6 tag were removed using SEC purification.

Results

IncC NTD has a high content of charged amino acids and the three-disorder promoting amino acids, Glycine, Serine and Proline, represent 35 % of the total sequence (19.5 %, 11.7%, and 3.9% respectively). The Protparam server (www.expasy.org/protpram) showed that IncC NTD is highly positively charged (5.1 charges at pH 7.5 and a pI of 10.1). Inc NTD has an extinction coefficient of 6,990 M^{-1} cm^{-1} (at 280 nm).

3.11 Computational analysis of IncC NTD

3.11.1 Protein structure disorder prediction

All disorder prediction servers show that IncC NTD is a predominantly disordered protein (Figure 3.6a). The predictions from VSL2 and IUPRED servers show that the regions between amino acids 1 to 10 (MGVIHEETAY) and 40 to 50 (SRWEATGDVRNVAGT) may be partially structured (Figure 3.6a). PONDR-FIT Metaserver provides a consensus disorder prediction using predictions from various servers (Figure 3.6a) and shows a similar prediction. The charge-hydropathy plot (CH Plot) shows IncC NTD to be among the disordered proteins when different proteins, known to be either natively folded or unfolded, are plotted on the basis of absolute mean charge *vs* mean scaled hydropathy (Figure 3.6b).

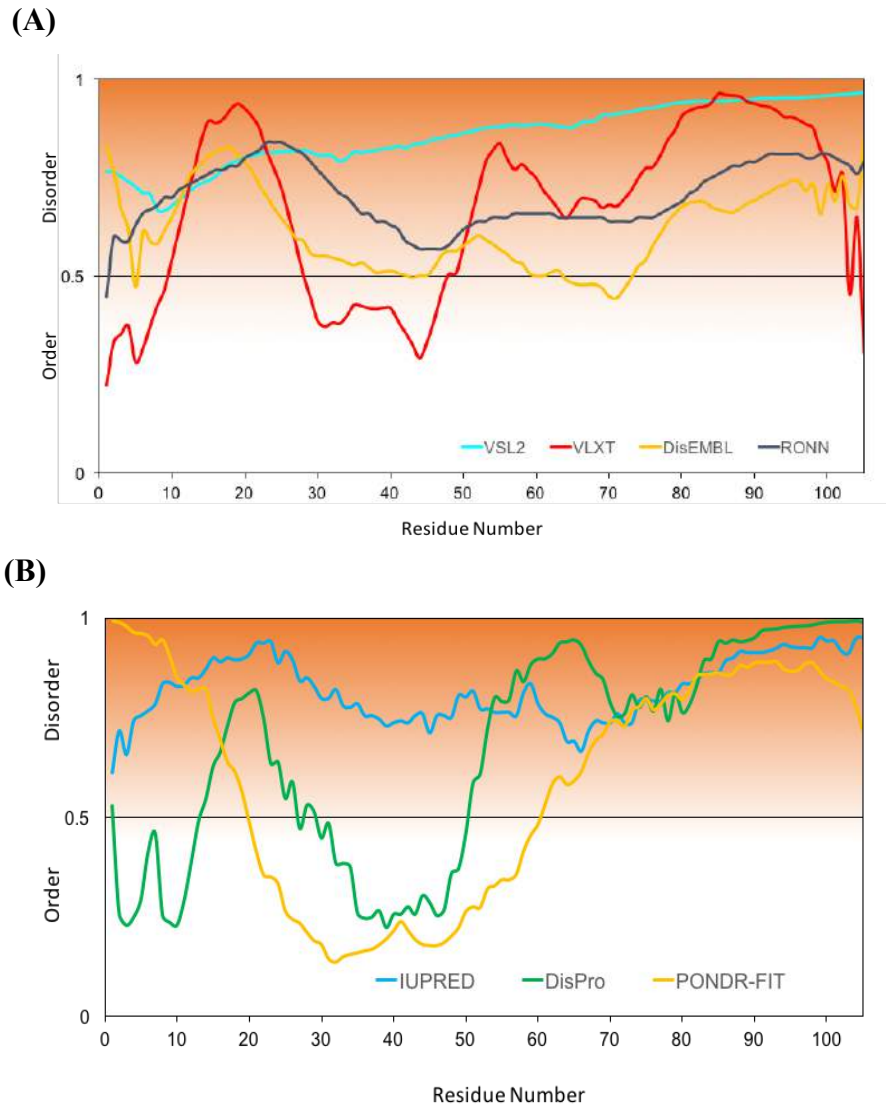


Figure 3.6a Disorder prediction for IncC NTD using various servers. (A) The VSL2 (blue), DisEMBL (yellow) and RONN (grey) server show that IncC NTD is a largely disordered protein. Although, VLXT (red) server shows that amino acid 30-45 may have some secondary structure elements. **(B)** IUPRED (Blue) server has shown IncC NTD is completely disordered but DisPro (green) and PONDR-FIT (yellow) servers have shown a similar prediction as VLXT that amino acid residues, 35-50 may have some structure propensity.

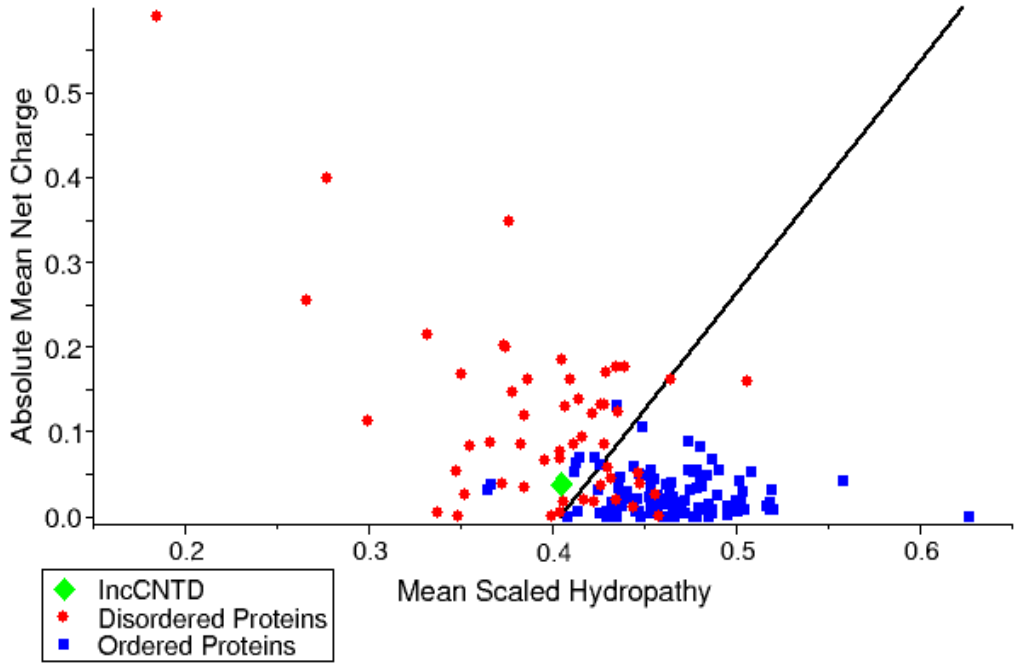


Figure 3.6b Charge-Hydropathy plot for IncC NTD. IncC NTD is a highly positively charged protein and positioned among the natively unfolded proteins in the Charge-Hydropathy plot (green diamond). The CH plot was obtained by plotting the mean scaled hydropathy vs absolute mean net charges of known folded and unfolded proteins. Folded/ordered proteins are shown in blue while unfolded/disordered proteins are shown in red. The black line shows the approximate boundary between folded and unfolded proteins.

3.11.2 Secondary structure prediction

The IncC NTD protein sequence was submitted to various servers (JPRED4, PsiPRED, SOMPA, YASPIN, SSPRO4, HMMSTR, SABLE and PASSpred) for secondary structure prediction. IncC NTD was predicted to be largely disordered by all servers. PsiPred, SOPMA, YASPIN, HMMSTR, and PSSpred showed formation of a short helix between residues 6-10 while, in the same region, Jpred4 and SSPRO4, showed the presence of a β-strand (Figure 3.7). In the region between 43-50 residues, a β-strand has been predicted by Jpred4, SOPMA, YASPIN, SSPRO4 and HMMSTR servers.

Although, from the computational predictions, IncC NTD has been found to be a largely disordered protein, the presence of some transient secondary structures or stable conformations upon interaction with other proteins or DNA cannot be discounted.

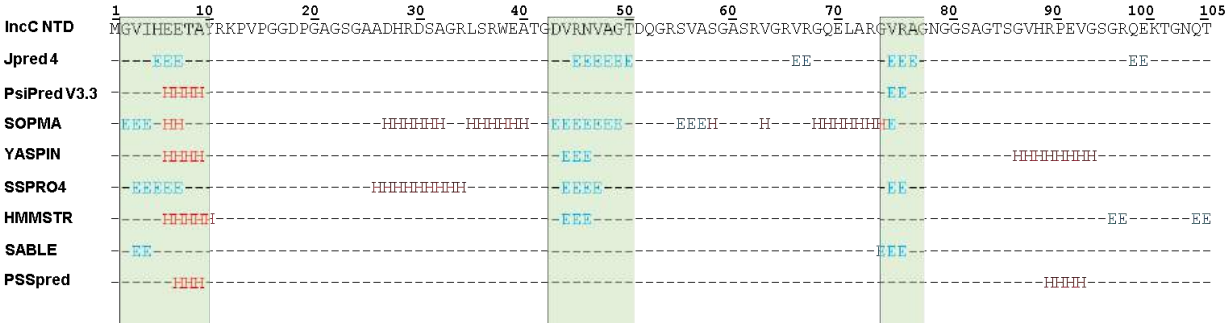


Figure 3.7 IncC NTD secondary structure prediction. All server predictions show that IncC NTD is largely disordered. Some secondary structure elements were predicted in IncC NTD. The light green shading represents the regions with predicted secondary structure elements. Although a consensus for secondary structure cannot be made from the predictions, the predictions show the presence of some α-helices or β- strands in IncC NTD.

3.11.3 Comparison of IncC NTD with N-terminal regions of ParA proteins

Sequence logos for IncC NTD and 16 closely related ParA N-terminal regions (top16 BLAST results) show that the region is not very identical although, Arginine, Serine, Histidine and Glycine residues were highly conserved at certain positions. Interestingly, a Cysteine residue (at position 45), found in many ParA N-terminal regions is not present in IncC NTD. The amino acid stretch HEETA at positions 5-9 is the most conserved among the protein sequences. This region has also been predicted to have some secondary structure in IncC NTD (Figure 3.7).

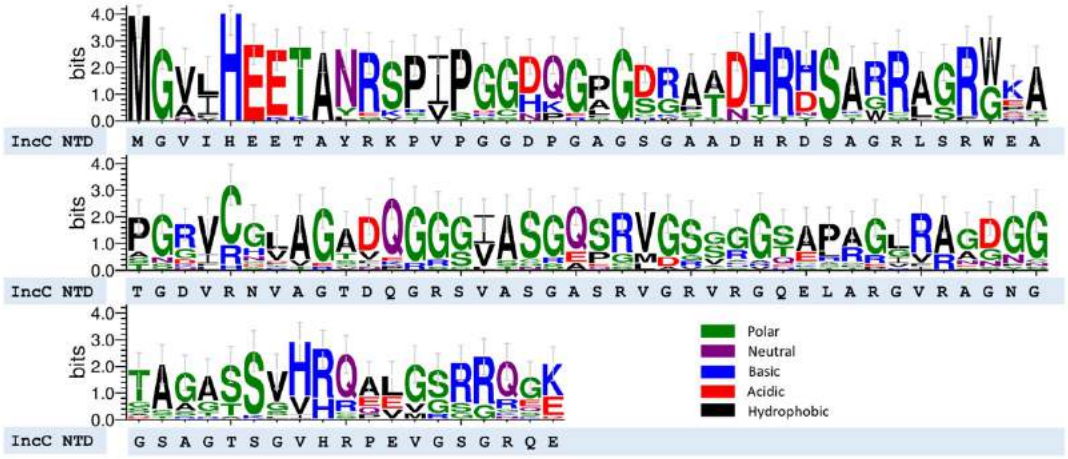


Figure 3.8 Sequence logos for N-terminal 100 residues of ParA proteins and IncC NTD. Sequence logos for IncC NTD and 16 other ParA N-terminals show that amino acids are well conserved in the region and a short amino acid stretch HEETA was found highly conserved. Amino acids are coloured by their chemical properties and error bars have been based on Bayesian calculations errors.

3.12 *incC-NTD* transformation in *E. coli* strains

The *incC-NTD* gene was previously cloned into a modified pET21a vector by Professor Thomas group (Batt, 2008, Appendix 1). 20 ng of the plasmid DNA was transformed in the *E. Coli* DH5α competent cells as described in methods (Chapter 2). The plasmid DNA was purified from *E. coli* DH5α and analysed for size and purity using agarose gel electrophoresis. The DNA concentration was determined using DNA Fluorimeter and Nanodrop. DNA sequencing showed no mutations in the *incC-NTD* gene. The pET28a-*incC_ntd* vector was then transformed into the *E. coli* BL21 (λDE3) strain to overexpress IncC NTD under the control of a T7 promoter. The transformants were isolated using Kan^R selection.

3.13 IncC NTD overexpression and purification

IncC NTD was expressed from an *incC_ntd* vector, using *E. coli* strain BL21 (λDE3). IncC NTD expression was studied at two different temperatures (37°C and 25°C), after induction at mid-log phase. IncC NTD-His_6 expressed well at 4 hr after induction at 37°C and, after this, no increase in bacterial growth (OD_{600}) was observed. Indeed, protein expression was lower in the overnight sample (Figure 3.9a). IncC NTD-His_6 expressed well overnight at 25°C. A good protein expression was observed, but the expression was not as high as after 4 hr incubation at 37°C (Figure 3.9b).

3.14 Analysis of IncC NTD solubility after protein expression

To check the solubility of the overexpressed IncC NTD-His_6 protein, the bacterial cell pellets from overnight cultures were resuspended in sodium phosphate lysis buffer (NaP). The bacterial cells were then lysed by sonication, followed by centrifugation at 32,000x *g* for 20 minutes to obtain clear cell lysate (soluble fraction, S) and insoluble pellet (P). Both fractions were suspended in SDS-containing stacking gel buffer (Appendix 2) and analysed by 12% SDS PAGE. IncC NTD-His_6 Protein was found in the soluble fractions obtained from protein expression at 37°C and 25°C (Figure 3.10). The temperature 37°C was chosen as the optimum temperature to get soluble protein overexpression in four hours.

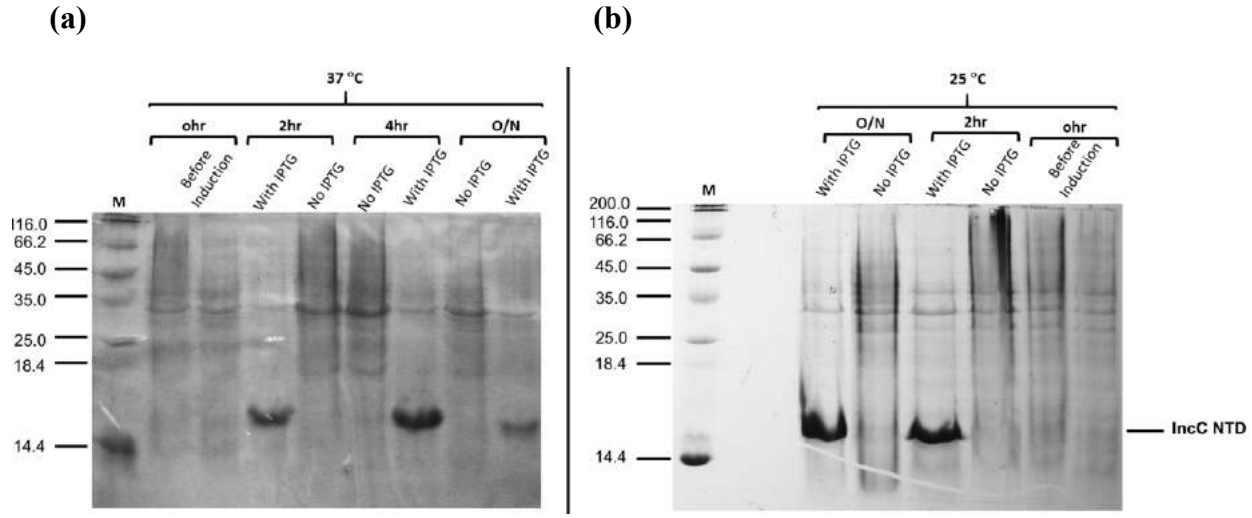


Figure 3.9 IncC NTD overexpression at different temperatures. (a) IncC NTD protein expression was observed at 37°C and monitored after 0, 2, 4 hr and overnight. The protein expression after 4 hr was found higher than overnight expression, maybe some of the protein was lost due to proteolysis. **(b)** IncC NTD protein expression at 25°C was obtained and monitored after 0, 2 hr and overnight. The protein expression in 2 hr and overnight samples was found almost similar.

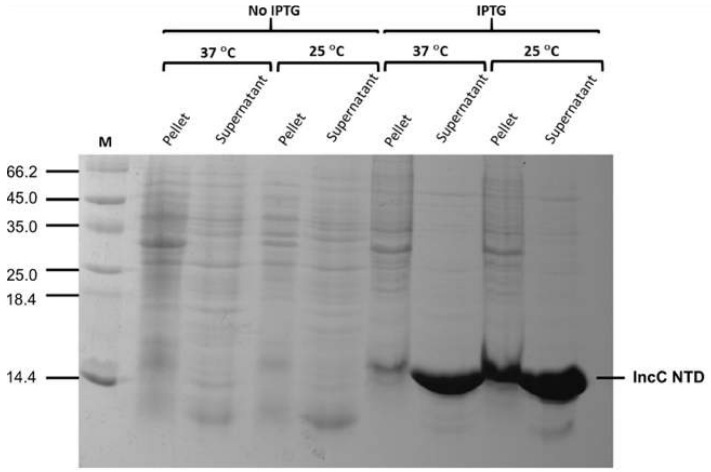


Figure 3.10 IncC NTD protein solubility at 25°C and 37°C. The expression of IncC NTD was observed in soluble (supernatant) and insoluble (pellet) fractions. The protein was found in the soluble fraction at both temperatures but the soluble protein expression was higher in at 37°C.

On SDS PAGE, IncC NTD-His_6 shows an apparent molecular weight of ~14 kDa (Figure 3.10). That is more than its theoretical molecular weight (13.14 kDa) determined from the amino acid sequence (Appendix 3). This is consistent with it being an IDP. IDPs show lower mobility while running in SDS PAGE, in comparison to molecular weight determined by Mass spectrometry, due to their abnormal charge distribution, amino acid composition and shape, (Tompa, 2002). Some reports show that the molecular weight of IDPs estimated by SDS PAGE

is ~1.2-1.8 times higher than their calculated molecular weight (Tompa, 2002).

3.15 Mass Spectrometry analysis of IncC NTD

The accurate subunit mass of the purified IncC NTD-His_6 protein was determined using mass spectrometry (ESI-MS), as described in Chapter 2 (section 2.29). The molecular mass of IncC NTD-His_6 from mass spectrometry was determined to be 13016.5 Da (Figure 3.11) while the molecular mass from the amino acid sequence was calculated as 13149 Da (Appendix 1). The two molecular masses might differ be due to loss of C- or N- terminal amino acid.

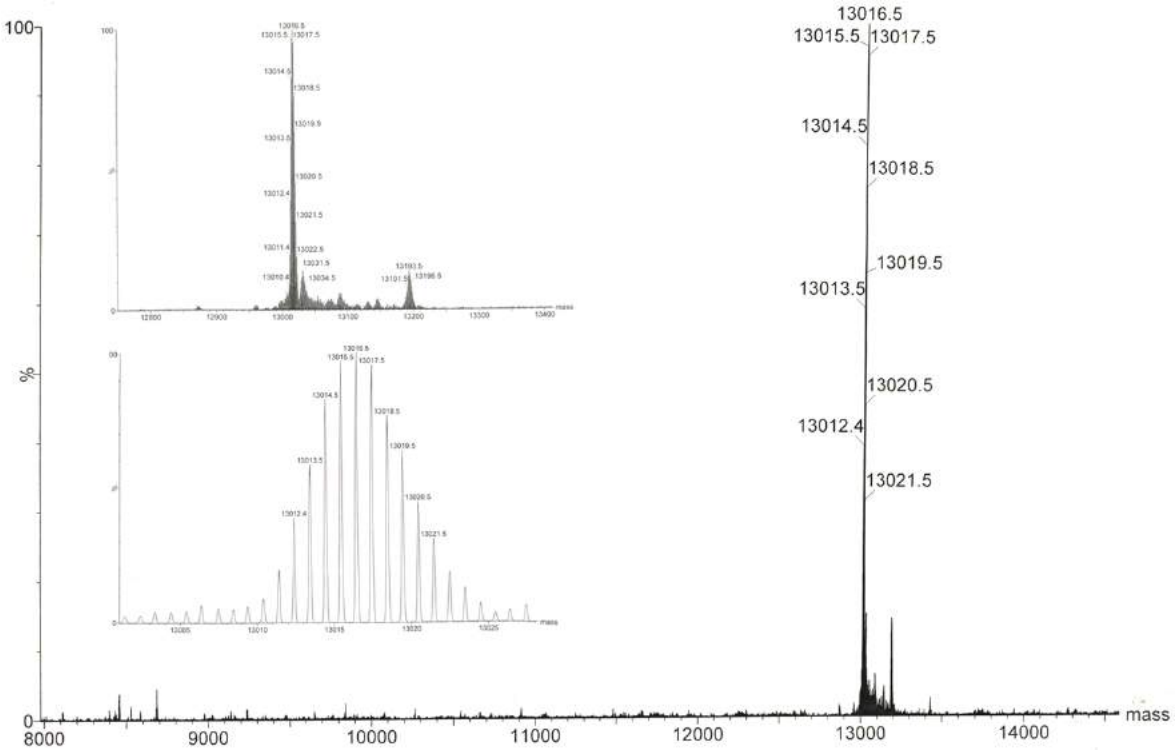


Figure 3.11 Mass spectrometry analysis of IncC NTD. The IncC NTD-His_6 protein was subjected to ESI-MS to calculate the accurate molecular mass. The purified protein was given to the Mass Spectrometry facility in School of Chemistry, University of Birmingham. IncC NTD-His_6 showed the molecular mass of 13016.10 with ESI-MS while molecular mass calculated from the amino acid sequence was found to be 13140 Da. The inset shows the monoisotopic masses for IncC NTD. The two molecular masses might differ be due to loss of C- or N- terminal amino acid.

3.16 IncC NTD protein purification (Large-scale production)

IncC NTD protein was expressed with a His_6 tag at the N-terminal and purified, from 2 L bacterial culture, by affinity column chromatography using Ni-NTA Affinity chromatography and size exclusion chromatography.

3.16.1 Ni-NTA chromatography

The clear lysate was obtained as described in Chapter 2 and loaded onto a Ni- NTA column. The column was washed to remove non-specifically bound proteins. The fractions with bound IncC NTD-His_6 protein were eluted with a linear gradient of imidazole and all fractions were assayed using Bradford reagent. The SDS PAGE shows that IncC NTD purified as a single band (Figure 3.12).

3.16.2 Size exclusion chromatography (SEC)

To analyse the dispersity of IncC NTD protein in the solution as well as to further purify it, the protein (2.5 mL, 2-10 mg/mL) was loaded onto a size exclusion Superdex G75 column. A single peak of absorbance at 280 nm was observed in the chromatogram, eluting at 71 mL (Figure 3.13). The SEC column was calibrated with proteins of known molecular mass and a standard molecular weight curve, as well as a standard Stokes radius curve was obtained. IncC NTD was eluted at 71 mL that corresponds to a spherical protein of ~29 kDa, while a monomeric IncC NTD-His_6 has a molecular mass of 13.1 kDa. This suggests that the protein is either elongated in shape or forms dimers. IDPs show anomalous behaviour in size exclusion chromatography due to their elongated shape and elute earlier from SEC column and show a higher apparent molecular mass than their calculated mass. The Stokes radius (R_s) for IncC NTD was found to be 23.1 Å (Figure 3.13).

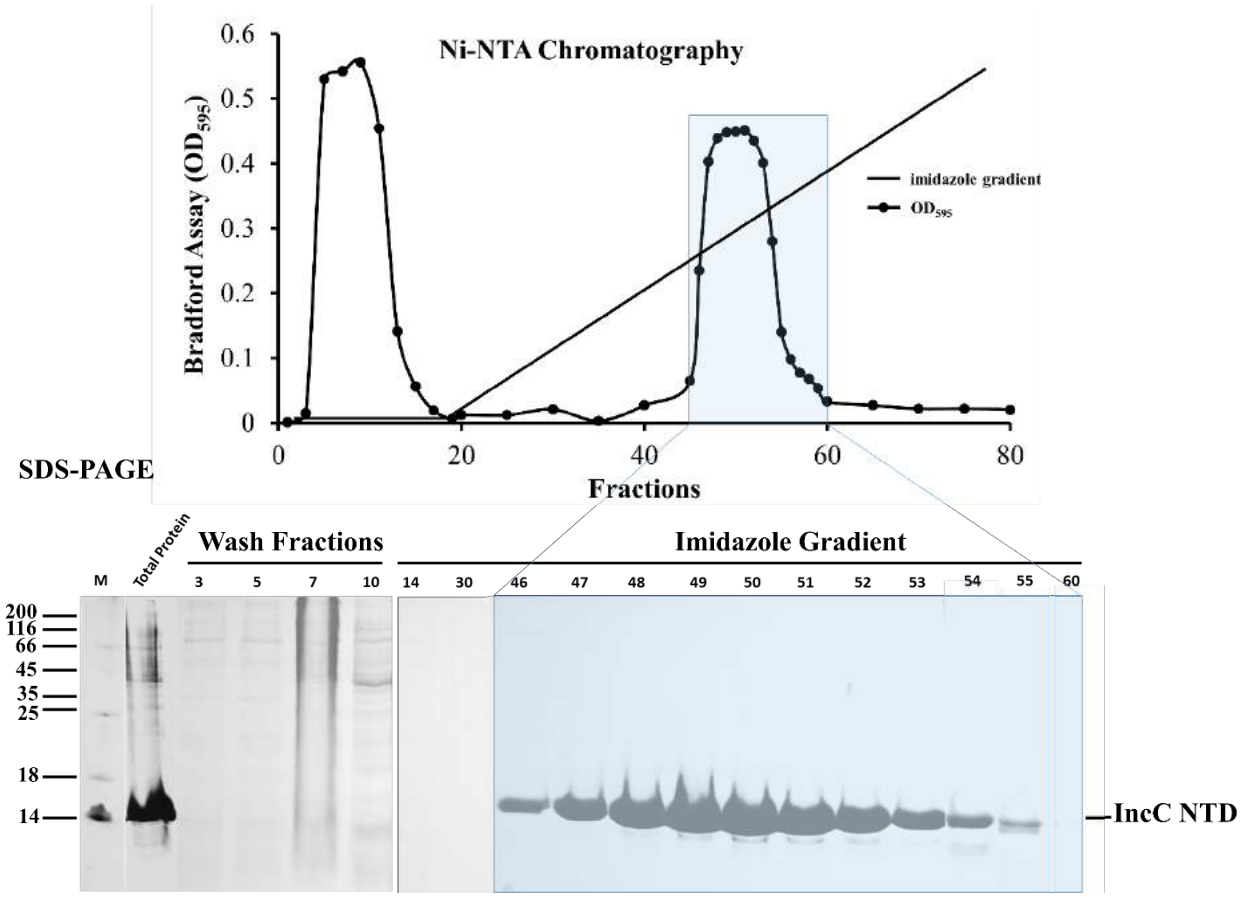


Figure 3.12 IncC NTD-His$_6$ protein purification by Ni-NTA chromatography. IncC NTD was expressed using BL21 (λDE3) cells at 37°C using 1 mM IPTG as the inducer. The clear lysate (onto) was loaded on a Ni-NTA column (15 mL).). The column was washed using Tris buffer pH 7.5 (20 mM Tris, 300 mM NaCl and 20 mM imidazole) and non-specifically bound proteins were removed. A linear imidazole gradient (20-300 mM) was used for protein elution (using the same buffer). 1.5 mL fractions were collected and proteins were assayed using Bradford reagent. The graph shows the results from the Bradford assay of the eluted protein fractions. The protein fractions 45-60 contained high protein concentrations and were analysed by SDS PAGE, that shows IncC NTD purified as a single band. The IncC NTD- His$_6$ protein showed a molecular mass of ~ 14 kDa that was little higher than molecular mass from calculated amino acid sequence (13.1 kDa).

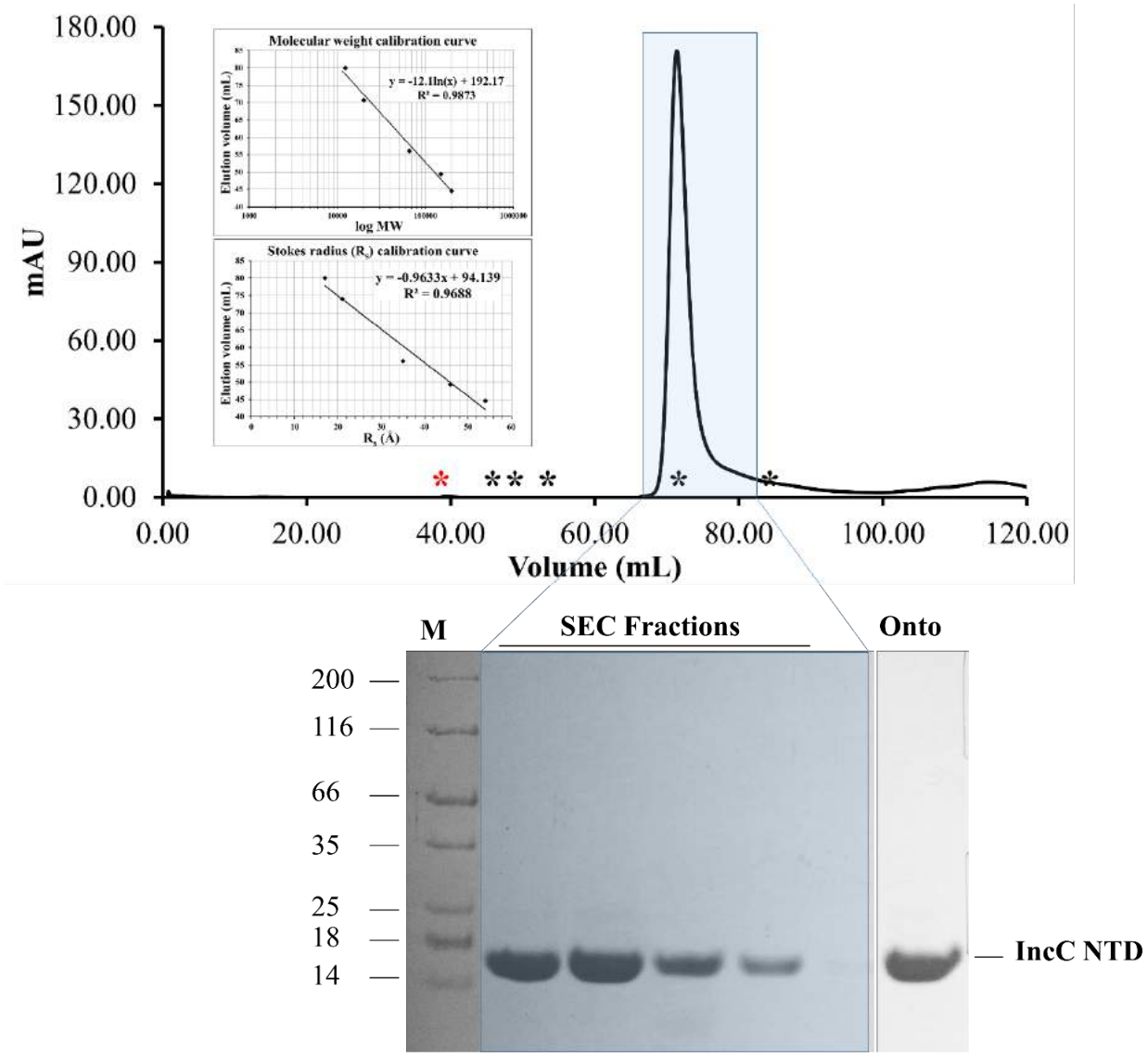


Figure 3.13 Size exclusion chromatography of IncC NTD. IncC NTD-His_6 was purified using a Superdex G75 column. The purified protein fractions from Ni-NTA chromatography were loaded onto a SEC column (120 mL, HiLoad 16/600 Superdex 75 column). Tris buffer pH 7.5 (20 mM Tris, 150 mM NaCl and 0.1 mM EDTA) was used to elute 2.5 mL fractions. Protein fractions with high absorbance (mAU_{280}) were loaded on SDS PAGE. IncC NTD was observed as a single band on SDS PAGE. The standard molecular weight and Stokes radius curves were obtained using the standard proteins (β-Amylase, Mol. wt. 200 kDa, R_s 54 Å, Alcohol Dehydrogenase mol. wt. 147 kDa, R_s 46 Å, Bovine Serum Albumin, Mol. wt. 66 kDa, R_s 35 Å, Carbonic Anhydrase, Mol. wt. 29 kDa, R_s 21 Å and Cytochrome C, Mol. wt. 12.4, R_s 17 Å (La Verde et al., 2017). IncC NTD eluted at 71 mL that corresponds to a molecular weight of ~29 kDa for a globular protein. A Stokes radius of 23.1 Å was obtained for IncC NTD. The protein either has an elongated conformation or forms dimers.

3.16.3 Ultrafiltration

The SEC purified fractions were concentrated using ultrafiltration (3 kDa cut off size membrane) and Vivaspin centrifugal concentrators (Millipore, MWCO 5,000 Da). A protein concentration more than 20 mg/mL was easily obtained. Most of the IDPs are prone to aggregation at high concentrations but IncC NTD was found to be soluble at high concentration (~50 mg/mL).

3.17 SEC-MALLS

Size exclusion chromatography with multi-angle laser light scattering (SEC-MALLS) showed that IncC NTD-His_6 consists of a single, monodisperse species. The SEC-MALLS was performed to measure the Refractive Index, light scattering at multiple angles and QELS (Quasi-Elastic Light Scattering) of the sample, and analysed using Astra® ver 6. The molecular weight was determined using the light scattering and refractive index. QELS (Quasi-Elastic Light Scattering) was used to measure the hydrodynamic radius (R_h), which was found to be 23.6± 3.6 Å, while the molecular mass calculated from MALLS was 12.11±1.5 kDa (Figure 3.14), showing that the protein is monomeric under these conditions.

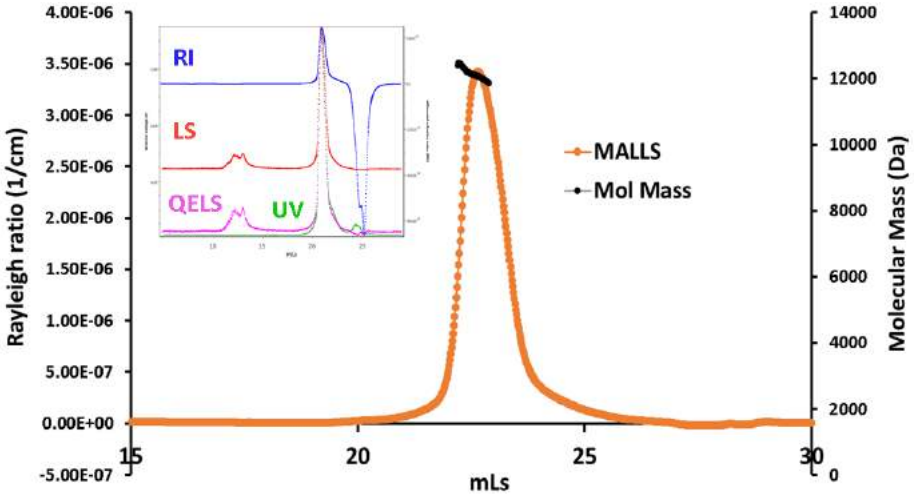


Figure 3.14 SEC-MALLS of IncC NTD protein. Protein sample (1-2 mg/mL) was loaded onto a Superdex 200 SEC column (flow rate 0.0.7 mL/min, Tris pH 7.5, 300 mM NaCl and 0.1 mM EDTA) and the eluted fractions were monitored for Refractive Index and MALLS. SEC-MALLS data analysis show that IncC NTD-His_6 is a monomeric protein with molecular mass of 12.1±1.5 kDa. The QELS was used to determine hydrodynamic radius (R_h), which was found to be 23.6 Å. The Rayleigh ratio (plotted) is proportional to the scattering intensity, RI is Refractive Index, LS is light scattering and QELS is Quasi-Elastic Light Scattering that measures time-dependent fluctuations in the scattered light. UV absorbance was monitored at 280 nm. Astra ver. 6 was used for data analysis.

3.18 His_6 Tag cleavage from IncC NTD

As stated in section 3.11, the IncC NTD protein was expressed with a His_6 tag at the N-terminal. The His_6 tag has been fused to IncC NTD via a linker containing a Thrombin cleavage site (LVPR/GS). The His_6 tag was removed by thrombin cleavage. The cleavage conditions were optimized by incubating the reaction mixtures for 1-18 hr and His_6 cleavage was monitored every hour for 1-6 hr and overnight. SDS PAGE showed that in 3 hr, the His_6 tag was removed from all IncC NTD in the reaction mixture (Figure 3.15a). The cleaved His_6 tag and Thrombin were removed from IncC NTD by a subsequent SEC purification (3.15b).

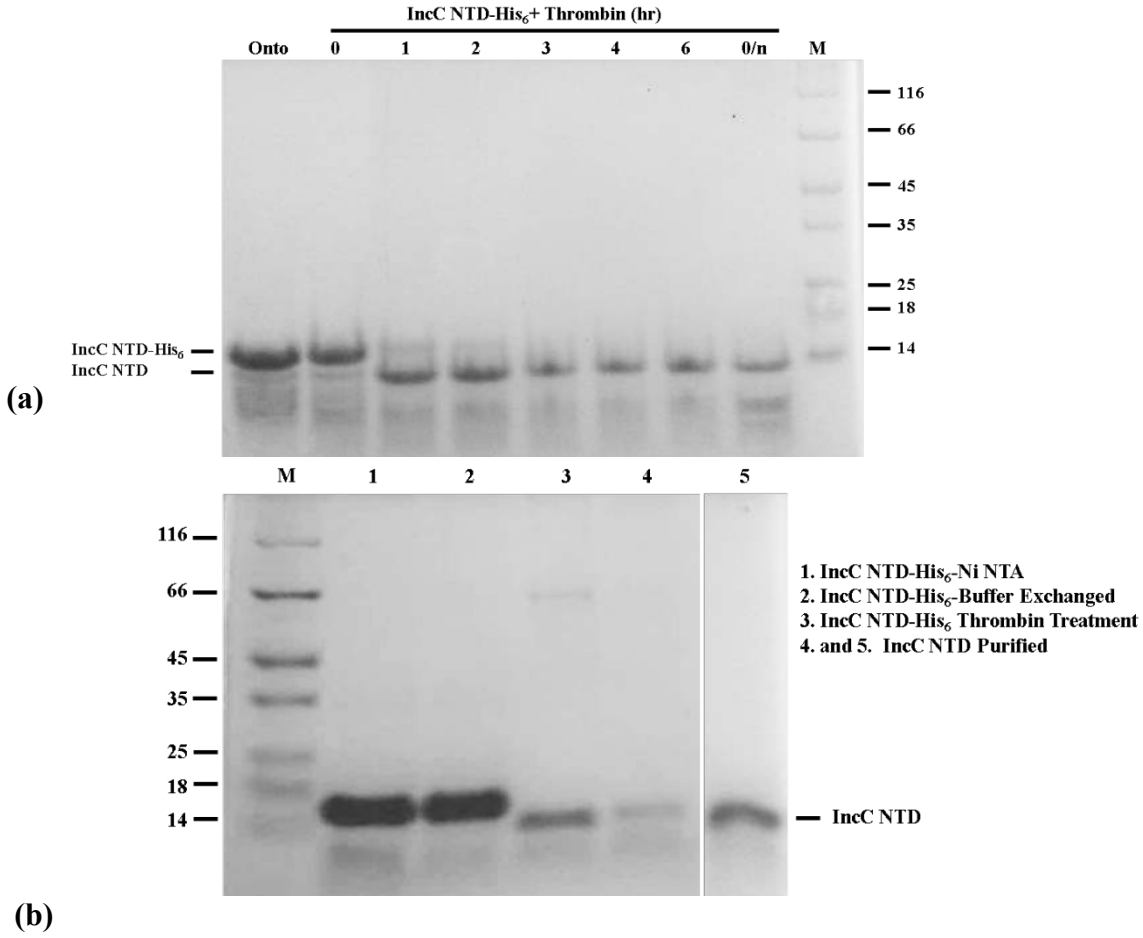


Figure 3.15 Removal of the His_6 tag from IncC NTD by Thrombin cleavage. IncC NTD (1-5 mg/mL) was treated with 1 units of Thrombin/ 2 mg of the IncC NTD protein (GE Healthcare) using the cleavage buffer (50 mM Tris pH 8.0, 150 mM NaCl, 2.5 mM $CaCl_2$). **(a)** The reaction mixture (1 mg protein + 1 U thrombin) was incubated at room temperature for 1-6 hr and overnight to optimize the thrombin cleavage. The His_6 tag for all IncC NTD present in the reaction mixture was removed in 3 hr **(b)** IncC NTD (5 mg/mL), treated with Thrombin and was purified using size exclusion chromatography to remove free His_6 tag and Thrombin. SDS PAGE shows (3, 4 lanes) His_6 tag was successfully cleaved and Thrombin, as well as free His_6 tag, was removed.

3.19 Biophysical characterization of IncC NTD Protein

3.19.1 Analytical Ultra Centrifugation (AUC)

To check the mono-dispersity of IncC NTD in solution, AUC analysis was performed. The time lapses, in recorded AUC data files, can restrain the precision of the sedimentation coefficient (Zhao et al., 2013). To overcome this problem, REDATE (Rectifying Emender of Data Acquisition Time Errors) was applied to IncC NTD velocity sedimentation data to overcome data acquisition time errors. REDATE was applied to both absorbance and interference data. Three IncC NTD samples (with protein concentration 1.20, 0.84 and 0.25 mg/mL) were used for sedimentation velocity AUC analysis. Scans at 280 nm were obtained every 5 minutes and show the progress of the protein as it sediments down the sample cell. The absorbance data were fitted to the Lamm equations and residuals were calculated (Figure 3.16). The analysed AUC data show that the sedimentation coefficient of IncC NTD increases slightly with protein concentration. The sedimentation coefficients for the three different IncC NTD concentrations (0.25, 0.84 and 1.20 mg/mL) are 0.963 S, 1.053 S, and 1.116 S (Figure 3.16). The sedimentation coefficients correspond to the molecular masses of 10.9 kDa, 11.5 kDa and 12.3 kDa while the expected molecular mass of the monomeric IncC NTD protein is 13.1 kDa (Table 3.2).

Table 3.2 The c(s) distribution, f/f_o ratios and mol. wt. obtained from AUC for IncC NTD

Sample	Concentration	c(S)[a]	Sw, 20, w[b]	f/f_o[r]	Mol. wt.[d]	RMSD
1	**0.25 mg/mL**	**0.963**	**1.036±0.018**	**1.53**	**10.9 kDa**	**0.0085**
2	**0.84 mg/mL**	**1.053**	**1.133±0.012**	**1.71**	**11.5 kDa**	**0.0165**
3	**1.20 mg/mL**	**1.116**	**1.195±0.014**	**1.81**	**12.30 kDa**	**0.0209**

a Sedimentation coefficient
b Sedimentation coefficient at 20 °C
c Frictional ratio
d Molecular weight calculated from sedimentation coefficient

The changes in the c(S) of different IncC NTD samples show that either there was some aggregation of IncC NTD at a higher concentration or the protein forms dimers (Figure 3.16). The frictional ratios (f/f_o) for IncC NTD are higher than f/f_o for globular proteins. In IDPs f/f_o ratios are usually higher than 1.25 (a typical frictional ratio for globular proteins). The spherical radius (R_s) of IncC NTD was calculated to be 13 Å with a diffusion rate (D) of 1.64 nm^2/sec.

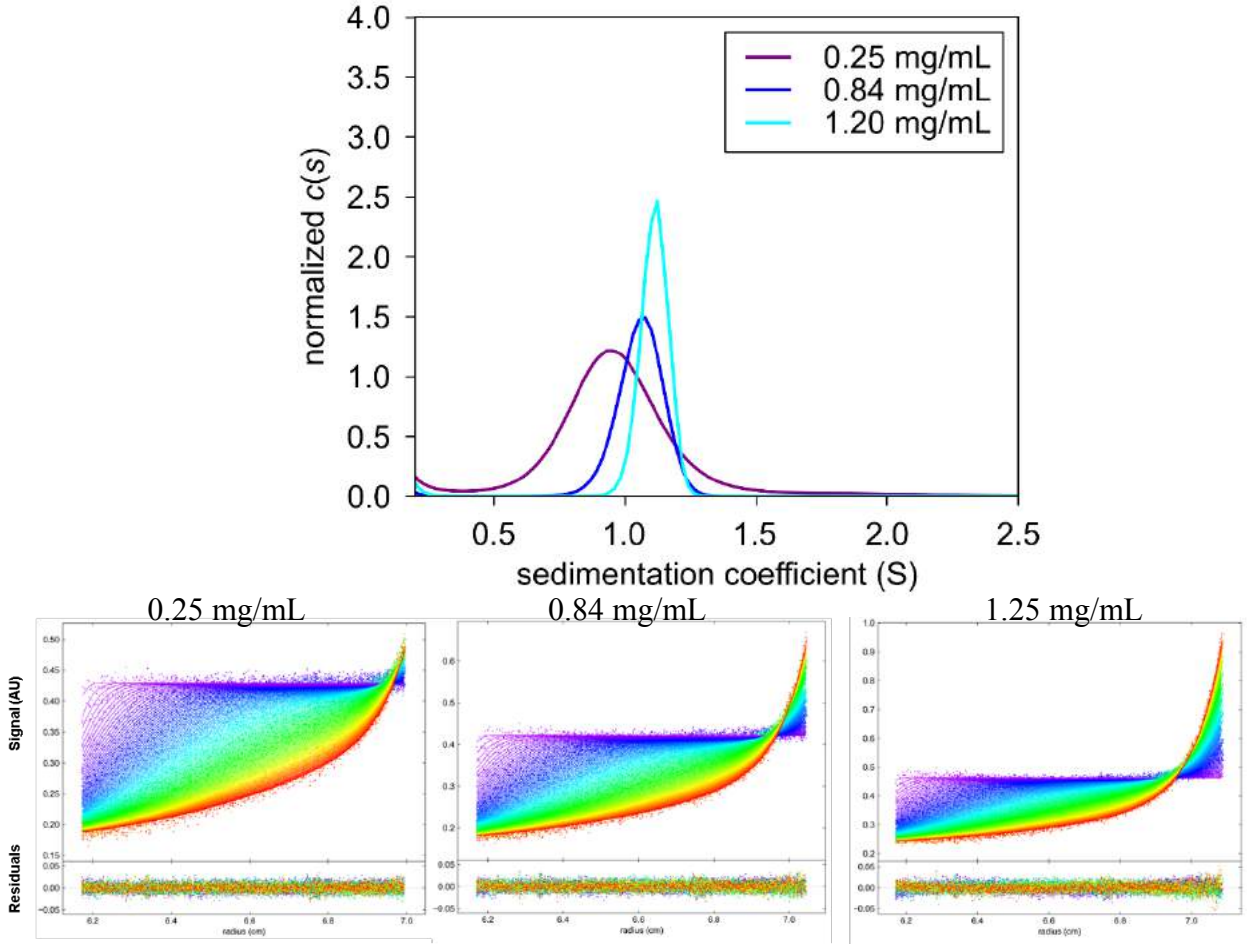


Figure 3.16 Analytical ultracentrifugation (AUC) analysis of IncC NTD. Three different samples of IncC NTD with different concentrations (0.25, 0.84 and 1.2 mg/mL) were used in AUC analysis. The velocity sedimentation run was performed for 16 hr by collecting scans every 8 minutes. The SEDFIT software was used to get the best fit of data. The upper and bottom meniscus used were in the range of 6.10-6.15 and 7.14-7.17 respectively. The frictional ratio (f/f_o) was calculated as 1.53, 1.71 and 1.81 for the three samples. The residuals plot shows the quality of fit. The sedimentation distribution [S] vs c(s) was obtained using the c(s) distribution model. The molecular mass was obtained from c(M) distribution model to be 10.9, 11.5 and 12.3 kDa for the three samples. The c(s) distributions from the fitted data were exported to GUSSI and compared with each other. The c(S) distribution shifted/increased at higher concentration showing that either the protein aggregates slightly or forms a dimer.

3.19.2 SAXS for IncC NTD samples

SAXS data for IncC NTD were collected at beamline B29 (ERSF, France) using two different sample concentrations (7-13 mg/mL). SCÅTTER ver. 3.1R from BIOISIS.net was used to analyse the SAXS data. X-ray scattering of buffer solutions and protein samples was measured. The scattering data for the protein samples were corrected by subtracting the buffer scatter from the protein scattering, then a Guinier plot ($Log_{10}I(q)$ vs q^2) was obtained to get information for the radius of gyration (R_g). The R_g values for samples with different concentrations of IncC NTD samples were calculated (Table 3.3, Figure 3.17). The same data was then used to obtain normalized Kratky plots ($I_q/I_o(q.R_g)^2$ vs $q.R_g$), where the absence of any bell-shaped curve shows that IncC NTD is an intrinsically disordered protein (Figure 3.18). A standard Kratky plot is given in Chapter 2 (Section 2.26.3). P(r) distribution functions were generated to calculate the particle's maximum dimension (D_{max}) and real-space R_g. The molecular mass of IncC NTD was determined using the following equation and Bovine Serum Albumin (BSA) was used as reference sample.

$$\text{Sample Mol mass} = \frac{\text{Sample } I_{(o)}}{\text{Reference } I_{(o)}} \times \text{Reference mol mass}$$

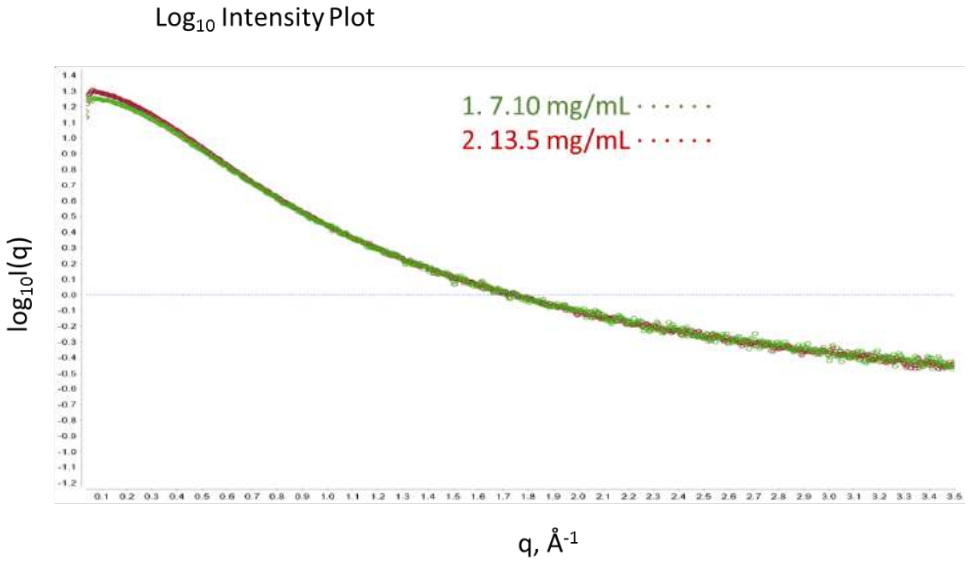
The molecular mass of IncC NTD-His_6 was calculated to be 13.5 and 14.6 kDa from the two IncC NTD samples with different concentrations (Table 3.3, Figure 3.18).

Table 3.3. SAXS analysis for IncC NTD

Sample (Conc)	R_g[a] (Å)	R_g[b] (Å)	Mol Mass[c] kDa	Porod Volume	D_{max} (nm)
1 (7.1 mg/mL)	31.4 (±0.05)	31.2 (±0.15)	13.5	28.9	10.5
2 (13.5 mg/mL)	32.1(±0.10)	32.5 (±0.25)	14.6	29.5	10.5

[a] Determined from Guinier analysis
[b] Determined from Distance Distribution
[c] Determined from zero angle intensity of scattering

(a)



(b)

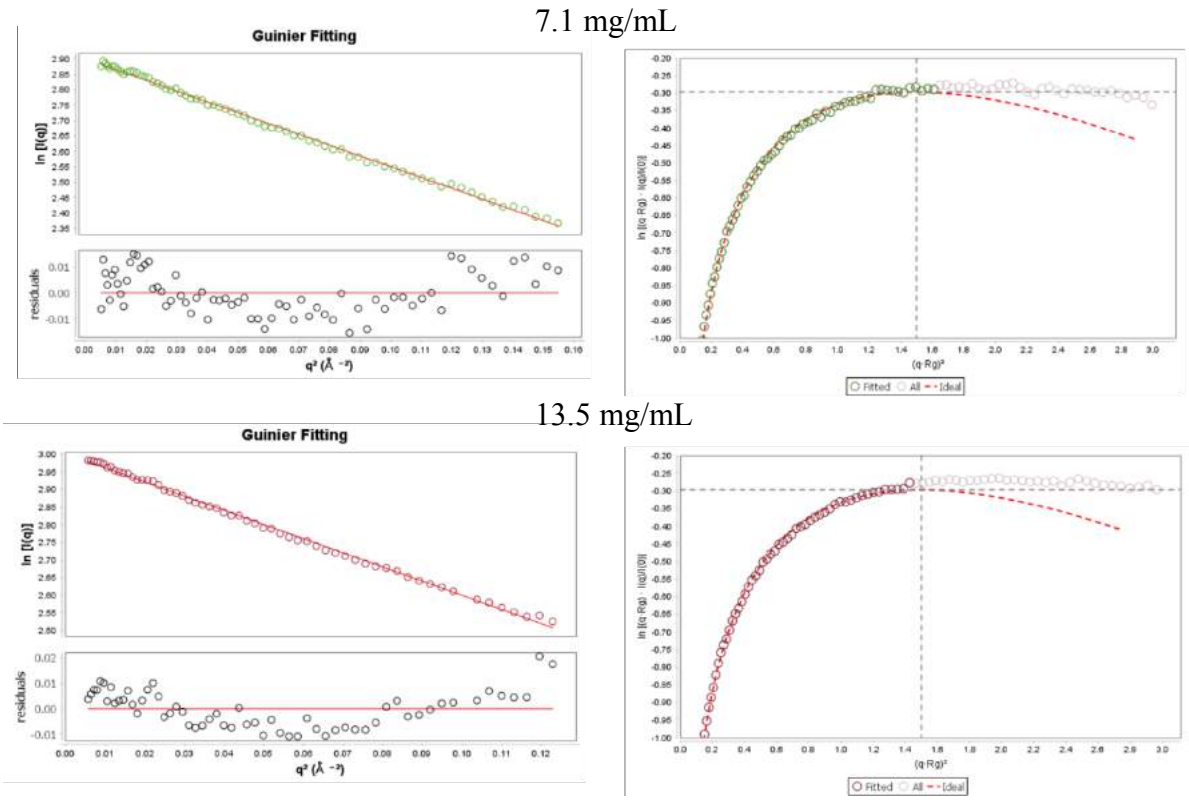


Figure 3.17 IncC NTD SAXS analysis at two different concentrations. **(a) A log intensity plot** obtained for the two IncC NTD samples, using SCÅTTER ver. 3.1R. Linear regions in the log intensity plot were used to determine the quality of data and **(b) Guinier plots.** The radius of Gyration (R_g) was determined using the slopes of the Guinier plots. For sample 1 (7.1 mg/mL), R_g was found to be 31.4±0.05 Å (an sRg range of 0.025-0.096 was used) and for sample 2 (13.1 mg/mL), R_g was found to be 32.1±0.10 Å (using sRg range of 0.55-1.30).

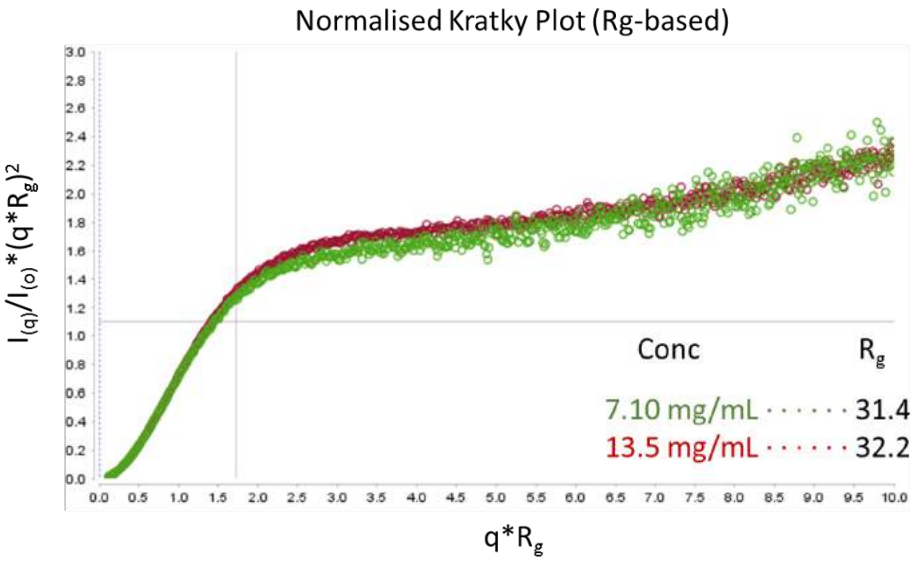
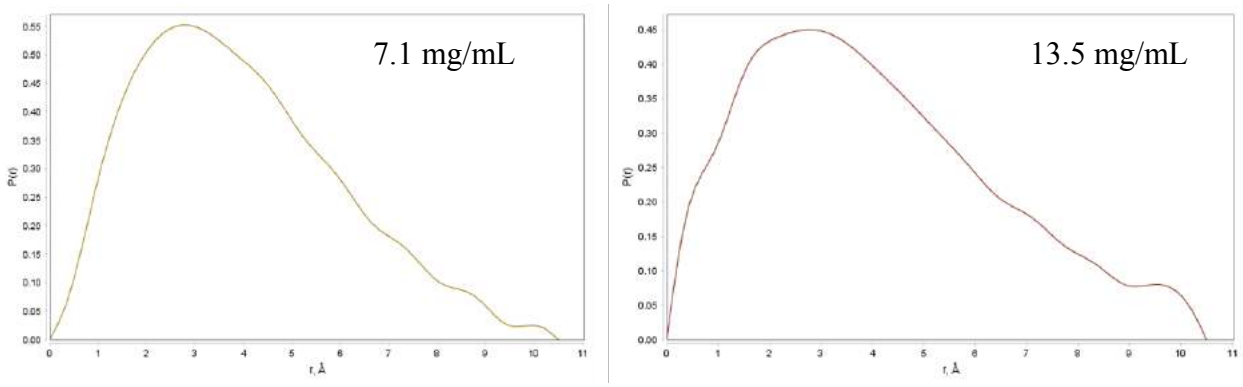


Figure 3.18 IncC NTD SAXS data analysis by Kratky plot. A Kratky plot obtained for IncC from the same data as used for Guinier plots. SCÅTTER ver. 3.1R was used to obtain Kratky plots. Kratky plots for folded, partially folded and unfolded proteins are given in chapter 2. The Kratky plots show that IncC NTD is a predominantly unfolded protein.



Conc	R_g	$I_{(o)}$	D_{max}
7.10 mg/mL · · · · · ·	31.2	15.0	10.5
13.5 mg/mL · · · · · ·	32.5	16.3	10.5

Figure 3.19 IncC NTD SAXS analysis for P(r) distribution. SAXS data were analysed using SCÅTTER ver. 3.1R to obtain the P(r) distribution, and hence D_{max} and R_g. The I_o was used to calculate IncC NTD molecular mass (Table 3.3). D_{max} for both IncC NTD samples was found to be 10.5 Å.

3.19.3 Circular Dichroism (CD) of IncC NTD

Circular Dichroism (CD) was performed to estimate the secondary structure composition of IncC NTD. CD spectra were collected using protein sample (0.2 mg/mL) in CD buffer (10 mM sodium phosphate pH 7.5, 150 mM sodium perchlorate, where sodium perchlorate was used as it shows less absorbance than NaCl). The CD spectra were obtained in the presence and absence of the His_6-tag to observe any effect of His_6 tag on the IncC NTD protein structure. IncC NTD was found to have a random coil-like structure in both cases (Figure 3.20) and thus the removal of the His_6-tag has no effect on the protein secondary structure. The secondary structure content estimated by Dichroweb using CD data, was found to be 4% α-helix and 19% β-sheet (Figure 3.20). These results are in agreement with computational prediction and SAXS calculations, where IncC NTD was found to be natively unfolded.

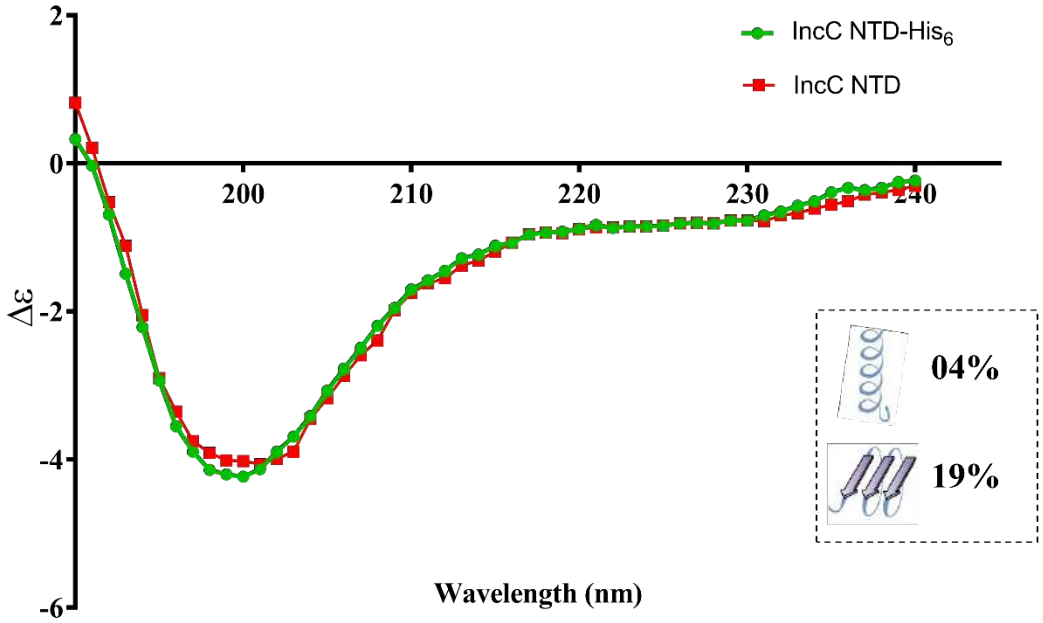
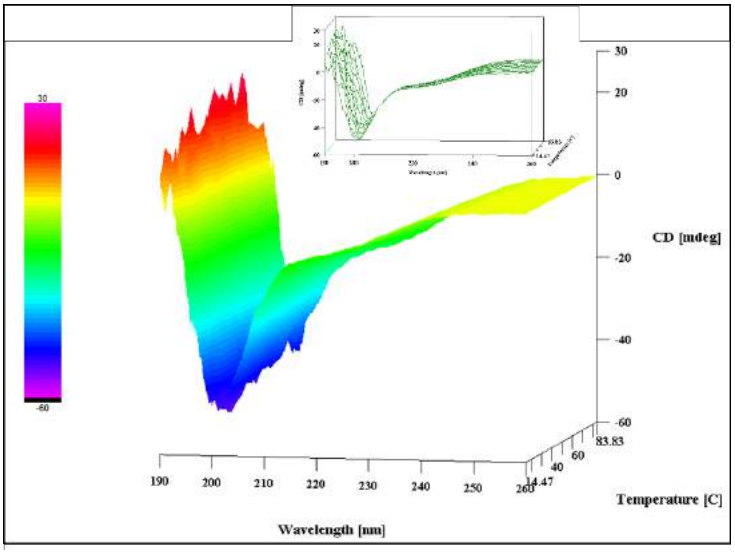


Figure 3.20 Circular dichroism (CD) spectra of IncC NTD. CD spectra were collected using JASCO J-1500 Spectrophotometer (using 0.1 mg/mL protein concentration and 1 mm CD cuvette). CD data were collected by obtaining eight scan spectra (260-185 nm) with the wavelength step of 0.2 nm, scanning speed of 25 nm/min and CD scale of 200 mdeg/1.0 dOD. The DichroWeb server and the CDSSTR (reference data set 3) program was used for CD data fitting and secondary structure prediction. The CD spectra for IncC NTD-His_6 (green) and IncC NTD without His tag (red) show that the overall protein secondary structure was not changed after removing the His_6 tag. Both spectra show IncC NTD as a random coil-like protein and the predicted secondary structures to be 4-5% α-helices and 19% β-sheets.

3.19.4 IncC NTD CD melting curves

The CD spectra of IncC NTD were used to monitor any change in its secondary structure as a function of temperature (Figure 3.21a). CD melting curves were obtained by plotting the change in CD signal at 200 nm and 220 nm wavelengths with respect to temperature (3.21b). The analysis of the CD data showed little change with temperature, because IncC NTD does not contain considerable secondary structure.

(a)



(b)

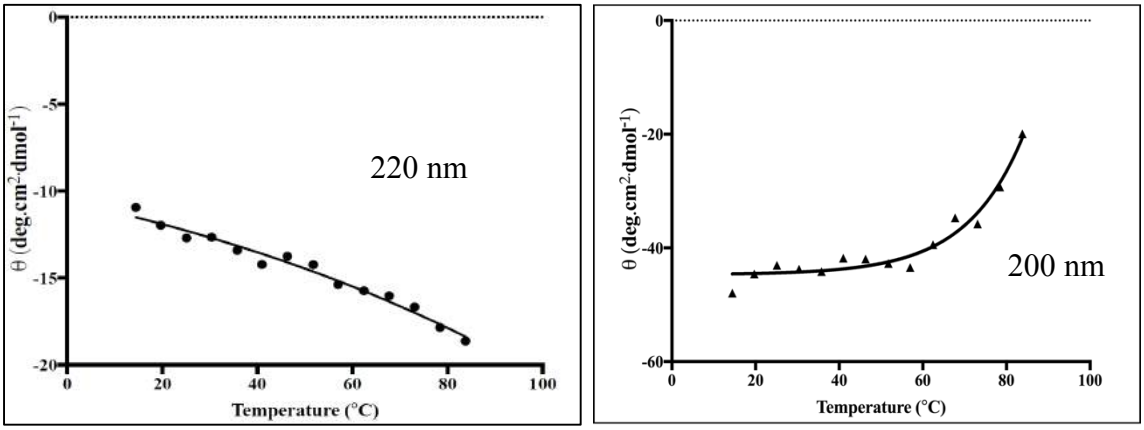


Figure 3.21 The melting curves of IncC NTD obtained using Circular dichroism (CD). CD spectra were collected using JASCO J-1500 spectrophotometer equipped with temperature Peltier. IncC NTD (0.5 mg/mL) in CD buffer (10 mM sodium phosphate pH 7.5, sodium perchlorate) was used in 1 cm cuvette to obtain CD melting curves. Three CD spectra were collected at varying temperatures (15-90°C). The temperature was increased by 0.2°C/min and CD-standard spectra were collected at an interval of 5°C. CD data were analysed using interval data analysis software in Spectra manager ver. 2.1. Three dimensional, as well as two-dimensional plots, were obtained and data was exported to the obtained melting curve at 200 nm and 220 nm. **(a)** 3D and 2D plots from interval data analysis as a function of temperature. Different colours represent different wavelength regions in the CD spectrum **(b)** The melting curve for IncC NTD obtained using the θ variation as a function of temperature increase, by monitoring 200 nm (Right panel) and 220 nm (left panel). The melting curve did not show a considerable change in CD signal at different temperatures.

3.19.5 IncC NTD Thermal Shift Assay (TSA)

IncC NTD stability, in various homemade buffers as well as in the buffers of the pH slice kit (Hampton Research) was monitored by TSA in the presence and absence of DNA (Buffer compositions are given in Appendix 6), TSA was performed as described in Chapter 2 (section 2.26.4). IncC NTD (2 mg/mL) was found to be stable over a wide range of pH (pH 5-9), although the protein was found to be most stable at neutral and slightly acidic pH (6-7) (Figure 3.22a). In the presence of DNA (10 μM, O_AO_B DNA), a biphasic TSA curve was obtained. This suggests that some of the IncC NTD protein was bound to DNA and was stabilised (Figure 3.22b).

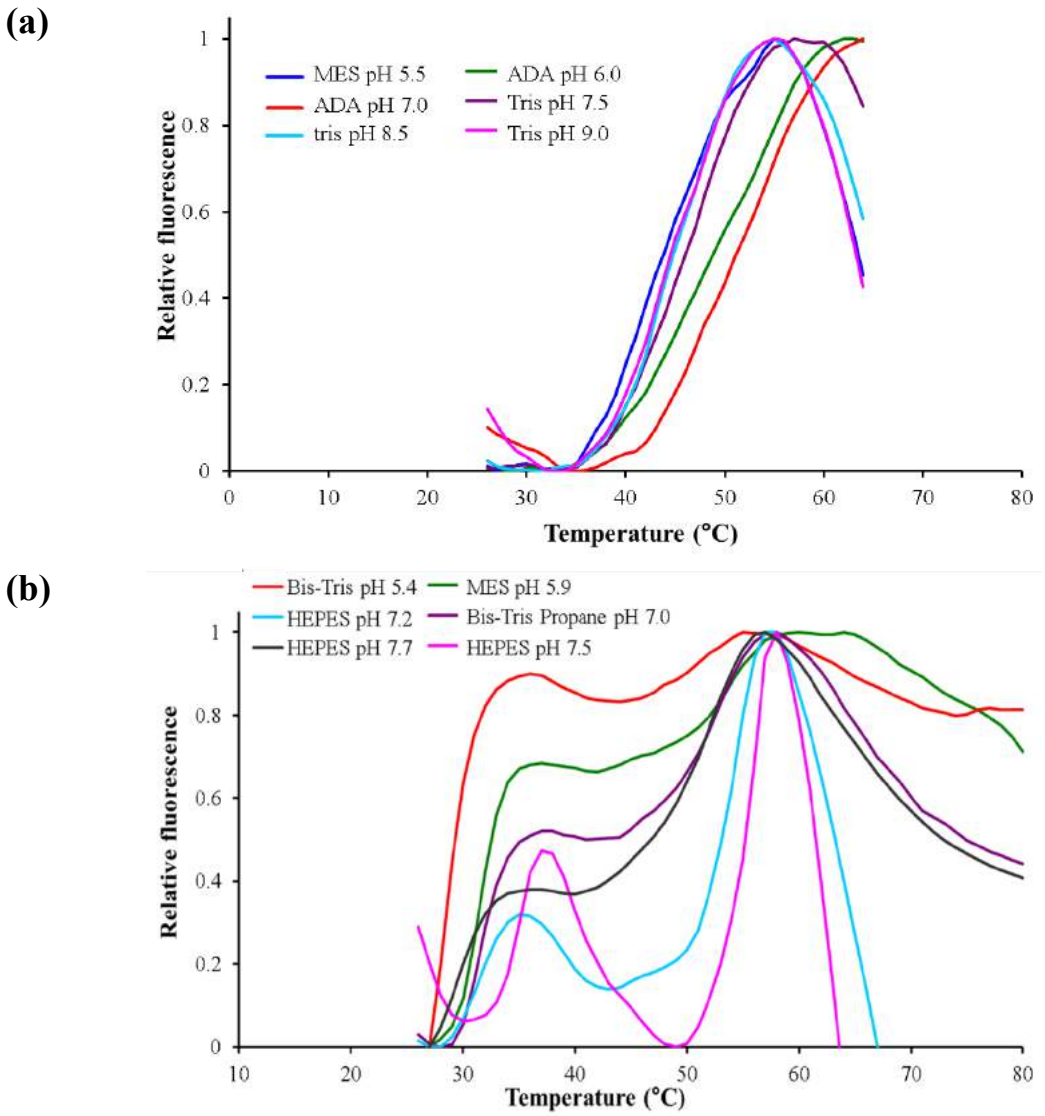


Figure 3.22 IncC NTD thermal shift assay. Without (a) and with (b) DNA The assay was performed using homemade buffers and also the pH slice kit of buffers (Hampton Research). The Mx3000P qPCR System (Agilent) was used with Sypro filter. The fluorescence was monitored after each degree rise in temperature, from 25-80°C. Data analysis was performed by exporting raw data to spreadsheets. Data were analysed using DSF Analysis Ver. 3.0 MS Excel script (ftp://ftp.sgc.ox.ac.uk/pub/biophysics). GraphPad Prism ver 7.0.4 was used to get Boltzmann fitting and melting temperature (T_m). The coloured lines represent the relative fluorescence of the dye and protein in different buffers vs temperature. The inset gives the buffer compositions and the melting temperature of the protein in each. **(a)** pH optimization for IncC NTD (2 mg/mL) show that it is stable over a wide range of pH (5-9), although at lower pH (6-7) the protein was found to be more stable **(b)** A biphasic melting curve was observed for IncC NTD (2 mg/mL) in the presence of DNA (10 μM, O_AO_B). This suggests that some of IncC NTD was bound to DNA.

3.19.6 EMSA for IncC NTD

Previous studies (Batt et al., 2009) showed that the N-terminal region of IncC1 protein enhances the DNA binding of the protein and it was suggested this region may have a DNA binding helix-turn-helix (HTH) motif. A similar motif has also been predicted in the N-terminal 100 amino acids of many other ParA proteins (Batt et al., 2009).

EMSA with PCR amplified fragments

Three DNA fragments (153 bp, 148 bp, 165 bp), PCR-amplified from the RK2 plasmid, were used to monitor IncC NTD DNA binding. The three fragments contain at least one KorB binding site were randomly chosen and used for EMSA (Figure 3.23). The fragments were radiolabelled as described in Chapter 2 (section 2.15.3), different concentrations of protein were added, and the samples loaded on a polyacrylamide gel. The gel (Figure 3.24) shows that as the protein concentration increases, the intensity of the band from the free DNA decreases, and, at the highest protein concentrations, there is a band near the top of the gel, due to bound DNA. IncC NTD bound to the DNA fragments only at very high protein concentrations (~ 400 uM). In the lanes with protein, there is a DNA smear on the gel, suggesting that IncC NTD binds DNA weakly and dissociates in the gel (Figure 3.24).

EMSA using longer DNA (PKK113 fragments) and short oligos (O_AO_B)

To obtain a longer DNA fragment, the pKK113 plasmid (pUC18, Appendix 1 containing an O_B site (TTTAGCGGGCTAAA) inserted within the MCS) was cut using the *Pvu*II restriction enzyme. Two linear fragments (291 bp and 2,364 bp) were obtained after restriction digestion. ESMA was performed using both the linear fragments as well as with only with 291 bp fragment (Chapter 2, section 2.15.4). EMSA showed IncC NTD binds to both DNA fragments non-specifically (Figure 3.25a and 3.25b), though DNA binding was very weak. DNA smears observed on the gel (Figure 3.25a) suggesting a dissociation of protein-DNA complex during electrophoresis. IncC NTD also showed binding to longer 2,364 bp DNA (Figure 3.25b), suggesting it interacts with DNA irrespective of DNA fragment length. The small O_AO_B oligonucleotide (Chapter 2, Table 2.5) was used for IncC NTD EMSA again showing the formation of a complex that the dissociates during electrophoresis (Figure 3.25c).

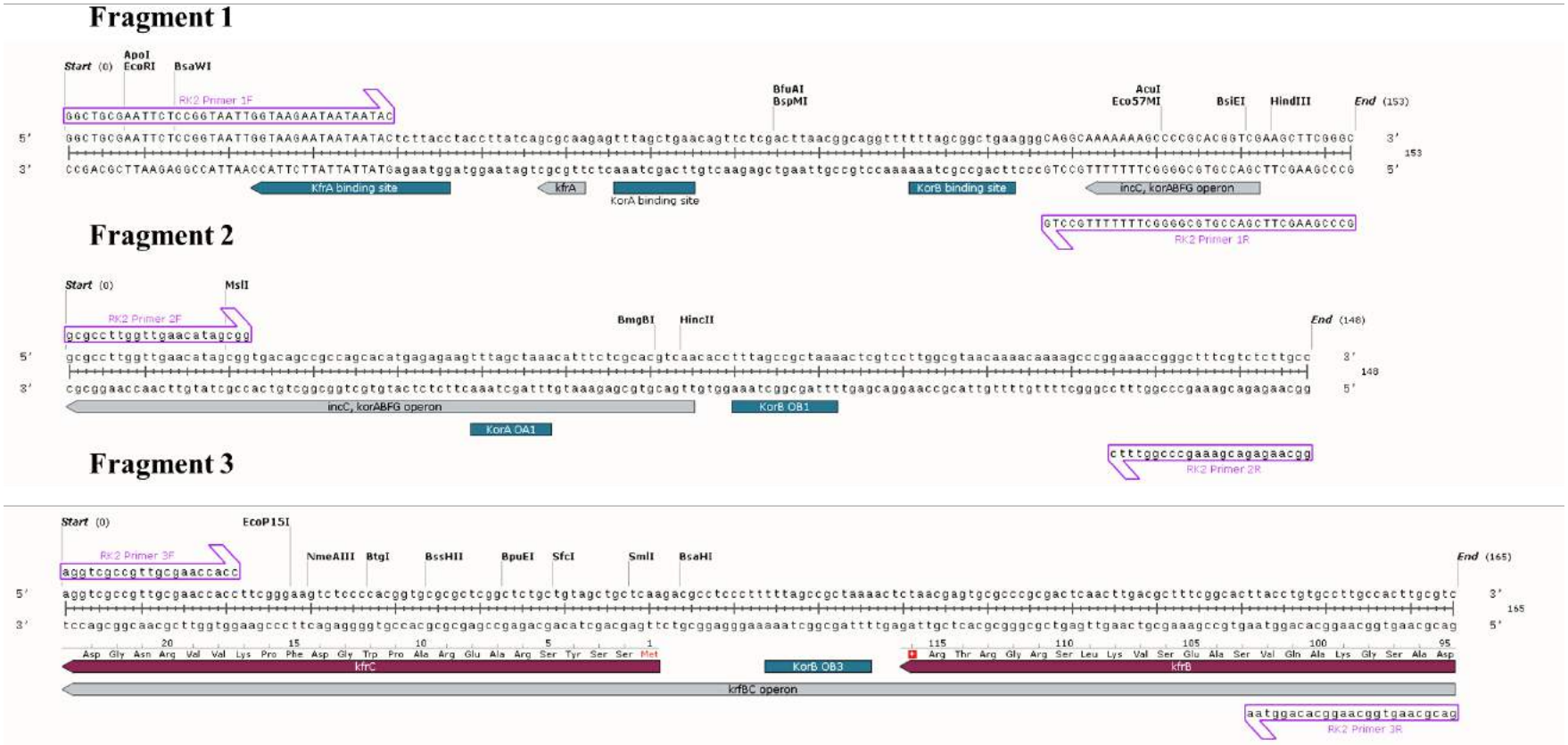


Figure 3.23 PCR amplified fragments from the RK2 plasmid. Three different fragments were PCR amplified from the RK2 plasmid for EMSA. Fragment 1 (153 bp) was amplified from the RK2 plasmid region 55,919-56,022 and has KorB and KorA binding sites. Fragment 2 (148 bp) was amplified from the RK2 plasmid region 59,383-59,509 and contained KorA OA1 and KorB OB1 binding site, while fragment 3 (165 bp) contained KorB OB3 binding site and amplified from RK2 plasmid region 54,375-54,519.

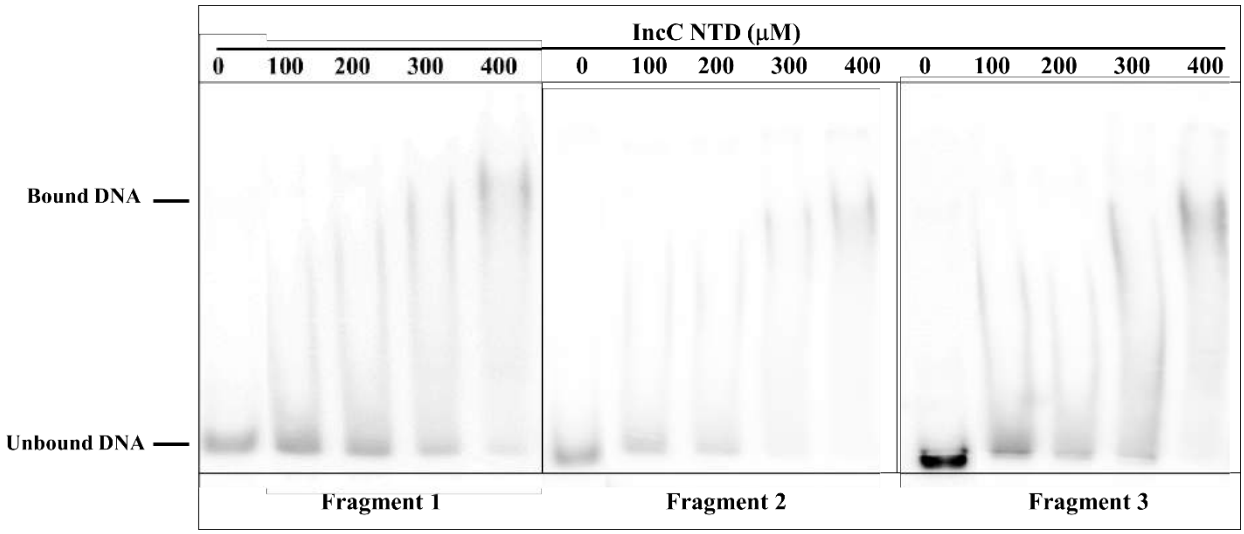


Figure 3.24 IncC NTD EMSA using PCR amplified fragments. EMSA was performed using three different PCR amplified DNA fragments. The fragments were radiolabelled using ^{32}P from y32P -ATP as described in chapter 2. Samples for EMSA were prepared using Tris buffer pH 7.5 (containing Tris 10 mM, NaCl 150 mM, 5 % glycerol, 0.1 mM EDTA, 10 mM $MgCl_2$, and 0.1 μg/mL BSA). The samples were incubated with increasing concentrations of IncC NTD, for 15 min at room temperature, and loaded onto a 5% polyacrylamide gel, electrophoresis was done at 2 V/cm (running buffer 0.5 % TBE). The gel was dried and placed on a film for an hour before developing using phosphorimager (BioRad). The image was visualised and saved using Quantity One software (BioRad). EMSA shows that IncC NTD binds to DNA very weakly, giving smeared DNA bands on the gel.

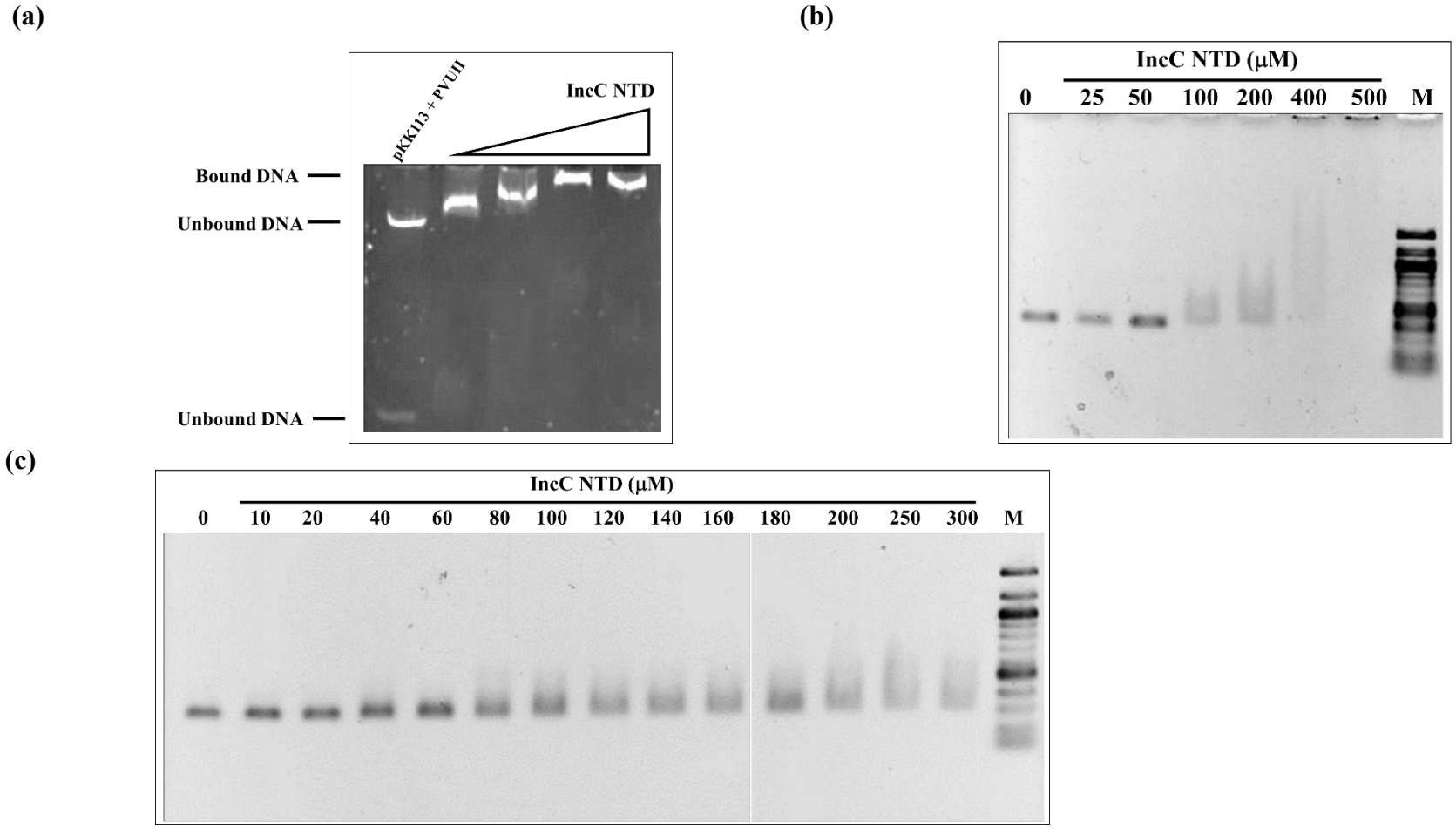


Figure 3.25 EMSAs for IncC NTD. IncC NTD-DNA binding was observed using longer and shorter DNA fragments. **(a)** EMSA was performed using 291 bp and 2,364 bp DNA fragments and using Agarose gel electrophoresis (1 % agarose and TBE running buffer). The protein-DNA sample mixture was prepared using DNA binding buffer (1x DNA binding buffer contained Tris pH 7.5, NaCl 150 mM, 5 % glycerol, 0.1 mM EDTA, 10 mM $MgCl_2$, and 0.1 μg/mL BSA). IncC NTD varying concentrations (50, 100, 150, and 200 μM, lane 2-5) were used in EMSA. The longer DNA fragment (2,364 bp) was retarded and some DNA smears for the shorter fragment were also observed **(b)** EMSA was performed using a 291 bp DNA fragment and varying IncC NTD (25-500 μM) to monitor DNA binding. At higher protein concentrations, DNA smears were observed **(c)** EMSA, performed using a short oligo (57 bp) shows IncC NTD binds to DNA very weakly. DNA smears were also observed on the gel, showing protein-DNA dissociation in the gel.

3.20 Discussion

IDPs are widespread among all domains of life, Eukarya, Bacteria and Archaea, but eukaryote proteomes have a much larger proportion of IDPs in comparison with others. One prediction shows, 4.2% of proteome disorder in Bacteria, while in Eukarya the disorder levels are even higher (up to 33%). In Drosophila and Humans, proteome disorder is predicted to be 36.6% and 35.2%, respectively. Interestingly, the extent of disorder in the proteins that are indispensable for cellular functions rises to 20.6% in *E. coli* and 52.4% in *S. cerevisiae* (Tompa et al., 2006). In particular, bacterial proteins involved in cell division and chromosome/plasmid partitioning often contain intrinsically disordered regions either in C-terminal or N-terminal regions. For instance, the polar organizing protein Z (PopZ) from *Caulobacter crescentus* has been found to be largely disordered (~75%) (Holmes et al., 2016). Intrinsically disordered regions (IDRs) in PopZ interact with many proteins during chromosome segregation, including ParA and ParB (Ptacin et al., 2010). The plasmid partitioning protein KorB also contains IDRs that are involved in interactions with partner proteins. In some ParB proteins, a flexible N-terminal region is involved in ParB-ParB (dimer) interactions (Funnell, 2016). Other cell division proteins, like ZipA and FtsZ, also contain intrinsically disordered regions (Lopez-Montero et al., 2013, Buske and Levin, 2013).

Usually, bacteria do not express numerous disordered proteins. Interestingly, many proteins from the RK2 plasmid contain long IDRs (>30 amino acids) (Figure 3.26). The *par* operon proteins are expressed with long IDRs regions (> 25% of amino acids residues are in present IDRs). Disordered regions are non-homologous and evolution of such regions cannot be easily traced. Previous studies suggest these regions originated or evolved with a requirement to interact with other partners, including DNA, RNA, proteins, ligands and metals ions. Therefore, IDRs may accommodate mutations and have more charged amino acids than folded regions (Brown et al., 2002).

IncC NTD contains more than 70 % small and charged amino acids. Residues like G, D, E, P, and S are considered to be disorder-promoting amino acids and have been found in abundance in IDPs (Uversky, 2013). It is also notable that IncC NTD contains few order-promoting amino acids (C, F, W, Y) and does not contain any Tyrosine (Y) or Cysteine (C) residues (Figure 3.27).

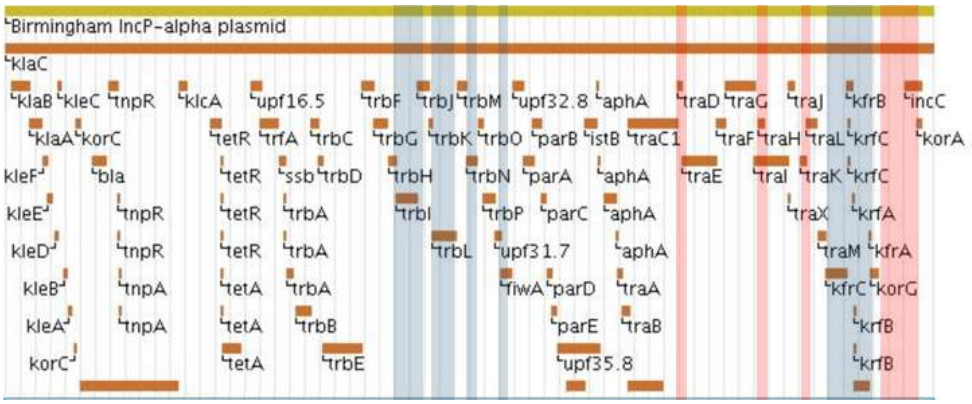


Figure 3.26 Disordered regions of protein expressed from the RK2 plasmid. IDRs predictions for all proteins expressed from the plasmid RK2 were obtained using DISPRO database (http://www.disprot.org). The greenish yellow line shows the linear ~60 kb RK2 plasmid. The brown lines show proteins express from the RK2 plasmid. Pink shaded areas show proteins with IDRs encompassing ≥25 % of the proteins, while blue shading shows proteins having IDRs encompassing ≥10 % of individual protein. IncC and KorB contain disordered regions in >25% of total protein length.

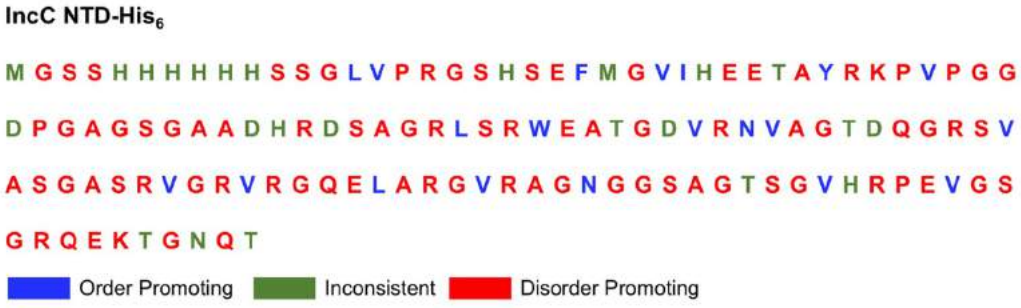


Figure 3.27 IncC NTD protein sequence analysis. Amino acids are coloured on the basis of their disorder propensities. Blue colour amino acids are ordered promoting, red is disorder-promoting while green amino acids are inconsistent towards the secondary structure. IncC NTD has been found to contain a large proportion of disorder-promoting amino acids.

The charge/hydropathy (CH) plot shows that IncC NTD lies at the borderline of ordered and disordered proteins (Figure 3.6b). Disordered proteins are either highly positively or highly negatively charged due to containing predominantly charged amino acids (Uversky, 2013). In previous studies, it was suggested that IncC NTD may contain an HTH motif that is involved in DNA binding (Batt, 2008). Here, secondary structure predictions show that IncC NTD is largely disordered and has only short stretches that may form secondary elements (Figure 3.7). The CD spectra for IncC NTD confirmed it is a predominantly intrinsically disordered protein and secondary structure contents were found to be 4% α-helix and 19% β-sheet (Figure 3.20). These

results were also in agreement with the Kratky plot where the absence of bell-shaped curve shows that IncC NTD as an intrinsically disordered region (Figure 3.18). It has been found that C- or N- terminal tails in many proteins undergo a disorder to order conformational change when they interact with other partners such as proteins, DNA or metals ions (Mendoza-Espinosa et al., 2009).

On SDS PAGE, the molecular weight of IncC NTD was a little higher than expected (~ 14 kDa instead of 13.1 kDa obtained from the amino acid sequence). This might be due to the high positive charge and elongated shape of the protein. Contrasting results were found when the molecular mass was calculated from SEC and other biophysical tools. The molecular mass from SEC was calculated as ~29 kDa (Figure 3.13), from AUC as 10.9-12.3 kDa (Table 3.2), from SEC-MALLS as 12.1 kDa and from SAXS as 13.5 and 14.6 kDa (Table 3.3). Mass spectrometry analysis showed the molecular mass to be 13.01 kDa (Figure 3.11), while the theoretical molecular mass (from the IncC NTD-His_6 amino acid sequence) was calculated as 13.14 kDa (https://web.expasy.org/cgi-bin/protparam/protparam) (Appendix 3).

The Differences in hydrodynamics radius were also observed when calculated from different experiments. SEC showed a Stokes radius (R_s) of 23 Å, while SEC-MALLS showed 23.6 Å. From AUC, R_s was calculated as 13 Å. SAXS was used to obtain the radius of gyration for IncC NTD and value was found to be 31.2-32.5 Å. The differences in the molecular masses and Stokes radius are probably due to the high charge and dynamic shape of the protein, as most of these biophysical tools base these calculations on the basis of protein size and shape.

Positively charged patches in IDRs were predicted to be involved in binding to negatively charged DNA. The location of the charged patches in IDPs determines the way hydrogen bonding or electrostatic interactions would be established. Non-specific DNA interaction usually occurs by low-affinity electrostatic interactions and allows rapid on-off association/dissociation. Proteins can diffuse along the DNA by hopping and sliding. EMSAs show that, IncC NTD interacts weakly and non-specifically with DNA (Figure 3.25).

The full-length IncC1 protein, with IncC NTD, may form dimers and tetramers, increasing the local concentration of IncC NTD. Therefore, the IncC NTD affinity for DNA in the full-length protein may be more than as an individual domain. Some IDRs show very weak binding affinity when interacting with their partners, and K_d values ranging from 10 μM to 1 mM have been reported in linker regions of a modular protein involved in DNA replication (Kearsey

et al., 2009). A cold shock domain of human Y-Box protein YB-1 also showed to have a very weak K_d (150 μM) using NMR when a 5-mer (ATTGG) oligo was titrated (Zeeb and Balbach, 2003).

CHAPTER 4

Determination of secondary structure of IncC NTD by NMR

Chapter 4

Determination of the secondary structure of IncC NTD by NMR

IncC NTD is important as a missing link between the chromosomal and plasmid ParA proteins as stated in Chapter 3. IncC NTD has been characterized as predominantly disordered by various biophysical methods and it binds to DNA non-specifically. Crystallography trials for the IncC NTD protein were not successful, so a structural characterisation of IncC NTD has been performed using NMR. Conventional proton detected NMR (^{1}H NMR) experiments are not suitable for the study of an intrinsically disordered protein (IDP) like IncC NTD due to a considerable overlap of the ^{1}H amide signals, so a series of carbon-detected NMR (^{13}C NMR) experiments were performed using a cryogenic probe optimised for direct ^{13}C-detection. Data were obtained and used to assign backbone resonances and to determine the secondary structure for IncC NTD. IncC NTD was overexpressed using ^{15}N, ^{13}C minimal medium (detailed in Appendix 5) and purified as described in Chapter 2 (Section 2.18).

A total of ~9% (12,108) structures, deposited in Protein Data Bank (PDB) by December 2017, have been solved by NMR (Statistics from https://www.rcsb.org/), showing that NMR is not as an efficient method to elucidate protein structures as compared to X-ray crystallography and Cryo-EM (Electron Microscopy). However, NMR is an important method to determine protein structure-function relationships, to map protein-ligand binding/interaction sites (Jeeves and Knowles, 2015, Abraham et al., 2008) and provides information about protein dynamics (Kleckner and Foster, 2011) and conformational heterogeneity (Zeymer et al., 2016).

4.1 NMR for IDPs

It is extremely difficult to get IDPs crystallised due to their dynamic structure; that makes NMR attractive to study IDPs dynamics and flexibilities at the atomic level (Bermel et al., 2006). When probing IDPs/IDRs with conventional NMR, problems include the lack of resonance dispersion in the proton dimension. Progress and developments of new NMR methods allow ^{13}C observed experiments to get better resolution spectra for IDPs.

4.2 One dimensional (1D) NMR

One dimensional NMR (1D ^{1}H NMR) is a very fast and basic NMR experiment. It

completes in very short time and does not require a highly-concentrated sample. The one-dimensional proton spectrum shows a signal (peak) for each unique hydrogen (^{1}H, proton) in a molecule. Different positions of peaks (chemical shifts) in a 1D spectrum, are due to different chemical environments. The 1D ^{1}H NMR spectrum (ranging ~ -1 to ~12 ppm) can be obtained to analyse protein conformation, aggregation stability (Kwan et al., 2011). A spectrum with well-dispersed backbone amine peaks in the range of 8.5-11 ppm, shows a folded protein, while disordered proteins show much less dispersed spectra, with intense signals at ~8.3 ppm (Figure 4.5). Such a spectrum is an indicator of a random coil-like conformation of a protein. The extent of disorder can be estimated by comparing a peak intensity at ~8.3 ppm with that of the other dispersed peaks (Rehm et al., 2002).

4.3 Two-dimensional (2D) NMR

Resolving the NMR spectrum using a second dimension gives better peak dispersion and allows the interpretation of more information than the 1D spectrum. It also provides the basis for higher order multidimensional experiments with which to gather more information about protein structure from chemical shifts, spin relaxation, scalar or dipolar coupling.

4.3.1 HSQC

The ^{1}H, ^{15}N HSQC (Heteronuclear Single Quantum Correlation) is a NMR experiment that correlates the shifts of amide nitrogen and the attached amide proton group. Each peak in HSQC represents one amide ^{15}N-^{1}H pair, so ideally HSQC gives one peak per residue. It has been widely used to collect basic structure information as well being used for ligand binding experiments. Magnetisation transfer between ^{1}H and ^{15}N in HSQC is shown schematically in Figure 4.1

Using NMR, it is possible to get information about fractional helix propensity in the disordered segments of proteins. Chemical shifts, coupling constants and subsequently, sequential assignments can be used to get backbone structure information of a protein. NMR is a powerful method that has made possible to obtain IDPs ensemble conformations. Post Translational Modifications (PTMs) can also be studied by HSQC as PTMs in a specific region may reduce local helicity and allow interacting partner to segregate (Baker et al., 2007).

In case of unfolded proteins, ^{1}H-^{15}N HSQC shows crowded signals around ~8.3 ppm leading to ambiguous peak identification for the most of individual residues in a spectrum (Arai

et al., 2009). To overcome poor spectral resolution, 2D HSQC can also be obtained using the $^{13}C'$–^{15}N or $^{13}C'$–$^{13}C_{\alpha}$ pairs but $^{13}C'$ is not as sensitive as the ^{1}H nucleus, therefore ^{1}H,^{15}N HSQC spectra are usually obtained in conventional NMR experiments.

4.4 Carbon-detected NMR

In 1988, Markley proposed a proton-less approach to design experiments for NMR, but at that time ^{13}C sensitivity was the most important issue. Recent advances in NMR instrumentation along with cryogenic-probes and sensitive coils incorporation, allows direct carbon detection to be suitable for protein assignment. This helps in solving IDPs structures by using the better peak dispersion found in both carbon and nitrogen dimensions to circumvent the poor proton dispersion in IDPs (Bermel et al., 2008, Novacek et al., 2012, Pantoja-Uceda and Santoro, 2013). It is critical to study Prolines in IDPs due to their abundance, a possible role in PTMs and protein-protein interactions. Carbon-detected experiments allow Proline assignments that are not possible in proton-directed experiments (as the Proline contain no amide proton). Salt concentration does not adversely affect sensitivity of the carbon-detected experiments in comparison to proton-detected experiments (Shimba et al., 2004). Bulky, charged Arginine side chains can be studied for protein interactions and protein dynamics using distinct, carbon-detected experiments, in which the Arginine side chains can be uncoupled from the distinct backbone environment (Mackenzie and Hansen, 2017). In carbon-directed experiments, magnetisation is started from the ^{13}C nucleus (Cα or C′) and is finally detected at same ^{13}C or another ^{13}C and provides better resolution than could be achieved using conventional proton detected methods (Bertini et al., 2004b). Resolution and sensitivity can be increased by incorporating different methods and approaches including the IPAP and S2E schemes and hetero-nuclei and inter-residue correlations depending on homonuclear ^{13}C decoupling (Bermel et al., 2005, Bermel et al., 2007, Bermel et al., 2013).

4.4.1 CON

The ^{15}N,^{13}C CON spectrum correlates carbonyl carbon and amide nitrogen resonances from peptide backbone and gives a basic spectrum in carbon-detected NMR experiments. Each peak in the CON spectrum, represents an individual residue in a protein, showing the shifts of the C′ of residue *i*, and the associated N group of residue *i*+1 (Figure 4.1). The ^{13}C detected approaches are better than traditional NH based experiments to allow sequential walking along a

protein backbone and to get secondary structure information for IDPs. High exposure to solvent (water) increases proton exchange rate in disordered protein segments leading to peak broadening and loss of intensity in a spectrum that eventually effects sensitivity and resolution of experiments (Kosol et al., 2013, Bermel et al., 2013).

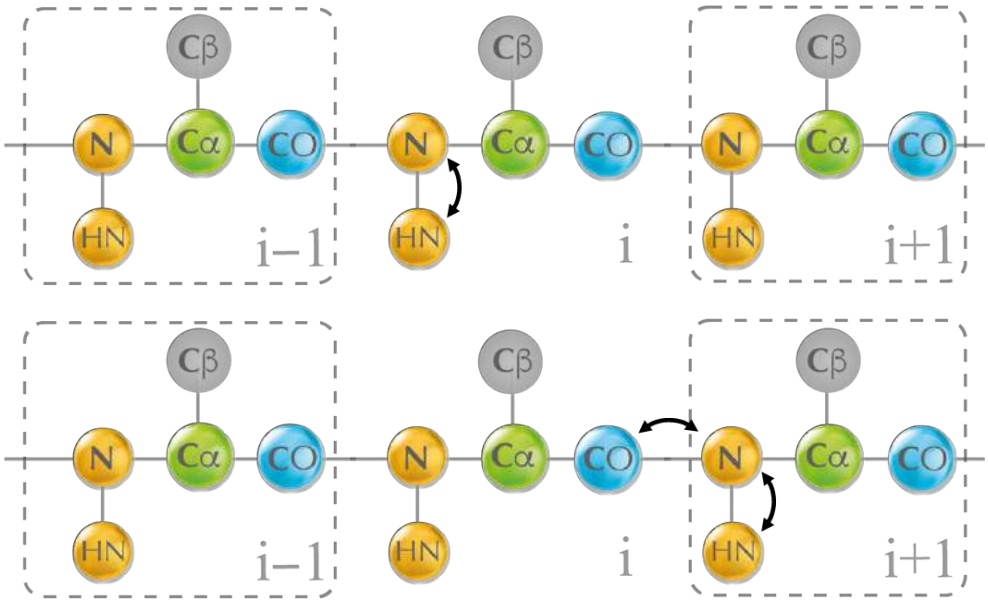


Figure 4.1 Magnetization transfer for the 1H 15N HSQC and triple resonance CON experiments. (a) In an HSQC experiment, magnetization transfer occurs from ^{1}H to ^{15}N via J-coupling. ^{15}N chemical shift is observed and magnetisation is detected after transferring back to ^{1}H. **(b)** In CON experiment, magnetization from ^{1}H is transferred to ^{15}N followed by transfer to ^{13}Co nuclei, chemical shift is observed at ^{15}N and detected at ^{13}Co.

4.5 Multi-dimensional NMR

Multi-dimensional NMR helps in resolving peak overlaps and makes it possible to assign all chemical shifts. Backbone and side chain assignments can be determined, as well as 3D structures of proteins can be obtained, using Nuclear Overhauser Effect (NOE). Multi-dimensional approaches involve scalar connectivity and correlations between hetero-nuclei, through NH, CH, CC, and CN couplings. At least three dimensions (3D) experiments are required for backbone assignments, to resolve signal overlaps and to determine protein secondary structure. 3D experiments (Figure 4.2) help to increase spectrum resolution and to resolve overlapping peaks.

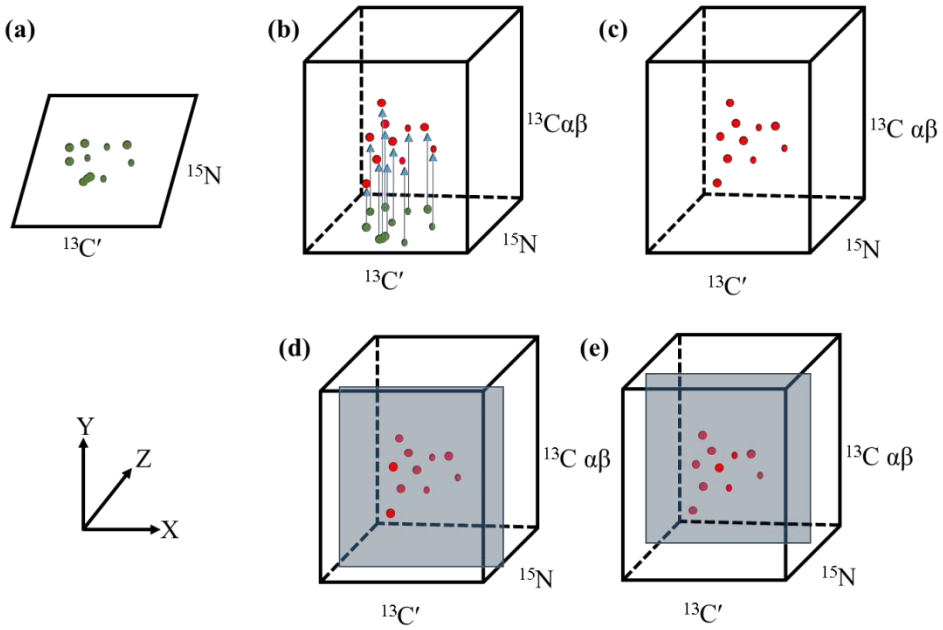


Figure 4.2 An illustration of three-dimensional (3D) carbon directed NMR experiments. (a) A typical 2D peak arrangement (2D CON spectra), while Co is in the x-axis and ^{15}N in the y-axis. **(b)** In three-dimensional experiment, the Z dimension is introduced. Here in illustration, Co and Cαβ are in x, y dimensions while N is in Z dimension. **(c)** In third dimension, now it is possible to analyse correlation of three heter-onuclei. **(d, e)** The CON peaks can be observed in different planes, due to side chain couplings, and each plane can be individually analysed and assigned. So, a 2D C′ – Cαβ spectrum is obtained with varying ^{15}N dimension.
(Figure adapted from https://sites.google.com/site/nmrgenerator/assignment-theory/visualizing-3d-spectra)

4.6 Backbone assignment

Carbon-directed NMR experiments can be used to assign backbone resonances of proteins. In carbon-detected experiments, it is recommended to use 2D ^{15}N, ^{13}C-CON as a root experiment. Complete backbone assignments for a disordered protein are possible using a set of carbon-detected experiments performed on ^{13}C-^{15}N labelled samples. Some proton-detected experiments can be helpful to complete and to confirm the backbone assignments. Carbon-directed triple resonance experiments used in this study are CANCO, CBCACON and CBCANCO. The proton-detected experiments HNCO, HNN and HNCN were used to complete and confirm protein NMR assignments.

4.6.1 CANCO experiment

The CANCO experiment is a carbon-detected NMR experiment that provides two Cα peaks, one each for the *i* and *i*+1 residue linked to the Ni and C′$_{i+1}$. Magnetization is transferred from ^{1}H^{Cα} to ^{13}Cα and from ^{13}Cα to ^{15}N (both N$_i$ and N$_{i+1}$). Chemical shift is evolved at ^{13}Cα and ^{15}N while signal is read at ^{13}C′ (Figure 4.3), subsequently showing peaks for ^{13}Cα, ^{13}Co and ^{15}N

by a 3D spectrum (Bermel et al., 2006, Bermel et al., 2009). A strip from the CANCO spectrum of IncC NTD has been shown in Figure 4.7.

4.6.2 CBCACON experiment

In the CACBCON experiment, information about the *i* residue are collected. Magnetization is transferred from $^{1}H^{C\alpha}$ and $H^{C\beta}{}_{i}$ to $^{13}C\alpha/C\beta i$, $^{13}C\beta$ to $^{13}C\alpha_{i}$ and from $^{13}C\alpha_{i}$ to C'_{i} and then to $^{15}N_{i+1}$ (Figure 4.3). This experiment provides chemical shifts for $^{13}C\alpha/C\beta$ of *i* residue (Bermel et al., 2006, Bermel et al., 2009). A strip from the CBCACON spectrum has been shown in Figure 4.7.

4.6.3 CBCANCO experiment

This experiment is similar to CBCACON, however magnetisation is transferred directly to ^{15}N from $C\alpha$, and hence information for $C\alpha_{I}$, $C\alpha_{i+1}$, ^{15}N (i+1) attached to carbonyl carbon (i) and intra-residue $C\alpha/\beta$ (both i and i+1) are obtained (Figure 4.3) (Bermel et al., 2006, Bermel et al., 2009). A strip from the CBCANCO spectrum has been shown in Figure 4.7.

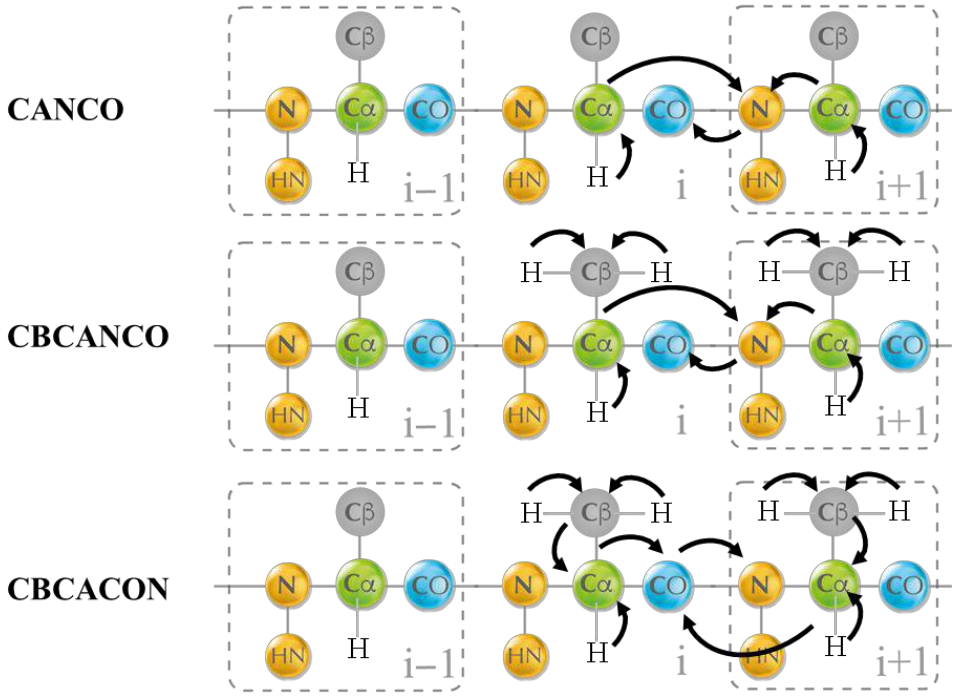


Figure 4.3 Magnetization transfer for carbon directed experiments. CANCO experiment supplement CBCACON and CBCANCO experiments to get $^{13}C\alpha/C\beta$ peaks for both *i* and *i*+1 residues. In the CANCO experiment, Magnetization is transferred from $^{1}H^{C\alpha}$ to $^{13}C\alpha$ and from $^{13}C\alpha$ to ^{15}N (both N_{i} and N_{i+1}). Chemical shift is evolved at $^{13}C\alpha$ and ^{15}N while signal is read at $^{13}C'$. In CBCACON, Magnetization is transferred from $^{1}H^{C\alpha}$ and $H^{C\beta}{}_{i}$ to $^{13}C\alpha/C\beta i$, $^{13}C\beta$ to $^{13}C\alpha_{i}$ and from $^{13}C\alpha_{i}$ to C'_{i} and then to $^{15}N_{i+1}$, while in CBCACON, magnetisation is transferred to ^{15}N from $C\alpha$, and hence information for $C\alpha_{I}$, $C\alpha_{i+1}$, ^{15}N (i+1) attached to carbonyl carbon (i) and intra-residue $C\alpha/\beta$ (both i and i+1) are obtained.

4.6.4 HNCO experiment

HNCO is a proton-based triple resonance (3D) experiment (Kay et al., 2011, Grzesiek and Bax, 1992). This experiment provides information about ^{1}H coupled to the ^{15}N and C′. From ^{1}Hs, magnetisation is transferred to ^{15}N, followed by selective transmission to the $^{13}C'$ (i-1) by $^{15}N^{H}$ $^{13}C'$ J-coupling. After magnetization evolving at hetero-nuclei, detection occurs at ^{1}H after transferring magnetisation back via ^{15}N (Figure 4.4). As a result, a single peak is obtained for each amino acid, at the frequencies of H_i, N_i, and C'_{i-1}. This experiment correlates i and i-1 information.

4.6.5 HNNH and HNCN experiments

HNNH is a proton-based triple resonance experiment that allows confirmation of sequentially assigned residue stretches and help to obtain assignments for ^{15}N of i+1 and i-1 sequential residues (Panchal et al., 2001) (Figure 4.4). The HNCN experiment is an accessory experiment to HNNH. Since two (3 peaks) peaks are obtained in HNN experiment for the i, i+1 and i-1, NH groups; in the HNCN experiment only one (2 peaks) peak to i and i-1. The two spectra allow distinguishing i+1 and i-1 peaks (Figure 4.9).

The HNCO experiment is used to correlate HNNH and HNCN with CON and to confirm backbone assignments. A schematic showing an extract of HNNH and HNCN assignments is given in Figure 4.9.

4.7 Secondary structure prediction using chemical shifts

The TALOS (Torsion Angle Likeliness Obtained from Shift and Sequence Similarity) server can be used to predict torsion angles of protein backbone on the basis of chemical shifts obtained from protein backbone assignments (Cornilescu et al., 1999). TALOS+ (https://spin.niddk.nih.gov/bax/software/TALOS/) uses a neural network based algorithm and a database of 200 proteins from BioMagResBank (BMRB) (Markley et al., 2008) to predict torsion angles of protein backbone using chemical shift data (Shen et al., 2009a, Shen and Bax, 2013).

4.8 Prediction of a 3D structure using chemical shift form CS-Rosetta server

CS-Rosetta (Chemical-Shift-Rosetta) is a web-based protocol that provides *de novo* protein structure prediction (Shen et al., 2009b) based on chemical shifts obtained from

backbone assignments. CS-Rosetta provides several structures (3,000 to 40,000) and ranks them with reference to RMSD values and energy scores.

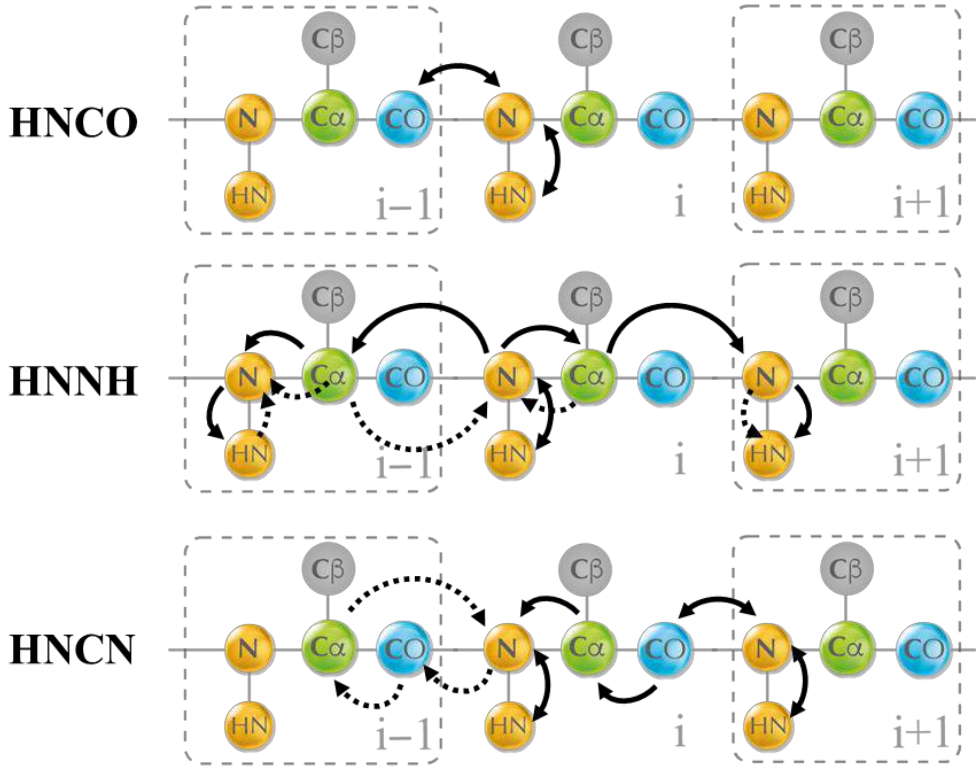


Figure 4.4 Magnetization transfer for proton-detected experiments. The HNCO experiment supplement ^{1}H information to the CON assignments. HNNH and HNCO experiments provide ^{15}N information of corresponding *i* and *i*+1 (and *i*-1) residues. In the HNCO experiment, magnetization is transferred from ^{1}H to ^{15}N and selectively to $^{13}C'$. In HNNH, magnetization is transferred from $^{1}H^{N}$ to^{15}N and then to $^{13}C\alpha$, this is then transferred to ^{15}N of *i* residue, *i*+1 and *i*-1, while detection happens by transferring magnetization back to $^{1}H^{N}$ of *i*, *i*+1and *i*-1. In the HNCN experiment, signal is read at *i* and *i*+1 $^{1}H^{N}$as the transfer from Cα is to the C′.

Methods

4.9 IncC NTD protein labelling

The ^{13}C ^{15}N labelled IncCNTD-His_6 protein is overexpressed using minimal cell growth media (Appendix 5) as described in Chapter 2 (Section 2.16). Protein is purified using the Ni-NTA chromatography and Size exclusion chromatography as described in Chapter 2 (Section 2.18 and 2.24).

4.10 1D and 2D NMR for IncC NTD

The ^{13}C ^{15}N labelled IncC NTD protein (300 μM, 10 % D_2O, 10 mM sodium phosphate, 150 mM NaCl, 0.1 mM EDTA, pH 6.5) is used to obtain NMR spectra. Initially, a 1D ^{1}H NMR spectrum for IncC NTD was obtained using 900 MHz, Bruker NMR spectrometer, equipped with TCI cryoprobe, at 298K. The water peak was suppressed by pre-saturation during the inter scan delay. The 2D ^{1}H-^{15}N HSQC spectrum for IncC NTD was obtained using the same ^{13}C-^{15}N labelled IncC NTD protein sample and NMR spectrometer. Data were processed and analysed using the programs TOPSPIN 3.2 (Bruker BioSpin, Germany) and CcpNmr Analysis (Collaborative Computational Project for NMR) ver. 2.4.2.

The IncC NTD peaks were poorly dispersed and there was considerable peak overlap in the HSQC. To overcome this problem carbon-detected NMR experiments were planned. For carbon-detected NMR experiments, the ^{13}C detected CON spectrum is obtained for IncC NTD (300 μM, 10 % D_2O, 10 mM Sodium Phosphate, 150 mM NaCl, 0.1 mM EDTA, pH 6.5) using a 600 MHz, Bruker spectrometer at 298K and using a carbon optimised TXO CryoProbe. Data are processed using the MddNMR (Orekhov and Jaravine, 2011) and NMRpipe (Delaglio et al., 1995) software and analysed using CcpNmr Analysis ver. 2.4.2.

The ^{15}N labelled IncC1-His_6 protein was overexpressed and purified using Ni-NTA chromatography. HSQC for the ^{15}N labelled IncC1-His_6 protein (50 μM, 10 % D_2O, 20 mM Tris pH 7, 150 mM NaCl, 50 mM sodium Glutamate, 0.1 mM EDTA, 10 % glycerol) was obtained, processed and compared to IncC NTD HSQC using CcpNmr Analysis ver. 2.4.2.

4.11 Backbone assignments for IncC NTD

A list of experiments used for IncC NTD backbone assignments, is given in Table 4.1. The carbon-detected triple resonance experiments CANCO, CBCACON and CBCANCO were used for IncC NTD backbone assignments. Proton-detected experiments HNCO, HNN and HNCN were also performed to complete and to confirm rest of assignments. The ^{15}N, ^{13}C labelled IncC NTD protein (300 μM, 10 % D_2O, 10 mM sodium phosphate, 150 mM NaCl, 0.1 mM EDTA, pH 6.5) is used to obtained 3D spectra (Table 4.1).

The HNCO experiment is used to correlate HNNH and HNCN experiment with CON and to confirm backbone assignments. ^{15}N, ^{13}C labelled IncC NTD protein sample (mentioned above) is used to obtain protein-detected spectra using a 900 MHz Bruker spectrometer and a TCI probe at 298K. Data were collected using the program TOPSPIN 3.2 (Bruker BioSpin, Germany) and

processed using the NMRPIPE software (Delaglio et al., 1995). Processed data were analysed using CcpNmr Analysis ver. 2.4.2. A schematic showing an extract of HNNH and HNCN assignments is given in Figure 4.10.

HSQC spectra were collected before each 3D experiment to monitor any change in signals due to protein instability. Automated assignment option was not available for carbon detected experiments in CcpNmr Analysis ver. 2.4.2. All peaks were picked and assigned manually. Peaks were sequentially assigned by locating the *i*+1/*i*-1 residues.

Table 4.1 NMR experiments used to assign the IncC NTD protein backbone. Nuclei correlated in the experiments have been given in "information obtained" column.

Experiments	Dimensions	Information obtained	References
HSQC	2D, (H, N)	One peak for each residue is expected for each residue, linking H (*i*) to N (*i*)	(Palmer et al., 1991, Kay et al., 1992)
CON	2D, (C′, N)	One peak for each residue is expected for each residueC′ (*i*) to N (*i*+1)	(Duma et al., 2003, Bermel et al., 2005)
CANCO	3D, (C′, Ca, N)	Two peaks are expected for each peak in CON, linking Cα (*i*) and Cα (*i*+1) to N (*i*+1), and C′(*i*)	(Bertini et al., 2004a, Bermel et al., 2005)
CBCACON	3D, (C′, Ca, N)	Two peaks are expected for each peak in CON, linking Cα (*i*), and Cβ (*i*) to N (*i*+1) and C′(*i*)	(Bermel et al., 2006)
CBCANCO	3D, (C′, Ca, N)	Four peaks are expected for each peak in CON, linking Cα (*i*), Cα (*i*+1), Cβ (*i*), and Cβ (*i*+1) to N(*i*+1) and C′ (*i*)	(Bermel et al., 2006)
HNCO	3D, (H, C′, N)	One peak for each residue is expected to link C′ (*i*-1) to N (*i*) and H (*i*). This spectrum allows spectra detecting C′ and those detecting H to be correlated.	(Duma et al., 2003, Bermel et al., 2005)
HNNH	3D, (H, N, N)	Two peaks are expected at N (*i*+1), and N (*i*–1) at each H (*i*) plane, together with a diagonal peak at N (*i*).	(Weisemann et al., 1993)
HNCN	3D, (H, N, N)	One peak is expected at N (*i*+1) for each H (*i*) plane, together with a diagonal peak at N (*i*) $H^N(i)$	(Bracken et al., 1997, Panchal et al., 2001)

4.12 Secondary structure prediction for IncC NTD using chemical shifts

Chemical shifts obtained from backbone assignments are converted to TALOS+ format using file converter option in CcpNmr Analysis ver. 2.4.2. For IncC NTD, the TALOS+ server (https://spin.niddk.nih.gov/bax/software/TALOS/) is used to predict secondary structure. Result file is viewed with jRAMA+ (Java-based TALOS+ file viewer). Predicted secondary structure is

compared with computational predictions obtained from various servers (SOMPA, JPred, PsiPred 3.3 and SABLE server, Chapter 3 section 3.11.2).

4.13 Prediction of the IncC NTD 3D structure using chemical shift

Chemical shifts from IncC NTD backbone assignments, are used to predict its 3D structure. Chemical shifts are exported in the TALOS+ format and submitted for modelling to the CS-Rosetta web-based protocol server, using Rosetta version 3.8, CS-Rosetta Toolkit version 3.3 (https://csrosetta.bmrb.wisc.edu/csrosetta/submit). CS-Rosetta set to generate 5,000 structure models on the basis of chemical shifts.

4.14 IncC NTD DNA binding

The HSQC spectra for ^{15}N ^{13}C IncC NTD protein (200 μM, 10 % D_2O, 10 mM Tris, 150 mM NaCl, 0.1 mM EDTA, pH 6.5) were obtained in the presence and absence of double-stranded dodecamer DNA (5` CGCGAATTCGCG 3`). Different ratios of IncC NTD: DNA were used (1:1 to 1:4) to monitor protein-DNA binding using HSQC spectra. Processed spectra were analysed using CcpNmr Analysis ver. 2.4.2 and peak shifts monitored by comparing the HSQC spectra.

Results

4.15 1D and 2D NMR of IncC NTD

The one-dimensional spectrum for IncC NTD shows that it is an predominantly unfolded protein with a dominant peak around ~8.3. Such peak is a characteristic of a random coil-like protein. Peak dispersion is very low in other regions (8.5-10 ppm) suggesting that there is little ordered structure (Figure 4.5).

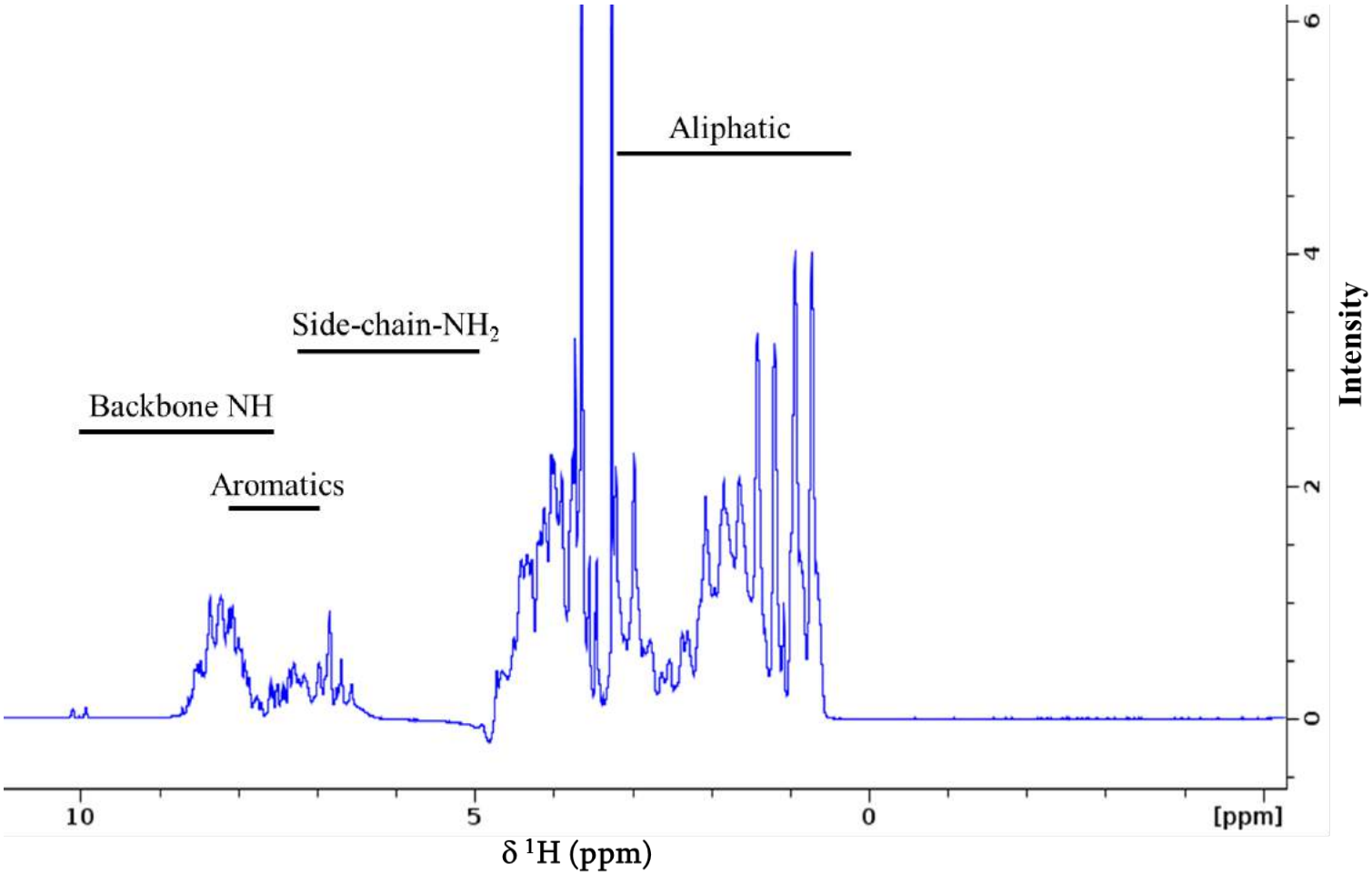


Fig. 4.5 The 1D NMR spectrum for IncC NTD. IncC NTD shows a 1D spectrum of an unfolded protein. A peak can be observed around ~8.3 showing random coil-like protein conformation, while peak dispersion is very low in other regions (8.5-10 ppm).

4.16 2D NMR for IncC NTD

The ^{1}H-^{15}N HSQC spectrum for IncC NTD was collected (Figure 4.6a) and a total of 121 peaks are observed in HSQC. As expected for an IDP, peak dispersion is very poor for IncC NTD HSQC in the proton dimension. Many overlapping peaks are found and almost all peaks are crowded in 1 ppm area (7.7–8.7 ppm region) (Figure 4.6a). To overcome this problem carbon- detected NMR experiments were planned and carbon-detected CON spectrum obtained as a root spectrum (Figure 4.6b). The CON spectrum shows better peak distribution in the ^{13}C dimension than the ^{1}H dimension. Peak dispersion can be observed over 5 ppm (173–178 ppm). Five prolines from IncC NTD can be observed in CON (135-140 ppm) (Figure 4.6b).

4.17 Backbone assignments for IncC NTD

4.17.1 Carbon-detected experiments

Collected data from the CANCO, CBCANCO and CBCACON experiments were processed and analysed in order to enable assignment of the IncC NTD protein backbone. Automated assignment was not viable so all peaks were picked and assigned manually. Peaks were sequentially assigned by locating the *i*+1/*i*-1 residues. For each peak in CON, two peaks are observed in the CANCO spectrum, linking the Cα (*i*) and Cα (*i*+1) to N (*i*+1), and C′(*i*) while in CBCACON, two peaks are observed linking Cα (*i*), and Cβ (*i*) to N (*i*+1) and C′(*i*). For CBCANCO, four peaks are observed for each peak in CON, linking Cα (*i*), Cα (*i*+1), Cβ (*i*), and Cβ (*i*+1) to N(*i*+1) and C′ (*i*). Information from the CANCO, CBCANCO and CBCACON spectra are collated and Cα/β chemical shifts for *i* and *i*+1 residues were obtained (Figure 4.7). Chemical shifts of Cα/β peaks were compared to reference shifts tables and amino acid residue type was tentatively assigned. Sequential assignments are obtained by comparing a stretch of sequentially assigned residues with protein sequence (Figure 4.8).

About 60% of the backbone assignments were obtained using carbons directed experiments. Residues sharing similar Cα/β shifts could not be assigned with confidence. To overcome this problem, proton-based experiments *i.e.* HNN, HNCN and HNCO were performed. These spectra provided confirmation of sequential assignments as well as removed any ambiguity resulting from assignments derived from the ^{13}C-detected experiments.

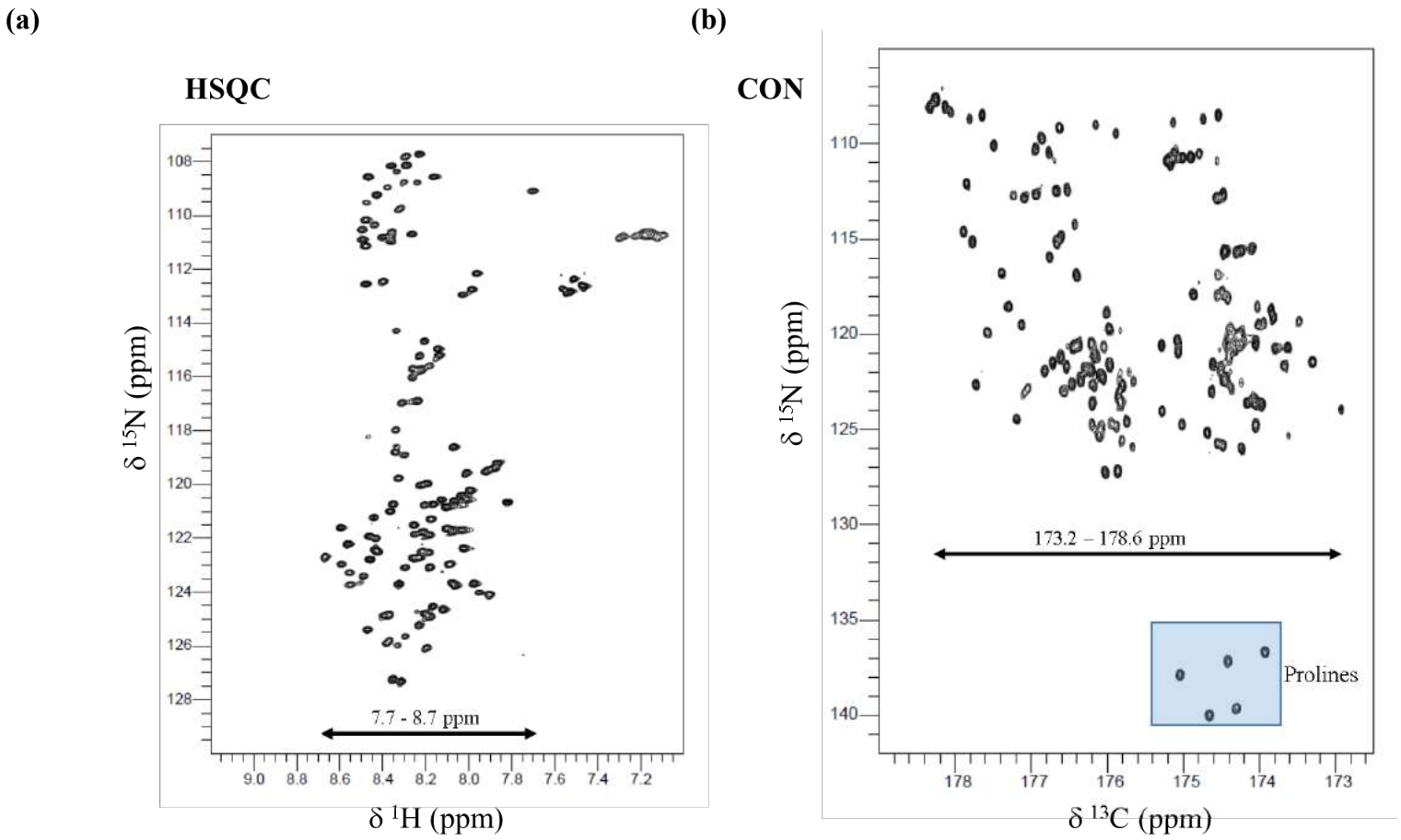


Figure 4.6 ^{1}H-^{15}N HSQC and ^{13}C-^{15}N CON of IncC NTD. (a) The ^{1}H-^{15}N-HSQC spectrum for IncC NTD obtained using a 900 MHz Bruker spectrometer at 298K and using a CryoProbe TCI. In HSQC, 121 peaks are observed. Ideally, in the HSQC, there should be at least one peak for each residue in a protein, plus 2 peaks from each sidechain of Asn, Gln and 1 peak from the sidechain of Trp (with exception of prolines). In HSQC of IncC NTD, all peaks are crowded in ~1 ppm (7.7–8.7 ppm region) and many peaks overlap each other. **(b)** The ^{13}C detected CON spectrum for IncC NTD obtained using a 600 MHz Bruker spectrometer at 298K and using a carbon specific CryoProbe TXO. In the ^{13}C detected CON spectrum, peak distribution is much better and peak dispersion can be observed over 5 ppm (173–178 ppm). Proline peaks can be observed and assigned using CON. For IncC NTD, five proline peaks are observed in CON (135-140 ppm).

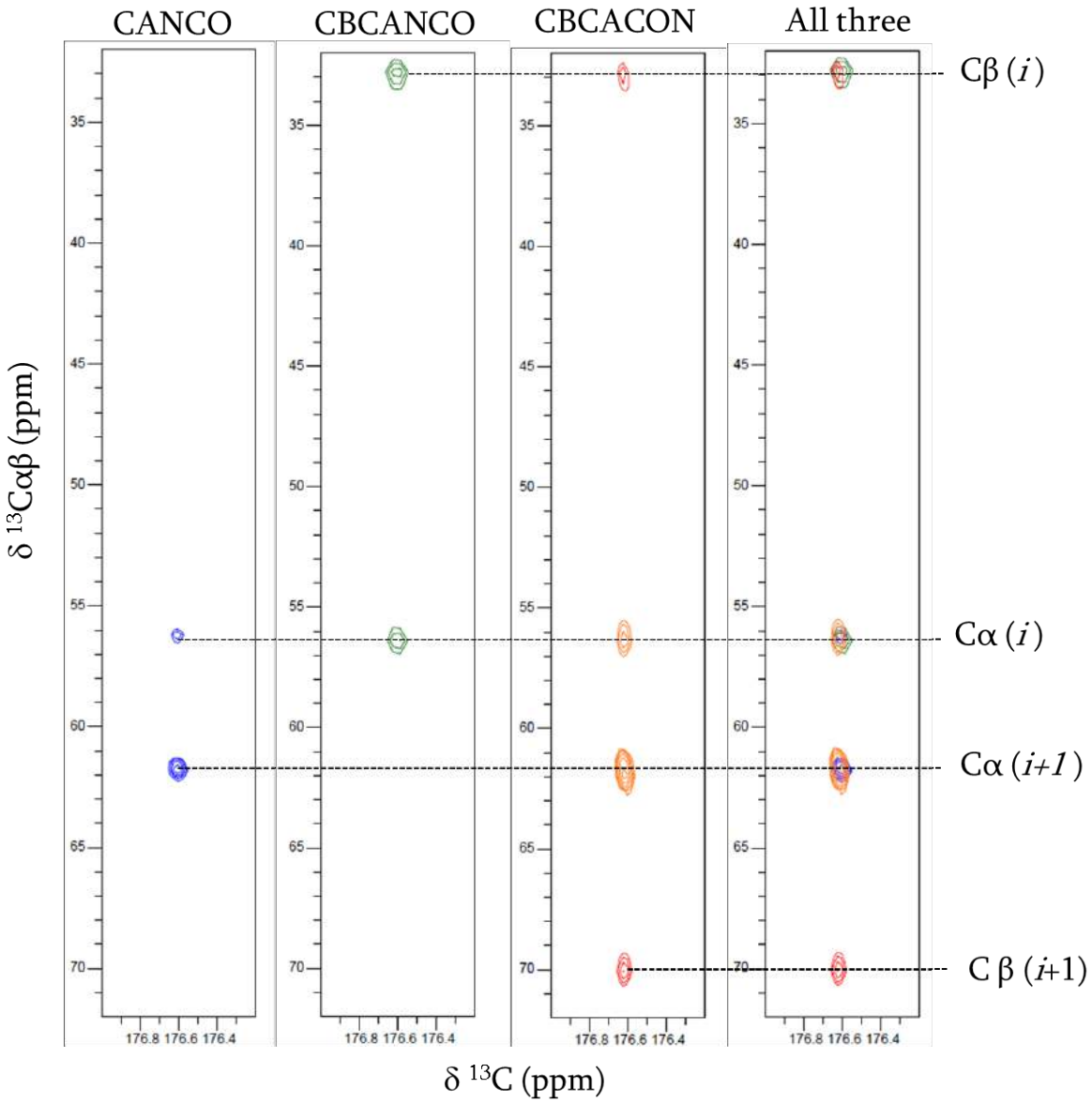


Figure 4.7 Strip extracts from the CANCO, CBCANCO, and CBCACON spectra. Strips have been obtained as single 2D planes from 3D spectra (15 N, 114.90 ppm) showing *i* residue as 7 Glu (E) by comparing Cα (*i*) and Cβ (*i*) chemical shifts to reference shift table, while next residue (*i*+1) is Threonine (T) as shown by its unique chemical shifts for Cα (~62 ppm) and Cβ (~ 70 ppm). In a same way, all planes of 3D spectra are analysed to obtain information for *i* and *i*+1 residue.

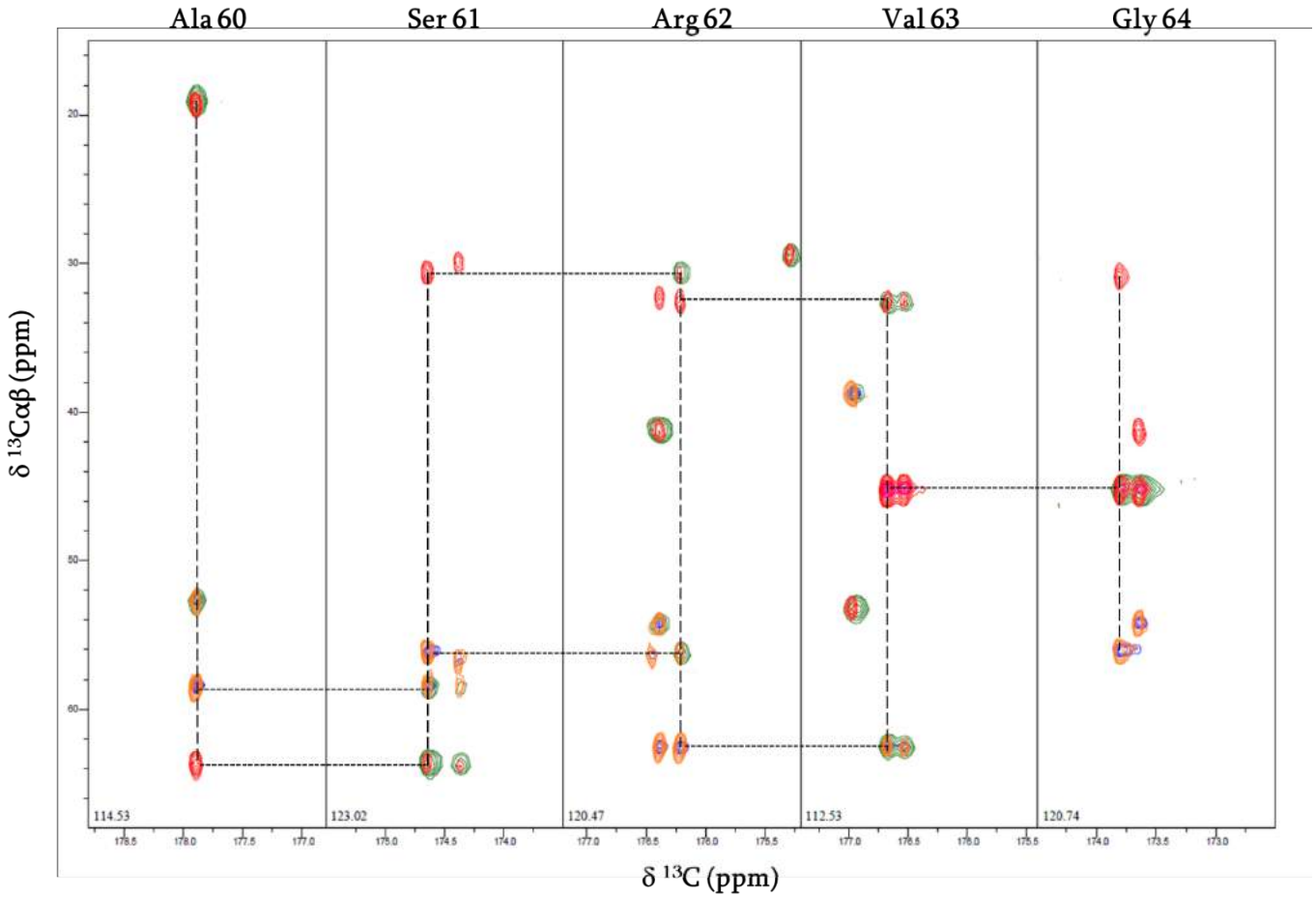


Figure 4.8 Schematic of sequential assignments obtained from the CANCO, CBCACON and CBCANCO spectra. The *i*+1 residues are looked for using corresponding Cαβ (*i*+1) peaks by going forward in sequential assignment as well as back to look corresponding *i* residue peaks. These strips show sequential assignments from Glu40 to Asp43 and colours of peaks correspond to Figure 4.7 where blue peaks show Cα of *i* and *i*+1 from CANCO, green coloured peaks for Cα and Cβ peaks of *i* residue from CBCACON. Orange colour shows Cα of i and *i*+1, red Cβ of *i* and *i*+1 from CBCANCO. By combining the spectra Cα and Cβ peaks for *i*+1 residues are obtained and peaks with same shifts in ^{13}Cα/β dimensions in other ^{15}N planes.

4.17.2 Proton-detected experiments for IncC NTD backbone assignment

HNCO, a proton-based triple resonance (3D) experiment, was used to sequentially assign residue stretches (Figure 4.9). Most of time, three peaks are observed for each observed amide proton in the HNN experiment one each for *i*+1 and *i*-1 and one for *i* (a diagonal peak). The HNCN experiment shows only one peak (*i*+1) and used to distinguish *i*+1 and *i*-1 peaks (Figure 4.9).

The HNCO experiment correlates the HNNH and HNCN experiments with CON and confirms backbone assignments (Figure 4.9). Using three spectra (HNCO, HNN, and HNCN), sequential assignments were made (Figure 4.9 and 4.10). These were then correlated with the analysis performed on the 13C-detected data and further sequence specific assignments were made. Approximately 85 % of backbone assignments are completed (Figure 4.11). The CON root spectrum is annotated using data from all backbone assignments (Figure 4.12 and 4.13). Some peaks for Glycine, Alanine and Serine cannot be assigned due to peak overlapping. Proline peaks are found in CON and assigned successfully. The HSQC is annotated with the assignments with the help of HNCO and CON spectra (Figure 4.14). In HSQC, most of Glycine peaks are found between 108-113 ppm but due to peak overlapping, few peaks could not be assigned confidently and thus are not labelled. Assigned chemical shifts are given in Appendix 5.

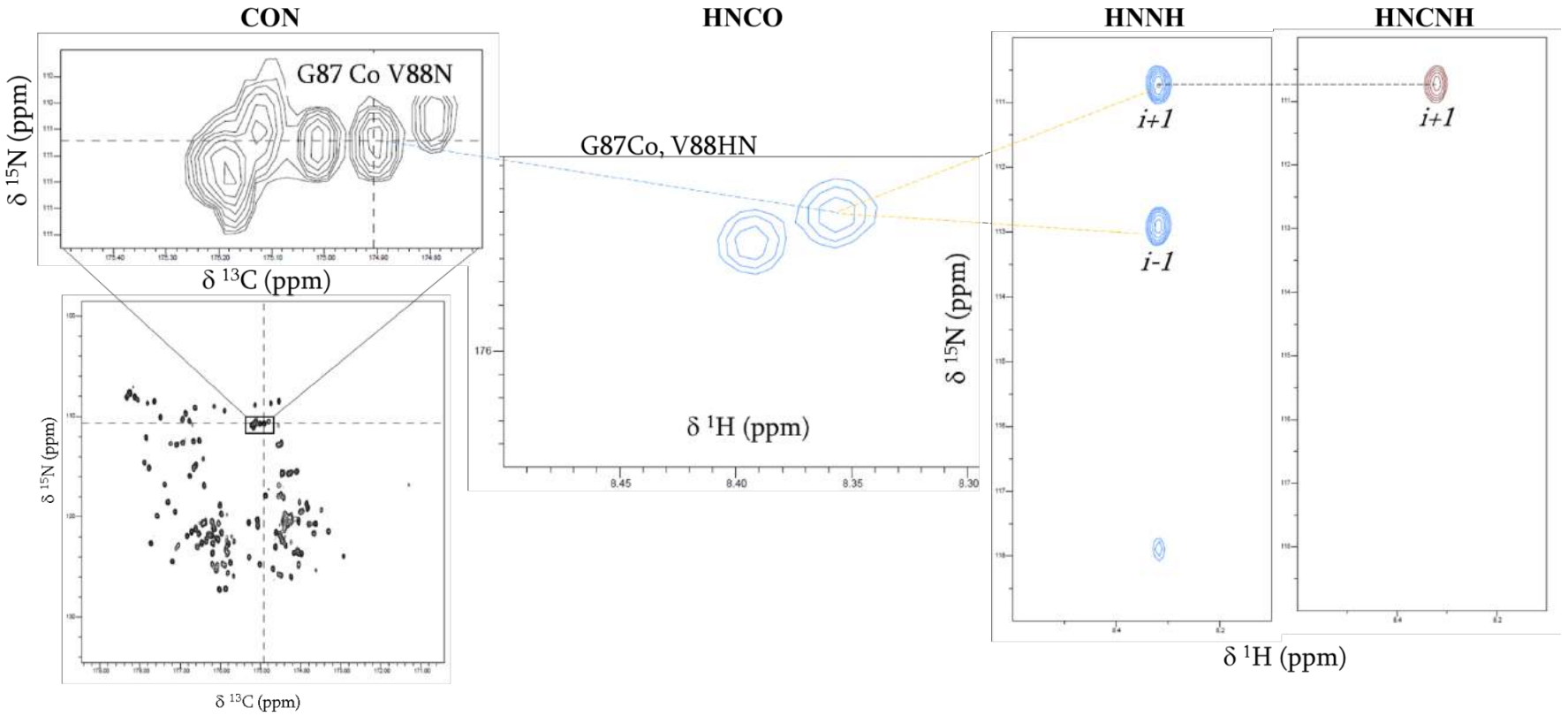


Figure 4.9 A schematic of correlation in the CON, HNCO, HNNH and HNCN experiments. ^{1}H information were obtained for each residue assigned in CON experiment. The CON assigned peak is navigated in the HNCO spectra to look for the ^{1}H chemical shift. The *i*+1 and *i*-1 peaks in HNNH and HNCN, for a corresponding ^{15}N shift in HNCO, are navigated. The ^{15}N shifts for *i*+1 and *i*-1 peaks are located in carbon directed experiments and sequential are assigned and confirmed.

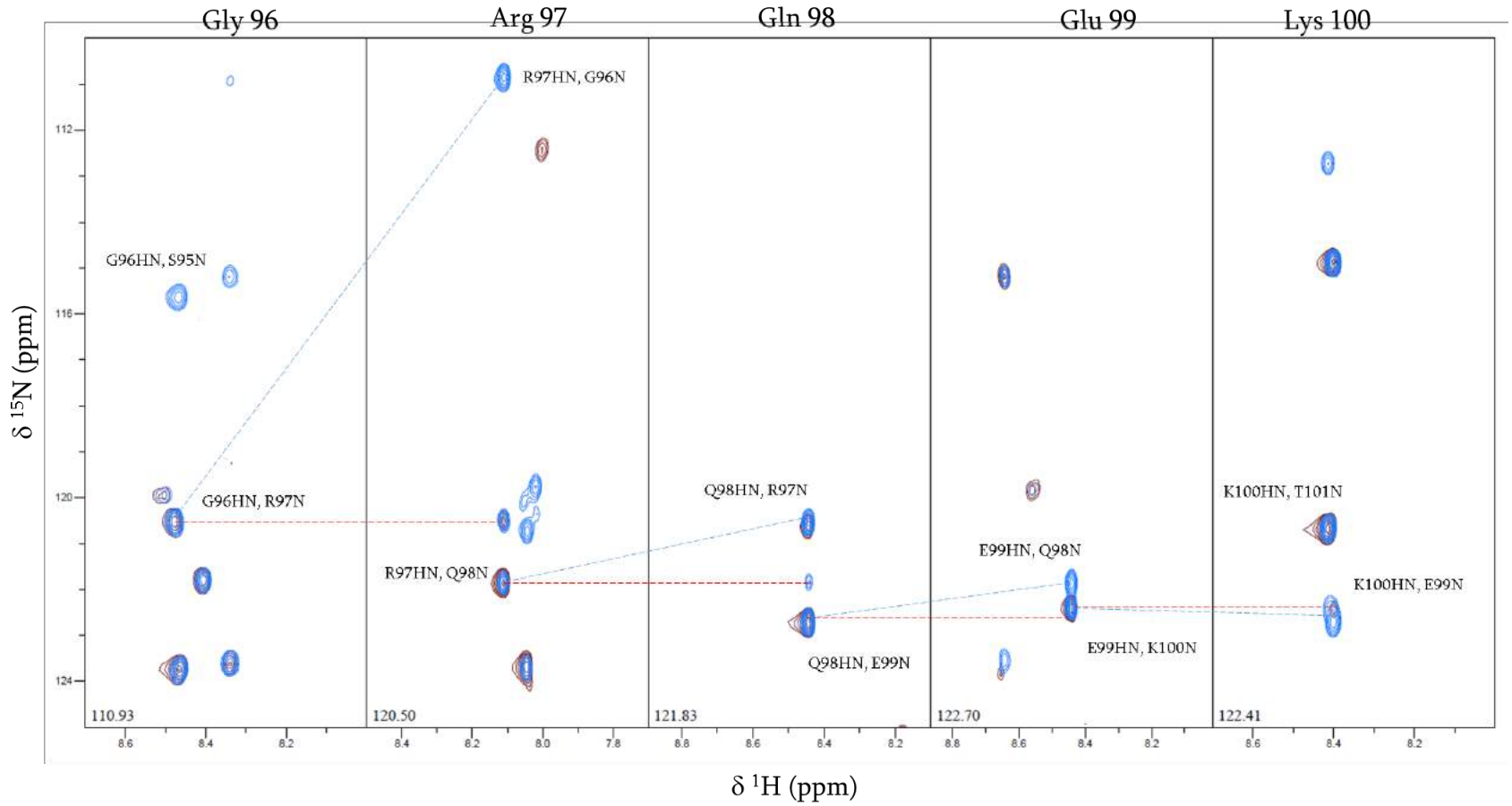


Figure 4.10 A schematics of sequential assignments for 96Gly to 100Lys in HNNH (blue) and HNCN (brown). From 96Gly, a ^{15}N chemical shift for 97Arg is obtained, the ^{15}N shift is 110.93 ppm (as shown in left corner of strip window) that corresponds to 97Arg ^{15}N in next strip window. So as a ^{15}N shift for 98Gln is observed at 120.50 ppm (as shown in the left corner of strip window). In a similar way, all sequential are correlated and confirmed

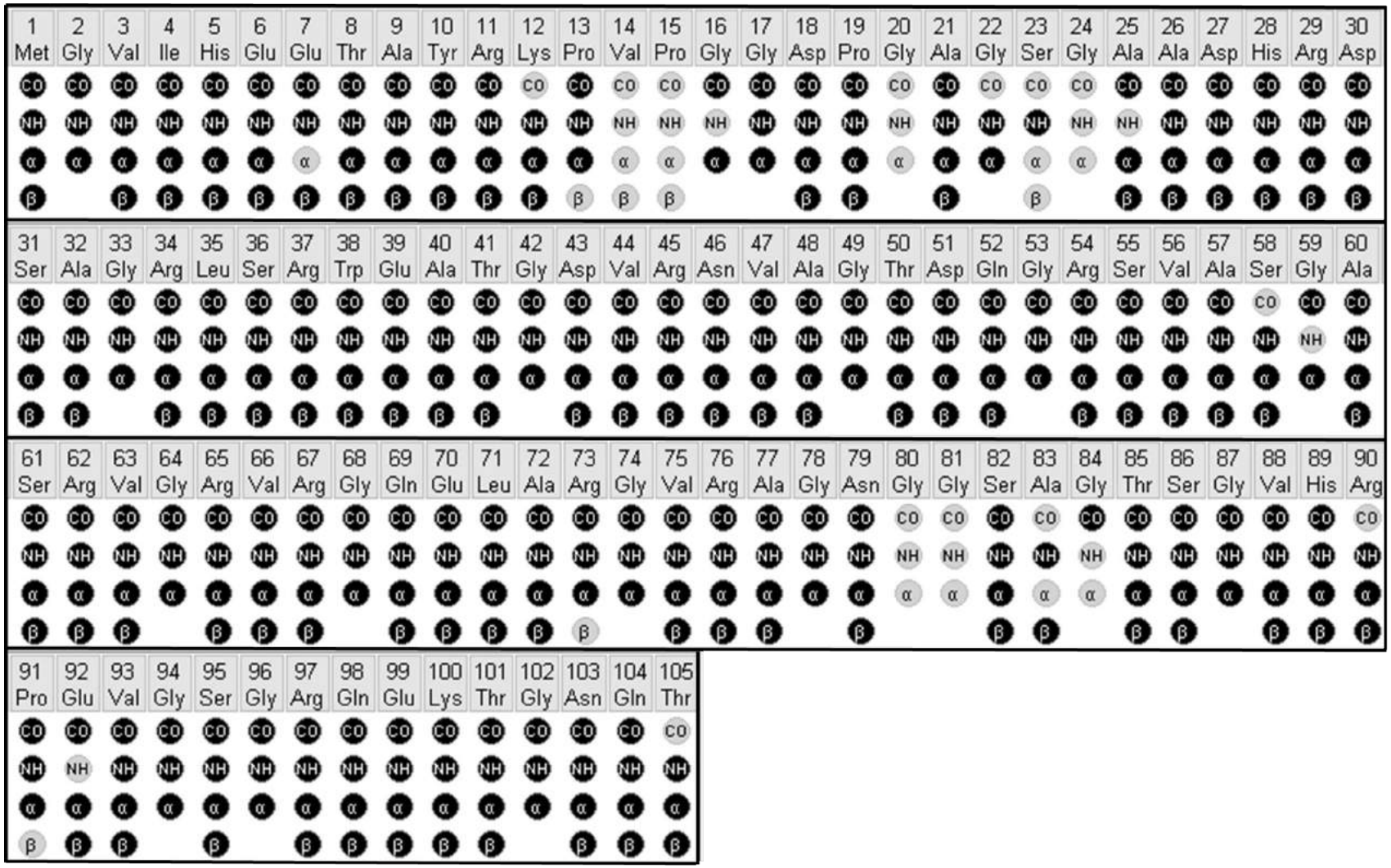


Figure 4.11 Extent of backbone assignments of IncC NTD. ~85 % of IncC NTD C′, Cαβ, N^H and H^N were assigned using CANCO, CBCACON, CBCANCO, HNCO, HNNH and HNCN spectra. Circles with black background show the presence of assignments for that nucleus.

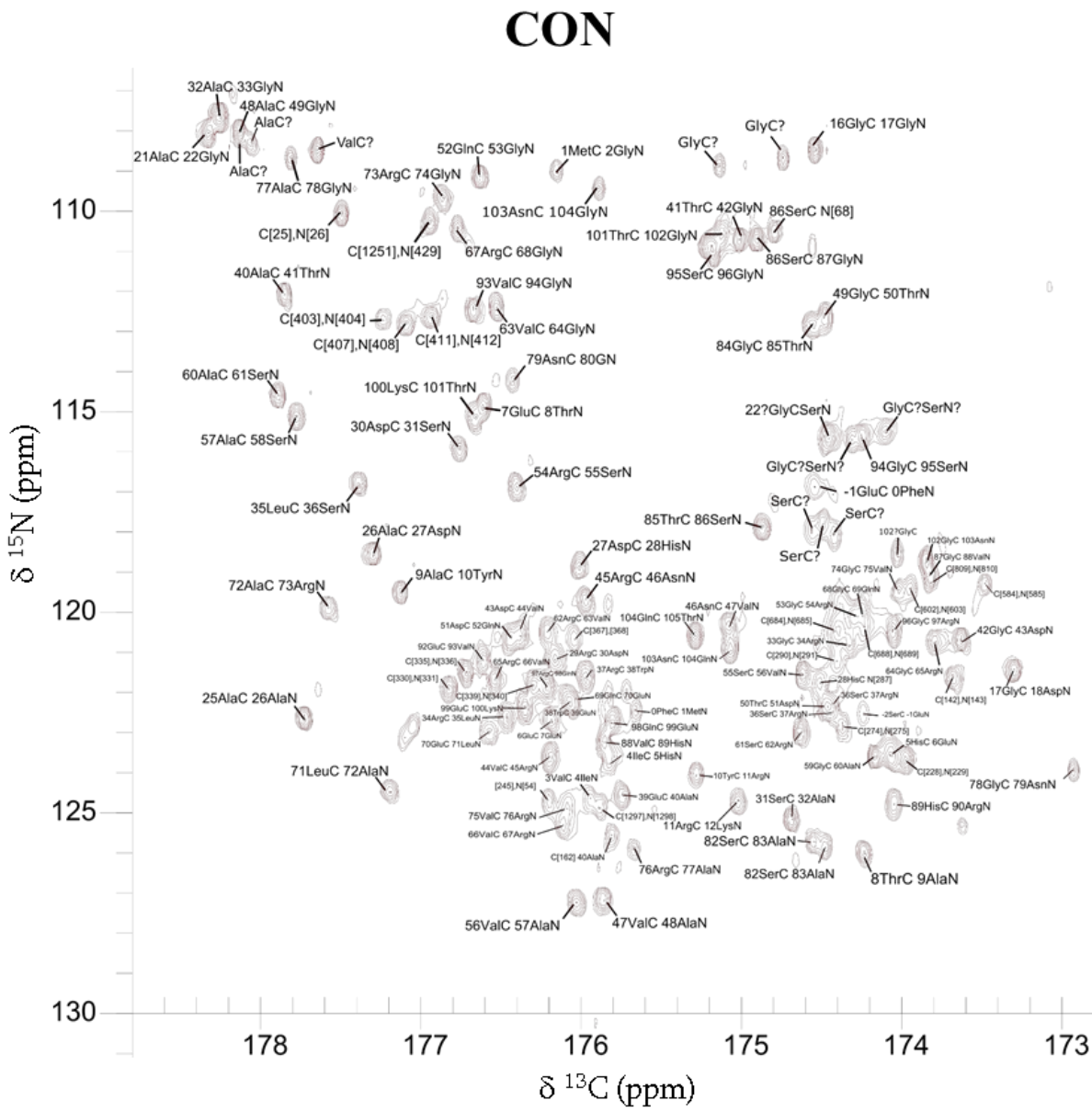


Figure 4.12 The assigned ^{13}C, ^{15}N CON spectrum for IncC NTD. The assigned spectrum shows Co (*i*), N^H (*i*+1) backbone assignments. The CON spectrum is used as the root experiment/spectra for protein backbone assignments using ^{13}C-detected spectra. This spectrum contains peaks from IncC NTD (105 amino acid) and some of His-tag peaks. Some Glycine, Alanine and Serine amino acids, due to overlapping peaks, could not be unambiguously assigned (labelled with '?'). Some peaks are located in multiple spectra but due to similar Cα/β shifts cannot be assigned and no definitive residue type could be determined (Spin system numbers are given). In the CON spectrum, more peaks are observed than expected. Some multiple peaks belong to a single residue, *i.e.* 82Ser, showing that multiple conformations are present.

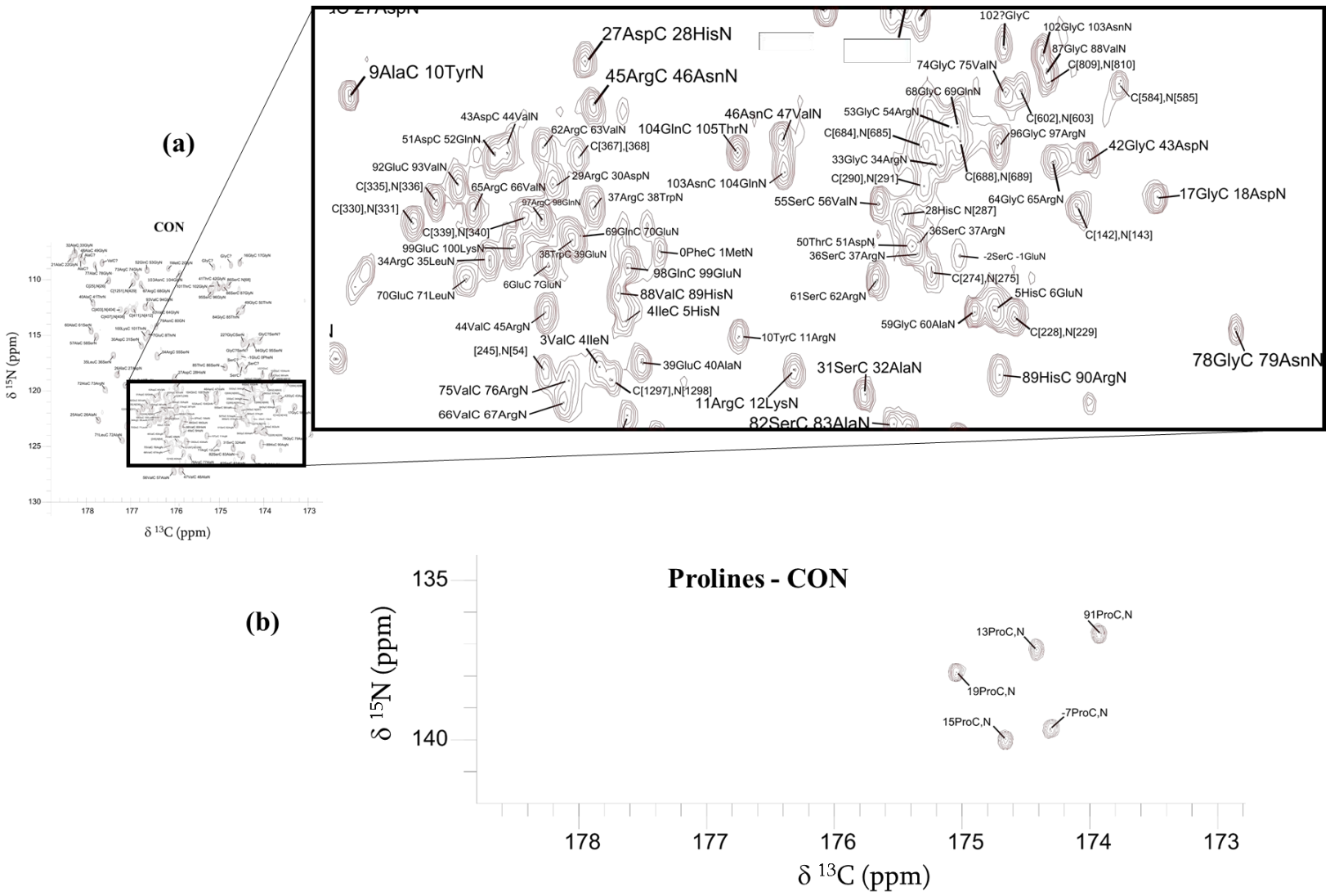


Figure 4.13 (a) The assigned ^{13}C, ^{15}N CON spectrum for IncC NTD. Inset shows zoomed area with overlapping assigned peaks. Most of the peaks in overlapping region are assigned but a few peaks are too ambiguous to assign **(b) The Proline assignments in the CON spectra.** Proline chemical shifts can be observed in carbon detected experiments, five proline amino acids of IncC NTD can be observed in CON around 136-142 ppm in ^{15}N dimension.

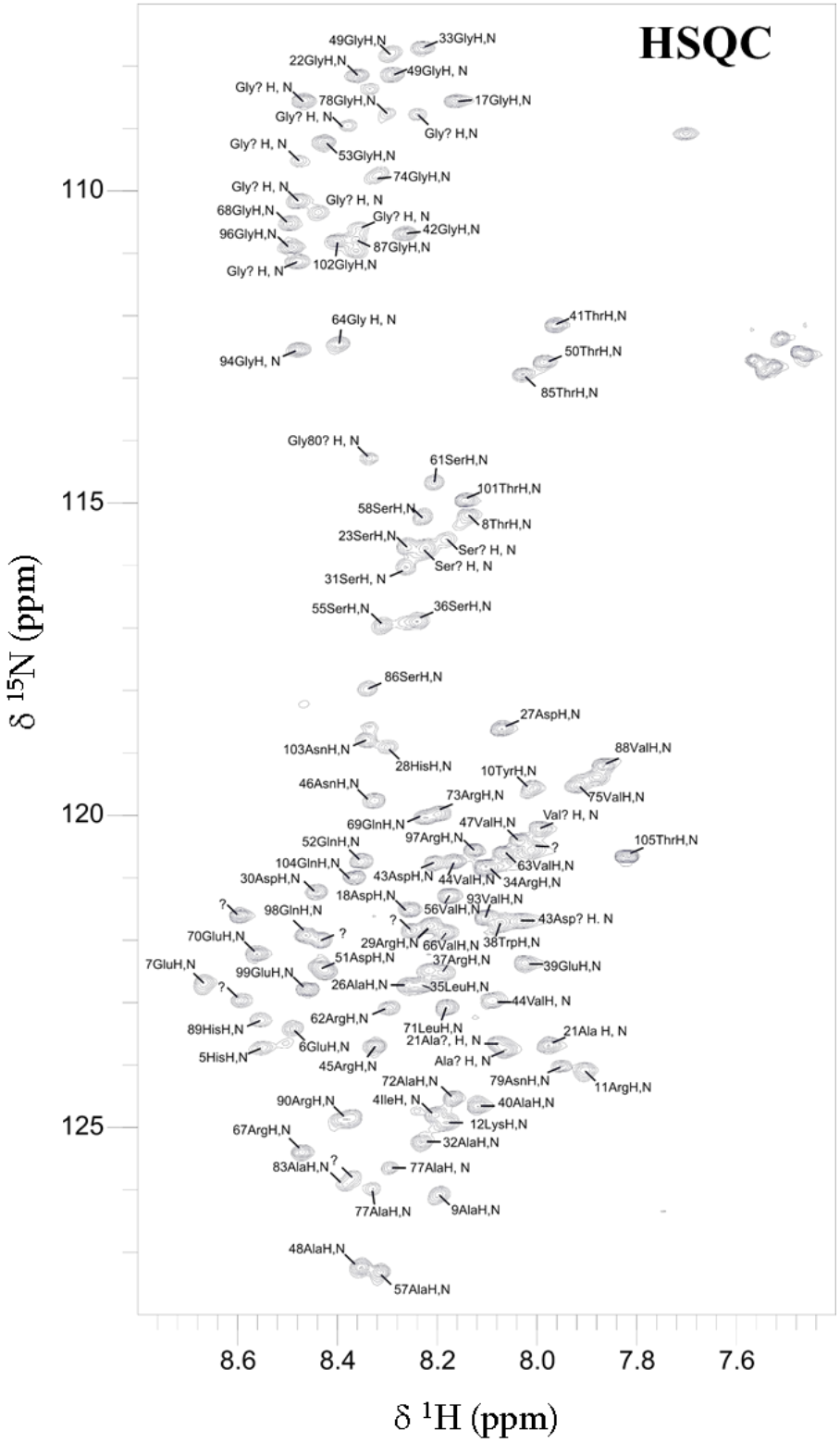


Figure 4.14 The assigned ^{1}H, ^{15}N HSQC spectrum for IncC NTD. HSQC spectrum showing ^{1}H, ^{15}N (*i*) assignments. Protein assignments are obtained from carbon-based experiment (CON) and by assigning the HNCO spectrum. Some peaks cannot be assigned (shown with '?'). Most of the His_6 tag related peaks, seen in CON, are not observed in the HSQC.

4.18 Secondary structure prediction for IncC NTD using chemical shifts

TALOS+ predicted that IncC NTD as a predominantly random coil with a small stretch of β-strand between 2-4 residues (Figure 4.15 and Figure 4.16). Regarding this stretch JPred, PsiPred and SABLE server predictions are in agreement with the TALOS+ server prediction. The TALOS+ prediction scores are shown in Figure 4.15. Shaded areas in the graph show some tendency towards secondary structure. β- strands may be present in regions with residue numbers 2-5, 14-16, 23-24, 36-40 and 76-81. A short region with high helical probability is seen in the region of residues 69-72 (Figure 4.15). The structure of IncC NTD was predicted to be highly flexible and very dynamic in nature by the TALOS+ server (Figure 4.17).

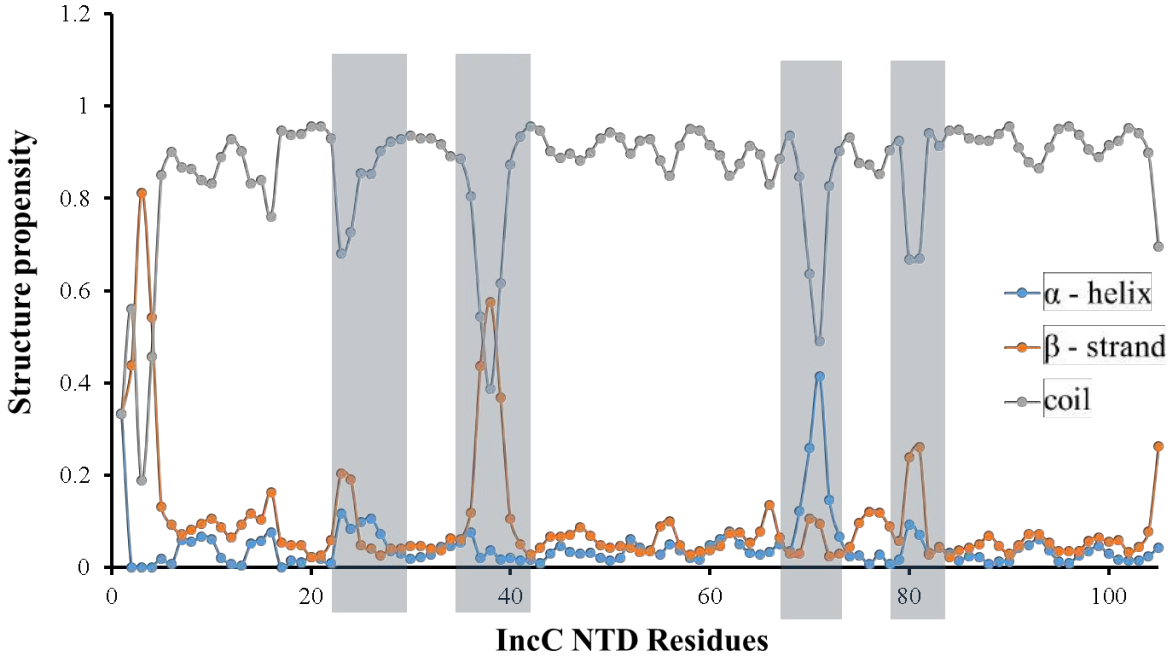


Figure 4.15 TALOS+ secondary structure prediction. The propensity for α- helix, β-strands and random coil in the IncC NTD structure. The shaded areas show that IncC NTD might have some secondary structure in that region. The TALOS+ server predicted the IncC NTD protein structure as a predominantly random coil with a small stretch of β-strand between 2-4 residues. β- strands may be present in regions with residue numbers 2-5, 14-16, 23-24, 36-40 and 76-8.

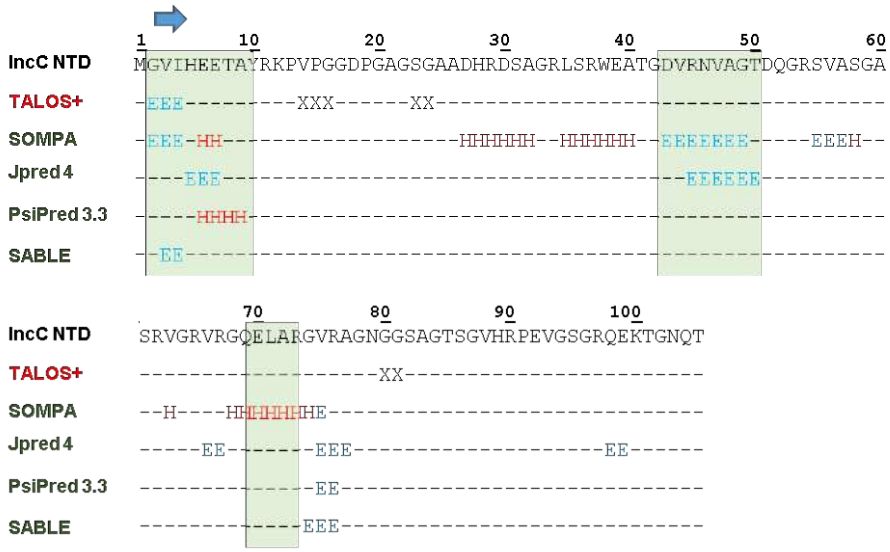


Figure 4.16 Comparison of secondary structure predictions of IncC NTD from the TALOS+ Server on the basis of NMR chemical shifts and predictions from web-based servers. E (blue) show β-strand , while H (red) shows α-helix. Prediction for a small stretch of β-strand is common among TALOS+. SOMPA, and SABLE.

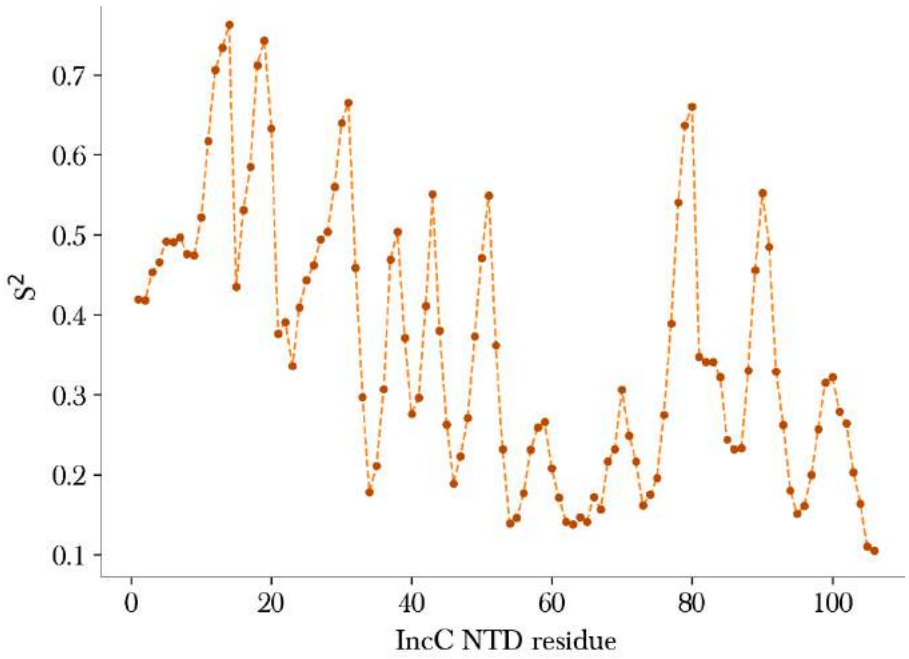


Figure 4.17 Generalized order parameter (S2) for IncC NTD. The IncC NTD protein was found to contain highly dynamic and flexible structure. A consistent S^2 value shows presence of stable secondary structure, while for IncC NTD values for S fluctuate throughout protein structure.

4.19 Prediction of the 3D structure of IncC NTD from the CS-Rosetta server

The CS-Rosetta server generated 5,000 *de novo* protein models with different RMSD scores. Ten best models out of 5,000 are selected on the basis of lowest energy score (RMSD). Score for selected models have been given in Table 4.2. Based on the lowest score the best IncC NTD model is selected.

Table 4.2 The RMDS scores for lowest energy models

Structure	Score
S_00374.pdb	7.87
S_02490.pdb	10.29
S_04103.pdb	9.18
S_02539.pdb	11.95
S_00868.pdb	11.29
S_01224.pdb	10.60
S_00347.pdb	12.05
S_00153.pdb	10.8
S_04437.pdb	9.36
S_04408.pdb	9.16

The IncC NTD model S_00374.pdb is selected and visualized using Chimera v 1.12. Protein structure contains two small stretches of α-helix; α1 from 59Gly to 63Val and α2 from 69Gln to 73Arg (Figure 4.18). Region for the α1(69Gln to 73Arg) also has been predicted in several other models with different RMSD (Figure 4.19). The Ramachandran plot of S_00374.pdb shows all residues are in favourable regions (Figure 4.20).

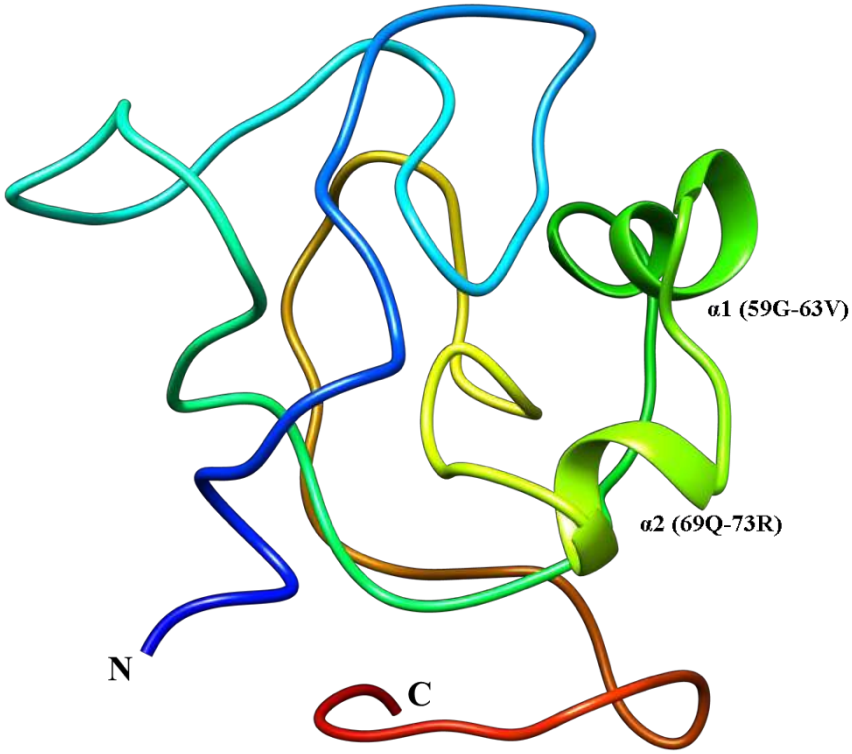


Figure 4.18 *De novo* IncC NTD structure model based on NMR chemical shifts. The IncC NTD model is obtained using the NMR chemical shifts. Two small stretches of α-helix have been observed α1(59G-63V) and α2 (69Q-73R) in protein structure.

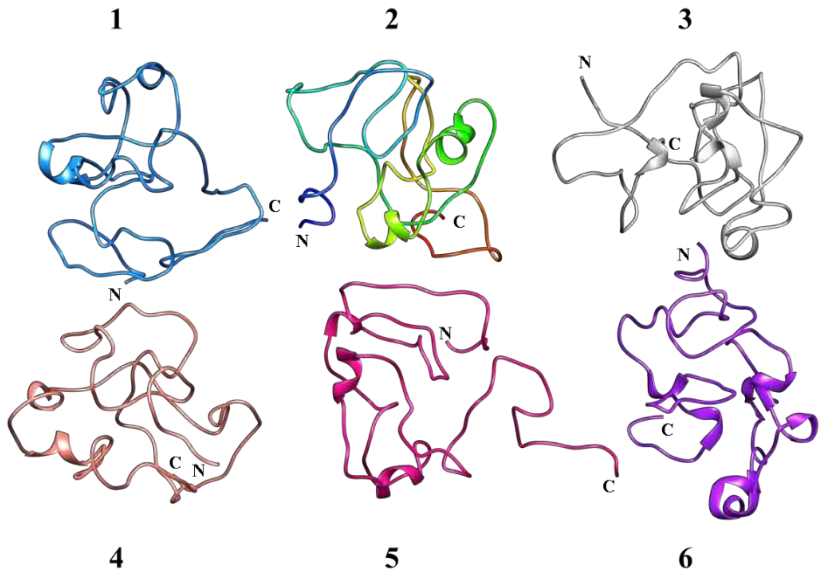


Figure 4.19 Different structures obtained from CS-Rosetta. The lowest energy model is # 2. The structure # 1, 3, 4 and 6 share α-helix in the region of 69Q-73R. Due to highly dynamic protein structure, a consensus model could not be obtained. The displayed six structures are selected with respect to lowest Rosetta energy score.

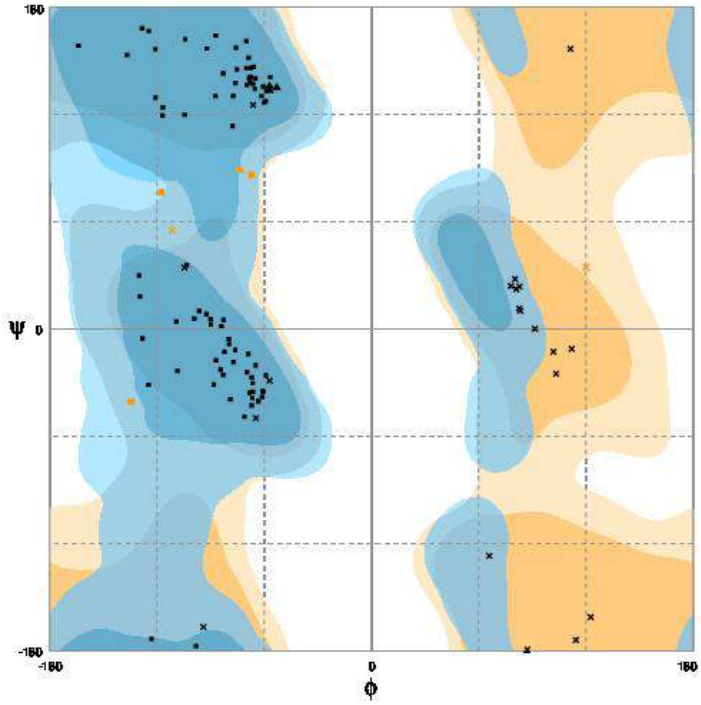


Figure 4.20 Ramachandran plot of IncC NTD best model (S_00374.pdb). The plot was obtained from the web-based server Rampage (http://mordred.bioc.cam.ac.uk/~rapper/rampage2.php) by plotting phi angles on the x-axis and Psi angles on y-axis. All phi and psi values are found in favourable or allowed regions. The blue shows favourable regions, while orange shows allowed regions.

4.20 HSQC of IncC1 and IncC NTD

The HSQC spectrum for the IncC1 protein is compared to HSQC of IncC NTD. Peaks in the two spectra are found to be almost identical, when overlaid each other except that some peaks were shifted due to use of different sample buffers for two proteins (Figure 4.21). Buffer used for IncC1 is optimised for its solubility and is different from the IncC NTD buffer, so peak shifts can be expected. IncC1 is a 40 kDa protein and may form multimers (dimers, tetramer and octamers) in solution. Most of HSQC peaks observed for IncC1, are seen in the IncC NTD HSQC spectrum (Figure 4.21) showing IncC NTD remains flexible even in the full-length IncC1 protein. The peaks corresponding to rest of IncC1 are not seen due to slow tumbling and large size of the IncC1 protein.

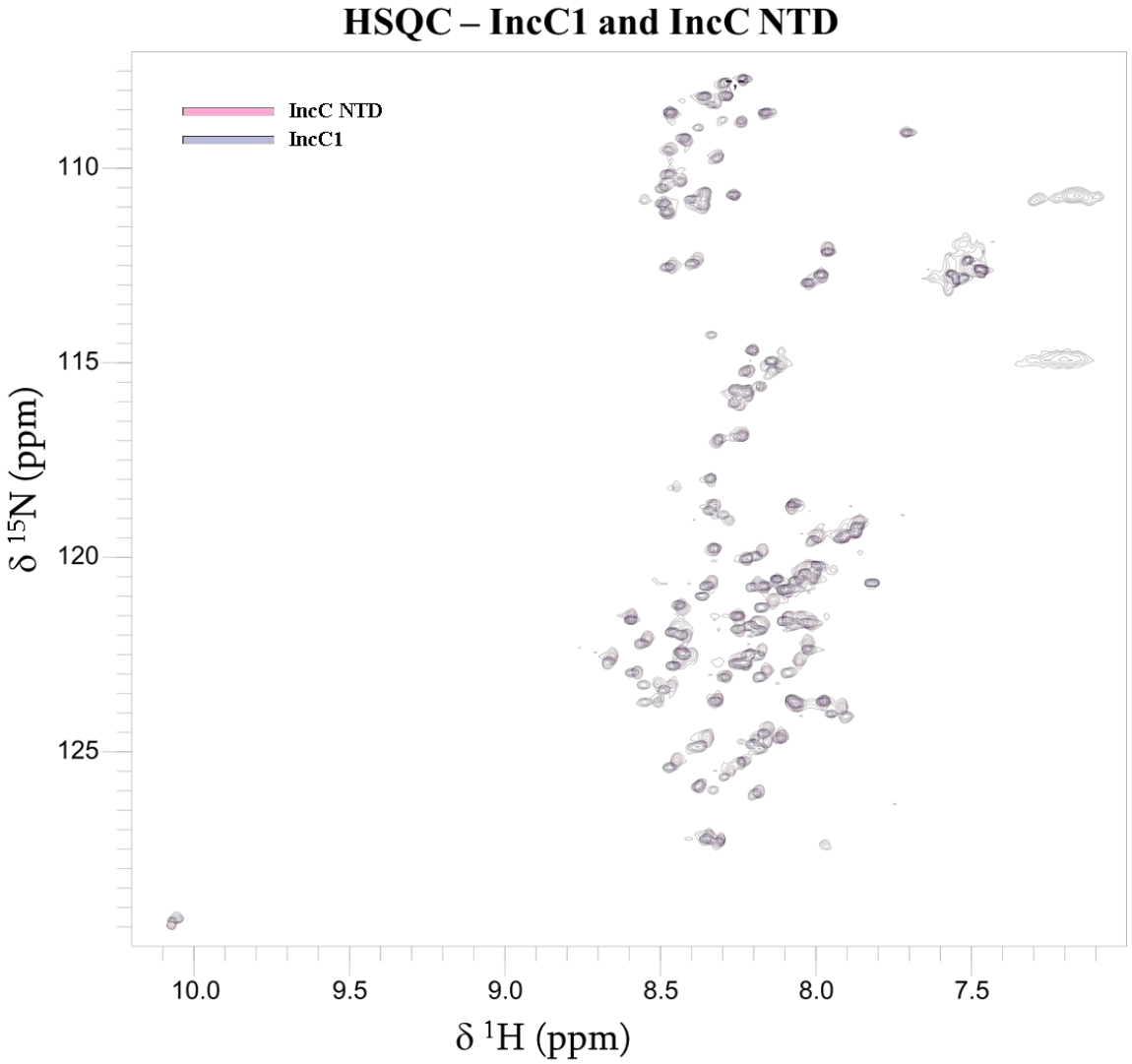


Figure 4.21 Comparison of HSQC from IncC1 and IncC NTD. Most of peaks in HSQC from IncC1 are same for the IncC NTD, showing it is most flexible part of full-length protein and remains unfolded. Some peaks are found to be shifted in IncC1 HSQC probably due to use of two different buffers for these two proteins.

4.21 IncC NTD DNA binding

Dodecamer double-stranded DNA was used to monitor DNA-protein interactions. HSQC spectra of the Apo IncC NTD were compared to IncC NTD spectrum in the presence of DNA (Protein: DNA, 4:1 ratio) (Figure 4.22). Some peaks, in presence of DNA, were slightly perturbed and have been labelled (Figure 4.22). The perturbed peaks are found to come from segments of the IncC NTD protein, rather than random individual amino acids. Peaks of residues 4-6 have been shifted. Another region where peak shifts are observed, include amino acids 37-40. Peaks from residues 64, 67-68, 70-74 also shifted, while intensity of a Ser61 peak is decreased. 90Arg and 99Glu peaks are also shifted. The comparison of HSQC spectra suggests that IncC NTD binds to DNA although binding is very weak, even in the 4:1 ratio of DNA to protein, so only slight peak shifts are observed.

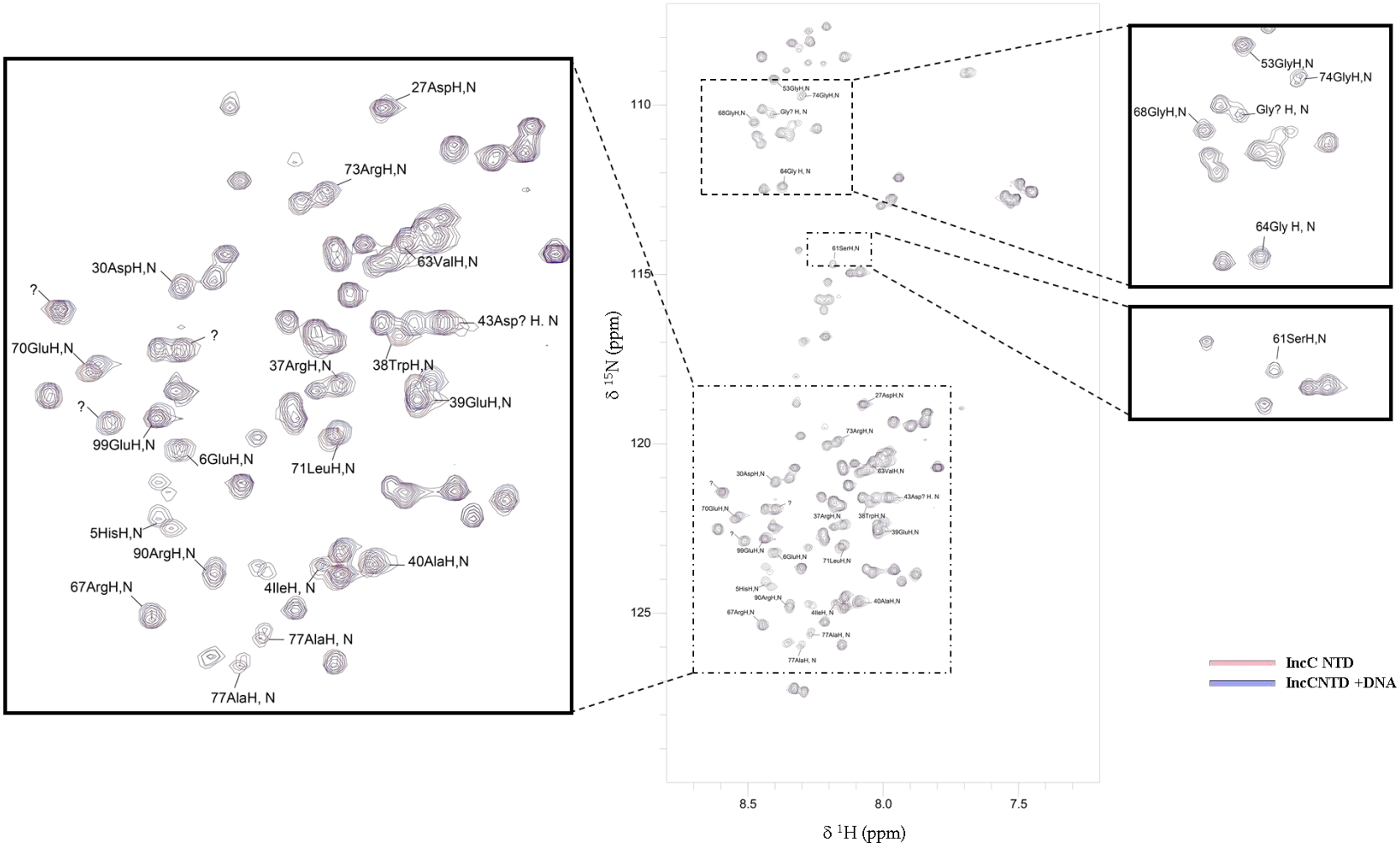


Figure 4.22 An overlay of the HSQC spectra of IncC NTD in the presence or absence of dodecamer DNA. Pink colour shows only IncC NTD (200 μM), while blue shows IncC NTD (200 μM) in the presence of Dodecamer DNA (800 μM). Both spectra were acquired using a 900 MHz Bruker NMR spectrometer at 298K.

4.22 Discussion

IncC NTD is an N-terminal region (105 a.a) present in the full-length IncC1 protein and may play an important role in plasmid partitioning. To elucidate the role of IncC NTD, biophysical methods have been used and it was found that IncC NTD is largely disordered (Chapter 3). Due to the predicted dynamic nature of protein, the backbone structure of the IncC NTD protein was probed using NMR. The ^{1}H ^{15}N HSQC spectrum of the IncC NTD showed overlapping peaks and overall poor peak dispersion. Disordered proteins always show very low chemical shift dispersion in the HSQC proton dimension and obtained proton observed spectra are very difficult to assign (Sahu et al., 2014). Carbon-detected NMR experiments can be used to circumvent poor spectrum quality and to get greater peak dispersion in carbon dimension (Gray et al., 2012). It also has been observed that carbon detected experiments are not affected by exchange of amide protons with water (Felli and Pierattelli, 2012). There is one limitation of carbon-detected protein NMR experiment and that is lower sensitivity in comparison with proton-detected experiments (Sahu et al., 2014). For the IncC NTD backbone assignments, carbon-detected experiments are performed instead of conventional proton-based experiments *i.e.* CON, CANCO, CBCACON and CBCANCO. To overcome the problem of low sensitivity, a triple-resonance probe optimized for carbon sensitivity, TXO CryoProbe, has been used in the carbon-detected experiments. CON has been used as a root spectrum. Previously, CON spectrum has been used to get excellent chemical shift dispersion for various IDPs *e.g.* Pdx1-C (Sahu et al., 2014) and ERD14 (Kosol et al., 2013).

Carbon detected experiments complement each other to get protein backbone assignments. In some previous studies, a 2D CON experiment has been combined with a 3D carbon-detected experiments to get more than 98% backbone assignments for the Oncomodulin protein (Babini et al., 2004). For the IncC NTD backbone assignments, unique shifts of certain Cα and Cβ peaks, from CANCO, CBCACON and CBCANCO experiments, are used to assign amino acids with respect to the reference chemical shifts. Most of the time, Ala, Ser, Thr and Leu are easy to predict from unique Cβ chemical shifts. But sometimes, especially when one of the sequential residues is Glycine, sequence specific assignment becomes more difficult. In IncC NTD, some of the N-terminal His_6 tag peaks are also observed, but due to high flexibility and high similarity of amino acids, many overlapping peaks, with very similar chemical shifts are observed and could not be assigned unambiguously. Some of the side chains carbonyl peaks

from Asn and Gln) are also observed in CON. In some cases, two different peaks are observed for a single residue, suggesting conformational changes in protein on a slow timescale.

In the assigned ^{15}N, ^{13}C-CON spectrum, each peak is correlated with position of amino acid in a protein sequence. In the CON spectrum, most of the glycine peaks are clustered around ^{15} N 108-115 ppm, while Serine peaks are grouped in ^{15}N 115-119 ppm. Around ^{15}N 120-125 ppm, many overlapping peaks observed and are difficult to assign. About 60 % of the IncC NTD backbone assignments are obtained using these 13C-detected experiments and rest of backbone assignment (to ~85%) are completed using HNNH and HNCN experiments. Chemical shifts derived from carbon-detected experiments are combined with chemical shifts from HNNH and HNCN using the HNCO experiment as an intermediate. The HNN and HN(C)N experiments have been considered as very successful experiments to study various unfolded protein (Panchal et al., 2001). After adding information from these experiments, some IncC NTD amino acids including, stretches with Serine, Glycine and Alanine are still found to be very ambiguous and are impossible to assign in a reliable way. Proline residues are abundant in IDPs (Dunker et al., 2008) and protein-detected experiments do not give signals for these residues. It is an additional advantage of carbon-detected experiments that Proline-rich proteins can be assigned by NMR (Dziekanski et al., 2015). In IncC NTD, five Proline amino acids are found in CON around ^{15}N 135 ppm and have been assigned successfully.

The HSQC spectrum for IncC NTD has been assigned with the help of HNCO and has been used in further DNA binding experiments. The results for secondary structure prediction from TALOS+ show a small beta strand in the the region of residue 2-4 that agrees with SABLE server secondary structure prediction. CS-Rosetta is used to obtain De novo model IncC NTD on the basis of chemical shift assignments. Two small stretches of α-helix are observed in de novo structure; an α1-helix from 59Gly to 63Val and α2-helix from 69Gln to 73Arg (Figure 4.18). The secondary structure precision from the SOPMA server predicts helix in the α2 region (69Gln to 73Arg) (Figure 4.16). In the TALOS+ prediction (Figure 4.15), this region shows some tendency towards α-helix. The same secondary structure is also observed in several other predicted models with different RMSD obtained from CS-Rosetta (Figure 4.19). The other regions with probable secondary structure in IncC NTD (Figure 4.15) might contain some short stretches of β-strand as they were predicted by all three secondary structure prediction servers (Figure 4.16) or it may be suggested that these regions might go into some transition when IncC NTD interact with other partners (DNA, other proteins).

The HSQC spectrum of the full-length IncC protein (IncC1) shows that IncC NTD is unfolded, even as N terminal region of full-length protein. The chemical shift differences observed between the IncC1 and IncC NTD HSQCs are likely due to different buffer being used (Figure 4.21).

The IncC NTD binding to DNA is studied using the dodecameric DNA. The DNA is palindromic and is restricted to small size to prevent the binding of large numbers of protein molecules to the DNA. IncC NTD binds to DNA weakly and likely non-specifically, it may suggest that protein backbone or some arginine side chains are involved in DNA binding. In the secondary structure predictions, by NMR and other methods, no structured DNA binding motif (for instance helix turn helix) was found. For IncC NTD-DNA HSQC spectrum, peaks in various regions are shifted. Though these regions are predicted predominantly as random coil they also contain some propensity to form secondary structures. It may be suggested that these regions may form a secondary structure when interacting with some partner *e.g.* other proteins and DNA. Peaks from the amino acids 4-6 and 37-40 are shifted with the addition of DNA (Figure 4.22). In the TALOS+ predictions, this region shows some propensity to form a beta strand. The peaks for residues 63Val, 64Gly, 67Arg, 68Gly, 70Glu, 71Leu, 74Gly, 77Arg are also shifted with DNA addition. In the TALOS+ predictions, residues 69-72 have a propensity to form an α helix, while the region 76-81 amino acids show β-strand propensity (Figure 4.15). The 90Arg and 99Glu peaks are also shifted in the presence of DNA.

CHAPTER 5

Optimization of expression, purification and functional characterization of IncC1 and IncC2 proteins

Chapter 5

Optimization of expression, purification and functional characterization of IncC1 and IncC2 proteins

5.1. IncC1and IncC2 proteins

Low copy number plasmids, as stated in chapter 1, require ParA and ParB proteins for plasmid partitioning and stable maintenance in daughter bacteria. IncC proteins are ATPases and belong to the ParA protein family. Low copy- number RK2 plasmid expresses two IncC proteins (IncC1 and IncC2) simultaneously in the cell. The IncC proteins interact with KorB (a ParB protein) and DNA during the partitioning process.

In this chapter, IncC proteins have been expressed and purified to determine their ATPase activity and interactions with KorB and DNA. The IncC2 protein consists of 259 amino acids (105 amino acid shorter than IncC1) and is highly positively charged with a calculated pI of 9.47 (amino acid composition and properties are given in Appendix 3). IncC2 protein was overexpressed using an overexpression vector (pET21a) with a His_6 tag. The protein was highly unstable and insoluble. Initially, the IncC2 protein aggregated after bacterial cell lysis or affinity purification. To enhance protein stability and solubility, several factors, described in Table 5.1, were optimized.

5.2. Protein solubility and stability

Protein aggregation is typically induced by protein self-association and by the formation of altered conformations rather than the native one. Protein expression and purification conditions play an important role in obtaining proteins in soluble form. Some bacterial strains have insufficient protein folding machinery and fail to keep proteins (most of the eukaryotic proteins) in a folded form during overexpression *e.g*. lack of chaperones, lack of post-translation modifications (PTMs) and no compartmentalisation.

While purifying proteins, several conditions should be controlled, including buffer composition, pH, ionic strength and concentration of the proteins. Some proteins are highly prone to aggregation at a higher concentration than present in their native environment. Many strategies have been developed to avoid protein aggregation during expression and purification.

Table 5.1: Factors contributing to protein stability and solubility during protein expression and purification

Protein Expression	
Effectors	Functions
Promoters	Choice of various promoters to control protein overexpression levels
Fusion tags	Use of different tags to enhance protein solubility during overexpression and purification
Bacterial strains	Use of different bacterial strains with specific properties
Growth and induction conditions	Optimization of growth media (composition), incubation time, temperature, pH and inducer concentration
Protein Purification	
Effectors	Functions
Lysis and purification buffers	Screening of different buffer components (for pH and ionic strength) to avoid protein aggregation during cell lysis and protein purification
Buffer Additives	Additives like Glycerol, Sucrose, PEG, Potassium Glutamate, Glutamine, DTT, β-Mercaptoethanol, TCEP increase the protein solubility and stability at an optimized concentration
Ligands	Ligands, including metal ions or other biomolecules, provide stability to proteins
Non-ionic detergents	Proteins with surface hydrophobic patches or membrane-associated proteins can be stabilized and solubilized using non-ionic detergents

5.3. Factors affecting soluble protein expression

5.3.1. Choice of promoters

Tightly regulated promoters play a vital role in the soluble protein overexpression. These promoters provide reproducible experimental conditions for large-scale protein production. Protein expression can be regulated and improved using low copy number plasmids. T7-promoter-based vectors (pMB1 ori, medium copy number) are the most widely used promoters to express proteins, in bacterial strains with a chromosomal copy of T7 RNA polymerase (λDE3) (Studier and Moffatt, 1986). Other promoters include the *ara*C promoter (present in pBAD vectors), induced by arabinose (Guzman et al., 1995) and the cold shock promoter, *csp*A, that is induced by a temperature change (pCold vectors) (Qing et al., 2004). The AraC protein plays the dual role of repressor and activator while controlling the *ara*C promoter. While expressing proteins using pCold vector and *csp*A promoter, temperature shift to low temperatures suppresses the expression of cellular proteins and temporarily halts cell growth (Qing et al., 2004).

5.3.2. Solubility tags and fusion proteins

Unstable and poorly soluble proteins can be expressed as fusion proteins. The solubility tag is present at either the C- or N-terminal of the fusion proteins. Although the fused tags can affect the properties of partner proteins, they show high success rates in stabilizing and solubilizing protein expression in *E. coli* strains (Esposito and Chatterjee, 2006). High throughput screening (HTS) platforms can be used to determine the best fusion tag for the proteins. Properties of some fusion tags are given in table 5.2.

Table 5.2: Some commonly used fusion tags and their properties

Tag	Protein	Size (a.a)	Organism	Reference
MBP	Maltose binding protein	396	*E. coli*	(Costa et al., 2014)
NusA	N-utilizing substance	495	*E. coli*	(Davis et al., 1999)
Trx	Thioredoxin	109	*E. coli*	(LaVallie et al., 2000)
GST	Glutathione-*S*-transferase	211	*Schistosoma japonicum*	(Smith and Johnson, 1988)
SUMO	Small ubiquitin modifier	~100	*Homo sapiens*	(Butt et al., 2005)
GB1	IgG domain B1 of Protein G	56	*Streptococcus sp.*	(Zhou and Wagner, 2010)
SET	Solubility enhancer peptide	<20	Synthetic	(Zhou et al., 2001)
Halo	Mutated dehalogenase	~300	*Rhodococcus sp.*	(Ohana et al., 2009)
ZZ	IgG repeat domain ZZ of Protein A	116	*Staphylococcus aureus*	(Rondahl et al., 1992)

The MBP tag was found to enhance protein stability by several folds but its large size (44 kDa) makes it unsuitable to fuse with various proteins. In many cases, after cleaving MBP, partner proteins aggregate in the solution. MBP works well only for small proteins (≤ 40 kDa) as the bacterial machinery is insufficient to express monomeric proteins >90 kDa (Lebendiker and Danieli, 2014). MBP can be combined with a His tag to purify proteins using IMAC chromatography. NusA is another fusion tag of 55 kDa. Sometimes, it is more efficient than MBP, but its size is a problem and. as with MBP, the fused protein may aggregate after tag cleavage (Nallamsetty and Waugh, 2006). In contrast to these tags, Trx (Thioredoxin) is a small fusion tag (~11 kDa) and helps in the stability and solubility of disulphide-containing proteins. It assists in refolding of proteins that require a reducing environment (Costa et al., 2014). The GST tag is basically considered as an affinity tag and most of the times does not contribute to protein solubility. The GST tag can be used in pull-down assays and in protein-protein interaction experiments. The SUMO (small ubiquitin-related modifier) tag has an advantage over other tags due to its smaller size (12 kDa) and having a specific protease (SUMO protease) cleavage site. The SUMO tag increases the solubility of proteins to almost same extent or more than the MBP

tag (Malakhov et al., 2004). GB1 is the immunoglobulin-binding domain B1 of streptococcal protein G. It consists of only 56 residues and, due to its small size (6.2 kDa) and high solubility, this GB1 domain can be used as solubility/stability tag. The NMR spectra for the GB1 domain have been assigned; that makes it easy to study GB1 fusion proteins by NMR and identify GB1 protein peaks without cleaving the tag (Zhou and Wagner, 2010).

5.3.3. Bacterial strains

Protein expression varies, in different *E. coli* strains, even under same conditions. BL21-based strains and K12-based strains may give a completely different solubility and stability to the same overexpressed protein due to some poorly understood reasons (Lebendiker and Danieli, 2014). The BL21-based strains like BL21 (λDE3) lack Lon and Omp-t proteases. These strains are highly suitable for protein expression. BL21(λDE3)-pLysS is the best strain for toxic protein expression as T7 lysozyme, expressed by pLysS plasmid, reduces the expression of target genes by inhibiting T7 RNA polymerase (Studier and Dunn, 1983). BL21 Star and BL21 Star (DE3) strains allow higher protein expression due to having a mutated RNaseE protein and hence reduced RNA degradation. T7 Express strain is a BL21 derivative. In this strain, the T7 RNA polymerase gene is expressed under the control of the wild-type *lac* promoter and provides lower basal expression of the gene of interest than in in BL21 (DE3) strains comparison, where the expression of T7 RNA polymerase is controlled by the lacUV5 promoter. A T7 Express strain derivative, T7 Express-LysY, expresses T7 lysozyme to control the T7 promoter and subsequently toxic gene expression. Some bacterial strains, like Origami™ strains, possess mutated thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes that facilitate the formation of disulphide bonds in the cytosol (Lobstein et al., 2012). This shows that the selection of appropriate bacterial strain is very important for stable and soluble protein expression.

5.3.4. Protein expression conditions

Protein aggregation can be prevented during protein expression by varying the conditions *i.e.* temperature, induction time, inducer concentration, and media composition. Lowering temperature, during the gene induction, allows slow peptide chain elongation and proper domain folding. Slowing down the process provides chaperones more time to fold proteins properly; that also increases protein solubility (Thomas and Baneyx, 1996). Induction duration determines the amount of overexpressed protein in the cell. Lowering the inducer concentration will reduce the production of mRNA and may prevent aggregation of elongating peptide chains. Media

composition also plays a crucial role in bacterial cell growth that subsequently effects proteins production. Adding osmolytes to growth media *e.g.* glycerol, sorbitol, proline, Potassium glutamate and benzyl alcohol, helps in proper protein folding (de Marco et al., 2005). Use of an autoinduction media and the media composition, *e.g.* different ratios of peptones: yeast extract: sodium chloride, also affects protein solubility (Lebendiker and Danieli, 2014). Various approaches can be combined to get the best results for protein stability and solubility.

5.4. Protein stability and solubility during purification

5.4.1. Buffer choice and additives

Both pH and salt concentration during purification effects protein aggregation and polydispersity (Churion and Bondos, 2012). The pH of lysis buffer can be adjusted according to the theoretical pI (isoelectric point) of the desired protein. After pilot protein production and purification, high throughput buffer screens can be used to get lysis buffer optimized for stable and soluble protein purification (Boivin et al., 2013).

5.4.2. Protein concentration

After cell lysis, some proteins are highly vulnerable to aggregation. The ratio of lysis buffer per gram of cell pellet can be varied between 3-10 mL of buffer/g of cell pellet. Mechanical or non-mechanical stresses, such as during sonication, or temperature, can be minimised to avoid protein denaturation during cell lysis. After obtaining the clear cell lysate, a quick/optimized purification step may also help to avoid protein aggregation. Soluble protein can be purified by avoiding column overloading and by diluting purified protein fractions using the optimized buffer.

5.4.3. Stabilizing agents

Many buffer additives can be used to increase protein stability and solubility. These include osmolytes *e.g.* amino acids (glutamine, glutamic acid), sugars alcohols (glucose, sucrose), polyols (glycerol, PEG). Kosmotropes (*e.g.* trehalose, glycine, betaine) and non-ionic detergents (*e.g.* DDM, Nonidet P40) also stabilize proteins by preventing aggregation. Ligands, including co-factors, and reducing agents (*e.g.* DTT, TCEP) can change protein conformation and increase or decrease protein stability/solubility.

Protein purification at low temperatures reduces the chances of protein aggregation. Protein purification under various conditions can be optimized by pilot protein production. The

thermal shift assay provides a good way to monitor protein unfolding at different pH and buffers. Ligand and detergent screening and optimization can be performed with this assay. Protein aggregates can be monitored by linear light scattering or dynamic light scattering.

5.5. IncC2 computational analysis

Before the experimental work, the secondary structure of the proteins was predicted using various online servers.

5.5.1. Secondary structure prediction

The IncC2 protein secondary structure was predicted using the same servers used for IncC NTD and also the PredictProtein server (https://www.predictprotein.org) (Rost et al., 2004). The PredictProtein server (PROF-sec) predicts protein secondary structure as well as solvent accessibility of protein residues. The server uses a set of neural networks and the multiple sequence alignment obtained from a BLAST search (Altschul et al., 1990) and MaxHom alignment (Dodge et al., 1998). For IncC2 protein structure prediction, a multiple sequence alignment of 33 different proteins was used by PredictProtein server. IncC2 was found to be predominantly α helical with few extended β sheets (Figure 5.1a, and d). Analysis of the buried and exposed amino acid stretches shows that the IncC2 is a globular protein with a hydrophobic core (Figure 5.1b and c). Two neural networks from the PredictProtein server, *i.e.* TMSEG and PHDhtm (Rost et al., 1996), predict transmembrane proteins (TMP) and transmembrane helices (TMH). These neural networks use position-specific scoring matrices (PSSM), generated by PSI-BLAST. TMSEG has predicted a transmembrane region in the IncC2 protein (Figure 5.1c). This predicted region may be a surface exposed, hydrophobic amino acid patch involved in protein-protein interactions. The multiple sequence alignment and motif predictions from the PredictProtein server, show that the IncC2 protein sequence contains Walker-like motifs *i.e.* Walker-like motif A (10-16 a.a) and Walker-like B motif (122-126 a.a) (Figure 5.2). The Walker-like A motif is present in the P-loop region in the NTPase domain (2-44 a.a). Other conserved regions were also observed, including an α-β-α motif (214-245 a.a).

5.5.2 Protein disorder prediction

The disordered regions in IncC2 were predicted using the various servers as described in Chapter 3 (section 3.11). The disorder predictions were consistent with the secondary structure predictions. All the prediction servers showed that IncC2 is a natively folded protein (Figure 5.3a). The charge-hydropathy plot (CH Plot) (Huang et al., 2014) showed IncC2 to be among the

ordered proteins when different proteins, known to be either natively folded and unfolded, were plotted on the basis of absolute mean charge *vs* mean scaled hydropathy (Figure 5.3b).

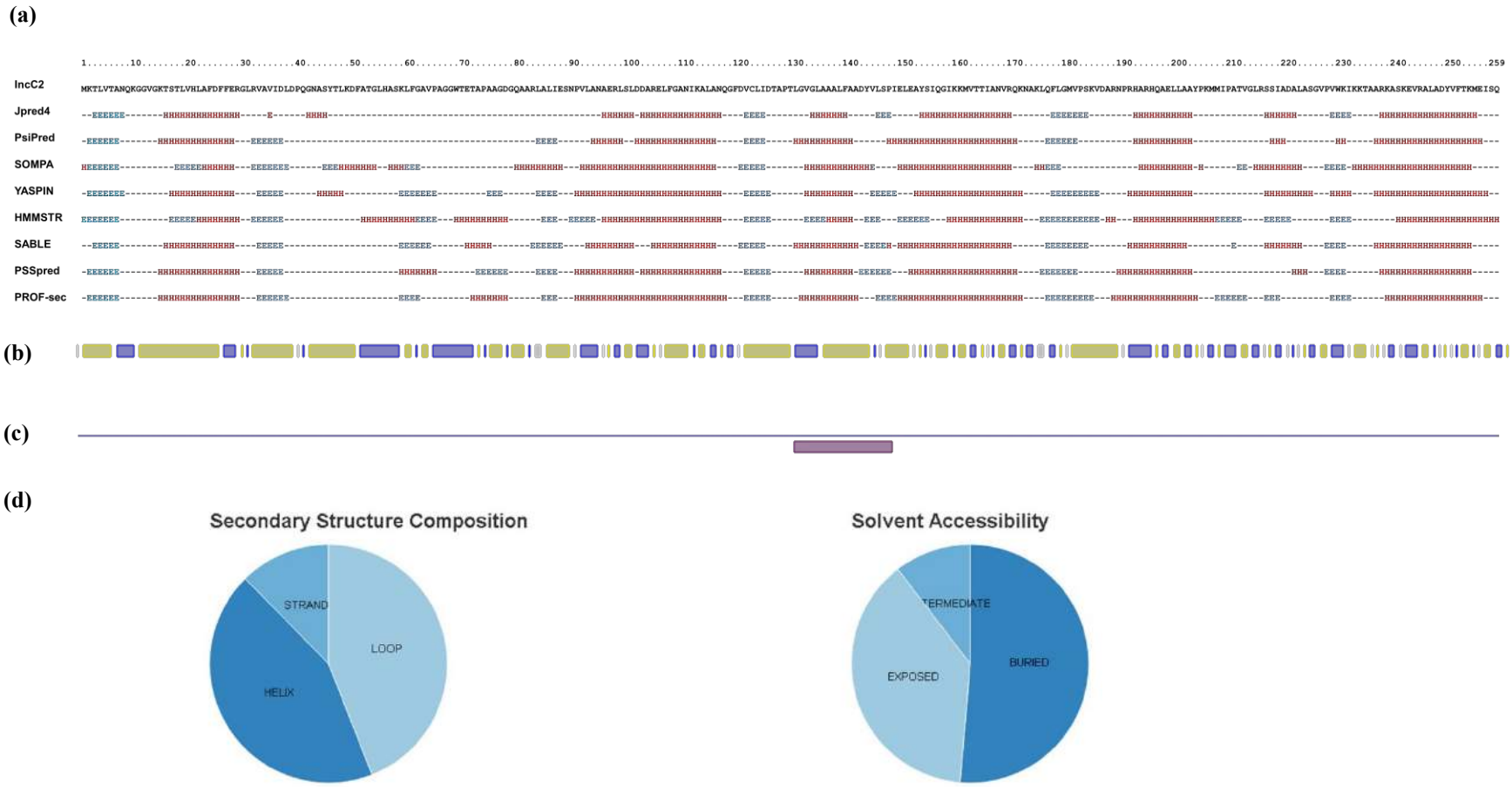


Figure 5.1 Secondary structure prediction and solvent accessibility for IncC2. (a) Secondary structure predictions for the IncC2 protein sequence shows it is a natively folded and predominantly α-helical protein. All servers have shown similar predictions, except SOMPA and HMSSTR have predicted more helices than others. 'E' represents predicted β-strand, while 'H' represent α-helices. **(b)** The PROFsec server predicts exposed (blue) and buried (yellow) amino acid stretches, these are evenly distributed across the sequence. **(c)** The PredictProtein server shows the presence of an exposed hydrophobic region (purple box) that may be membrane associated or involved in protein-protein interactions. **(d)** Approximate secondary structure composition and solvent accessibility from the PredictProtein server (Rost et al., 2004) show that IncC2 protein is predominantly α-helical and > 60% of the amino acids are buried.

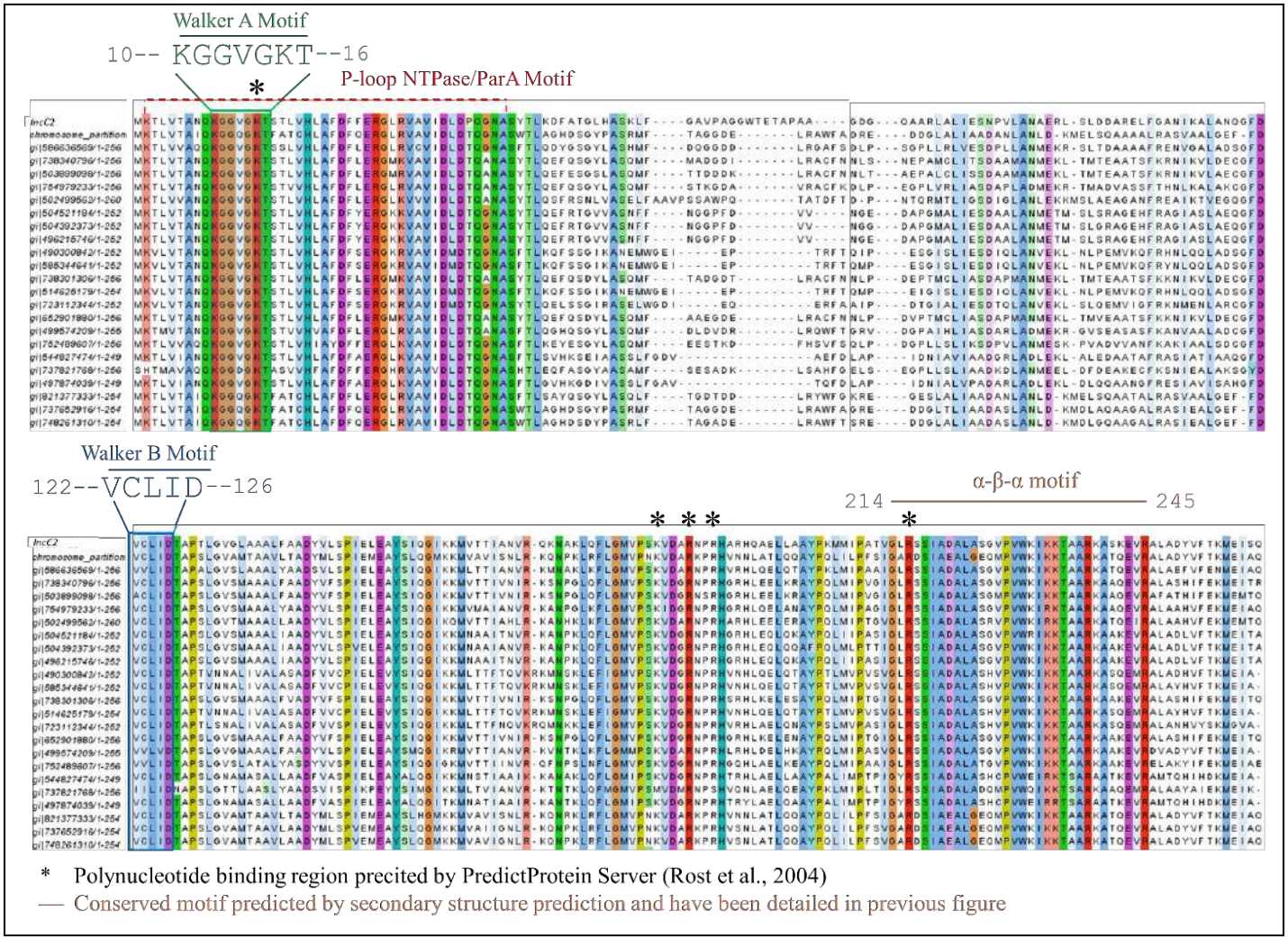
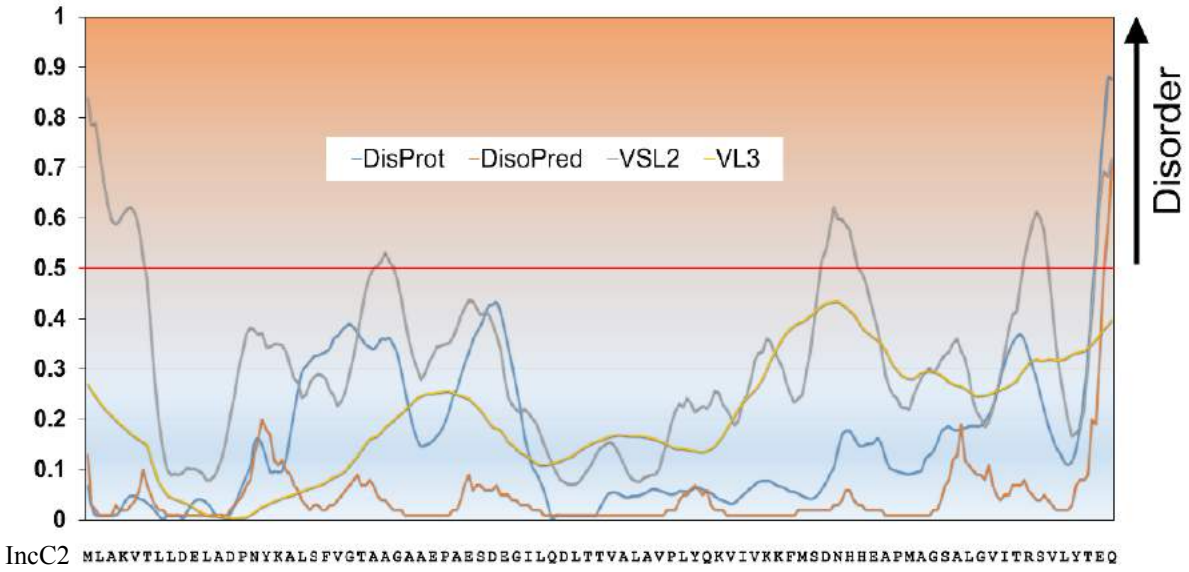
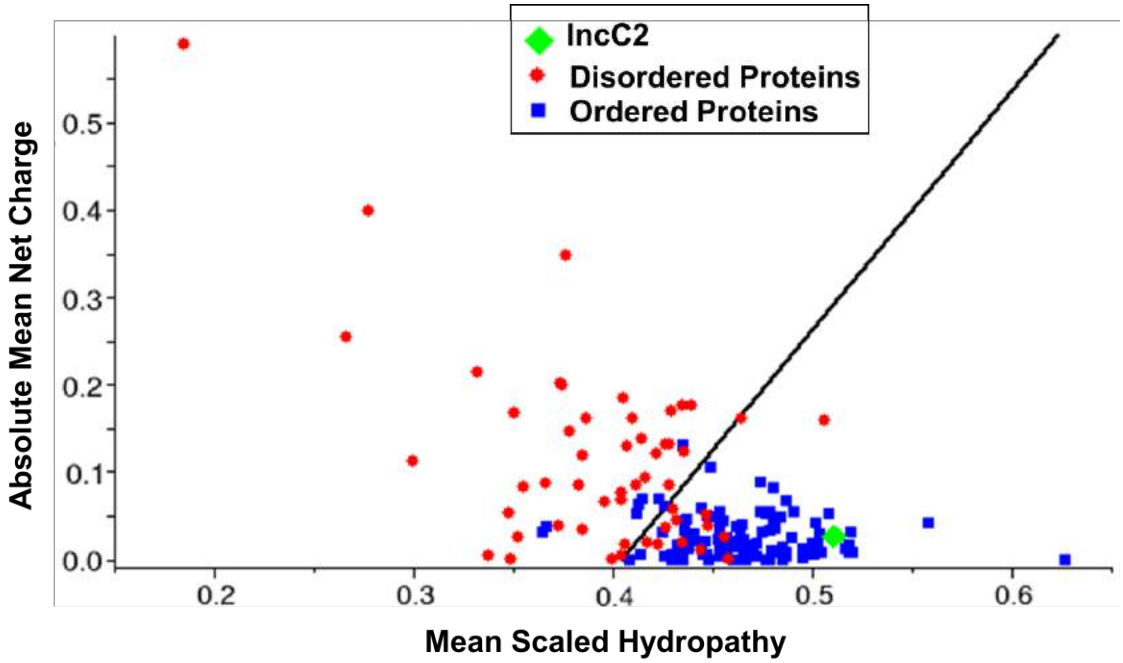


Figure 5.2 Multiple sequence alignment of IncC2 and other ParA proteins. BLAST was performed for IncC2 protein sequence, against the UniProt database (http://www.uniprot.org/), and a multiple sequence alignment was obtained to predict conserved regions and motifs present in IncC2. An alignment of the most similar 23 proteins is shown, from the Clustal W programme. The gene number for each sequence (from Uniprot database) is given on the left. The conserved amino acids are coloured by their side chain properties, as in this programme. Walker-like A (10-16 a.a) and Walker-like B motifs (122-126 a.a) have been found as the most conserved regions in IncC2. The Walker-like A and B motifs are boxed and shown in green and blue colours respectively. These motifs are involved in nucleotide binding and may form the nucleotide binding pocket (Koonin, 1993, Leonard et al., 2005). The asterisk (*) represents amino acids that are probably involved in nucleotide binding. A region (214 to 245 a.a) has been found to be conserved across ParA proteins and forms an α-β-α motif.



(a)



(b)

Figure 5.3 Prediction of disorder in IncC2 protein. (a) Different servers including DisProt (blue), DisoPred (yellow), VSL2 (grey) and VL3 (brown) were used for disorder prediction. In consensus, IncC2 is predicted as a folded protein. The disorder threshold level was set as 0.5 (defined by Neural Network score) and regions with a score more than 0.5 were predicted as disordered. VSL2 server (Xue et al., 2010) has shown some disorder probability in the C-terminal and other random stretches, not seen in the other predictions **(b)** IncC2 is a highly positively charged protein and positioned among the natively folded proteins in the Charge-Hydropathy plot (green diamond). The CH plot was obtained by plotting the mean scaled hydropathy vs absolute mean net charges of known folded and unfolded proteins. Folded/ordered proteins are shown in blue while unfolded/disordered proteins are shown in red. The black line shows the approximate boundary between folded and unfolded proteins.

5.6. IncC2 overexpression and purification

IncC2 protein was overexpressed using the same pET21a vector used for IncC NTD overexpression (Appendix 1). The protein was overexpressed with a His_6-tag to ease the protein purification by affinity chromatography. Pilot-scale protein production (using 100 mL LB media) was used to optimize the conditions for soluble protein expression. IncC2 protein was highly unstable. The protein precipitated either during cell lysis or protein purification. To get the protein in soluble form, protein expression, as well as purification conditions, were optimized.

5.6.1. Temperature optimization

For the bacterial cell growth and temperature optimization, *incC2* was expressed at 25°C and 37°C using BL21(λDE3) cells. The IncC2 protein was expressed at both temperatures but higher protein expression was observed at 37°C (Figure 5.4a). Cell lysis was performed as mentioned in Chapter 2 (Section 2.17). After cell lysis and centrifugation, the clear lysate (soluble fraction) and pellet (insoluble fraction) were analysed by SDS PAGE and protein was present only in the insoluble fraction at both temperatures (Figure 5.4b). It was suggested that the high expression rate might be causing the IncC2 protein to aggregate in the cell. A different protein expression pilot test was done at 16°C in parallel with 37°C and 25°C. Protein expression at 16°C did not improve protein solubility and most of the protein was still found in the insoluble fraction (Figure 5.5).

5.6.2. Strain optimization

To get the best bacterial strain for soluble protein expression, pET21a vector was transformed into three different *E. coli* strains *i.e.,* BL21(λDE3)pLysS, T7 Express and BL21(λDE3). IncC2 protein was expressed in all strains as shown by SDS PAGE (Figure 5.6a). Then cell lysis was performed, as in Chapter 2 (Section 2.17). Sonicated clear lysate (soluble fraction) and pellet (insoluble fractions) were compared using SDS PAGE. Little IncC2 protein was found in the soluble fraction (Figure 5.6b). The protein was aggregated either during overexpression or was precipitated during cell lysis. The effect of varying the concentrations of IPTG (0.1-1 mM) and Kanamycin (20-100 μg/mL) on IncC2 protein expression was then tested. The cells were lysed using different sonication times, or using mechanical homogenization or

lysozyme treatment, but protein solubility was not improved (data not shown). From these experiments T7 Express cells were chosen to express IncC2 at 16°C.

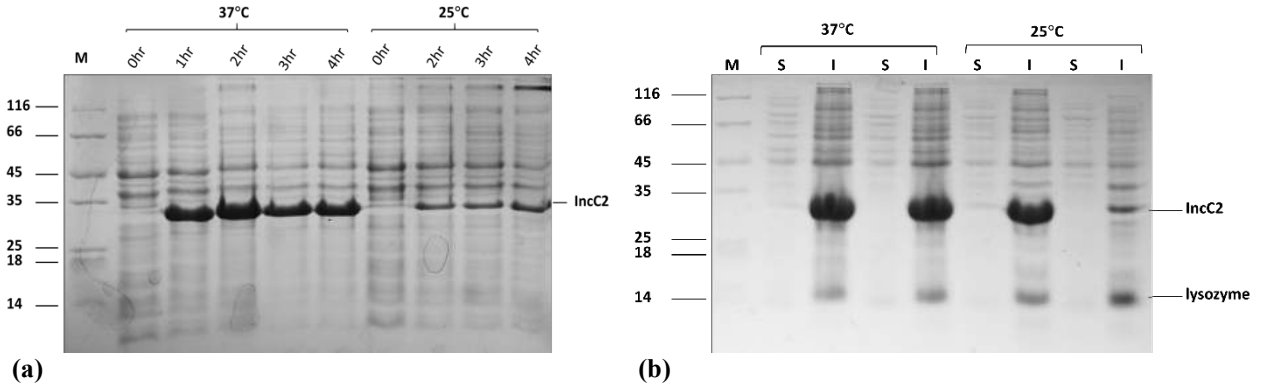


Figure 5.4 IncC2 protein expression at 25°C and 37°C. BL21(λDE3) cells, transformed with *incc2-pet21a,* were grown (50 mL LB media, in duplicates) at two different temperatures; 25°C and 37°C (using 1 mM IPTG as the inducer). 1 mL aliquots were obtained to monitor IncC2 protein expression at 0, 2, 3 and 4 hr. After 4 hr, bacterial cells were centrifuged and sonicated **(a)** The aliquoted samples at different times after induction were analysed by SDS PAGE, (loading equal amounts of samples, normalised by OD_{660}). SDS PAGE analysis shows that IncC2 was overexpressed at both temperatures, although protein expression at 37°C was more prominent. **(b)** Analysis of the cells 4 hr after induction showed that the IncC2 protein was found only in the insoluble fraction (pellet) after bacterial cell lysis and centrifugation. 'S' shows soluble protein fraction (clear lysate) and 'P' shows insoluble fraction (pellet obtained after cell lysis and centrifugation).

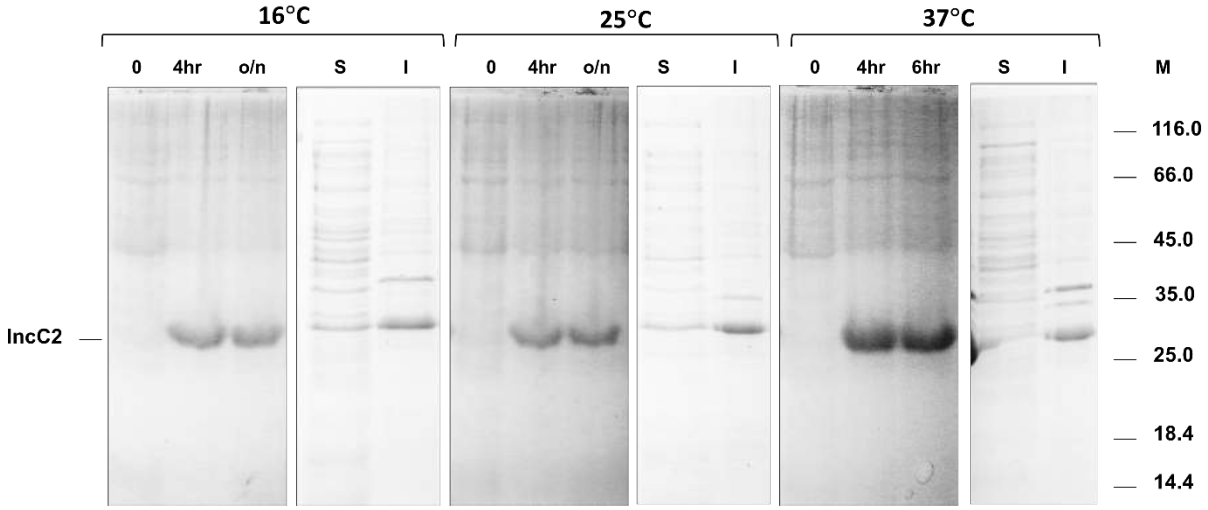
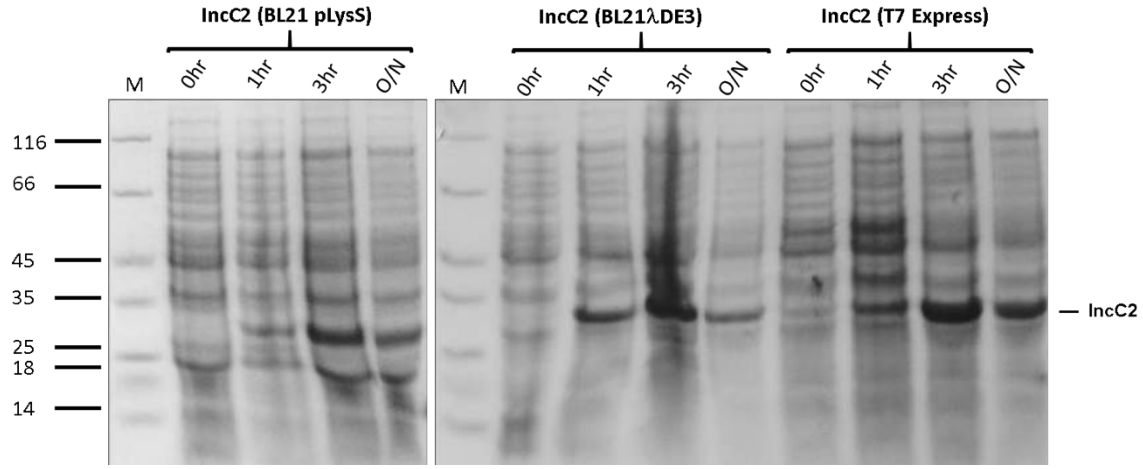
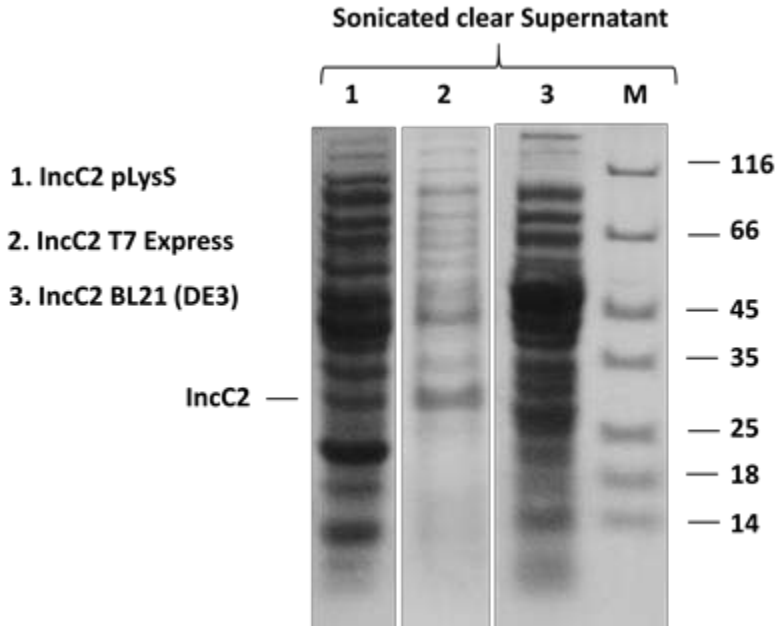


Figure 5.5 IncC2 protein expression at 16°C, 25°C and 37°C. BL21(λDE3) cells, transformed with *incc2-pet21a,* were grown at three different temperatures; 16°C, 25°C and 37°C (using 0.5 mM IPTG as the inducer). 1 mL aliquots were obtained to monitor IncC2 protein expression at different intervals (0 hr, 4 hr and overnight). The aliquots were analysed by SDS PAGE (loading equal amounts of samples, normalised by OD_{660}). IncC2 was expressed at all three temperatures in comparable amounts. SDS PAGE analysis of the clear lysate and pellet from the overnight sample, after cell lysis and centrifugation. 'S' shows the soluble protein fraction (clear lysate) and 'P' shows the insoluble fraction (pellet obtained after cell lysis and centrifugation). IncC2 protein was predominantly present in the insoluble fraction at all 3 temperatures.



(a)



(b)

Figure 5.6 IncC2 protein expression using different *E. coli* strains. The overexpression vector *incc2-pet21a* was transformed into three different *E. coli* strains BL21(λDE3) pLysS, T7 Express and BL21(λDE3). The bacterial cells were grown (50 mL LB media, in duplicates) at 16°C (using 0.5 mM IPTG as the inducer). 1 mL aliquots were obtained to monitor IncC2 protein expression at 0, 1, 3 hr and overnight. **(a)** SDS PAGE of bacterial cells at different times after induction. IncC2 protein was expressed in all strains but protein expression more prominent while using T7 Express strain. **(b)** The analysis of soluble fraction (sonicated supernatant) from the overnight cultures of different strains with SDS PAGE shows the presence of some soluble protein.

5.6.3. Lysis buffer optimization

To get the IncC2 protein in the soluble fraction, 20 different lysis buffers were varied in pH, glycerol, ionic strength, ion composition, presence of detergents and are listed in appendix 6. All lysis buffers were ice cold and contained protease inhibitors (Pierce™ Protease Inhibitor Mini Tablets). Soluble and insoluble protein fractions were analysed using SDS PAGE (as in Figure 5.7). The glycerol (10%) containing buffers were found to give more soluble protein, but the addition of detergents did not help protein solubility much. Figure 5.7 shows the results from eight of the buffers tested. The Buffer B2 (20 mM Tris pH 7.5, 300 mM KCl, 1 mM DTT, 50 mM Sodium glutamate, and 10 % glycerol) and Buffer B6 (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT, 50 mM Sodium glutamate, DNAse 50 μg/mL and 10% glycerol) were found to give the most protein in the soluble fraction after cell lysis (Figure 5.7). Buffers 3, 5, 7 and 8 were found not to help protein solubility.

5.6.4. IncC2 overexpression and purification, without DNase treatment

IncC2 protein expression using the T7 Express strain at 16°C was tested. The clear lysate was obtained using the lysis buffer (Tris 20 mM pH 7.5, 300 mM NaCl, 20 mM Imidazole and protease inhibitors). DNAse was not added to the lysis buffer. The IncC2 protein was expressed (Figure 5.8a) and purified by Ni-NTA column chromatography using Imidazole gradient elution (20-300 mM) (Figure 5.8b). IncC2 protein purified in a large amount but most of the protein precipitated after few hours.

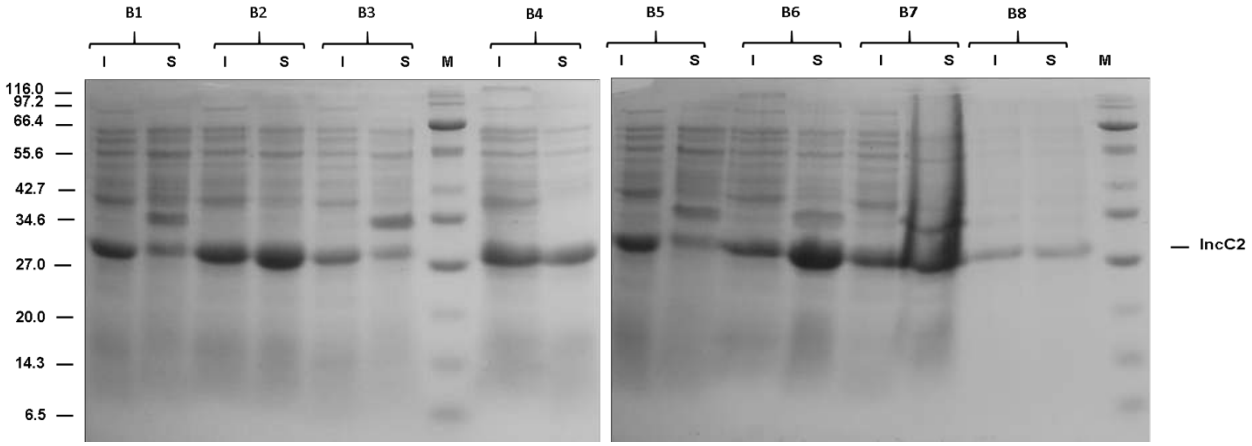
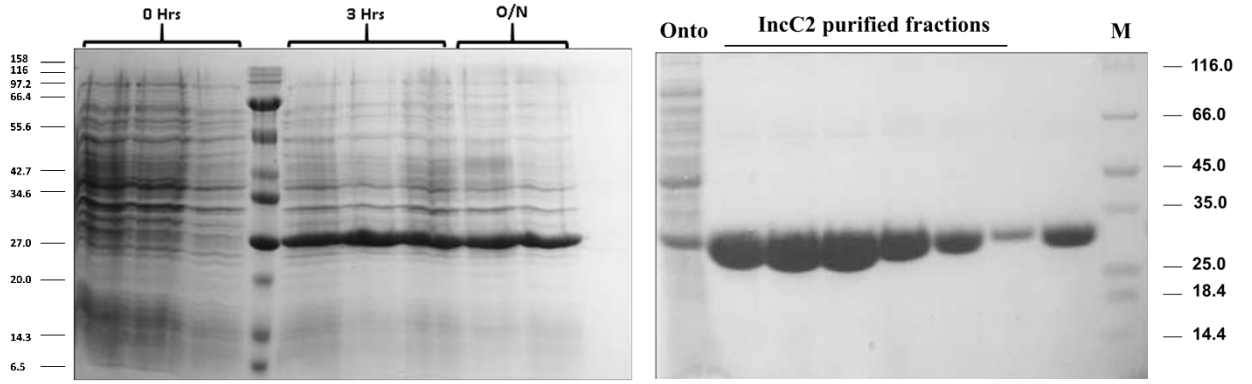


Figure 5.7 IncC2 protein buffer optimization. The bacteria were lysed using different buffers (Buffer 1-8, Appendix 6) and centrifuged to obtain soluble (S) and insoluble (I) fractions. The fractions were analysed using SDS PAGE. The buffers B2 (20 mM Tris pH 7.5, 300 mM KCl, 1 mM DTT, 50 mM Sodium glutamate, and 10 % glycerol) and B6 (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT, 50 mM Sodium glutamate, and 10 % glycerol) were found to be the most suitable buffers to get protein in soluble form. Buffer B4 (RIPA buffer) also provided the good solubility, but contained detergents.



(a) **(b)**

Figure 5.8 Pilot scale IncC2 protein overexpression and purification. (a) IncC2 protein was overexpressed using T7 Express bacterial cells at 16°C (200 mL LB media, 0.5 mM IPTG). 1 mL aliquots were used to monitor protein expression at 0, 3 hr and overnight. Good IncC2 protein expression was obtained **(b)** Bacteria were lysed using the buffer (20 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole, 10 % glycerol, and protease inhibitors). The same buffer was used for the protein purification. A linear imidazole gradient (20-300 mM) was used in the same buffer for protein elution. Protein eluted at 100-150 mM imidazole.

5.6.5. Solubility tag optimization

Although IncC2 was expressed and purified as a soluble protein, but a protein concentration of more than 2 mg/mL was never achieved. In attempts to get higher concentrations, IncC2 and IncC1 were overexpressed as fusion proteins with either the GB tag or the SUMO tag (Figure 5.9). The *incc1* and *incC2* genes were cloned into a GB vector (pOP3BT) and a Pet-SUMO vectors. The GB vector (pOP3BT) was provided by Dr Marko Hyvonen, Department of Biochemistry, University of Cambridge and the pet-SUMO vector was provided by Dr Muhammad Jamshad, School of Biosciences, University of Birmingham (Appendix 1 for plasmid maps). The *incC1* and *IncC2* genes were PCR amplified using the primer sets given in Table 5.3 and cloning was performed as described in Chapter 2 (Section 2.11). The fusion proteins IncC1/IncC2-GB and IncC1/IncC2-SUMO were overexpressed and purified as described in previous section 5.6. Unfortunately, GB and SUMO fusion tags did not help to increase the IncC1 and IncC2 protein concentration above 2 mg/mL. So, for the next protein preparations and biochemical characterization IncC-His_6 protein was used.

Table 5.3 Primers used for IncC1 and IncC2 cloning in GB tag vector and pet-SUMO vector. 'R' shows reverse prime, 'F' shows forward primer, and restriction sites are shown in red.

Cloning Oligo	Sequences
IncC1/2-R-HindIII	5`-GCCGCAAGCTTGTCGACTCATTGGGAAAT-3`
IncC2–F-BamHI	5`-CATAGTGGATCCATGAAAACTTTGGTCACGGC-3`
IncC1–F-BamHI	5`-CATAGTGGATCCATGGGTGTTATCCATGAAGAAAC-3`

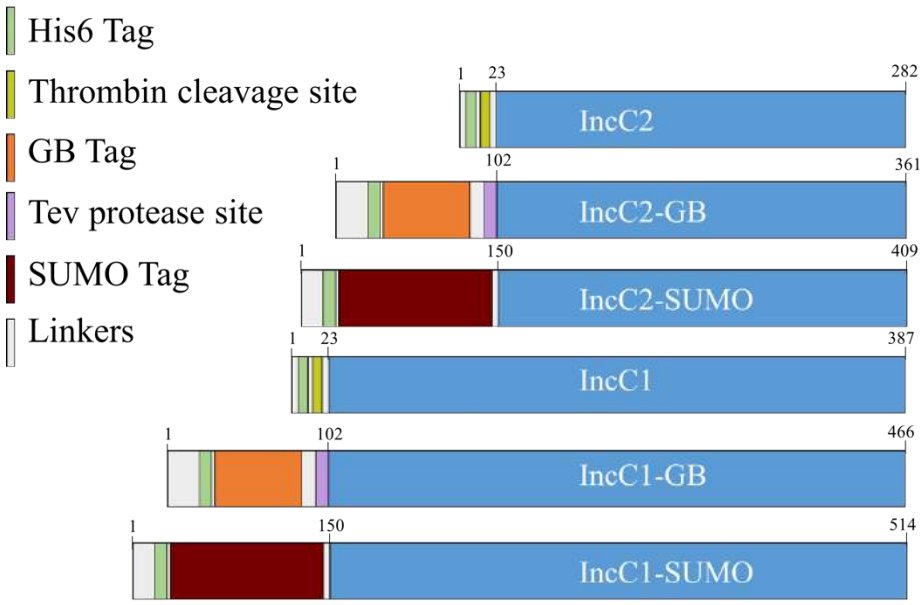


Figure 5.9 IncC1 and IncC2 as fusion proteins with GB and SUMO tag. The IncC1 and IncC2 proteins were overexpressed as fusion proteins. The region for GB tag (102 amino acids) consists of a His_6 tag, GB1 domain (55 amino acid) and TEV protease cleavage site. The SUMO tag region (150 amino acids) consists of His_6 tag, SUMO tag (93 amino acid) and SUMO protease cleavage site.

5.6.6. Thermal shift assay (TSA)

The IncC2 protein stability, in various homemade buffers as well as in the buffers of the pH slice kit (Hampton Research) was monitored by TSA. (Buffers compositions are given in Appendix 6), TSA was performed as described in Chapter 2 (section 2.26.4). The TSA curves for some of the buffers are given in Figure 5.10. IncC2 was more stable in buffers in the pH range 7.0-8.0. The IncC2 protein was most stable in Tris pH 7.5 (with 50 mM sodium glutamate, 10% glycerol) (Figure 5.10a), buffer B6, similar to the previous buffer optimisation experiment (section 5.6.3). IncC1 was found more stable at high temperature in comparison to IncC2 though the best buffer for IncC1 was the same as for IncC2 (Figure 5.10b).

This optimised buffer was used to purify IncC2.

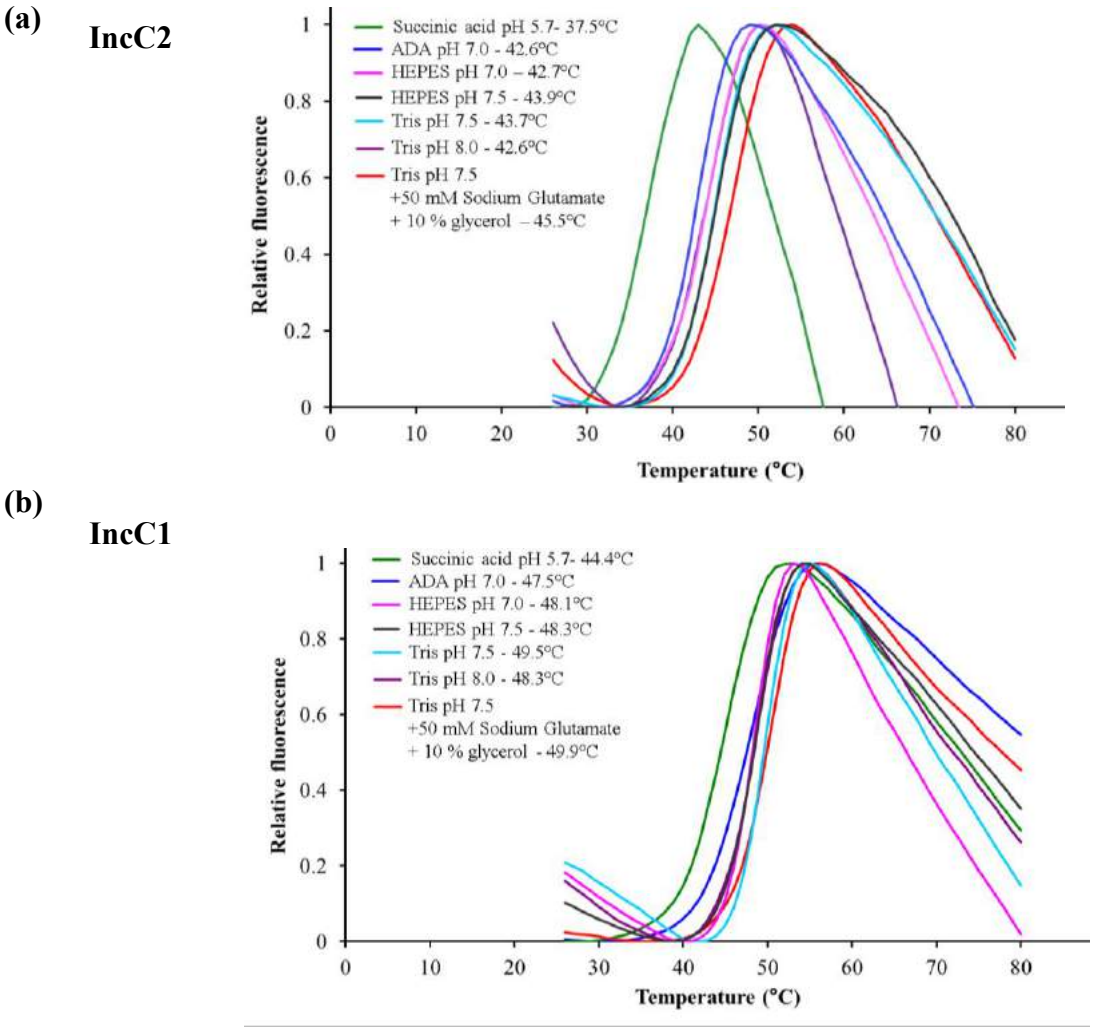


Figure 5.10 Thermal shift assay to optimize the pH and buffer for IncC2. The assay was performed using homemade buffers and also the pH slice kit of buffers (Hampton Research). The Mx3000P qPCR System (Agilent) was used with Sypro filter. The fluorescence was monitored after each degree rise in temperature from 25-80°C. Data analysis was performed by exporting raw data to spreadsheets. Data were analysed using DSF Analysis Ver. 3.0 MS Excel script (ftp://ftp.sgc.ox.ac.uk/pub/biophysics). GraphPad Prism ver 7.0.4 was used to get Boltzmann fitting and melting temperature (T_m). The coloured lines represent the relative fluorescence of the protein in different buffers vs temperature. The inset gives the buffer compositions and the melting temperature of the protein in each. **(a)** IncC2 has a T_m of 45.5°C in Tris pH 7.5 buffer (50 mM) 50 mM Sodium glutamate, 10 % glycerol. (b) IncC1 has a T_m of 49.9°C in Tris pH 7.5 buffer (50 mM) 50 mM Sodium glutamate, 10 % glycerol.

5.7. IncC2 protein purification

IncC2 protein was overexpressed using T7 Express at 16°C. The protein was purified by Ni-NTA affinity chromatography.

5.7.1. Ni-NTA chromatography

The T7 Express (16°C, 0.5 mM IPTG) cells were grown overnight and centrifuged to the obtained cell pellet. Clear lysate was obtained as described in Chapter 2 (Section 2.17), with. Tris buffer pH 7.5 (Buffer B6, with additional protease inhibitors, DNase, 20 mM imidazole) as the lysis buffer. The clear supernatant was loaded onto a Ni-NTA affinity column (15 mL) and a linear imidazole gradient was used to elute the protein. The Bradford assay was used to estimate the protein concentration in the eluted fractions (Figure 5.11, Ni-NTA chromatography). The purified IncC2 protein gave a single band on SDS PAGE (Figure 5.11). The purified protein showed a very high 260/280 ratio (1.20-1.50) that suggested that the protein may contain some nucleotides. In some IncC2 preparations, high salt concentrations (0.5-1.5 M) were added to the lysis buffer to get rid of nucleotides but the 260/280 ratio was still found to be above 1.0. In the initial protein preparations, the IncC2 protein precipitated during purification. In the later preparations, the eluted fractions were diluted two-fold in B6 buffer containing 0.1 mM EDTA. In later work, a 5 mL IMAC (immobilized metal ion affinity chromatography) column (with fast flow Ni-sepharose) was used for faster protein purification than the larger Ni-NTA column and to avoid protein aggregation on the column with time. A step-wise imidazole gradient was used (50, 100, 300 mM Imidazole) to elute the protein fractions (Figure 5.13). IncC2-containing fractions were buffer exchanged, to remove imidazole, using PD-10 desalting columns (GE Healthcare). The protein was flash frozen in liquid nitrogen before storing at -80°C.

5.7.2. Size exclusion chromatography (SEC)

Initially, the IncC2 protein was subjected to a second step of purification, size exclusion chromatography, to get rid of trace impurities. High salt buffer (B6 buffer with 500 mM NaCl) was used to elute protein fractions and to remove any bound nucleotides. The eluted fractions, with high absorbance at 280 (mAU), were analysed on SDS PAGE (Figure 5.12). IncC2 protein eluted earlier from the SEC column than expected. This shows that either IncC2 forms a high molecular weight species or it assumes a non-globular shape. The SEC was calibrated with a series of proteins of known molecular mass and Stokes radius (section 2.24). The Stokes radius

for IncC2 was estimated from the calibration curve to be 48 Å, while for a globular protein with molecular weight 29.5 kDa, as the IncC2 monomer, the Stokes radius is expected to be 36 Å.

The 260/280 ratio for SEC purified protein fractions was still >1.0. Hence the column did not remove any nucleotides. In addition, the IncC2 protein aggregated on the SEC column during purification and eluted as a very broad peak. The IncC2 protein from the Ni-NTA column was pure enough to carry out its structural characterization using various biophysical techniques and the SEC column was not used subsequently.

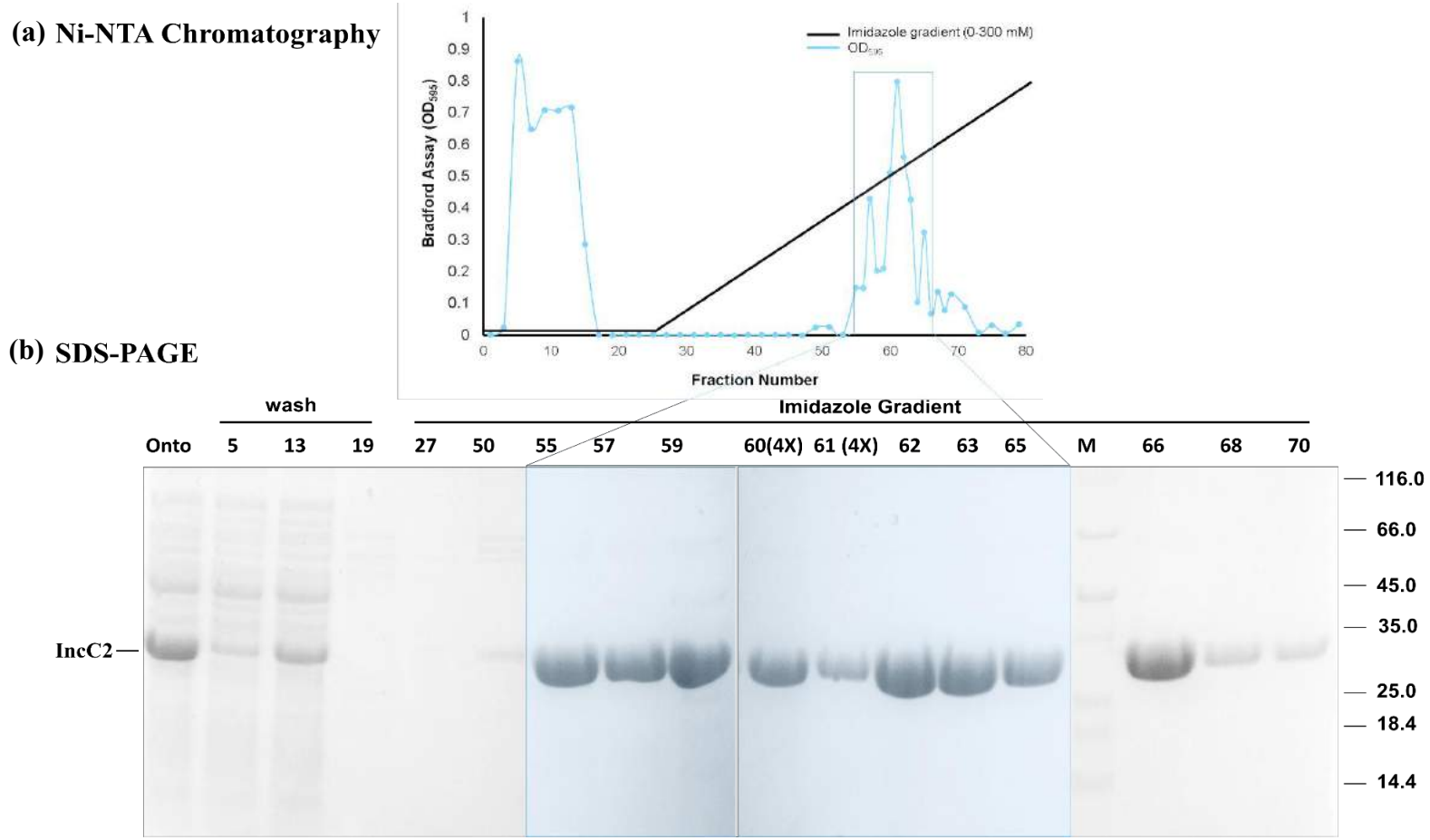


Figure 5.11 Purification of IncC2 protein by Ni-NTA chromatography. IncC2 protein was expressed using T7 Express cells at 16°C (2 L LB medium, 0.5 mM IPTG). The clear lysate (onto), obtained after cell lysis and centrifugation, was used for IncC2 purification. A 15 mL Ni-NTA column, equilibrated with Buffer B6 (plus 20 mM Imidazole), and a linear Imidazole gradient (20 -300 mM) was used for protein purification (using B6 buffer). 1.5 mL fractions were collected and protein contents were estimated using Bradford Assay. **(a)** The graph shows the results of the Bradford assay of the eluted protein fractions. The black line shows the Imidazole gradient (20-300 mM). Protein fractions 55-65, show high protein concentrations. These fractions eluted around 150 mM imidazole. **(b)** The protein fractions were analysed by SDS PAGE and showed IncC2 purified as a single band.

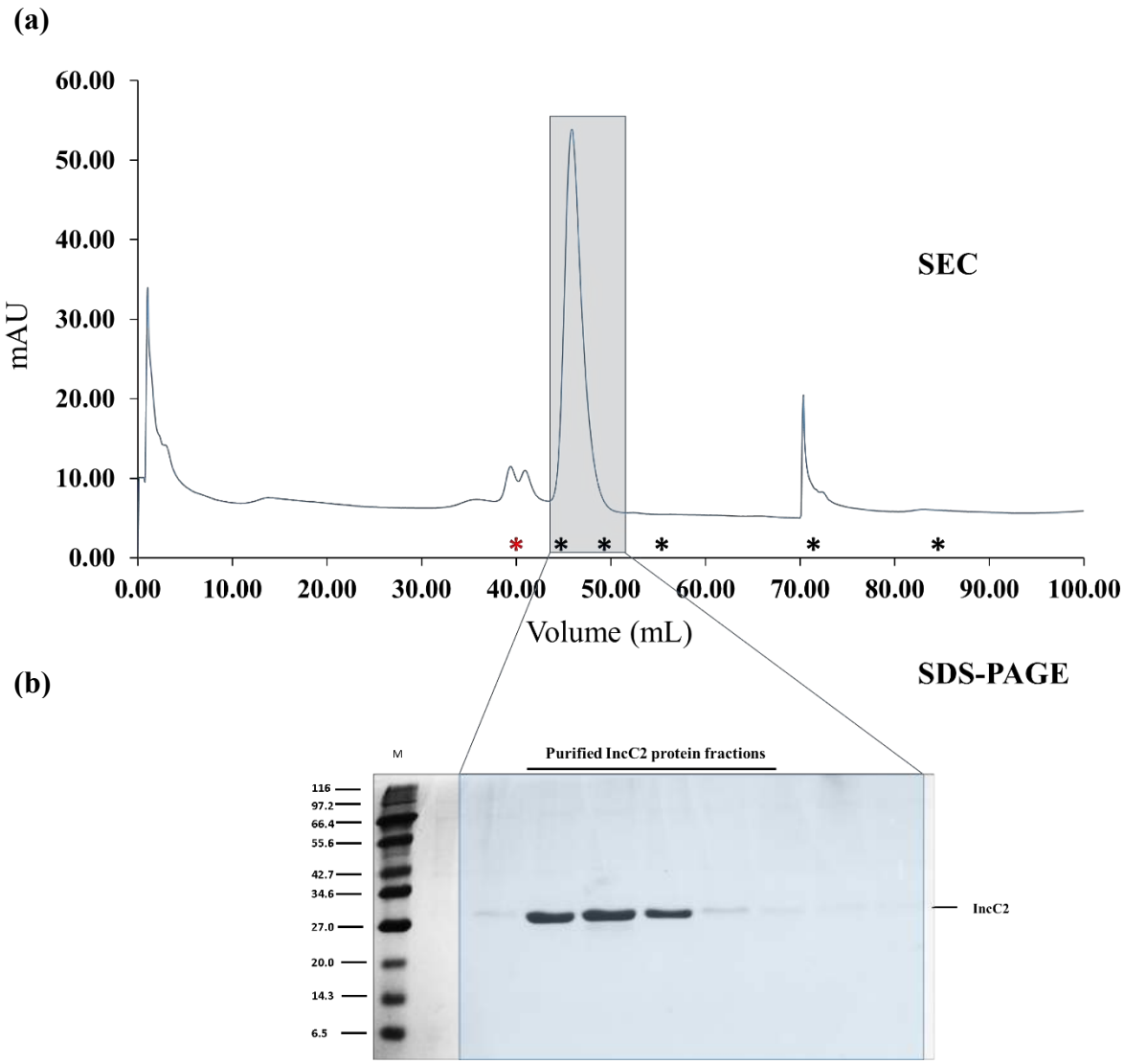


Figure 5.12 IncC2 protein purification using size exclusion chromatography (SEC). The purified IncC2 fractions, from Ni NTA chromatography, were subjected to SEC purification. **(a)** 2.5 mL protein was loaded on a 120 mL SEC column (HiLoad 16/600 Superdex 75 column). Buffer B6 with high salt (500 mM NaCl) was used to elute 2.5 mL fractions. The red asterisk '*' shows the column's void volume while black asterisks show the elution volumes of standard proteins that were used to obtain a molecular weight curve (section 2). Standard proteins include β-Amylase (mol. wt. 200 kDa, R_s 54 Å), Alcohol Dehydrogenase (mol. wt. 147 kDa, R_s 46 Å), Bovine Serum Albumin, (mol. wt. 66 kDa, R_s 35 Å), Carbonic Anhydrase (mol. wt. 29 kDa, R_s 21 Å) and Cytochrome C (mol. wt. 12.4, R_s 17 Å) (La Verde et al., 2017). The protein was eluted much earlier than expected from its monomeric mass. This suggests IncC2 either forms multimers or has a non-globular conformation **(b)** Protein fractions, with high absorbance (mAU at A_{280}), were analysed by SDS PAGE. The IncC2 protein was observed as a single band.

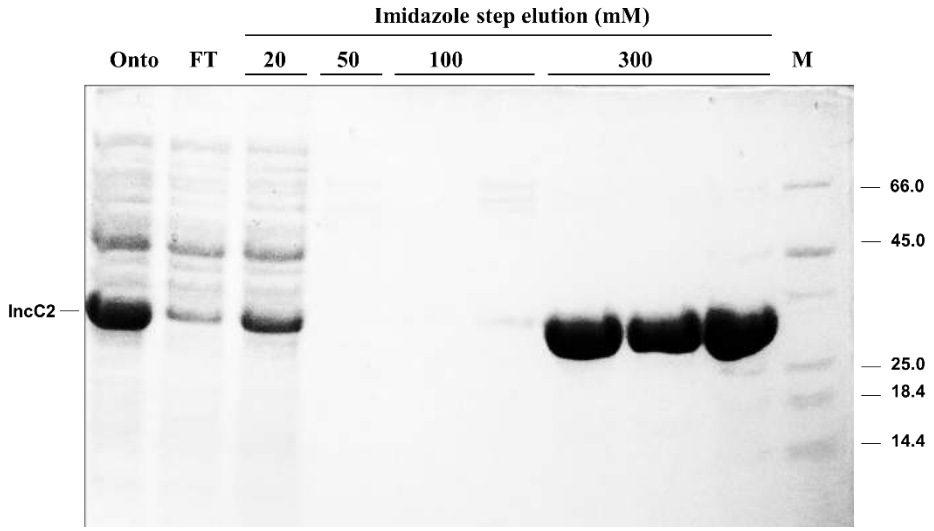


Figure 5.13 IncC2 purification using IMAC Ni-NTA chromatography. A 5 mL IMAC Ni-NTA column (GE Healthcare) was used for quick IncC2 protein purification. After loading clear lysate, the column washed with 3-5 column volumes 20 mM imidazole containing buffer (Buffer B6). After washing, His_6 tagged protein was fractionated using step imidazole elution (50, 100, 300 mM Imidazole). The protein fractions from each step were analysed by SDS PAGE. IncC2 protein was purified to a single band in 300 mM Imidazole eluted fraction.

5.8. Ultrafiltration and protein concentration

The purified protein fractions were concentrated using centrifugal concentrators (Vivaspin, MWCO 10,000 Da). The IncC2 protein precipitated at a concentration more than 3 mg/mL. At this protein concentration, protein crystallization was tried but good crystals were never obtained. Protein NMR was also not feasible at this concentration.

5.9. Mass Spectrometry analysis for IncC2

The accurate subunit mass of the purified IncC2-His_6 protein was analysed using mass spectrometry (ESI-MS) as described in Chapter 2 (section 2.29). The molecular mass for IncC2-His_6 from mass spectrometry was determined to be 30,108 Da (Figure 5.14) while the molecular mass from the amino acid sequence was calculated as 30059.59 Da. Hence the protein is intact.

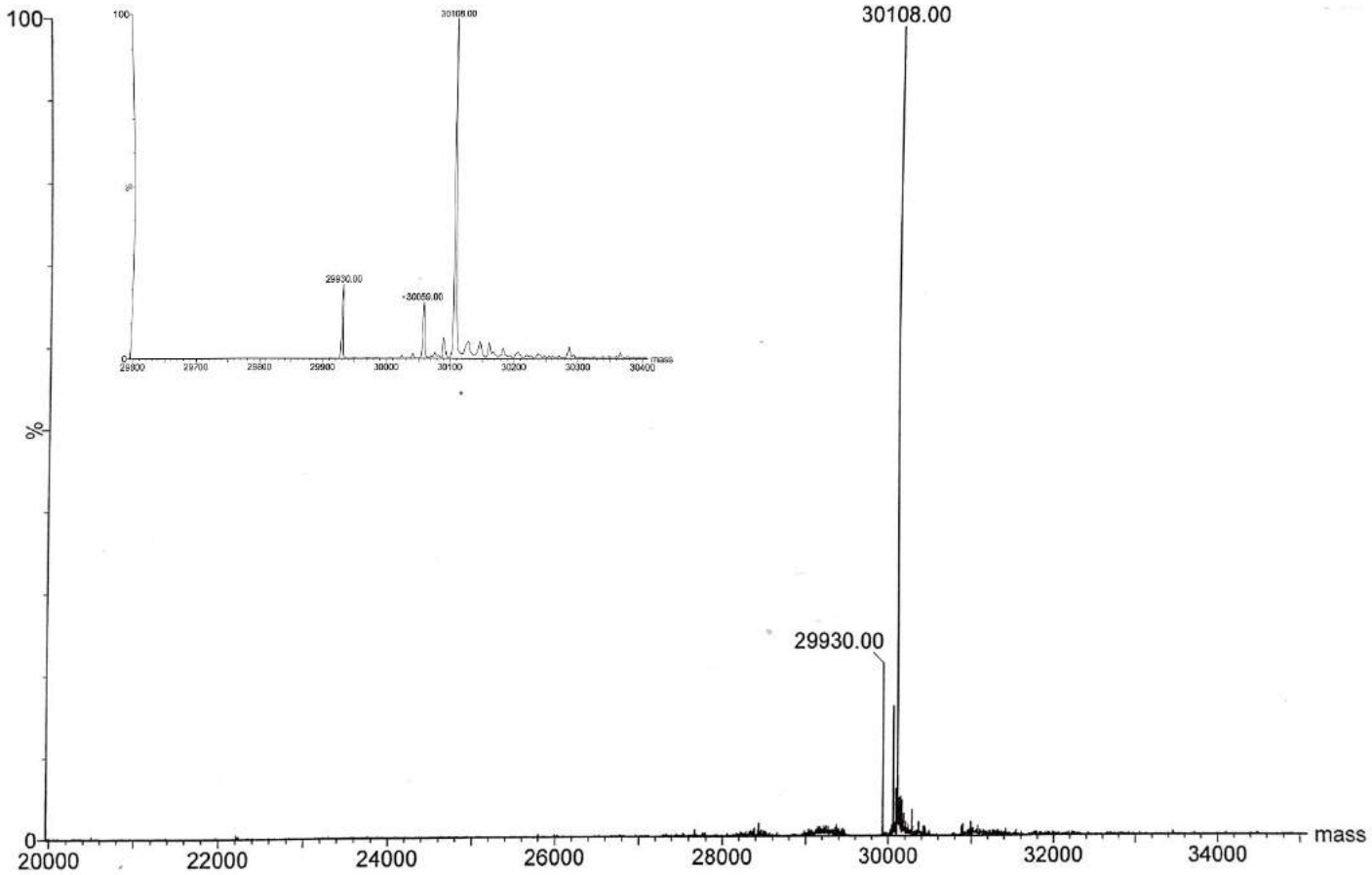


Figure 5.14 Estimation of IncC2-His$_6$ molecular mass using Mass Spectrometry (MS). The IncC2-His$_6$ protein was subjected to ESI-MS to calculate the accurate molecular mass. The purified IncC2 protein was given to the Mass Spectrometry facility in School of Chemistry, University of Birmingham. ESI-MS spectrum was obtained and molecular mass was calculated as 30,108 Da. Another ESI-MS peak for 29,930 Da was observed with a deviation of 178 Da from dominating peak, while the molecular mass calculated from the amino acid sequence is 30059.5. The difference in molecular mass between the two peaks might be due to loss of an amino acid followed by some PTM.

5.10. IncC1 overexpression and purification

To get the optimal protein overexpression for IncC1, growth, temperature and induction conditions were optimized in a similar way to the optimization procedure for IncC2 (this chapter, Section 5.6). IncC1 protein expression was monitored at four different temperatures; 16°C, 25°C, 30°C and 37°C using T7 Express cells and with 0.5 mM IPTG as the inducer. SDS PAGE analysis of the cells shows more IncC1 protein in the soluble fraction in cells grown at lower temperatures after induction. (Figure 5.15). For next IncC1 protein preparation, 16-25°C temperature was used for cell growth, to avoid any undesired aggregation.

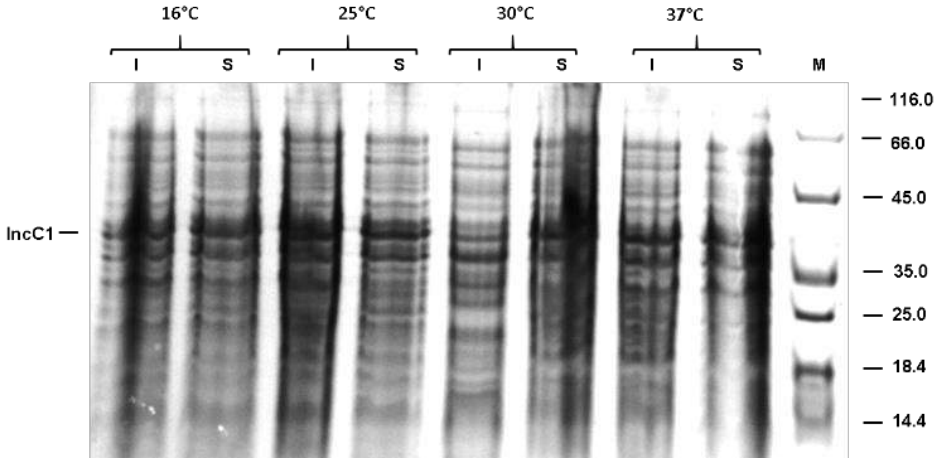


Figure 5.15 IncC1 protein expression at different temperatures. T7 Express bacterial cells, transformed with *incc1-pet21a,* were grown at four different temperatures; 16°C, 25°C and 37°C after induction with (0.5 mM IPTG). SDS PAGE of the cells shows that growth at lower temperatures (16°C and 25°C) gave more soluble protein. 'S' shows soluble protein fraction (clear lysate) and 'P' shows insoluble fraction (pellet obtained after cell lysis and centrifugation).

5.10.1. IncC1-His_6 purification Ni NTA chromatography

The IncC1 protein was expressed at 16°C using T7 Express (2 L LB medium, 0.5 mM IPTG). The clear lysate was obtained as described in Chapter 2 (section 2.17). Buffer B6 (plus 20 mM imidazole, protease inhibitors, DNase) was used as the lysis buffer. The clear supernatant was loaded on IMAC Ni-NTA affinity column. Protein fractions were collected using imidazole step elution. The eluted fractions, assayed by Bradford reagent, were analysed by SDS PAGE. IncC1 protein was purified to give a single band on SDS PAGE (Figure 5.16). The purified IncC1 protein, like IncC2, showed a high 260/280 ratio (1.0-1.5). The IncC1 protein was subjected to SEC column for further purification but the protein always aggregated on column.

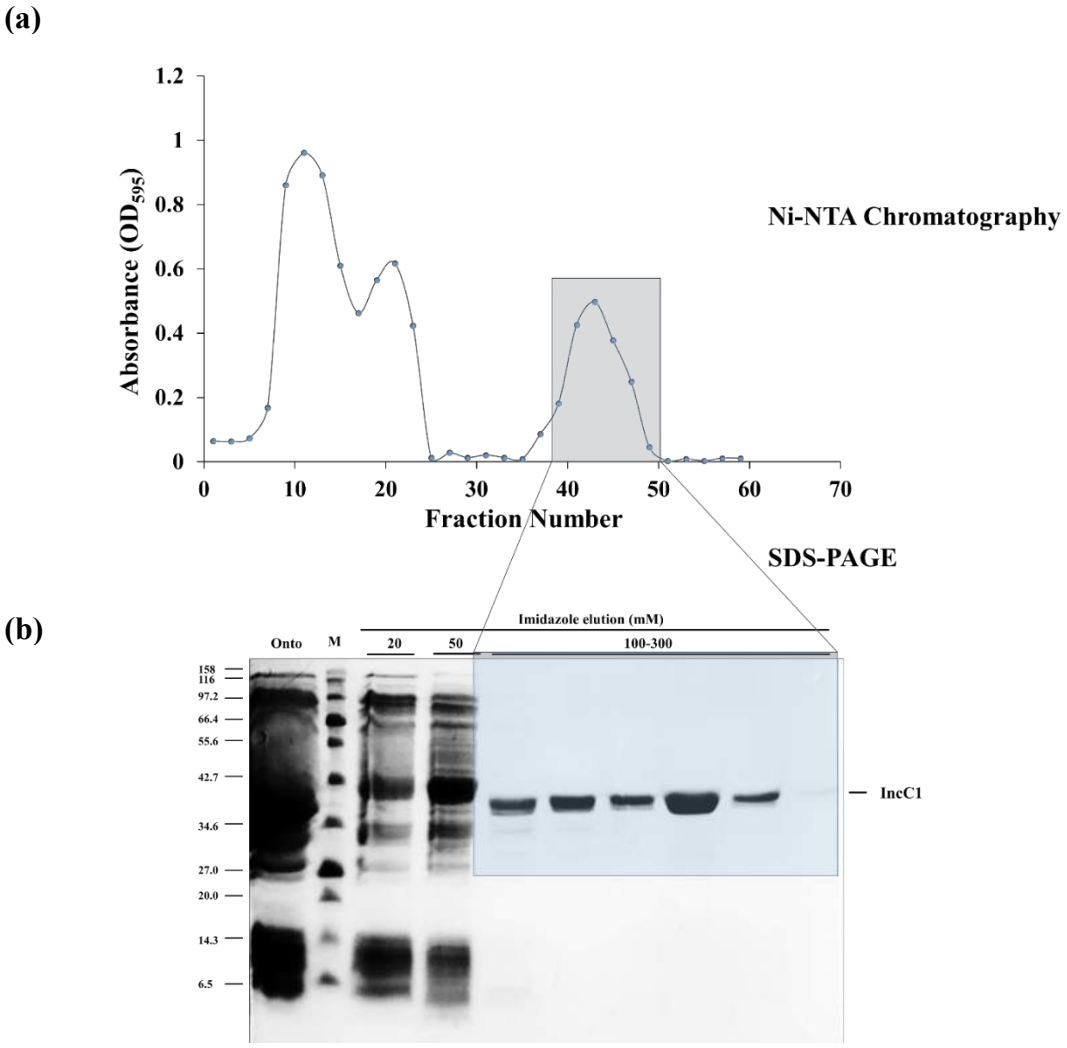


Figure 5.16 IncC1 protein purification using Ni-NTA chromatography. The IncC1 protein was purified using a 5 mL IMAC Ni-NTA column. T7 Express cells were used to express IncC1 at 16°C (0.5 mM IPTG, 2 L LB medium). **(a)** The clear lysate was loaded on the column and column was washed with 25 mL Buffer B6 (containing 20 mM and 50 mM imidazole). After washing, His_6 tagged proteins were eluted using a linear gradient of imidazole (100-300 mM Imidazole). **(b)** The purified IncC1 fractions were analysed by SDS PAGE and protein was found to be a single band.

5.10.2. SP Sepharose chromatography

To remove any bound nucleotide and to lower the 260/280 absorbance ratio, the IncC1 protein was further purified by SP Sepharose chromatography. The purified protein, from Ni-

NTA column chromatography, was buffer exchanged using a PD-10 column into SP buffer, containing Tris pH 7.5, NaCl 100 mM, EDTA 0.1 mM, 10 % Glycerol and 1 mM DTT. The protein (10 mL, at a concentration of 0.3 mg/mL) was loaded onto a 20 mL SP-Sepharose column pre-equilibrated using SP buffer. The protein fractions were eluted using SP buffer with a linear NaCl gradient (0.1- 1 M). The IncC1 protein was eluted from the column, at 250-300 mM NaCl. The IncC1 fractions had already been purified to homogeneity by the Ni-NTA column, and on analysis by SDS PAGE after SP Sepharose still gave a single band (Figure 5.17) of the correct size, so had not been degraded. The UV absorbance of the purified protein fractions was measured but the 260/280 ratio was still found to be >1.0. Hence this column did not remove nucleotides and was not helpful.

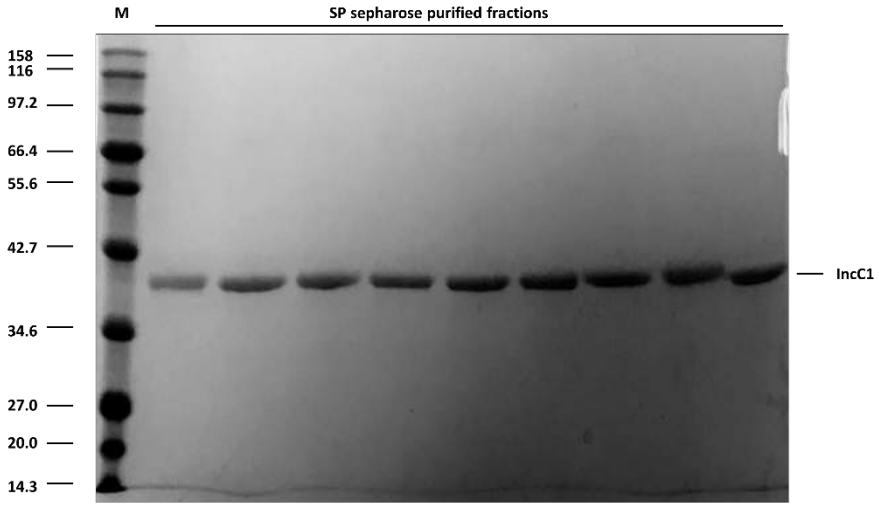


Figure 5.17 IncC1 protein purification using SP Sepharose chromatography. IncC1 protein was further purified using a 20 mL SP Sepharose column. The purified fractions from the Ni-NTA chromatography were loaded on the column. Protein was eluted using a linear salt gradient (0.1-1 M NaCl) in SP buffer (Tris pH 7.5, NaCl 100 mM, EDTA 0.1 mM, 10 % Glycerol and 1mM DTT). The IncC1 fractions eluted at 250-300 mM NaCl. Each fraction was analysed by Bradford assay to check the presence of proteins (not shown). IncC1 protein was found to be a single band on the SDS PAGE with an apparent subunit weight of ~40 kDa.

5.11. Mass Spectrometry analysis for IncC1

The molecular mass of the purified IncC1-His_6 protein was measured, using mass spectrometry (ESI-MS), as described in Chapter 2 (Section 2.29). The protein sample was given for Mass spectrometry analysis, to the Mass Spectrometry Facility, School of Chemistry, University of Birmingham. ESI-MS analysis was performed and molecular mass, for IncC1-His_6

protein, was measured as 40,710 Da (Figure 5.18), while the molecular mass calculated from amino acid sequence is 40661.04 The difference may be due to the calibration of the MS instrument. A second peak, with mass 40,533 Da is also observed, of greater intensity than the other. The difference in mass between the two peaks is not that of a single amino acid and so may be due to degradation followed by modification.

5.12. Biophysical characterization of IncC1 and IncC2 Proteins

After purification of IncC1 and IncC2 to homogeneity, the proteins were further characterized by different biophysical techniques.

5.12.1. SEC-MALLS

SEC-MALLS was employed to examine the association states of IncC1 and IncC2 in the presence and absence of nucleotides *i.e.* ATP and ADP. In the absence of any added nucleotide, IncC1 and IncC2 proteins were found to be in monomer and dimer equilibrium, in solution (Figure 5.19). For IncC2, a monomeric peak (~31 kDa), a dimeric peak (~60 kDa), a broad dodecamer peak (~ 370 kDa) and high molecular weight aggregate species were observed (Figure 5.19a). The high oligomeric species does not look monodisperse that suggests the presence of multiple very high molecular weight species. In the case of IncC1, a similar pattern was observed. Due to an error in calibration of the Refractive Index of the IncC1 samples, the molecular weights of the peaks for IncC1 cannot be calculated accurately, but the peak elution pattern suggests a monomer, a dimer peak along with oligomers and aggregates peaks (Figure 5.19b), similar to that of IncC2. In the presence of ADP, a similar pattern was observed for both IncC1 and IncC2 as in the absence of any added nucleotide; while, in the presence of ATP, homodimer species were dominant for both IncC1 and IncC2 proteins. Interestingly a new peak started to emerge in IncC1+ATP run, eluting after 24 min, suggesting something more compact than the monomer or protein degradation. The major conclusion from the SEC-MALLS is that monomeric protein population is shifted to a dimeric state in presence of ATP, for both IncC1 and IncC2. (Figure 5.19).

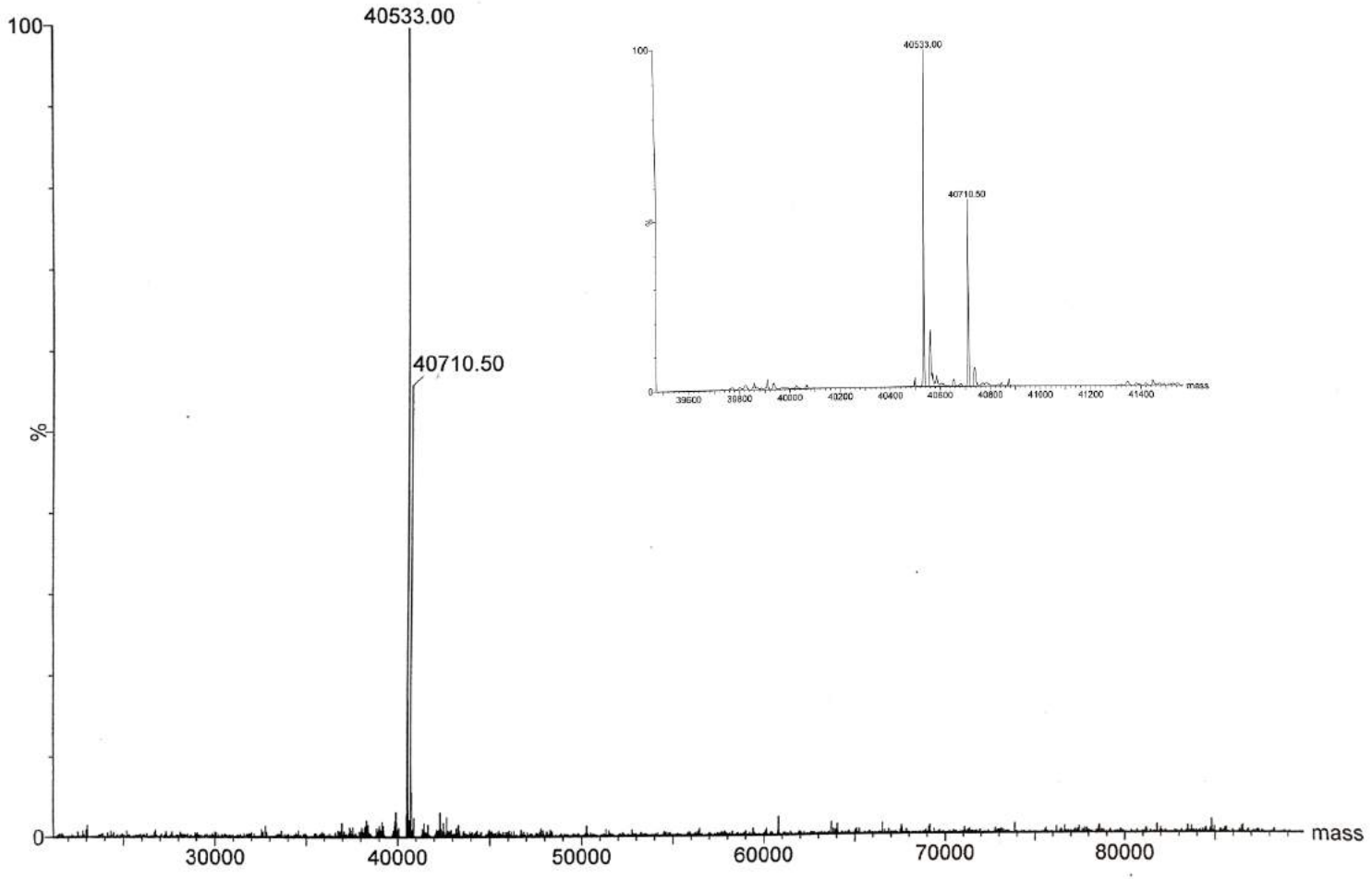
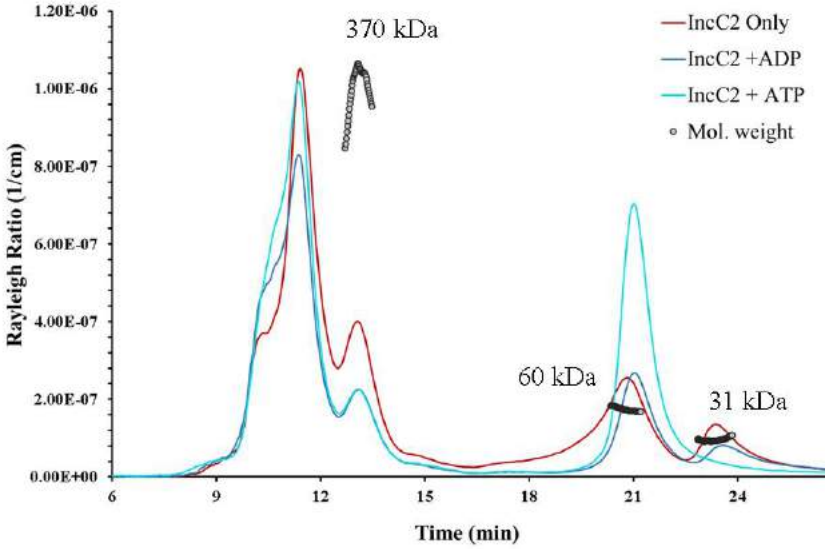


Figure 5.18 Estimation of IncC1-His$_6$ molecular mass using Mass spectrometry (MS). The IncC1 protein was subjected to MS for accurate molecular mass calculations. The purified IncC1 protein was given to the Mass Spectrometry facility in School of Chemistry, University of Birmingham. ESI-MS spectrum was obtained and IncC1-His$_6$ molecular mass was calculated as 40710 Da. Another ESI-MS peak of 40533 Da was observed with a deviation of 177.5 Da from the highest mass peak, while the molecular mass calculated from the amino acid sequence is 40661.04. The difference in mass is might be due to loss of N-terminal Methionine reside in addition to some PTM.

(a)



(b)

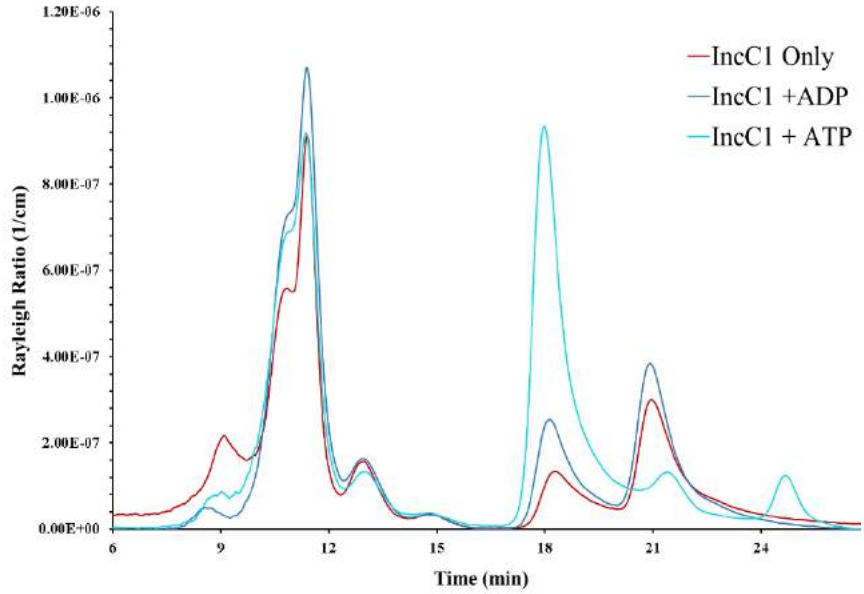


Figure 5.19 SEC-MALLS for IncC1 and IncC2. **(a)** In case of IncC2, a monomeric 31 kDa peak (23.5 mL) and a dimeric peak 60 kDa (21 mL) was observed, while oligomers species can also be observed as dodecamer and aggregate peaks. The IncC2 protein shows a shift of monomeric species to dimer in the presence of ATP, while, in the presence of ADP or without nucleotides, no conformational change was observed. **(b)** The IncC1 showed the same pattern as IncC2. Although molecular weights could not be calculated due to Refractive Index calibration problems, the peaks eluted at 21 mL, 18 mL may show monomer and dimer peaks. In the presence of ATP, a transition from monomer to dimer was observed.

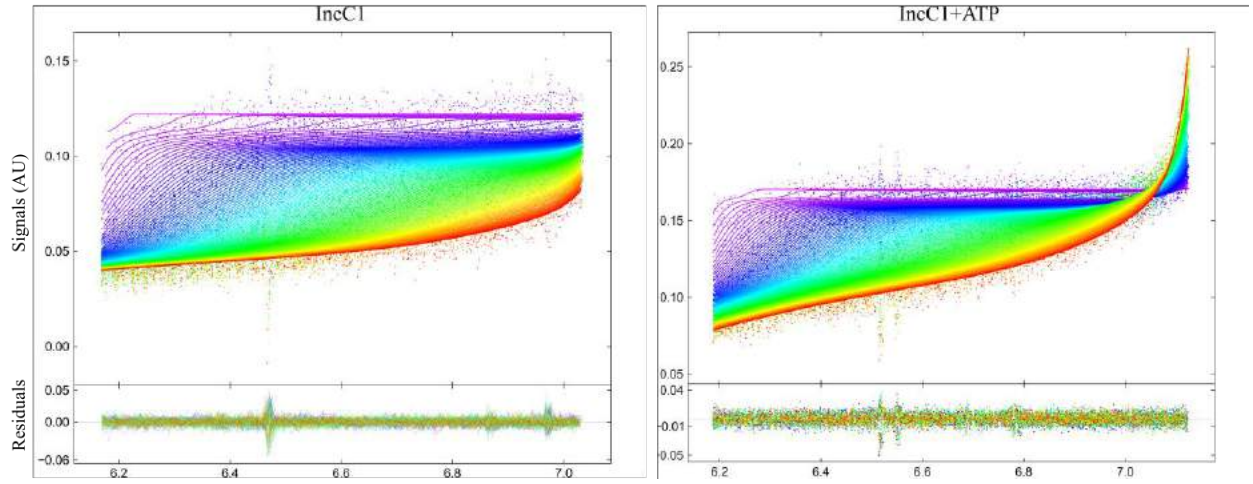
5.12.2. Analytical Ultracentrifugation (AUC)

AUC analysis was performed to check the monodispersity of IncC1 and IncC2 in solution. Two samples for each IncC1 and IncC2 protein (0.2 mg/mL), with and without ATP, were used for sedimentation velocity AUC (svAUC). The absorbance signal (AU) down the cell at different times of centrifugation was fitted to the Lamm equation using SEDFIT software and the data and the residuals were plotted for both proteins (Figure 5.20a, 5.21a). Plots of the sedimentation distribution c(S) vs S for IncC1 and IncC2 protein shows that both proteins are predominantly monomeric in the solution (Figure 5.20b and Figure 5.21b). The molecular weights for IncC1 and IncC2 from AUC were calculated as 41.5 kDa and 31.4 kDa respectively (Table 5.4), similar to the calculated monomeric masses (Table 5.4). The sedimentation coefficient for IncC1 and IncC2 was observed to increase in the presence of ATP. This may suggest some conformational change in the monomeric state of proteins, in the presence of ATP. The analysed data from AUC shows that the sedimentation coefficient of IncC1 increases slightly in the presence of ATP, suggesting a change in conformation, and less of the protein is monomeric. With the IncC2, a larger change in the value of S was observed in the presence of ATP but the mass is still that of a monomer.

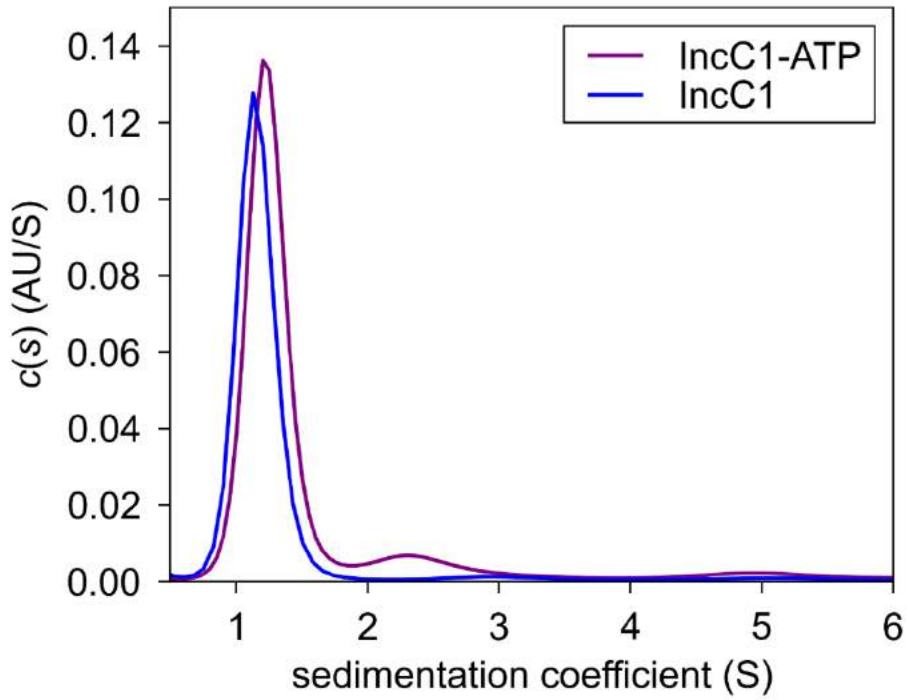
Table 5.4 The c(s) distribution, f/f_o ratios and mol. wt. obtained from AUC for IncC1 and IncC2

					Major Species				**Minor Species**			
Sample	**Conc mg/mL**	**f/f_o**	**rmsd**	**Mol. wt.[a]**	**$S_{20,w}$**	**c(S)**	**Mol wt.**	**%**	**$S_{20,w}$**	**c(S)**	**Mol wt.**	**%**
IncC1	0.20	2.65	0.0069	40.5	1.30	1.13	41.1	98.5	3.01	2.8	174	1.5
IncC1+ ATP	0.20	2.37	0.0064	40.5	1.70	1.27	38.6	94.0	2.40	2.5	113	6.0
IncC2	0.25	2.06	0.0130	29.9	1.75	1.47	31.4	100	-	-	-	-
IncC2+ATP	0.25	1.93	0.0148	29.9	2.11	2.08	34.6	99.5	4.50	4.10	270	0.5

[a]calculated from the amino acid sequence

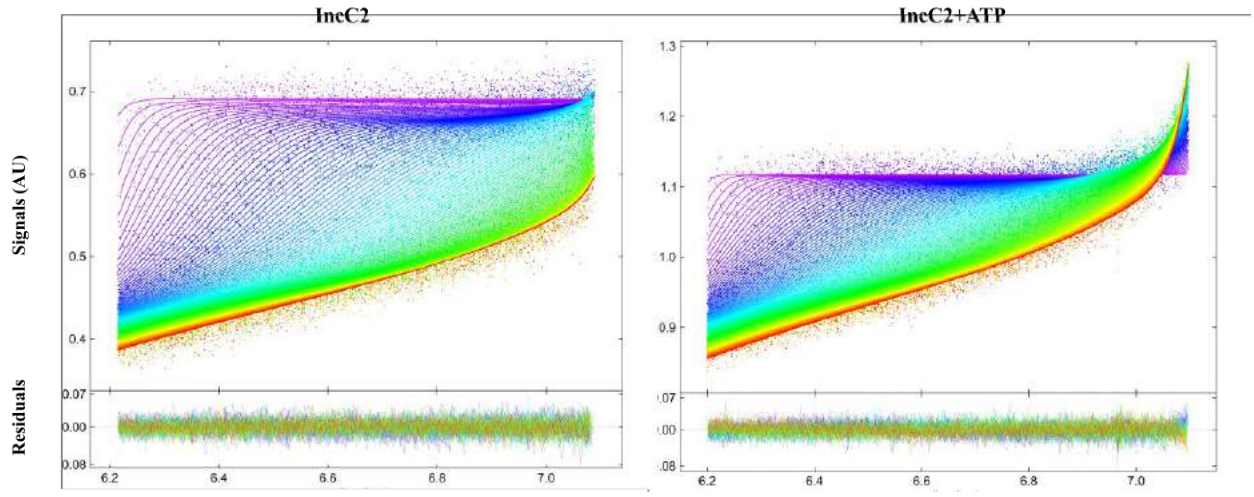


(a)

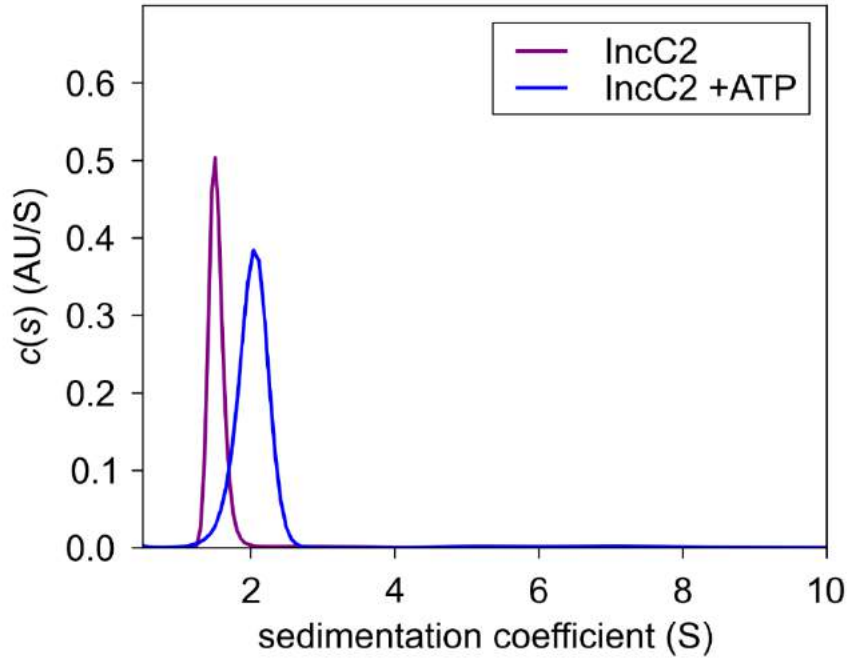


(b)

Figure 5.20 AUC analysis for IncC1. (a) IncC1-His_6 protein sample (0.20 mg/mL) was used with and without ATP (0.5 mM). AUC was performed at 40, 000 rpm and absorbance scans were collected every five minutes. The plot shows the absorbance at 280 nm down the cell at different times of centrifugation. SEDFIT software was used to analyse the AUC data. The upper and bottom meniscus was used as 6.12, 7.16 cm for IncC1 and 6.13, 7.19 for IncC1-ATP respectively. The frictional ratio (f/f_o) was calculated as 2.65 and 2.37 for IncC1 and IncC1-ATP respectively. The residuals show the quality of fit. **(b)** The sedimentation distribution [S] vs c(s) was obtained using the c(s) distribution model. The molecular mass was obtained from c(M) distribution model as 41.1 and 38.6 kDa for IncC1 and IncC1-ATP respectively. The c(s) distributions from IncC1 and IncC1-ATP fit were exported to GUSSI (a plugin to plot and compare data). In IncC1-ATP data, the peak from the major species is shifted. It might suggest a conformational change in monomeric protein state. A small amount of a minor species showing the molecular mass of a tetramer (174 kDa) was also found in the presence of ATP.



(a)



(b)

Figure 5.21 AUC analysis for IncC2. (a) IncC2-His_6 protein sample (0.20 mg/mL) was used with and without ATP (0.5 mM). AUC was performed at 40, 000 rpm and absorbance scans were collected every five minutes. SEDFIT software was used to analyse the AUC data. The upper and bottom meniscus was used as 6.16, 7.16 cm for IncC2 and 6.16, 7.19 for IncC2-ATP respectively. The frictional ratio (f/f_o) was calculated as 2.06 and 1.93 for IncC2 and IncC2-ATP respectively. The residuals show the quality of fit. **(b)** The sedimentation distribution [S] vs c(s) was obtained using the c(s) distribution model. The molecular mass was obtained from c(M) distribution model as 31.4 and 34.6 kDa for IncC2 and IncC2-ATP respectively. The c(s) distributions from IncC2 and IncC2-ATP fit were exported to GUSSI (a plugin for plotting and comparing data). IncC2 major species peak is shifted in the presence of ATP. This might suggest protein aggregation of protein conformational change in monomeric state. Minor species showing the molecular mass of an octamer (270 kDa) and even higher were also observed for IncC2.

5.12.3. SAXS for IncC1 and IncC2

The SAXS data were collected at beamline B29, ERSF, using different concentrations of IncC1 and IncC2 (~0.2-2 mg/mL), and in the presence/absence of nucleotide *i.e.* ATP and ADP. The software SCÅTTER ver. 3.1R (www.BIOISIS.net) was used to analyse the SAXS data. As described in chapter 2 and 3, a Guinier plot ($Log_{10}I(q)$ vs q^2) was calculated to get information about the radius of gyration (R_g). The first points of the Guinier plot were not linear, showing that the IncC1 and IncC2 proteins tend to aggregate in solution (Figure 5.22a and Figure 5.24a). The R_g values, for both proteins, in the presence and absence of ATP and ADP calculated from the linear portion of the Guinier plots are given in Table 5.5 (Figure 5.23 and Figure 5.24).

The same data were used to obtain normalized Kratky plots ($I_q/I_o(q.R_g)^2$ vs $q.R_g$) for all IncC1 and IncC2 samples. The P(r) distribution plots were obtained to calculate the maximum dimension (D_{max}) of the proteins, as well as the real-space R_g and I_o (Table 5.5) values. The molecular mass values for IncC1 were obtained using the I_o values from real space analysis, but due to the low intensity signal, were not accurate for IncC2.

The IncC1 and IncC2 Kratky plots, in the presence of ATP, are bell shaped, whereas in its absence they show an upward trend at the end, showing that the proteins might adopt a more globular protein conformation when bound to ATP (Figure 5.23). The BSA was used as a reference sample to calculate molecular mass using following equation.

$$\text{Sample Mol mass} = \frac{\text{Sample } I_o}{\text{Reference } I_o} \times \text{Reference mol mass}$$

Table 5.5. **SAXS analysis for IncC1 and IncC2.** The radius of gyration (R_g) values were calculated from Guinier Plots and Distance Distribution analysis/P(r) Plots. The qR_g values used for the Guinier analysis are also mentioned. The I_o (zero angle light scattering) from distance distribution plot was used to calculate the molecular mass.

Protein	Concentration (mg/mL)	R_g[a] (Å)	qR_g[a]	R_g[b] (Å)	qR_g[b]	I_o[c]	Mol Mass[d] kDa	Mol Mass[e] kDa	Porod Volume	D_{max} (Å)	χ^2
IncC1	1.5	26.37±0.40	0.59-1.29	24.06	51-999	48.0	40.5	44.6	67.41	68.00	0.37
IncC1-ADP	1.5	26.44±0.71	0.43-1.29	23.72	71-999	43.6	40.5	40.5	52.43	68.41	0.81
IncC1-ATP	1.5	27.46±0.47	0.77-1.24	25.61	46-999	59.1	40.5	55.5	117.61	80.51	0.31
IncC2	1.0	22.3±1.21	0.32-1.28	19.54	81-1071	0.27	29.9	-	25.00	52.50	1.33
IncC2-ATP	1.0	25.1±2.00	0.51-1.27	24.71	34-1000	0.29	29.8	-	43.77	52.28	0.88
IncC2-ADP	1.0	22.4±1.71	0.32-1.29	20.68	41-1000	0.32	29.8	-	85.81	70.77	1.71

[a]Determined from Guinier analysis

[b]Determined from Distance Distribution analysis

[c]Determined from Distance Distribution analysis

[d]Determined from the amino acid sequence

[e]Determined from zero angle intensity of scattering

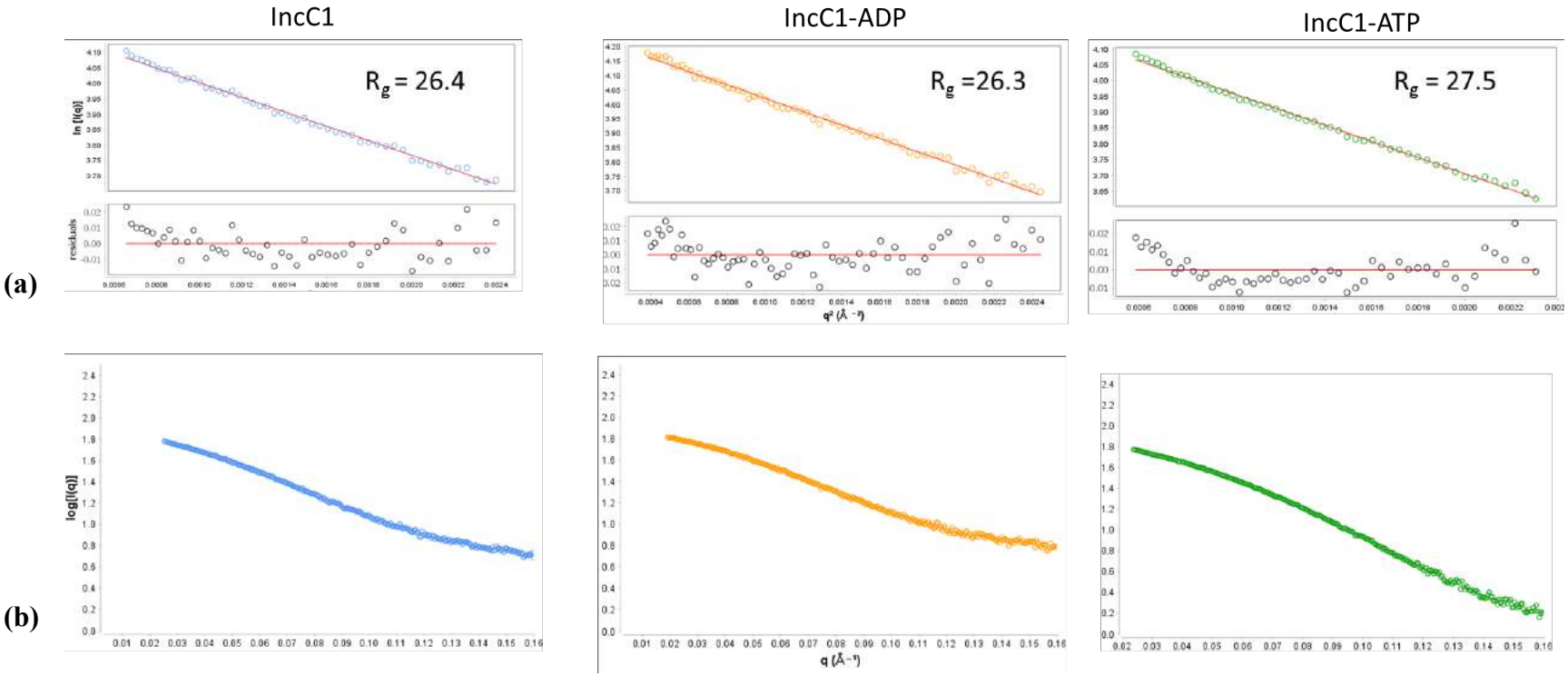


Figure 5.22 Guinier plots and Log-Intensity plot for IncC1. **(a)** The SAXS data were analysed using Scatter ver. 3.1R. Guinier plots were obtained from SAXS data after subtracting the buffer scattering form original data. The radius of gyration (R_g), for all samples, was calculated at very small scattering angles ($q<1.3/R_g$). The R_g, calculated for IncC1, was 26.4 Å using qR_g 0.59-1.29. R_g, for IncC-ADP, was 26.3 Å over qR_g 0.43-1.29 and R_g, for IncC1-ATP was 27.5 Å with qR_g 0.77-1.24. The residuals are shown in the panel below Guinier plots. **(b)** For the log-intensity plots, data have been plotted as the modulus of the momentum transfer (q) against Log(Intensity)

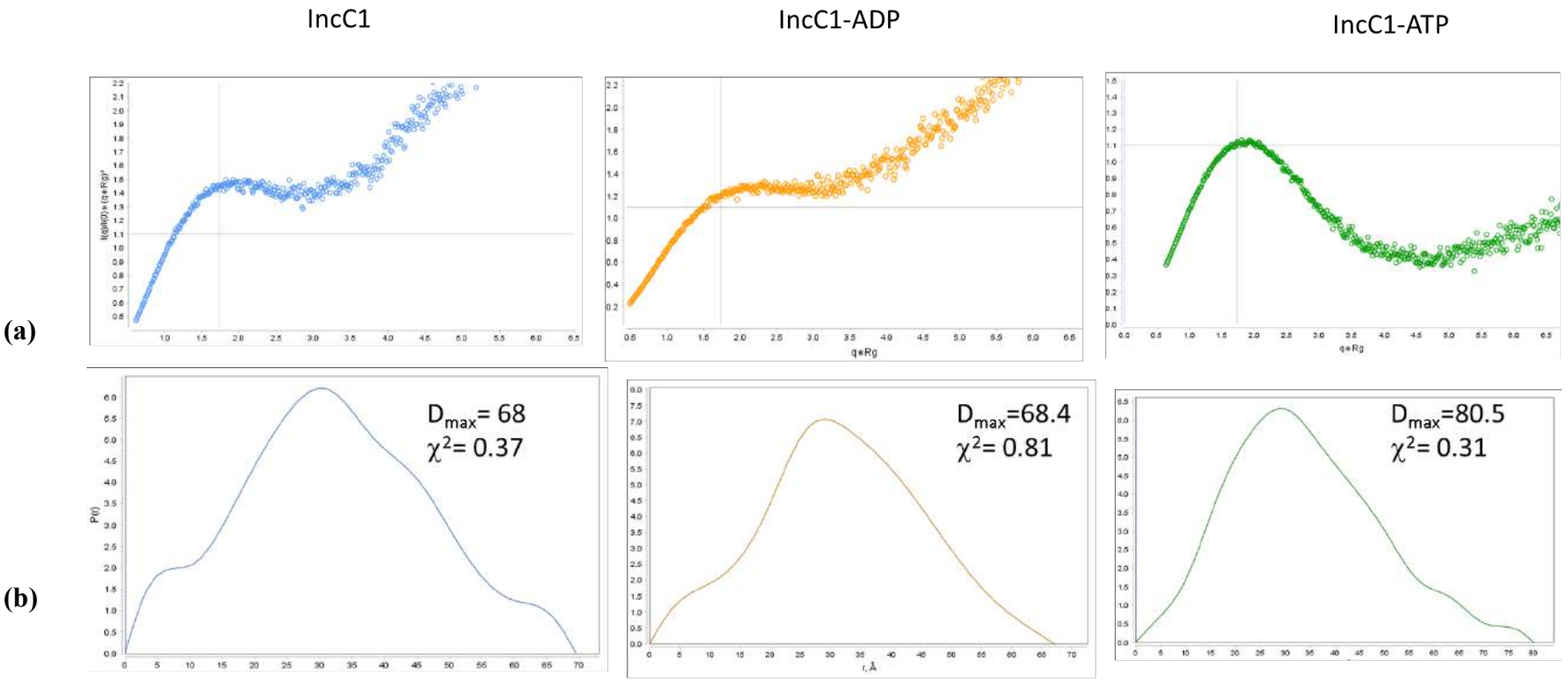


Figure 5.23 The normalised Kratky plots and Distance Distribution plots for IncC1. (a) The Kratky plot determines the protein folding. The standard Kratky plot curves for folded, partially folded and disordered proteins are shown in Chapter 2 (section 2.26.3). IncC1, in the presence of ATP, was found to be more globular in comparison to other samples. **(b)** The Distance distribution plots were obtained to calculate D_{max}. For IncC1, IncC1-ADP and IncC1-ATP D_{max} were found as 68, 68.4 and 80.5 Å respectively.

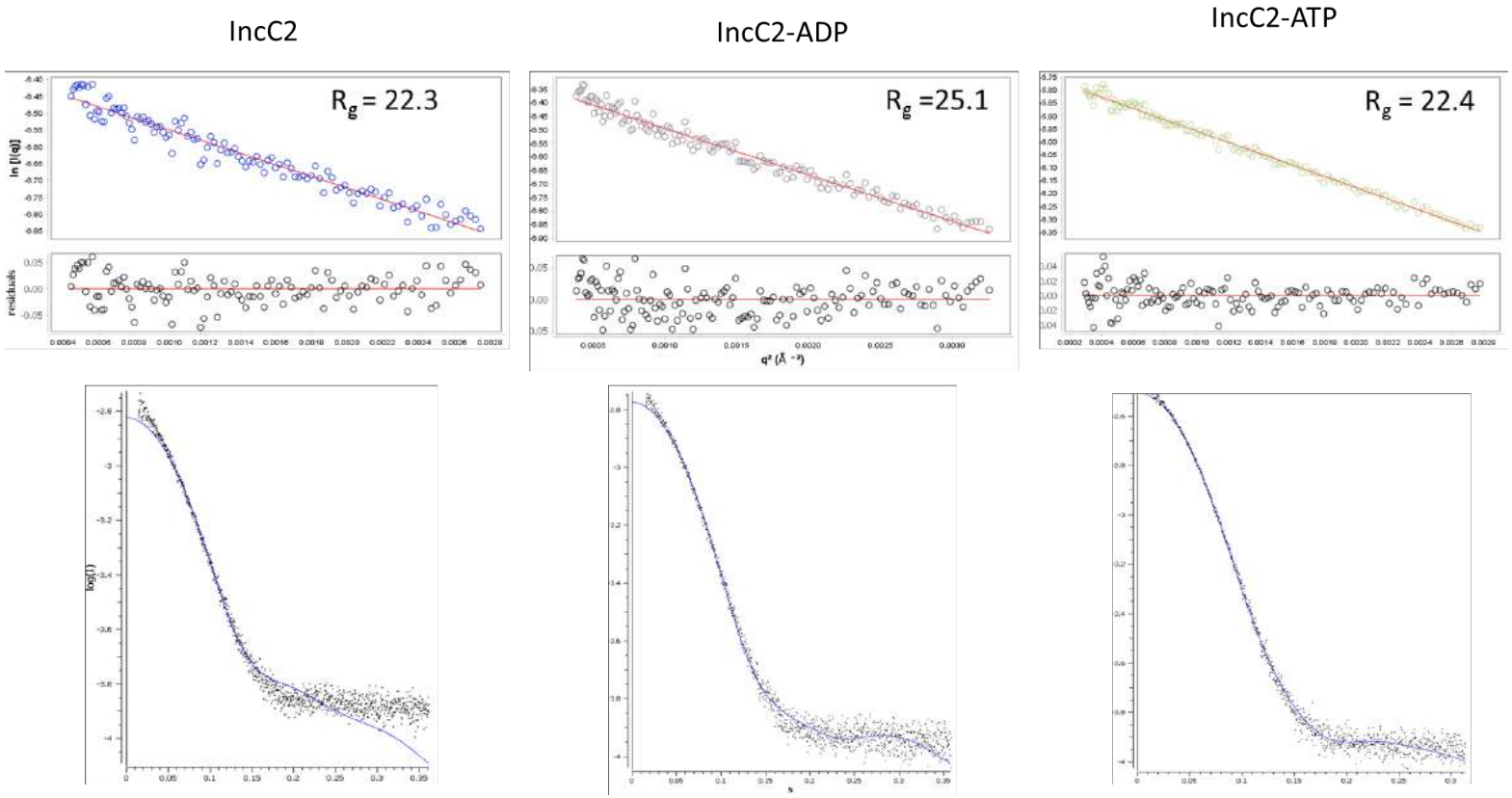


Figure 5.24 The Guinier plots and Log-Intensity plots for IncC2. (a) The Guinier plots were obtained from SAXS data after subtracting the buffer scattering from original data. The radius of gyration (R_g), for all samples, was calculated at very small scattering angles ($q<1.3/R_g$). The R_g, calculated for IncC2 was 22.3 Å using qR_g 0.32-1.28. R_g, for IncC2-ADP, was 25.1 Å over qR_g 0.51-1.27 and R_g, for IncC2-ATP was 22.4 Å with qRg 0.32-1.29. The residuals have been shown in the panel below Guinier plots. The residual fit shows that protein may be aggregating in the solution. **(b)** For the log-intensity plots, data have been plotted as (q) vs Log(Intensity) using Primus ver. 3.3 and I_o values were calculated.

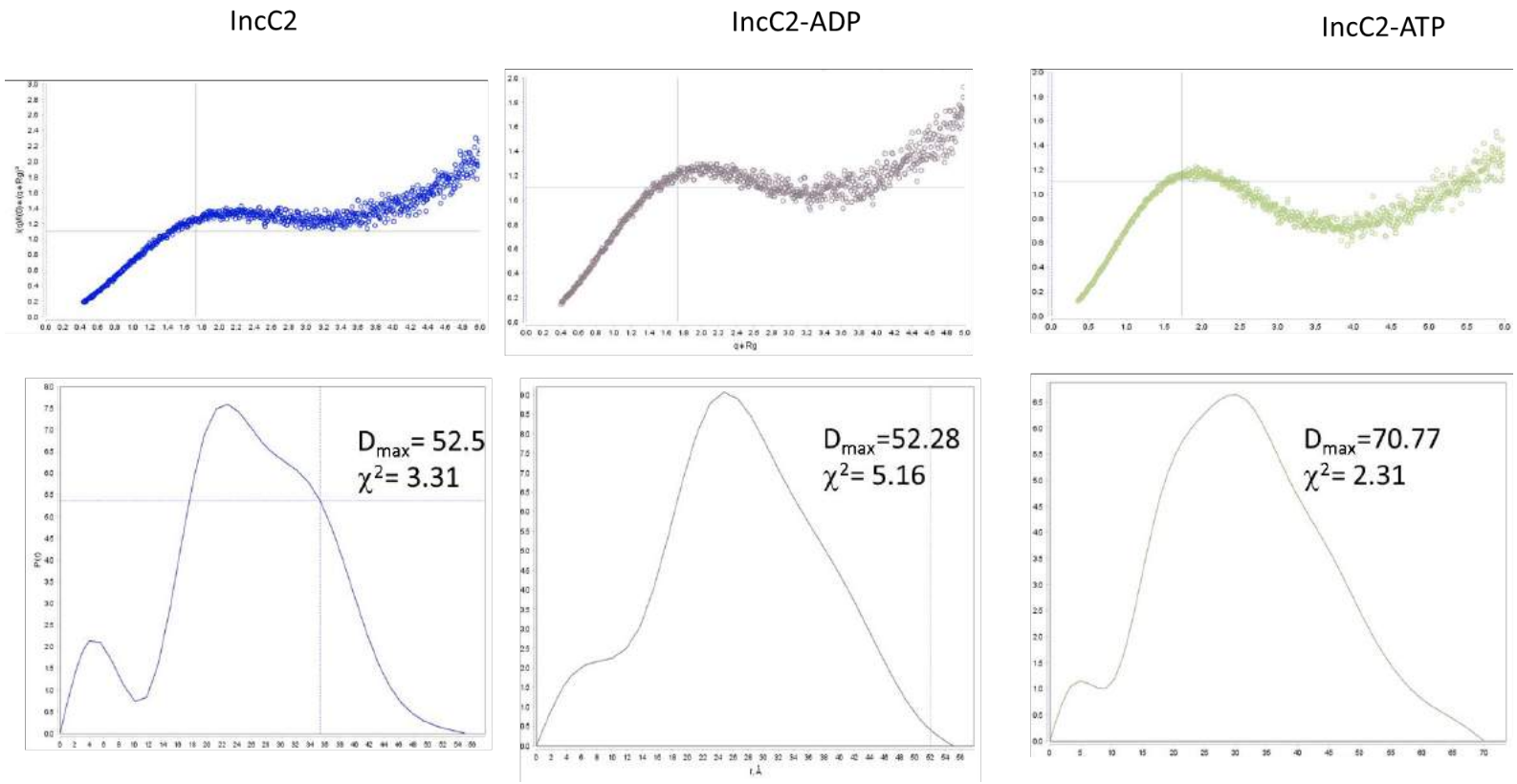


Figure 5.25 The Kratky plots and Distance Distribution plots for IncC2. (a) The Kratky plot determines the protein folding. The standard Kratky plot curves for folded, partially folded and disordered proteins have been shown in Chapter 2 (section 2.26.3). The IncC2 in the presence of ATP was found more globular in comparison to other samples or IncC2 seems to adopt a different conformation when binds to ATP **(b)** The Distance distribution plots were obtained to calculate D_{max}. For IncC2, IncC2-ADP and IncC2-ATP D_{max} were found to be 68, 68.4 and 80.5 Å respectively.

5.12.4. ATPase Assay

The IncC1 and IncC2, like other ParA proteins, contain Walker-like A and B motifs. IncC proteins were found to be weak ATPases. Using Malachite Green Assay and EnzChek® Phosphate Assay Kit (E-6646), the ATPase activity could not be determined in a reliable way, due to low activity. The ATPase activity was then determined by using High-Performance Liquid Chromatography (HPLC). Samples were taken at intervals after reaction and the amount of ATP and ADP was monitored after separation by HPLC, using the absorbance at 254 nm.

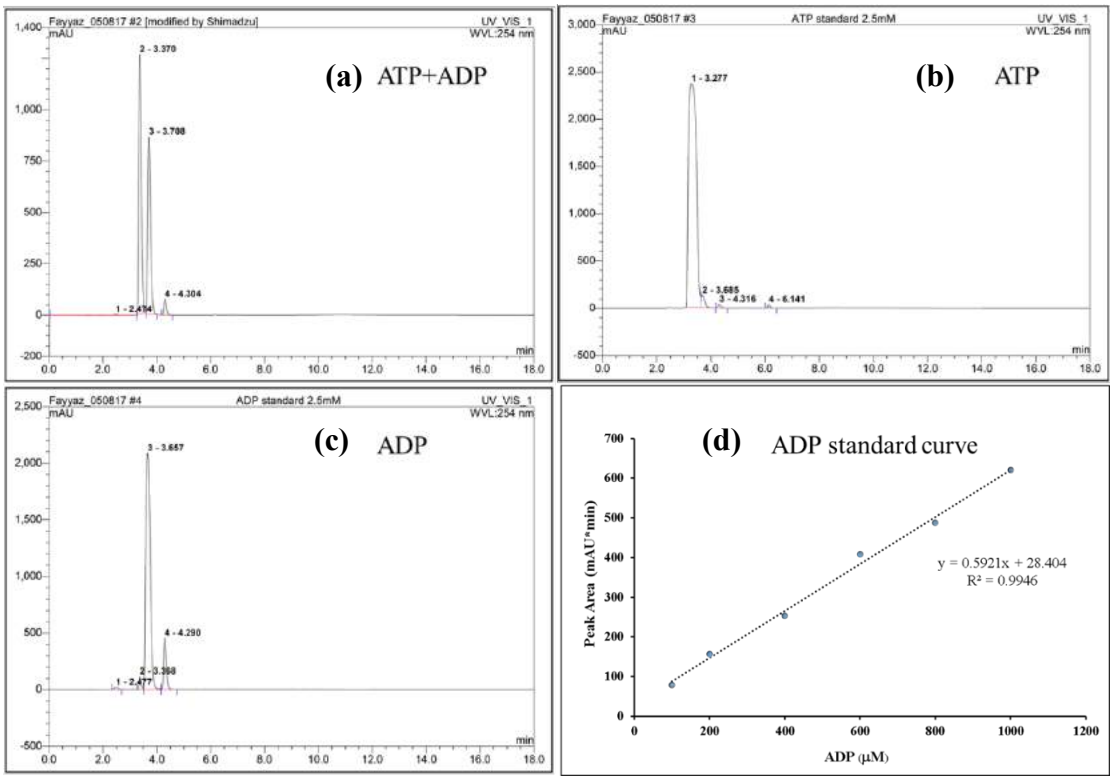


Figure 5.26. ADP and ATP separation using HPLC C18 column. (a) A sample with a mixture of ADP and ATP (2.5 mM, 2-4-fold diluted) was loaded on the C18 HPLC column (flow rate; 0.7 mL/min, 254 nm wavelength, 20 µL sample injection) using phosphate buffer pH 7.0 and methanol (95:5) as the mobile phase. The column was optimized for different run times and flow rates to get better separation of nucleotides. The ATP and ADP were quantified using ChemStation (B.04.03). **(b, c)** ATP and ADP samples were run individually to identify the peaks. **(d) An** ADP standard curve was obtained using ADP samples of varying concentrations (0.1-1.2 mM). The standard curve was used to calculate the ADP produced in the unknown samples.

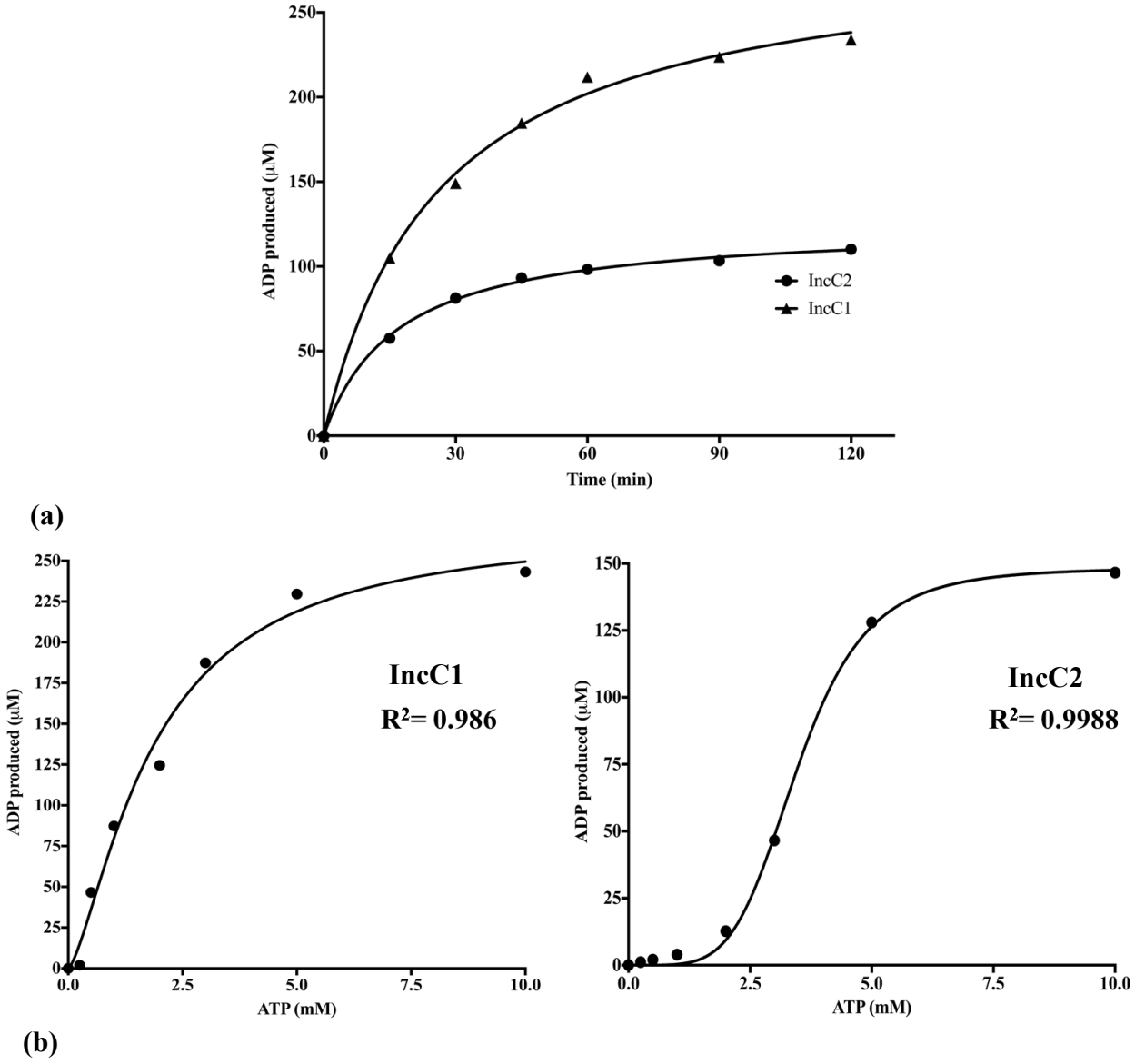


Figure 5.27 ATPase assays for IncC1 and IncC2 proteins. The ATPase activity of IncC proteins was determined by using 30 µM of protein in reaction mixtures containing 20 mM Tris-HCl buffer (pH 7.5), 20 mM Sodium Glutamate, 10 mM $MgCl_2$, 10% glycerol, 100 µg/ml bovine serum albumin. **(a)** In the time course assay, IncC1 and IncC2 protein samples were incubated for different time intervals (15-120 min) and perchloric acid was used to stop the enzymatic reaction. The samples were micro-centrifuged for 5 min at 10, 000 x *g* to remove any precipitate. The time course analysis shows that the ATPase activity for the IncC1 protein was higher than the IncC2 ATPase activity**(b)** The amount of ADP produced after 60 min of reaction is plotted vs ATP concentration in the assay. The equation for specific binding, with a Hill slope $\left(Y = \frac{B_{max}.x^h}{K_d^h.x^h}\right)$was used to fit the data where B_{max} is a maximum rate and K_d is a concentration of ATP required to get half-maximalrate , and 'h' is Hill slope. From the curve analysis, it can be observed that IncC1 may have a different ATP binding mode than IncC2. The affinity of IncC1(K_d=2.78±0.67), for ATP, is more than that of IncC2 (K_d=3.48±0.20). The Hill slope (h) is 4.8 and 1.78 for IncC2 and IncC1 respectively.

IncC1 shows 2-fold higher activity than IncC2 at the same protein and ATP concentration (Figure 5.27a and b). The apparent rate after 60 min reaction was plotted vs initial ATP

concentration and fitted to the Hill equation. The activity of IncC1 has different dependence on ATP than IncC2. The apparent K_m of IncC1(K_m=2.78±0.67), for ATP, is more than that of IncC2 (K_m=3.48±0.20), while the apparent Hill slope (h) is 4.8 and 1.78 for IncC2 and IncC1 respectively (Fig 5.27b). The apparent reaction rates for the ATPase activity, after 60 min, were calculated. For IncC1, the specific ATPase activity calculated was 58.33 nM ADP/min/μM protein while for IncC2, the specific activity was 27.50 nM ADP/min/μM. The ATPase activity in case of both proteins increased in the presence of the 57 bp O_AO_B oligodeoxynucleotide (1.6 fold for IncC2 and 1.9-fold for IncC1). IncC2 ATPase activity increased slightly in the presence of wild-type KorB (1.2-fold) but surprisingly, the ATP activity of IncC1 unchanged with the addition of wild-type KorB protein. In the presence of full-length KorB and DNA (57 bp O_AO_B oligos), ATPase activity increased for both proteins (3.25 fold for IncC2 and 2.5-fold for IncC1) (Table 5.6) and was higher than with either DNA or KorB alone.

The ATPase activity for IncC1 and IncC2 was also determined in the presence of several KorB mutants (Table 5.6), provided by Mr Anmol Gautam and Dr Eva Hyde. IncC2 ATPase activity was slightly decreased in the presence of some KorB mutants *i.e.* and KorB LΔ40 in comparison to wild-type KorB.

Table 5.6 Apparent ATPase activity (nM ADP/uM protein/min) for IncC1 and IncC2 (30 μM) in the presence of DNA, wild-type KorB and KorB mutants. An equal concentration of KorB and KorB mutants (20 μM) was used in the assay. The reaction was carried out for 90 min and stopped by adding perchloric acid. The samples were micro-centrifuged to remove any precipitate before injecting to C18 HPLC column. The amount of ADP produced was calculated using ADP standard curve.

Interacting Partner	**Apparent ATPase Activity (nM ADP/uM protein/min)**	
	IncC2	**IncC1**
None	27.50	58.33
KorB	35.28	55.00
DNA	43.06	113.60
KorB + DNA	88.06	149.44
KorB NΔ150	30.56	58.06
KorB CΔ100	43.89	52.22
KorB NΔ31CΔ100	61.67	83.61
KorB LΔ40	25.56	71.67
KorB DBD	42.22	-
KorB NTD	37.78	-
KorB CΔ105	33.89	-

- not measured

5.12.5. Circular Dichroism for IncC1 and IncC2

Circular Dichroism was performed to estimate the secondary structures of IncC1 and IncC2 in CD buffer (10 mM sodium phosphate pH 7.5, 100 mM sodium perchlorate). The CD data were analysed, using the DichroWeb online server, to predict the secondary structure (Figure 5.28). The two proteins IncC1 and IncC2 were found to be predominantly α-helical proteins (Figure 5.28). For IncC1, the secondary structure was estimated to contain 48% α-helices and 15 % β-sheets (Figure 5.28a) while IncC2, was estimated to contain 61% α-helices and 12% β-sheets (Figure 5.28b).

5.12.6. Melting curves for IncC1 and IncC2 DNA using Circular Dichroism

CD melting curves were obtained for IncC1 and IncC2 to monitor any change in secondary structure as the function of temperature. The melting temperature (T_m) for IncC1 and IncC2, was calculated by plotting the change in CD signal at wavelength 220 nm. This wavelength was selected as both proteins were predominantly helical. T_m, for IncC1 and IncC2, was observed as 42.7±1.2°C (Figure 5.29) 36.0±0.7°C (Figure 5.30) respectively. It shows that IncC1 protein is more thermostable than IncC2.

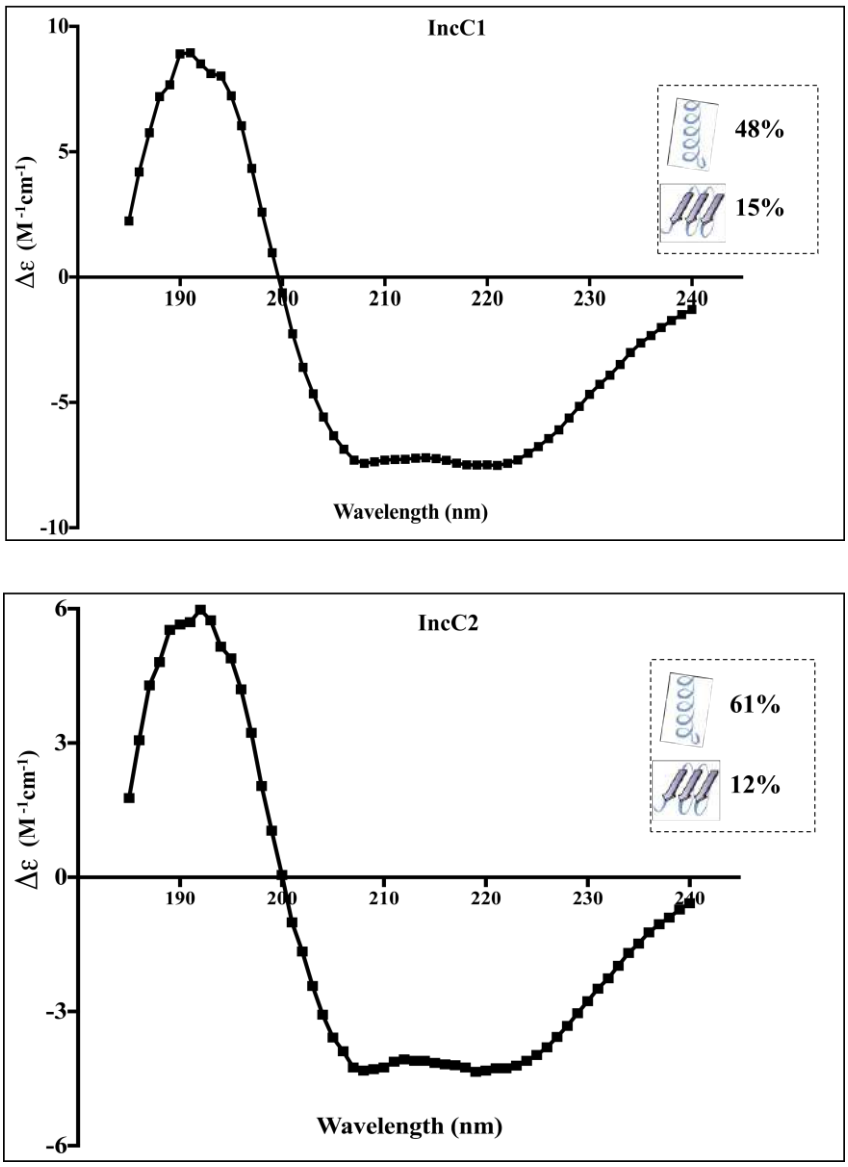
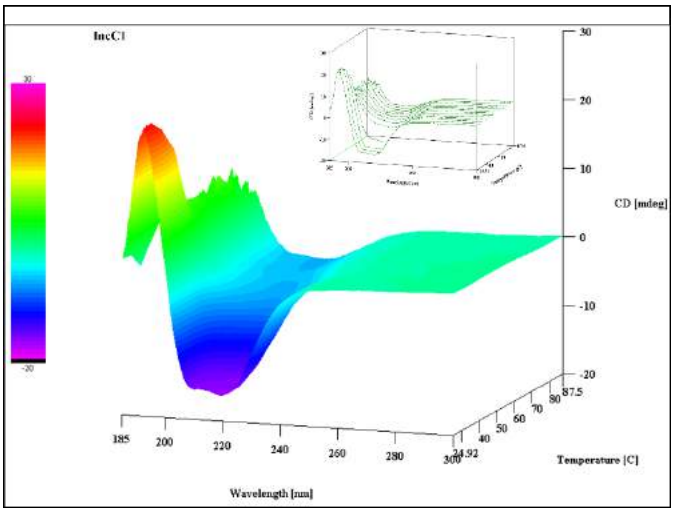


Figure 5.28 Circular dichroism (CD) spectra of IncC1 and IncC2. All CD spectra were collected using a JASCO J-1500 Spectrophotometer. CD spectra of IncC1 and IncC2 (0.2 mg/mL) were collected in CD buffer (10 mM Sodium Phosphate pH 7.5, 100 mM Sodium Perchlorate) using a 2 mm CD cuvette. For each protein, 8 scans were collected from 260-185 nm, scanning speed of 25 nm/min). The DichroWeb server and the CDSSTR (reference data set 3) program were used to analyse the CD data and estimate the secondary structure. The decrease of molar ellipticity around the 220 nm region, as well at 228 nm shows that the IncC1 and IncC2 proteins are predominantly α-helical proteins. **(a)** IncC1 secondary structure was predicted to be 48% α-helices and 15% β-sheet, while **(b)** IncC2 secondary structure was predicted to be 61% α-helical and 12% β-sheet.

(a)



(b)

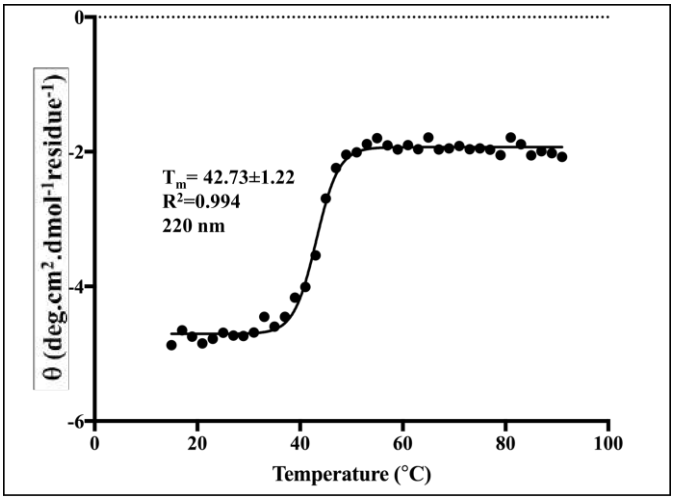
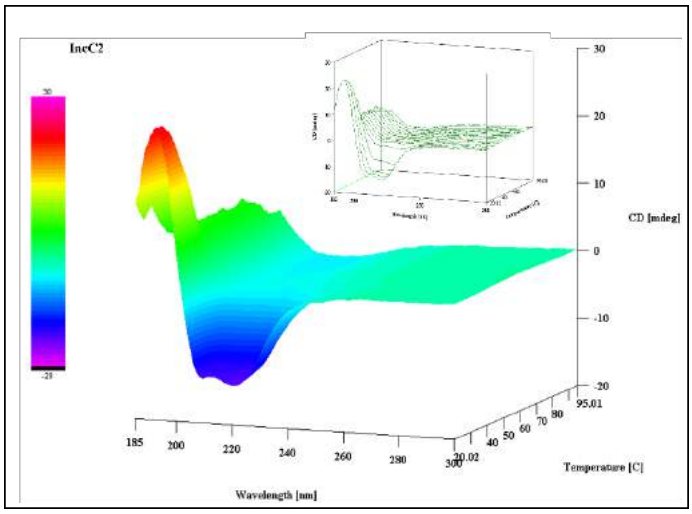


Figure 5.29 The melting curves of IncC1 obtained using Circular dichroism (CD). CD spectra were collected using a JASCO J-1500 spectrophotometer equipped with a Peltier temperature controller. The IncC1 (0.25 mg/mL) in CD buffer (10 mM sodium phosphate pH 7.5, 100 mM sodium perchlorate) was used in a 1 cm cuvette. Three CD spectra were collected at varying temperatures (15-90°C). The temperature was increased by 0.2°C/min and CD spectra were collected at intervals of 2°C. The data were analysed using the interval data analysis software in Spectra manager ver. 2.1. 2D, as well as 3D plots, were obtained. The data were exported to a spreadsheet and the melting curve was obtained by plotting the CD signal at 220 nm vs temperature. **(a)** 3D and 2D plots of CD spectra vs temperature **(b)** The melting curve for IncC1 protein was obtained by plotting the changes in CD signal (θ variations), as the function of temperature, at 220 nm wavelength. The melting curve was fitted using the Boltzmann sigmoid equation. The melting curve showed the T_m of 42.73±1.22°C.

(a)



(b)

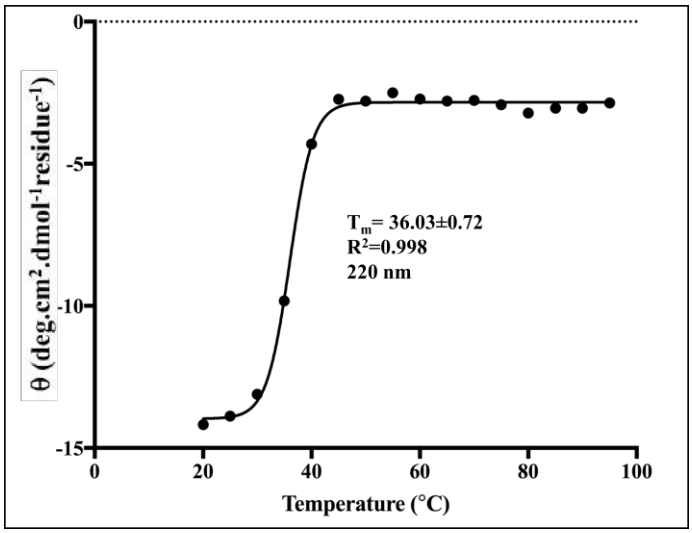


Figure 5.30 The melting curves of IncC2 obtained using Circular dichroism (CD). CD spectra were collected using JASCO J-1500 spectrophotometer equipped with temperature Peltier. The IncC2 (0.25 mg/mL) in CD buffer (10 mM sodium phosphate pH 7.5, sodium perchlorate) was used in 1 cm cuvette. Three CD spectra were collected at varying temperatures (15-90°C). The temperature was increased by 0.2°C/min and CD spectra were collected at an interval of 5°C. The data were analysed using interval data analysis software in Spectra manager ver. 2.1. **(a)** 3D and 2D plots plotted as changes in CD spectra vs temperature **(b)** The melting curve for IncC2 protein was obtained by plotting the CD signal (θ variations), as the function of temperature, at 220 nm wavelength. The melting curve was fitted using the Boltzmann sigmoid equation in Graphpad Prism ver. 7.1. The melting curve gave the T_m as 36.03±0.72°C.

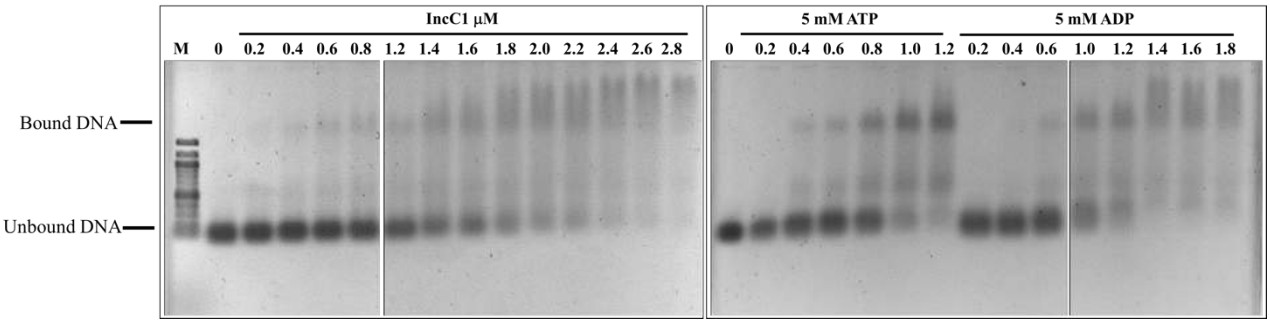
5.13. IncC1 and IncC2 DNA interactions

Many plasmid-encoded ParA proteins interact with the DNA to regulate their own *par* operon. Longer ParA proteins have an extension of ~100 amino acids and may contain a helix-turn-helix (HTH) motif. Short ParA proteins are mostly encoded by chromosomes and lack a HTH motif. The DNA binding of these proteins is either stimulated or inhibited by presence/absence of nucleotides *i.e.* ATP and ADP. The ADP-ATP switch enables ParA proteins to perform their functions during DNA segregation. ParB protein also affects the DNA binding property of ParA proteins.

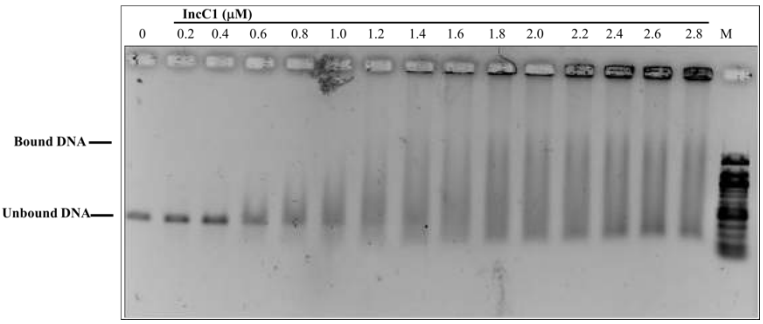
IncC1 and IncC2 proteins, like other ParA proteins bind to DNA. The IncC NTD, the N-terminal extension of IncC1 protein, also binds to DNA independently and non-specifically (as shown in Chapter 3).

5.13.1. EMSAs for IncC1 and IncC2

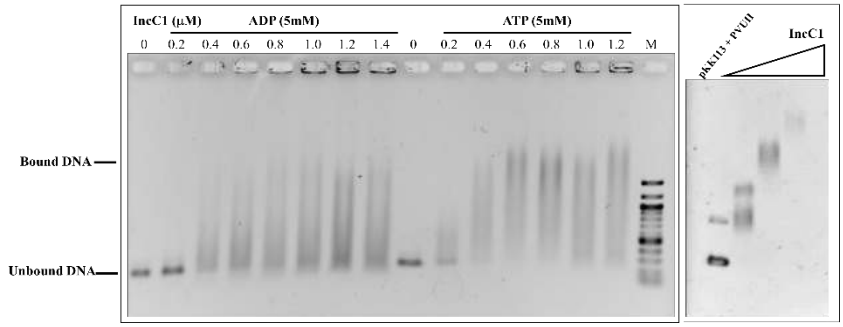
Both IncC1 and IncC2 have been found to bind DNA non-specifically (Batt et al., 2009). To determine the DNA binding activity of these proteins, EMSAs were performed as described in Chapter 2 (Section 2.15). EMSAs were performed, using longer DNA fragments (291 bp and ~3 kb) as well as using a shorter oligodeoxynucleotide (57 bp O_AO_B), in the absence and presence of nucleotides *i.e.* ATP and ADP. EMSA showed that IncC1 binds and retards the shorter DNA, giving a higher band, but some DNA smears were also found on the gel, showing that the protein dissociates in the gel. IncC1 protein binds to the shorter DNA in the presence or absence of nucleotides (ATP, ADP), although, the protein has higher DNA binding affinity in the presence of ATP or ADP (Figure 5.31a), as more distinct bound bands are seen. Similar results were obtained with the longer DNA, however here, an agarose gel was used rather than an acrylamide gel, and much more dissociation of the protein was observed, giving predominantly smears on the gel (Figure 5.31b). In contrast, it was observed that IncC2 does not bind DNA in the absence of ATP or ADP (Figure 5.32a) while, DNA binding was observed, with both long and short DNA, when either ADP, ATP or ATPγS was present (Figure 5.32b and Figure 5.33a-b).



(a)

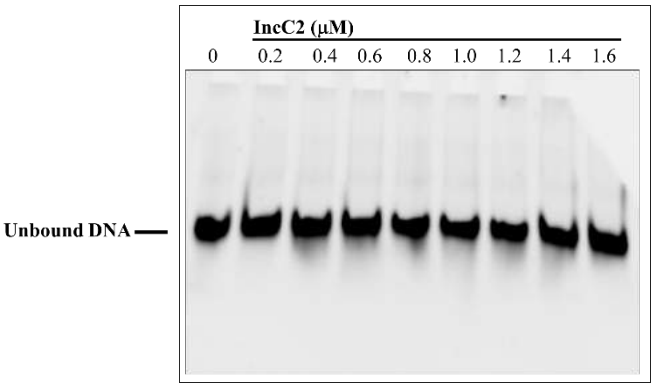


(b)

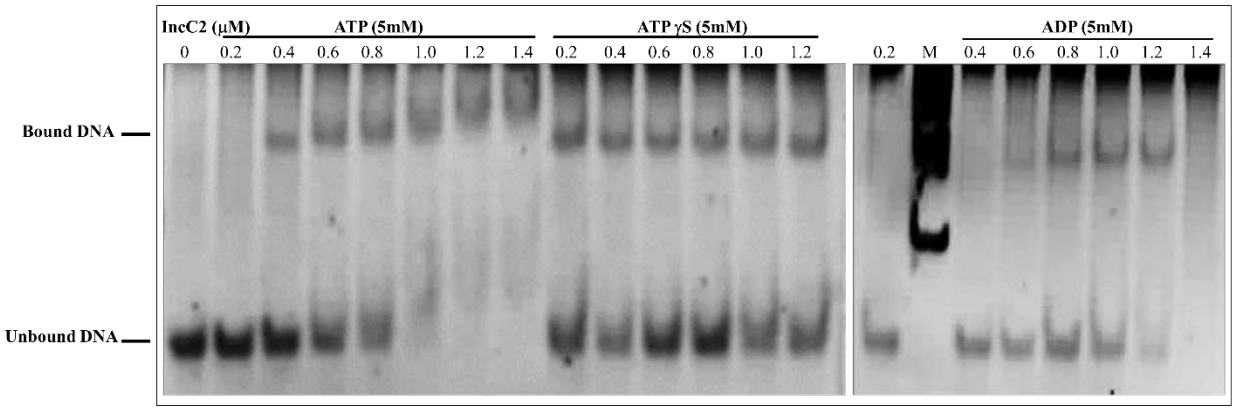


(c)

Figure 5.31 EMSA for IncC1. The DNA binding of the IncC1 was performed using longer DNA fragments of (291 bp and ~3kb) as well as a shorter DNA oligo (57 bp, O_AO_B oligo). **(a)** The EMSAs were performed using 57 bp O_AO_B oligos and 8 % PAGE (1X TBE buffer, 4 cm/V). DNA binding buffer (1x DNA binding buffer contained Tris pH 7.5, NaCl 150 mM, 5 % glycerol, 0.1 mM EDTA, 10 mM $MgCl_2$, and 0.1 μg/mL BSA) was used to prepare protein-DNA samples. IncC1 binds and retards the shorter DNA. DNA smears were also found on the gel. In the presence of ATP and ADP, IncC1 showed DNA binding with higher affinity. In the presence of ATP, the DNA binding affinity was even higher than ADP. **(b)** EMSA, using a 291 bp DNA fragment was performed by Agarose gel electrophoresis (1 % agarose, 1X TBE running buffer, 2-4 V/cm). IncC1 varying concentrations (0.2-2.8 μM) were used. The gel shows DNA retardation and DNA smears in the absence of any nucleotide **(c)** In the presence of ATP and ADP, IncC1 binds to DNA with higher affinity. At high protein concentration, IncC1-DNA complex did not enter in the gel, might be due to the formation of high molecular species. **(d)** The IncC1 was allowed to bind 291bp and ~3 kb linear DNA fragment. The DNA binding shows that IncC1 binds to DNA fragments irrespective of length.

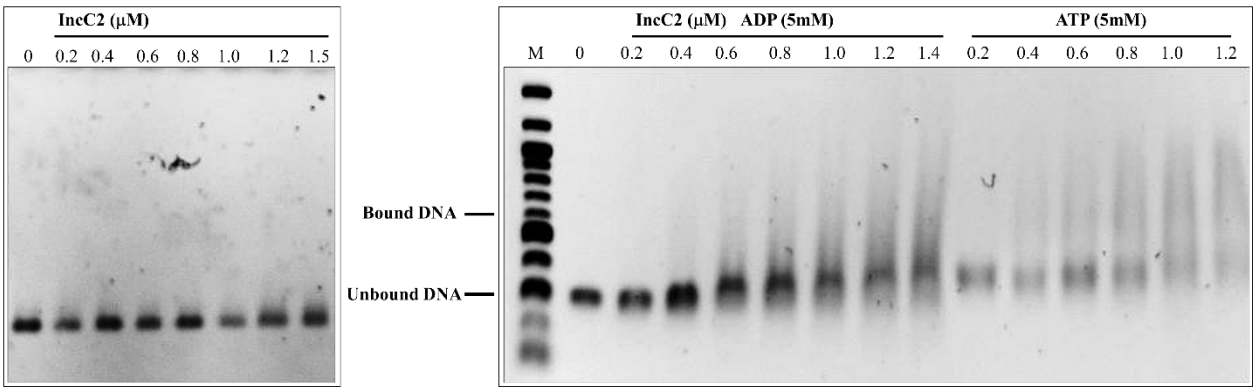


(a)

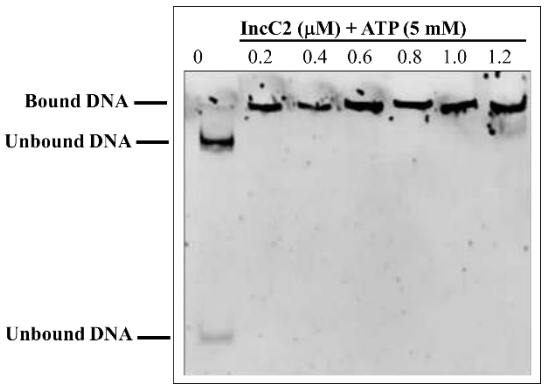


(b)

Figure 5.32 EMSA for IncC2. The DNA binding of the IncC2 was performed using the short DNA oligos (57 bp, O_AO_B oligo). **(a)** The EMSAs were performed using 8 % PAGE (1X TBE buffer, 4 cm/V), while DNA binding buffer (1x DNA binding buffer contained Tris pH 7.5, NaCl 150 mM, 5 % glycerol, 0.1 mM EDTA, 10 mM $MgCl_2$, and 0.1 μg/mL BSA) was used to prepare protein-DNA samples. IncC2 did not bind to the short DNA fragments in the absence of nucleotides *i.e.* ATP and ADP. **(b)** In the presence of ATP, ATPγS and ADP, IncC2 was found to interact with shorter DNA oligos. IncC2 at varying concentrations (0.2-1.4 μM) was incubated with DNA and either nucleotide. The gel shows DNA retardation. In the presence of ATP, all protein was bound to DNA at 1 μM concentration but protein was half bound to DNA in the presence of ATPγS at the same concentration.



(a)



(b)

Figure 5.33 EMSA for IncC2. The DNA binding of the IncC2 was performed using the longer DNA fragments (291 bp and ~3kb) **(a)** The EMSAs were performed using 8 % PAGE (1X TBE buffer, 4 cm/V). DNA binding buffer (1x DNA binding buffer contained Tris pH 7.5, NaCl 150 mM, 5 % glycerol, 0.1 mM EDTA, 10 mM $MgCl_2$, and 0.1 μg/mL BSA) was used to prepare protein-DNA samples. IncC2 did not bind to longer DNA fragments in the absence of nucleotides *i.e.* ATP and ADP. In the presence of ATP, ATPγS and ADP, IncC2 was found to interact with DNA oligos. IncC2 at varying concentrations (0.2-1.4 μM) was incubated with DNA and either nucleotide. The gel shows DNA retardation. In the presence of nucleotide, DNA smears were observed on the gel. **(b)** In the presence of ATP, IncC2 interacts with DNA fragments with and without an O_B site and all protein binds to DNA at 0.2 uM concentration in the presence of ATP.

5.14. IncC1 and IncC2 interactions

The interactions between IncC1 and IncC2 proteins were observed using Microscale Thermophoresis (MST). IncC2 protein was labelled using Red HNS dye (NT-647-NHS, Nanotemper Technologies). IncC1 (10 μM stock concentration) was serially diluted and added to MST reaction mixture as described in Chapter 2 (section 2.28). Figure 5.34a shows the normalised fluorescence for each sample during the experiment. The normalized fluorescence change was plotted as the function of IncC1 protein concentration to calculate its binding affinity for IncC2 (Figure 5.34b). The data analysis shows that IncC1 and IncC2 interact with each other with a K_d of 130±10 nM (Figure 5.34).

5.15. IncC-ATP/ADP interactions

A Thermal Shift Assay (TSA) was used to monitor the interactions of IncC1 and IncC2 with ATP and ADP. TSA was performed, and the melting temperatures measured, as described in Chapter 2 (section 2.26.4), with different concentrations of nucleotides. Data were analysed as described in Figure 5.10. The melting temperature for the proteins was plotted as function of temperature (Figure 5.35) and fitted to a non-linear sigmoidal curve. IncC2 was fully saturated with ~8.1 mM ATP and ~9.5 mM ADP (Figure 5.35a), while IncC1 was fully saturated with ~5.3 mM ATP and ~7.6 mM ADP (Figure 5.35b). It was found that both proteins were more thermostable in the presence of ATP than ADP.

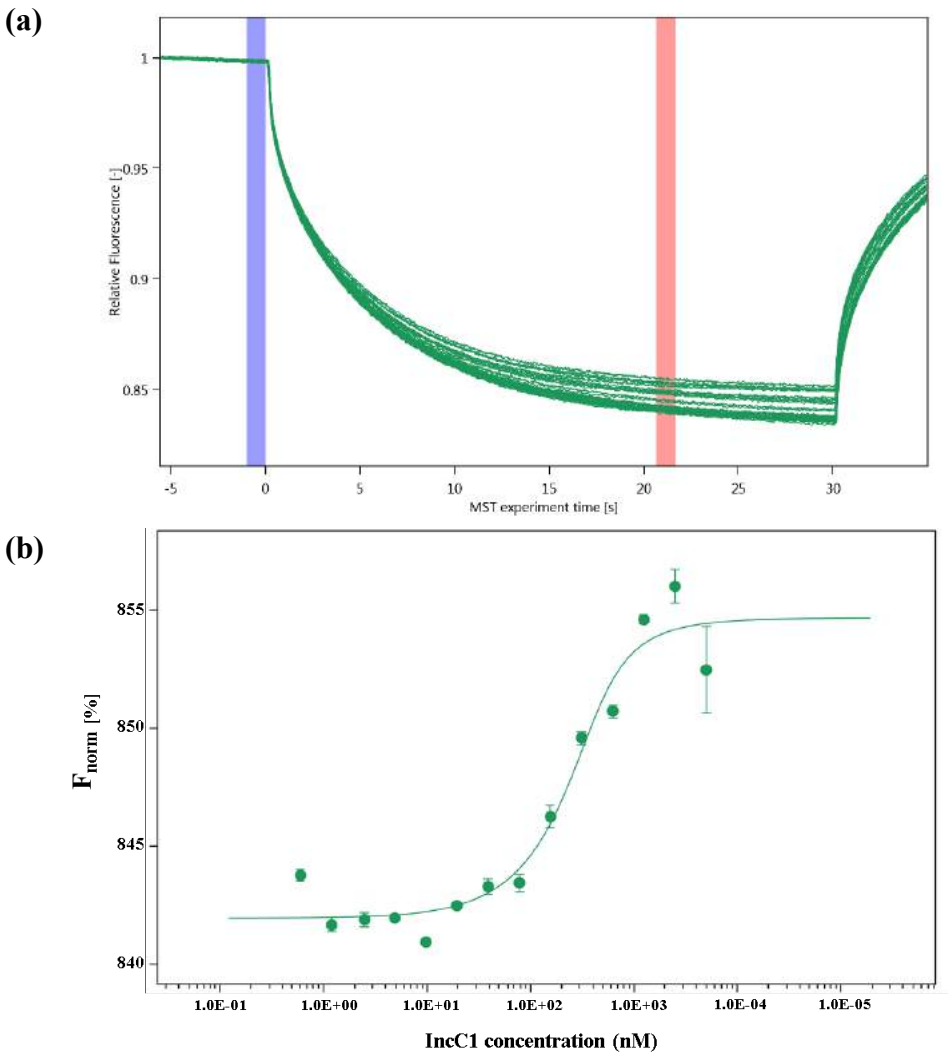


Figure 5.34 IncC1-IncC2 interaction analysis using MST. 14 serial two-fold dilutions were made from IncC1 in MST buffer (20 mM HEPES, pH 7.4, 100 mM KOAc, 10 mM MgSO4, and 2 mM DTT). An equal amount of fluorescently labelled IncC2 protein was added to all reaction mixtures **(a)** Raw data showing changes in relative fluorescence in different samples, during the experiment. The 40 % MST power and 15 % LED power was used, with Laser pulse 'on' time of 30 s, and Laser 'off' times of 5 s. The coloured bars show the regions of the curves that were taken to obtain the Fluoresence change (F norm). **(b)** A single K_d binding model was fitted to the data using MO Affinity Analysis ver. 2.1.5 and the K_d was calculated to be 130 ±10.7 nM. The error bars show the difference between three replicate experiments with new dilutions of the same protein preparation.

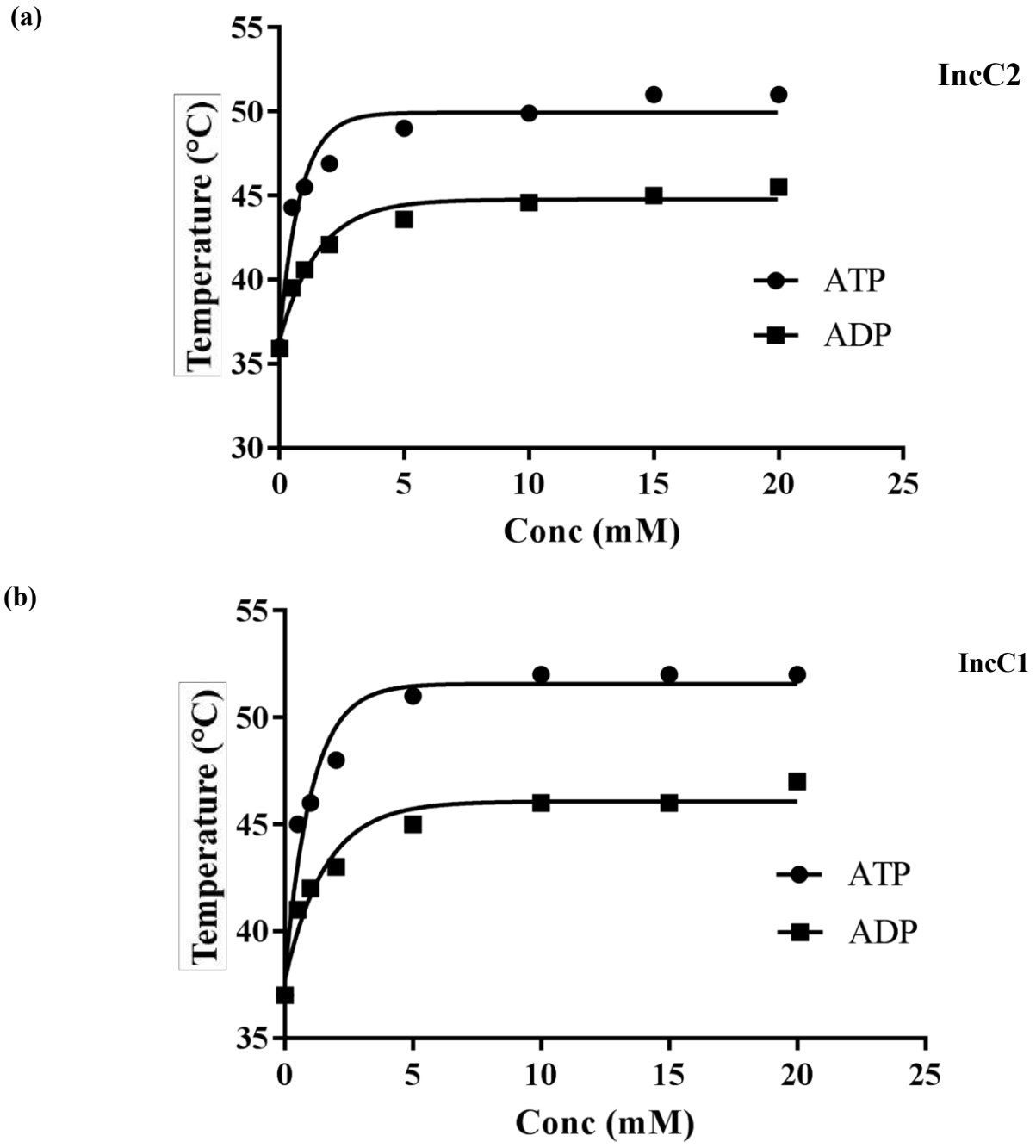


Figure 5.35 IncC1 and IncC2 thermal stability in the presence of ATP and ADP. TSA was performed to determine the thermostability of two proteins in the presence of ATP and ADP. 0-20 mM nucleotide was added to different samples while 20 μM of each protein was used. The T_m was calculated for each sample and plotted against nucleotide concentration. This was fitted to a single site, hyperbolic curve, to obtain the binding affinity **(a)** IncC2 was found to be more stable in the presence of ATP than ADP. IncC2 was fully saturated with ~8.1 mM ATP and ~9.5 mM ADP **(b)** Similar results were obtained for IncC1. IncC1 was fully saturated with ~5.3 mM ATP and ~7.6 mM ADP. In comparison, IncC1 protein is more thermostable than IncC2.

5.16. Discussion

There are two ParA proteins (IncC1 and IncC2) present in the RK2 plasmid; that is very unusual for a ParABS DNA partitioning system; although it has been demonstrated that only one IncC protein is sufficient for the segregation process (Williams et al., 1998). The number of IncC1 and IncC2 molecules varies significantly in different hosts. In *Pseudomonas putida* and *E. coli,* the ratio of IncC1 to IncC2 was found to be 1:1 and 10:1 respectively. This shows IncC1 might be a dominant protein in the plasmid partitioning process in a particular host. This also suggests the N-terminal domain (105 amino acids) in IncC1 might play an important role in IncC1 function as well as in its dominance in different hosts (Batt, 2008). The two IncC proteins have been found to make very high oligomers that were not soluble during protein purification (Batt, 2008).

IncC2 secondary structure prediction and disorder determination show that it is a natively folded protein with high α-helical content (Figure 5.1). The other plasmid and chromosomal ParA proteins, across different bacteria, have been found to have high helical contents (Zhang and Schumacher, 2017, Schumacher, 2008). The computational prediction results were in agreement with the secondary structure estimates from Circular Dichroism (Figure 5.28). CD shows that more than 60% of the IncC2 protein consists of α helices. The IncC2 protein was found to be highly unstable and aggregated during the cell lysis and protein purification. IncC2 protein was expressed, varying growth times, temperatures, inducer concentrations and using different bacterial strains (Figure 5.4-5.6). At low temperatures and using T7 Express *E. coli* strain, IncC2 protein expression was found to be better than other conditions but soluble protein was only obtained after optimizing the lysis buffer. Tris buffer pH 7.5, supplemented by Glycerol, sodium Glutamate and a reducing agent, was helpful in solubilizing the overexpressed protein (Figure 5.7). In other work, glycerol and Glutamate salts have been found to stabilize proteins by preventing protein aggregation (Lebendiker and Danieli, 2014). Lim et al., (2014) have also used 50 mM Potassium glutamate during successful ParA purification. Previously, Batt, et al. (2009) also used glycerol to maintain IncC1 and IncC2 protein solubility/stability.

The IncC2 protein was more stable in Tris pH 7.5 buffer and in high salt concentrations (500-1000 mM) as found by thermal shift assay and buffer optimization. IncC1 and IncC2 proteins were lysed and purified using the same buffer mentioned above but a soluble protein concentration > 2 mg/mL was not achieved. Even fusion to the solubility tags like the SUMO tag

and the GB tag (Figure 5.9) did not increase protein solubility by a considerable amount. This insolubility might be due to the formation of high molecular weight oligomers by these proteins at high concentrations and the fusion tags may not prevent the formation of these complexes. During IncC2 purification, by Ni-NTA chromatography, the protein precipitated either on the column or in the eluted protein fractions. This suggested that protein elution at high concentrations, or high Imidazole concentrations or interactions of the protein with free Ni leaching from the column were causing the protein to precipitate. To avoid some of these problems, column overloading was avoided, and eluted protein fractions were diluted with B6 Buffer pH 7.5 with 0.1 mM EDTA.

The accurate Stokes radius and molecular weight could not be determined for IncC2 protein using size exclusion chromatography, where the protein was eluted few fractions after the void volume (Figure 5.12). This suggests, either IncC2 protein does not have a globular conformation, or it aggregated during elution. In previous studies, the His tag from overexpressed IncC1 and IncC2 was removed to try to enhance protein solubility, but this did not affect the proteins' solubility/stability (Batt, 2008). The 260/280 Absorbance ratios for the purified proteins was found to be >1.0 and could not be lowered even after SEC or SP Sepharose purification. This suggests the proteins have some bound moiety (nucleotides or DNA) that absorbs at 260 nm wavelength. The presence of reducing agents *i.e*. DTT, also alters the 260 nm absorbance peak as DTT absorbs at 250 nm (Porterfield and Zlotnick, 2010); however this was not present in all samples. By Mass spectrometry (ESI-MS), IncC1 and IncC2 molecular masses were found as 40.6 kDa and 30 kDa respectively (Figure 5.14 and 5.18). These masses both differ from the molecular masses calculated using the amino acid sequences, by +49Da. Additional mass might be due to the presence of some sugar or Imidazolone derivative attached to the proteins during protein purification or due to some post-translational modification (PTM) or it may be due to mis calibration of the mass spectrometer.

SEC MALLS analysis for IncC1 and IncC2 show that protein may undergo some conformational changes when interacting with ATP. In the presence of ATP, both IncC proteins were predominantly dimeric (Figure 5.19). The SEC analysis for the Soj (a homolog of chromosomal ParA protein in *Bacillus subtilis*) showed that it is a monomer in the absence of any nucleotide or with ADP, while, a monomer-dimer equilibrium can be observed in the presence of ATP (Leonard et al., 2005). However, the AUC results for Soj protein did not agree with the SEC MALLS results for the same protein, and showed a monomer (Leonard et al.,

2005). IncC2 protein has shown similar results. IncC2 protein, like Soj protein, was predominantly monomeric in AUC analysis, in the presence or the absence of nucleotides (ATP, ADP) but dimeric with ATP in SEC-MALLS. It can be proposed that Soj and both IncC proteins form homodimers in the presence of ATP, but during the long AUC sedimentation run (18 hr), the majority of ATP added was hydrolysed. The AUC analysis also showed high frictional ratios (Table 5.4) for IncC1 and IncC2. This was might be due to non-globular shape or presence of aggregates in the protein samples. Although the molecular mass calculated from AUC was similar to the monomeric masses of both IncC proteins; in the presence of ATP, the sedimentation coefficient was changed. This might show protein conformational changes in the presence of ATP (Figure 5.20, 5.21).

The Kratky plot for IncC proteins, obtained from SAXS data, also shows that in the presence of ATP the IncC proteins undergo some conformational change and they are more globular in the presence of ATP (Figure 5.23 and Figure 5.25). Some other ParA proteins also have been found to be more stable in the presence of ATP (Bouet et al., 2007, Hester and Lutkenhaus, 2007).

The multiple sequence alignment of the IncC2 protein against other ParA proteins shows that the proteins are well-conserved and contain sequence motifs, including Walker-like A and Walker-like B motif (Figure 5.2). These motifs are involved in nucleotide (ATP, ADP) binding and hydrolysis. The Type I ParA proteins contain a GK rich segment (KGGXXK[T/S]) called a Walker-like A box (Koonin, 1993). This motif has been found to be different from conventional walker motifs found in other ATPases and kinases in prokaryotes and eukaryotes (Walker et al., 1982) which is GXXGXGKT. The flexible P-loop is formed by the Walker-like A motif amino acids and facilitates nucleotide binding by interaction with the phosphates (Hayashi et al., 2001, Leonard et al., 2005, Pratto et al., 2009). The lysine (K) and serine/threonine (S/T) are highly conserved across species in Walker-like motifs, and bind ATP and magnesium respectively (Hishida et al., 1999, Mitchell and Oliver, 1993, Panagiotidis et al., 1993, Seefeldt et al., 1992, Hayashi et al., 2001, Jang et al., 2000, Story and Steitz, 1992). Previously, it was shown that Lysine and Threonine residues present in the IncC Walker-like motif (K15 and T16) are involved in Mg^{+2} coordination and ATP hydrolysis (Siddique et al., 2002). In the current study, both IncC proteins were found to be weak ATPases. The apparent specific ATPase activity of IncC1 was found to be 58.33 nM ADP/min/μM and for IncC2 it was 2-fold less than IncC1 at 27.50 nM

ADP/min/μM. Figure 5.27a). Other ParA proteins are also found to contain very weak ATPase activity and need either ParB or DNA or both for stimulation (Bouet et al., 2007, Hester and Lutkenhaus, 2007). The ATPase activity for both IncC1 and IncC2 was stimulated in the presence of either DNA or KorB (Table 5.6). When both KorB and DNA were added to the ATPase reaction mixture, the activity was enhanced ~ 3-fold, greater than the effect of each addition suggesting some co-operativity between the partners.

IncC1 and IncC2 ATPase activities were also monitored in the presence of different KorB mutants. The ATPase activity of IncC2 was stimulated in the presence of KorB CΔ100, KorB NΔ31CΔ100, KorB DBD, KorB NTD and KorB CΔ105 while KorB NΔ150, KorB LΔ40, did not show any effect on ATPase activity of IncC2; KorB LΔ40 even inhibited the ATPase activity. For IncC1, the presence of DNA increased the ATPase activity about 2-fold. The KorB mutant KorB NΔ150 did not affect ATPase activity, while KorB N31CΔ100 and KorB LΔ40 increased the ATPase activity of IncC1 (Table 5.6). Previously, it was shown that IncC2 protein potentiated the repressor activity of wild-type KorB and its mutants (Lukaszewicz et al., 2002), so IncC must interact with these proteins and enhance their DNA binding.

The ParA proteins (P1, SopA) auto-regulate their own *par* operon by binding to the DNA and an N-terminal region (first ~100 amino acids) plays an important role in some ParA-DNA interactions (Friedman and Austin, 1988, Mori et al., 1989). In previous studies, IncC1 and IncC2 proteins were found to interact with DNA non-specifically and with different DNA binding affinities (Batt, 2008, Batt et al., 2009) but only in the presence of nucleotides. The EMSAs in this study show that IncC1 has a higher affinity than IncC2 for both shorter and longer DNA fragments (Figure 5.31a, b). IncC1 binds in the absence of nucleotides, but IncC2 does not, showing that the presence of N-terminal domain (IncC NTD) in IncC1 is required for DNA binding in the absence of nucleotides. Previously, the greater affinity of IncC1 for DNA over IncC2 was suggested to be due to the presence of a putative HTH in the NTD of IncC1 (Batt et al., 2009). In this study, it was found that IncC NTD binds to the DNA at very high concentrations (Chapter 3, section 3.19.6) but it does not have any HTH motif (Chapter 4). In the presence of ATP and ADP, IncC1-DNA binding affinity was enhanced (Figure 5.31 and Figure 5.33). EMSAs, with longer DNA fragment, show DNA smearing on the gel. This might suggest negative cooperative IncC1-DNA interaction (Figure 5.31b). The IncC2 did not interact with either longer or shorter DNA fragment in the absence of nucleotides (ADP, ATP) (Figure 5.32

and Figure 5.33). In the presence of ATP and ADP, IncC-DNA binding affinity was much greater. The IncC2-DNA binding pattern was different in the presence of ATP γS. That suggests, the conformational changes in the IncC2 protein, during ATP binding and hydrolysis, directly affects the DNA binding (Figure 5.32b). DNA binding for ParA proteins, from P1 and F plasmids, were more stimulated by ADP than by ATP (Bouet et al., 2007, Davey and Funnell, 1994). IncC proteins are different in a way that both, ATP and ADP, stimulated the DNA binding. ParA proteins like SopA and Soj have been reported to bind to DNA non-specifically (Bouet et al., 2007) in the presence of ATP-stimulated manner. The IncC proteins also interact with DNA non-specifically. The non-specific DNA interactions may help in potentiating KorB repressor activity. IncC1 remains bound to the non-specific DNA even in the absence of ATP or ADP and this may help in plasmid diffusion along the nucleoid. IncC1 and IncC2 protein form heteromers as shown by MST. The two proteins interact each other with a high binding affinity (K_d 130 nM). Such interactions between IncC proteins are likely to be required for plasmid partitioning.

CHAPTER 6

Mapping the structure and interactions of IncC and KorB proteins

Chapter 6

Mapping structures and interactions of the IncC and KorB proteins

Mapping protein-protein interactions is important to elucidate protein functions. Many approaches can be adopted to study protein-protein interactions, including the binary approach and the co-complex approach. Binary approaches do not tell us the composition of protein complexes but reveal binary interactions. Two-hybrid methods are usually used to get binary information in prokaryotes (Parrish et al., 2007) and eukaryotes (Uetz et al., 2000). Protein chemical-crosslinking followed by Mass spectrometry analysis, provides information about protein-protein interactions as well as giving the composition of multimeric complexes.

6.1. ParA-ParB protein interactions

ParA and ParB proteins interact with each other to allow efficient plasmid partitioning. The different partitioning models, described in Chapter 1, show different modes of ParA-ParB interactions. ParA either polymerises into filaments and interacts with DNA and ParB, or it forms a dimer and its dimeric form interacts with other partners.

IncC1, IncC2 (ParA proteins) and KorB (a ParB protein) are involved in the partitioning of the plasmid RK2. Previously, *in vitro* studies showed that IncC1, IncC2 and KorB each form dimers and interact with each other and with DNA (Batt et al., 2009). In this chapter, Bacterial Adenylate Cyclase Two-Hybrid (BACTH) assays have been used to study *in vivo* interactions between IncC1, IncC2, KorB, and KorA. A chemical crosslinking (CL) approach, using specific chemical cross-linkers, is adopted to determine protein-protein interactions *in vitro*. Cross-linked peptides are identified by Mass spectrometry. A homology model for IncC2 is obtained and the cross-links identified are mapped onto protein homology models for structure optimization.

6.2. Bacterial Adenylate Cyclase Two-Hybrid (BACTH) system

The BACTH system can be used to determine protein-protein interactions in a semi-quantitative way. To measure the interactions using BACTH system, two proteins are expressed as fusion proteins, each containing a reporter domain at either the C- or the N- terminal. These reporter domains are two halves of the catalytic domain of *Bordetella pertussis* Adenylate Cyclase (cyaA) *i.e.* T18 and T25 (Karimova et al., 1998). These two catalytic domain fragments interact with each other and give active Adenylate Cyclase, that converts ATP into cAMP. The

cAMP formed interacts with CAP (catabolite activating protein) and forms a CAP-cAMP complex. The complex, in turn, induces specific promoters and leads to overexpression of genes under the control of these promoters (Figure 6.1A). Importantly, the T18 and T25 domains do not interact with each other if they are both independently expressed in a cell (Ladant and Ullmann, 1999). Lack of interaction leads to no cAMP production (Figure 6.1B). If, two fusion proteins, one containing the T25 fragment and the other containing the T18 fragment at either the C-or N-terminus, interact with each other, they bring the two catalytic fragments into functional proximity (Figure 6.1C). This leads to production of cAMP and to expression of a reporter gene (*e.g.* β-galactosidase) (Figure 6.1C and 6.1D). Activity/signals from reporter gene (β-galactosidase) can be measured quantitatively. If the fusion proteins do not interact with each other, no signal/activity will be measured (Karimova et al., 1998). cAMP-CAP complex is proportional to interactions of fusion proteins at a binding equilibrium.

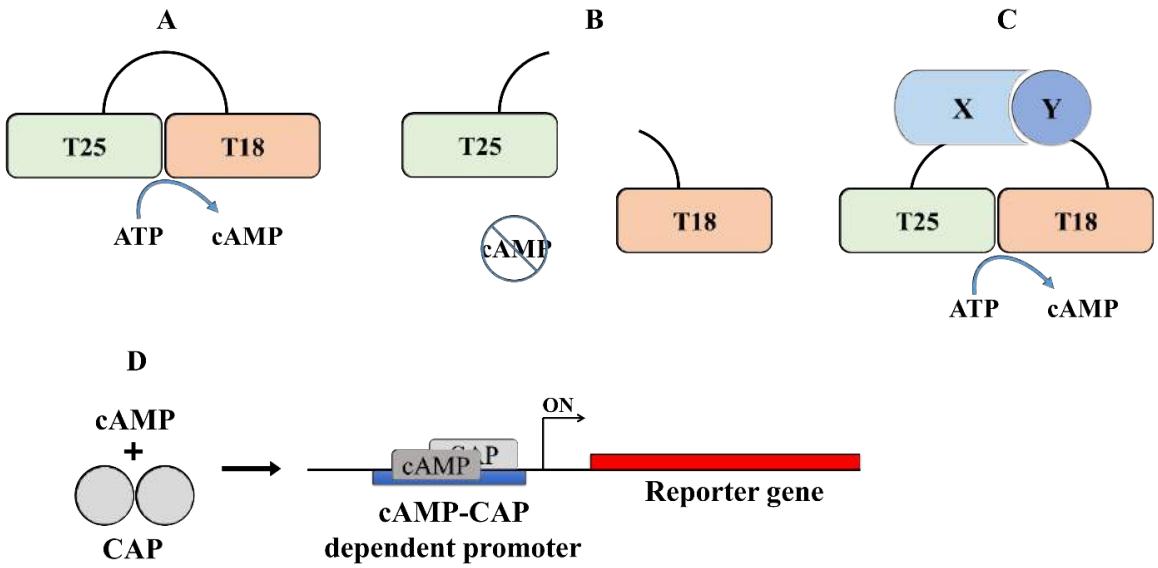


Figure 6.1 BACTH system. Interactions of two fusion proteins can be measured *in vivo* using the BACTH system. The two boxes, T18 (orange) and T25 (green) represent the catalytic fragments of Adenylate Cyclase. T25 represents 1-224 a.a of cyaA while T18 represents 225-399 a.a (**A)** When full-length Cya A protein (1-399) is expressed in cell, the two fragments (T25 and T18) interact with each other and convert ATP to cAMP. **(B)** When the catalytic fragments are expressed individually in cells, the two halves do not interact with each other and cAMP is not produced. **(C)** Two proteins X and Y are expressed in cell as fusion proteins; one protein fused to T18 and other fused to the T25 fragment. If the two proteins interact with each other and bring the two catalytic fragments into functional proximity, the Adenylate Cyclase activity reinstates to produce cAMP from ATP. **(D)** cAMP makes a complex with CAP protein, to form the cAMP-CAP complex that, in turn, activates transcription of some reporter gene (*e.g. lacZ*). Adapted from (Karimova et al., 1998).

β-galactosidase is an enzyme usually used as a reporter gene in the BACTH assay. The gene for β-galactosidase, *lacZ*, is controlled by the *lac* operon. This enzyme is fairly stable and remains active at a wide range of temperature, -10°C to 40°C (Rouwenhorst et al., 1989). The activity of β-galactosidase can be measured by its ability to degrade β-galactosides, including lactose and ONPG (O-nitrophenol-β-galactoside). ONPG gives a yellow colour upon hydrolysis to ortho nitro phenol, and galactose, and allows a semi-quantitative measure of enzyme activity by monitoring the absorbance of the ONP released at 420 nm (Lederberg, 1950).

6.3. Chemical crosslinking of proteins

Protein-protein interactions can be studied using a chemical crosslinking approach. Cross-linked proteins can be further analysed using electrophoretic separation and Mass spectrometry. Since the analysis of ribosomal subunits by chemical crosslinking in 1970s, many protein complexes including translocases, proteasomes, Hsp100 chaperone ClpA and RNA polymerases have been studied (Wilkinson et al., 1997, Karadzic et al., 2012, Hinnerwisch et al., 2005, Jennebach et al., 2012). Many efficient chemical cross-linkers have been developed with different arm lengths, from zero to more than 35 Å, and with specificity towards certain amino acids/functional groups.

6.3.1 Chemical crosslinking with Mass spectrometry

Advancements in Mass spectrometry have allowed high-throughput identification of cross-linked peptides/proteins. The distances between different amino acids can be deduced from cross-linked peptides and this information can be used to build/verify three-dimensional homology models of a protein. Site-specific crosslinking allows identification of particular peptides, while short and medium spacer arm-lengths can be used to obtain distance restraints and conformational space. Unlike NMR and X-ray crystallography, crosslinking-based protein modelling approaches can be applied to a heterogeneous mixture of several proteins, of any size or under different physiological conditions. Chemical cross-linkers with a selectivity tag, allow enrichment of cross-linked peptides with a strong cation exchange chromatography step (Fritzsche et al., 2012). Use of isotopically labelled linkers can enhance the accuracy of identification of the cross-linked peptides. Bis-Sulfo Succinimidyl Suberate (BS3) and Di-Succinimidyl Suberate (DSS) are homo-bifunctional amine cross-linkers (Figure 6.3). The two cross-linkers are identical to each other except in that BS3 is water-soluble due to presence of the sulfo-groups, while DSS is not. The isotopically-labelled BS3 cross-linker, BS3-d_4, (with 4

deuterons instead of protons) is also available commercially (Figure 6.2). Both, BS3 and DSS react with primary amine groups (-NH_2, pH 7-9) in proteins. Lysine is a frequent amino acid in many proteins and primary amino groups from two different Lysine amino acids, if present in close proximity (<34 Å), interact with either end of a cross-linker. The proximity threshold is defined by the Lysine side chain flexibility, the spacer arm length, and protein dynamics. During chemical crosslinking, the two Succinimidyl groups are replaced when primary amine groups from the Lysine amino acids reacting with cross-linker (Figure 6.3a). If a second Lysine is not present in close proximity to the already cross-linked Lysine, dead-end cross-links, to water or buffer, are produced. (Figure 6.3b).

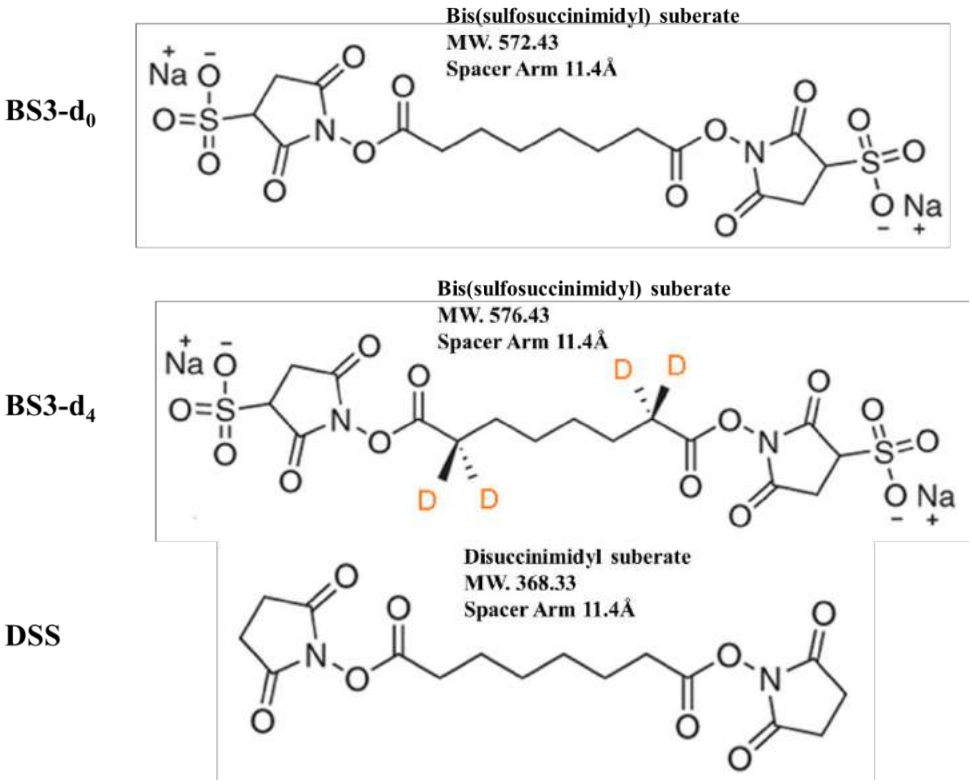


Figure 6.2 BS3-d_0, BS3-d_4 and DSS chemical crosslinkers. The BS3 and DSS cross-linkers are identical to each other and have a spacer arm length of 11.4 Å. Both are Di-Succinimidyl Suberate except BS3 contains the Sulfo-groups that makes it water soluble. DSS is a non-polar cross-linker and is required to be dissolved in DMSO before use. BS3-d_4 is an isotopically labelled version of BS3 and contains four deuterium atoms instead of hydrogens atoms (shown in orange colour).

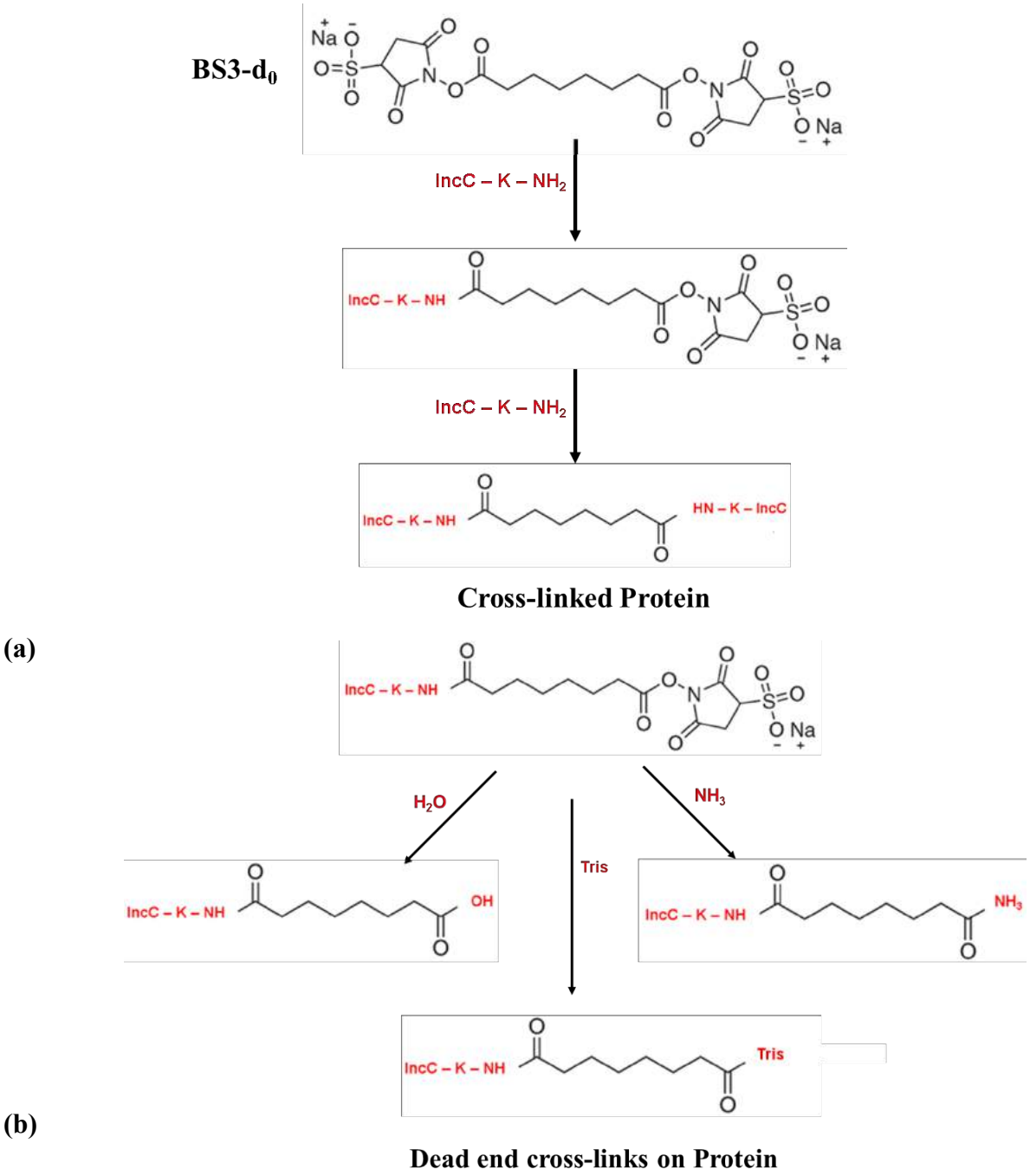


Figure 6.3. A schematic diagram showing protein (IncC)-BS3 crosslinking. (a) Primary amine groups present in the Lysine amino acids are involved in BS3 crosslinking. The Succinimidyl groups present in BS3, are replaced by Lysine's primary amine groups present in a close proximity (<34 Å). **(b)** Sometimes, a second primary amine group from another Lysine is not available. In that case, BS3 may react with water, NH_3 or Tris to produces dead-end crosslinks.

6.3.2. Mass spectrometry data analysis for crosslink identification

The accurate identification of cross-linked peptides is very important in the study of protein-protein interaction. Crosslinking data obtained from Mass spectrometry is complex and difficult to analyse as accurate crosslinks are present in low quantity in comparison to unmodified peptides and cross-linked false positives (mono-link crosslinks and loop-linked crosslinks) (Figure 6.5). In case of protein crosslinking in a cell lysate or inside a cell, the rate of false positives is even higher than crosslinking a purified protein. Use of isotopically labelled cross-linkers and cross-linked peptide enrichment lead to good sample preparation, while MS-MS methods such as collision-induced dissociation (CID) or higher energy collision dissociation (HCD) allow confident data analysis. Multiple software programmes are available to analyse Mass Spectrometry data to identify cross-linked peptides. These software programmes include Spectrum Identification machine (SIM-XL) (Lima et al., 2015), Mass Matrix (MM) (Xu et al., 2010), StavroX (Götze et al., 2012), xQuest/xProphet (Leitner et al., 2013), MeroX (Arlt et al., 2016), Kojak (Hoopmann et al., 2015), PepLynx (Zelter et al., 2010), Protein Prospector (Trnka et al., 2014), and MS2Assign (Schilling et al., 2003). Most of these software programmes use search algorithms that calculate all possible peptide-peptide crosslinks from protein sequence databases. Then, cross-linked peptides are identified by their mass differences from the non-cross-linked peptides. High confidence crosslinks with minimum false discovery rate (FDR) are extracted and distance restraints are calculated. Distance restraints can be used to build or optimize a protein homology model.

6.3.3. Homology modelling improved by chemical crosslinking

Some proteins, like IncC, are highly unstable at high concentrations. The structures of such proteins cannot be determined using NMR or X-ray crystallography. Homology models can be obtained for such proteins, using *ab initio*, threading or template-based protein modelling approaches. Many software programmes (standalone and web-based) are available to generate protein homology models. A homology model obtained from computational predictions is not necessarily the correct model (Xu et al., 2000) and needs verification/optimization. Crosslinking data can be used to evaluate the quality of predicted homology models. Crosslinks are mapped onto a protein model and the Cα-Cα distances restraints allow structure validation by limiting theoretical folds.

6.3.3.1. Comparative protein modelling

Template-based homology modelling is also called comparative protein modelling. Appropriate structure templates from the protein data bank (PDB) are used for sequence-template alignment (Figure 6.4, left panel). A good optimal alignment score is required to obtain a reliable protein model. After sequence-template alignment, the query sequence is superimposed /aligned on backbone of a target template and a homology model is made. The homology model is refined by modelling side chains and loops (Krieger et al., 2005). Side chains and loop structures can be further optimized by energy minimisation (Figure 6.4). The predicted model/models can be ranked with respect to RMSD or model scores (z-score, c-score). Template-based modelling software programmes include MODELLER (Eswar et al., 2006), MODWEB (Biasini et al., 2014), SWISS-MODEL server (Schwede et al., 2003), CPH-MODELS (Nielsen et al., 2010) and YASARA (Krieger et al., 2003). Protein homology models can be validated using online servers including ProSa Web (Wiederstein and Sippl, 2007), PROCHECK (Laskowski et al., 1993), PROCESS (Berjanskii et al., 2010), and Verify3D (Eisenberg et al., 1997).

6.3.3.2. Threading-based protein modelling

Protein models can be obtained using threading-based approaches that predict protein structure based on fold recognition (Xu, 2003) (Figure 6.4, centre panel). A database of various structural templates is generated, using low pairwise sequence similarity to predict the structure of the query protein. The query protein sequence is aligned with each template structure and energy scores are calculated. A final model is obtained by using superimposing the query sequence to the most probable target sequence-template alignment (Figure 6.4).

One of the best threading-based homology modelling software is called RApid Protein Threading by Operation Research technique (RAPTORX) (Kallberg et al., 2014). This software uses position-specific iterated BLAST (PSI-BLAST), secondary structure prediction, sequence homology, pairwise interactions and solvent accessibility to construct a scoring matrix. A final model is obtained using the scoring matrix. Other threading-based software and web servers include Phyre2 (Kelley et al., 2015), I-TASSER (Yang and Zhang, 2015), MUSTER (Wu and Zhang, 2008), HHpred (Söding et al., 2005) and LOMET (Wu and Zhang, 2007).

6.3.3.3. *Ab initio* protein modelling

Ab initio protein modelling involves prediction of the consecutive α-helix and β-sheet stretches from the amino acid sequence of the protein (Fig 6.4 right panel). These secondary elements are then assembled into low-resolution models. Several iterations of secondary structure predictions are performed by randomly picking different models. Predicted models are evaluated using the available information *e.g*. residue pair difference, the radius of gyration (R_g), NMR chemical shifts, amino acid contact predictions and energy calculations. A number of models can be predicted and a protein model with a lower energy score than the previous model is selected based on the Monte Carlo minimization (Figure 6.4). *Ab initio* methods are computationally expensive. Therefore, GPU-based computers or web servers are usually used to submits modelling jobs. The software used for *ab initio* predictions include ROBETTA (Kim et al., 2004) and QUARK (Zhang et al., 2016).

Information obtained from crosslinking data analysis (cross-linked peptides, cα-cα distance constraints) can be mapped on protein homology model to validate or to optimize it. The web-based, XWALK tool (Kahraman et al., 2011) can be used to map and visualize cross-links on protein structures. Some software programmes can export Chimera/PyMol scripts with cross-link distance information *e.g.* the *.pml scripts from SIM-XL, that can be viewed in PyMol. Homology models can be validated by online model validation servers including PROCHECK, Prosa-web and VERIFY3D.

In this chapter, BS3 and DSS cross-linkers have been used for *in vitro* protein crosslinking. A schematic illustration of protein crosslinking and analysis is given in Figure 6.5.

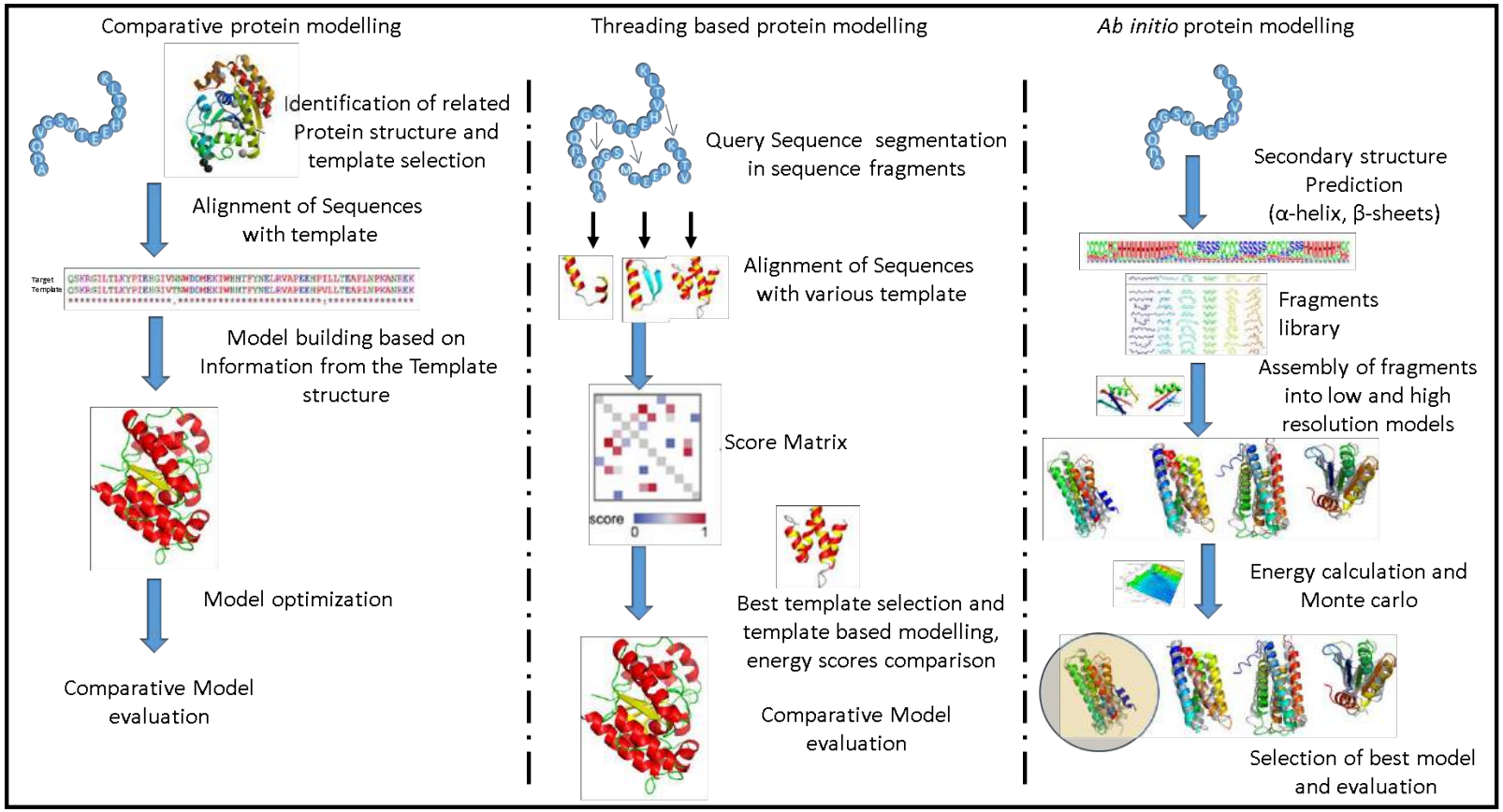


Figure 6.4 Different approaches for protein homology modelling. In comparative protein modelling, (left) protein template structures are obtained on the basis of sequence homology. The query sequence is aligned with template sequences. The alignments, as well as the secondary structure information, are used to build a protein homology model. The protein model is further optimized and evaluated. In threading-based modelling (centre), the query protein sequence is segmented and aligned with various templates. A score matrix is obtained using the information from position-specific iterated BLAST (PSI-BLAST), secondary structure prediction, sequence homology, pairwise interactions and solvent accessibility. The best template is selected using a scoring matrix and a final model is obtained. In *ab initio* modelling (right), secondary structure elements are predicted from the amino acid sequence and a fragment library is generated, based on these predictions. Fragments are then assembled into low-resolution models. The best model is obtained using these low-resolution models and Monte Carlo calculations. Models are evaluated using online structure validation servers.

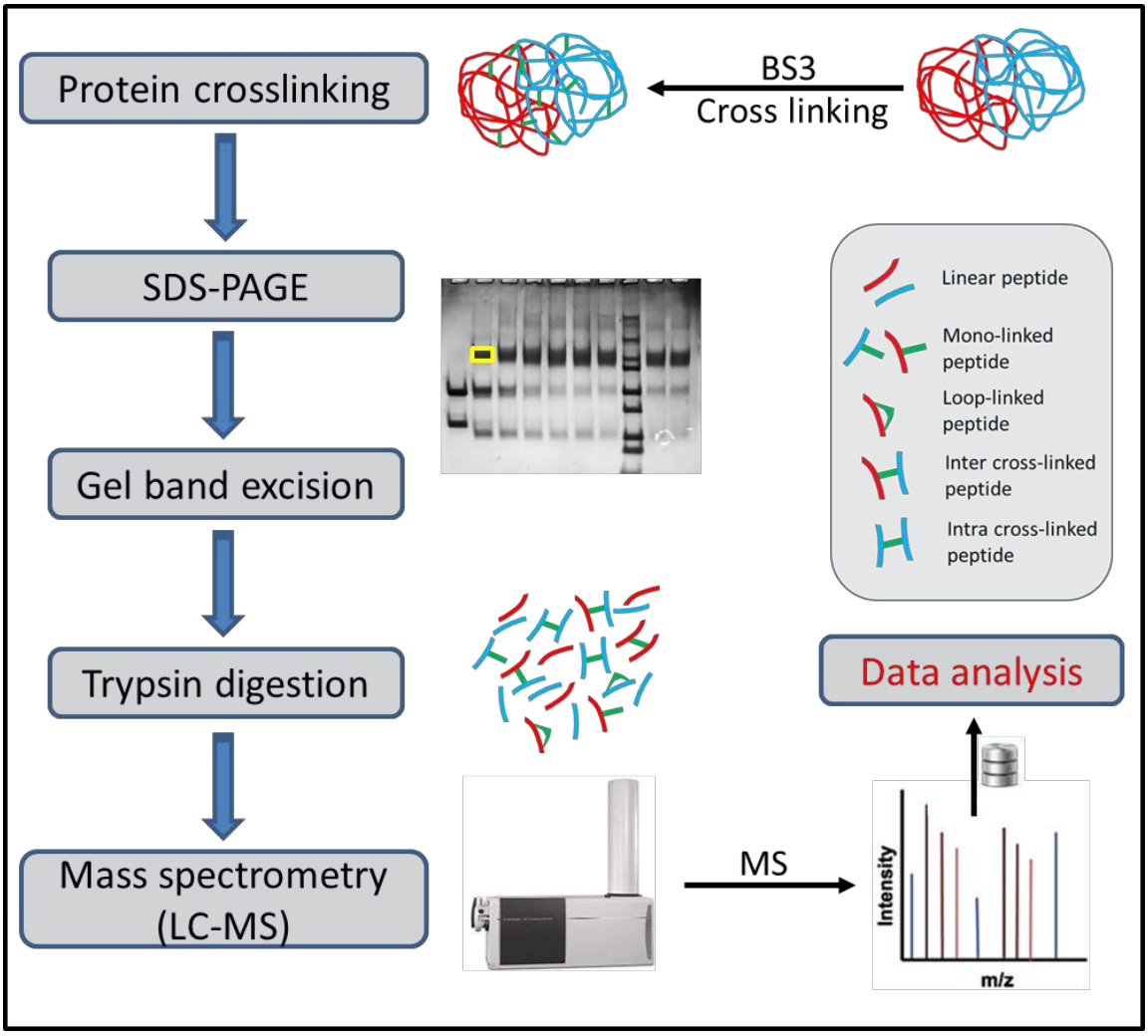


Figure 6.5 A schematic of protein crosslinking followed by Mass spectrometry. Two proteins (shown in red and blue) are cross-linked in the presence of the BS3 cross-linker. Cross-linked proteins are analysed by SDS PAGE. The purified protein bands are excised and trypsin-digested before being subjected to Mass spectrometry. Data obtained from Mass spectrometry is analysed using cross-linked peptide identification software and the presence of either of five types of cross-linked peptides (linear peptides, mono-linked peptides, loop-linked peptides, inter-cross-linked peptides and intra cross-linked peptides) is identified.

6.4. Methods

The gene cloning methods used in this chapter, are described in Chapter 2. BACTH plasmids (pUT18, pUT18c, pKT25, pKNT25, pUT18c-Zip, pKT25-zip) were obtained from Dr Prateek Sharma (Grainger Lab) (Battesti and Bouveret, 2012,). Plasmid maps are given in Appendix I and their properties are described in Table 6.1. Cloning and sequencing primers are given in Table 6.2.

6.5. Construction of protein fusions with T18 and T25

The full length *incC2*, *incC1*, *korB*, and *korA* genes were PCR amplified from the RK2 plasmid using the primer sets given in Table 6.2. The amplified genes contained *KpnI* and *PstI* restriction sites. Restriction digestion was performed and the gene inserts were gel purified. The pUT18, pUT18c, pKT25, pKNT25 vectors (Table 6.1) used for cloning were digested using *KpnI* and *PstI* restriction sites and gel purified, as described in Chapter 2 (section 2.6). The genes were cloned in such a way that T25 and T18 fragments are present at either C- or N- terminal of the cloned gene (Figure 6.6). The cloned genes were sequenced to rule out any unintended mutation.

For the BACTH assay, two different plasmid vectors, each containing a respective gene, were sequentially transformed in BTH101 and transformants were selected using a respective antibiotic (Kanamycin and Ampicillin).

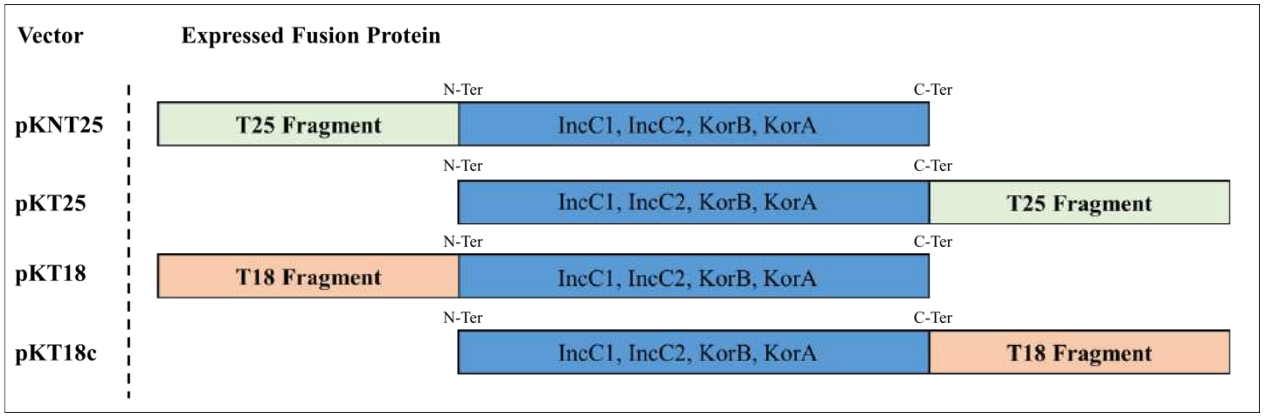


Figure 6.6. Fusion proteins expressed from the BACTH vectors. The fusion proteins expressed from pKNT25 and pKT18 vectors contained T25 and T18 fragments at their N-termini, respectively, while the fusion proteins expressed from pKT25 and pkT18c contained the T25 and T18 fragments at their C-termini. Green and orange colours represent the T25 and T18 fragments, respectively, while blue colours show the proteins (IncC1, IncC2, KorB, or KorA) fused to either fragment.

Table 6.1 Plasmids used in this chapter. Plasmid maps are given in Appendix 1.

	Plasmid	Selectable Marker	Replicon	Properties	Reference
1.	pUT18	Amp	pMB1	Derived from pUC19, encodes the T18 fragment of adenylate cyclase (CyaA)	Euromedex (Battesti et al., 2012)
2.	pUT18c	Amp	pMB1	Derived from pUC19, encodes the T18 fragment of CyaA	Euromedex
3.	pKT25	Kan	pMB1	Derived from pSU40, encodes the T25 fragment of CyaA	Euromedex
4.	pKNT25	Kan	pMB1	Derived from pSU40, encodes the T25 fragment of CyaA	Euromedex
5.	pUT18c-Zip	Amp	pMB1	pUT18c with leucine zipper inserted in frame to the T18 fragment	Euromedex
6.	pKNT25-Zip	Kan	pMB1	pKNT25 with leucine zipper inserted in frame to the T25 fragment	Euromedex
7.	Incc1_pUT18	Amp	pMB1	pUT18 with *incc1* inserted in *KpnI-pstI* restriction site.	This Study
8.	Incc1_pUT18c	Amp	pMB1	pUT18c with *incc1* inserted in *KpnI-pstI* restriction site.	This Study
9.	Incc1_pKT25	Kan	pMB1	pKT25 with *incc1* inserted in *KpnI-pstI* restriction site.	This Study
10.	Incc1_pKNT25	Kan	pMB1	pKNT25 with *incc1* inserted in *KpnI-pstI* restriction site.	This Study
11.	Incc2_pUT18	Amp	pMB1	pUT18 with *incc2* inserted in *KpnI-pstI* restriction site.	This Study
12.	Incc2_pUT18c	Amp	pMB1	pUT18c with *incc2* inserted in *KpnI-pstI* restriction site.	This Study
13.	Incc2_pKT25	Kan	pMB1	pKT25 with *incc2* inserted in *KpnI-pstI* restriction site.	This Study
14.	Incc2_pKNT25	Kan	pMB1	pKNT25 with *incc2* inserted in *KpnI-pstI* restriction site.	This Study
15.	KorA_pUT18	Amp	pMB1	pUT18 with *korA* inserted in *KpnI-pstI* restriction site.	This Study
16.	KorA_pUT18c	Amp	pMB1	pUT18c with *korA* inserted in *KpnI-pstI* restriction site.	This Study
17.	KorA_pKT25	Kan	pMB1	pKT25 with *korA* inserted in *KpnI-pstI* restriction site.	This Study
18.	KorA_pKNT25	Kan	pMB1	pKNT25 with *korA* inserted in *KpnI-pstI* restriction site.	This Study
19.	KorB_pUT18	Amp	pMB1	pUT18 with *korB* inserted in *KpnI-pstI* restriction site.	This Study
20.	KorB_pUT18c	Amp	pMB1	pUT18c with *korB* inserted in *KpnI-pstI* restriction site.	This Study
21.	KorB_pKT25	Kan	pMB1	pKT25 with *korB* inserted in *KpnI-pstI* restriction site.	This Study
22.	KorB_pKNT25	Kan	pMB1	pKNT25 with *korB* inserted in *KpnI-pstI* restriction site.	This Study

Table 6.2 PCR and sequencing primers used in this chapter. The restriction sites in cloning primers are underlined.

	Primer Name	PCR/Seq	Sequence (3`-5`)
1.	IncC2_Fwd_pUT	PCR	GCATGCCTGCAGGATGAAAACTTTGGTCACGGCC
2.	IncC2_Rev_pUT	PCR	ACTTAGGTACCCGTTGGGAAATCTCCATCTTCGTG
3.	IncC2_Fwd_pKT	PCR	GCATGCCTGCAGGGATGAAAACTTTGGTCACGGCC
4.	IncC2_Rev_pKT	PCR	ACTTAGGTACCCGTTGGGAAATCTCCATCTTCGTG
5.	IncC1_Fwd_pUT	PCR	ATGCCTGCAGGATGGGTGTTATCCATGAAGAAACG
6.	IncC1_Rev_pUT	PCR	ACTTAGGTACCCGTTGGGAAATCTCCATCTTCGTG
7.	IncC1_Fwd_pKT	PCR	ATGCCTGCAGGGATGGGTGTTATCCATGAAGAAAC
8.	IncC1_Rev_pKT	PCR	ACTTAGGTACCCGTTGGGAAATCTCCATCTTCGTG
9.	KorB_Fwd_pUT	PCR	GCATGCCTGCAGGATGACTGCGGCTCAAGCC
10.	KorB_Rev_pUT	PCR	TACTTAGGTACCCGGCCCTCGATGAGCGC
11.	KorB_Fwd_pKT	PCR	GCATGCCTGCAGGGATGACTGCGGCTCAAGCC
12.	KorB_Rev_pKT	PCR	TACTTAGGTACCCGGCCCTCGATGAGCGC
13.	KorA_Fwd_pUT	PCR	ATGCCTGCAGGATGAAGAAACGGCTTACCGAAAGC
14.	KorA_Rev_pUT	PCR	AGGTACCCGTCGTTTGGTTTCCTGTTTTTTCTTG
15.	KorA_Fwd_pKT	PCR	ATGCCTGCAGGGATGAAGAAACGGCTTACCGAAAG
16.	KorA_Rev_pKT	PCR	TAGGTACCCGTCGTTTGGTTTCCTGTTTTTTCTTG
17.	pUT18_seq	Seq	GATCCAGGCCGCCCG
18.	pKNT25_seq	Seq	GCCTGATGCGATTGCTGC
19.	pKT25_Frwd	Seq	CCAGGGTTTTCCCAGTCACG
20.	pKT25_Rev	Seq	GATTAAGTTGGGTAACGCCAGGG
21.	pUT18c_Fwd	Seq	CTGGAAACGGTGCCGG
22.	pUT18c_Rev	Seq	CAGATTGTACTGAGAGTGCACCATATTACTTAG

6.6. Bacterial Two-Hybrid Assay

Protein-protein interactions were detected by the BACTH System. Two proteins, one fused to T18 and the other fused to the T25 fragment, were co-expressed in BTH101 (a $CyaA^-$ *E. coli* strain). 25-50 ng of each vector (T18, T25) was sequentially transformed into BTH101 (100 μL) competent cells. The antibiotic-selected transformants were grown in a deep well block (96 wells block, 0.6 mL LB media each with appropriate antibiotic) at 37°C, overnight. The next day, 20 μL of each overnight culture was inoculated into 1 mL fresh LB media in another deep well block and the cells were grown, in triplicate, to mid-log phase (OD_{660} ~ 0.3) using 0.5 mM IPTG and appropriate antibiotics.

6.6.1. X-gal Plate Assay

For screening of β-galactosidase activity, one drop (~10 μL) from each culture was plated on LB agar plates containing 5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside (40 μg/ml, X-gal) and isopropyl-β-D-thiogalactopyranoside (0.5 mM, IPTG). Blue colonies on LB agar/X-Gal plates show protein-protein interaction within the colonies, while white colonies show that co-expressed proteins do not interact with each other. The pKT25-zip and pUT18-zip vectors were used as positive controls. The zip vectors contained Leucine zipper domains that interact strongly with each other. Maltose-MacConkey agar plates (supplemented with 0.5 mM IPTG, 1% Maltose) were also used to screen for β-galactosidase activity. Red colonies show protein-protein interactions, whereas white ones do not.

6.6.2. β-galactosidase (β-gal) assay

A modified Miller assay (β-galactosidase assay) (Miller, 1972) was used for semi-quantitative estimation of protein-protein interactions. The Miller assay is based on quantitation of β-galactosidase activity, that catalyses the hydrolysis of β-galactosides into galactose. In the assay, O-nitrophenol-β-galactoside (ONPG) is used as a substrate (colourless) for β-galactosidase. The ONPG hydrolysis results in O-nitrophenol (Yellow) and β-galactose. The O-nitrophenol production is monitored by absorbance at OD_{420} (Figure 6.7). Absorbance values are converted in Miller units by using the Miller equation.

The Miller assay and the Miller equation is often used to measure the β-galactosidase activity (Griffith and Wolf, 2002). The Miller equation uses OD_{420} (for ONPG hydrolysis),

OD_{600} (for bacterial cell density) and provides a measure of enzyme activity in Miller units. The Miller equation is as follows

$$\text{Miller Units} = \frac{1000 \times OD_{420}/min}{v \times OD_{600}}$$

Where 'v' is the volume of bacterial cells/cultures in millilitreIn this chapter, the assay was performed using 96 wells plates. Cells were lysed at mid-log by adding 20 µL of cells to 80 µL permeabilization solution (100 mM Na_2HPO_4, pH 7.5, 20 mM KCl, 2 mM $MgSO_4$, 0.8 mg/mL CTAB (hexadecyl trimethyl ammonium bromide), 0.4 mg/mL sodium deoxycholate and 5.4 µL/mL beta-mercaptoethanol). Cells were lysed for 10 minutes at room temperature and 150 µL of substrate solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/mL O-nitrophenyl-β-D-galactoside (ONPG) and 2.7 µL/mL β-mercaptoethanol) was added to each well. The OD_{660} was measured, and then the change in OD_{420} was monitored using a BioTek Micro-well plate reader, at 30°C, at 1 min intervals, over 60 minutes. Data was exported to the MS Excel spreadsheets for analysis, as graphs of OD_{420} *vs* time (Figure 6.7). The rate of change in absorbance with time was calculated from the linear part of each graph. Some graphs are shown in Appendix 7. The ZipT18 and ZipT25 vectors, transformed in BTH101, were used as positive controls, while BTH101cells, transformed with either single vector, were considered to be negative controls.

6.7. Chemical crosslinking

The BS3 and DSS cross-linkers were used to monitor homo- and hetero- protein interactions. For IncC1 and IncC2 crosslinking, the BS3 concentration was optimized to generate accurate crosslinking data. The protein concentration was checked by the Bradford assay before crosslinking, as described in Chapter 2 (Section 2.23.2). The IncC proteins (0.5-10 µM) were mixed with varying concentrations of BS3 (1:30 to 1:62000 ratio of protein to crosslinker) to optimize the crosslinking reaction (crosslinking buffer includes 20 mM Potassium phosphate, 150 mM NaCl, 10 % glycerol and 1 mM DTT). The BS3 crosslinking was also optimized for crosslinking reaction time (5-30 min at room temperature). The optimal protein:BS3 ratios (1:1600) and reaction time (10 min) was used for further experiments. KorB and KorA proteins are also analysed by BS3 chemical crosslinking for homo- and hetero- protein complexes.

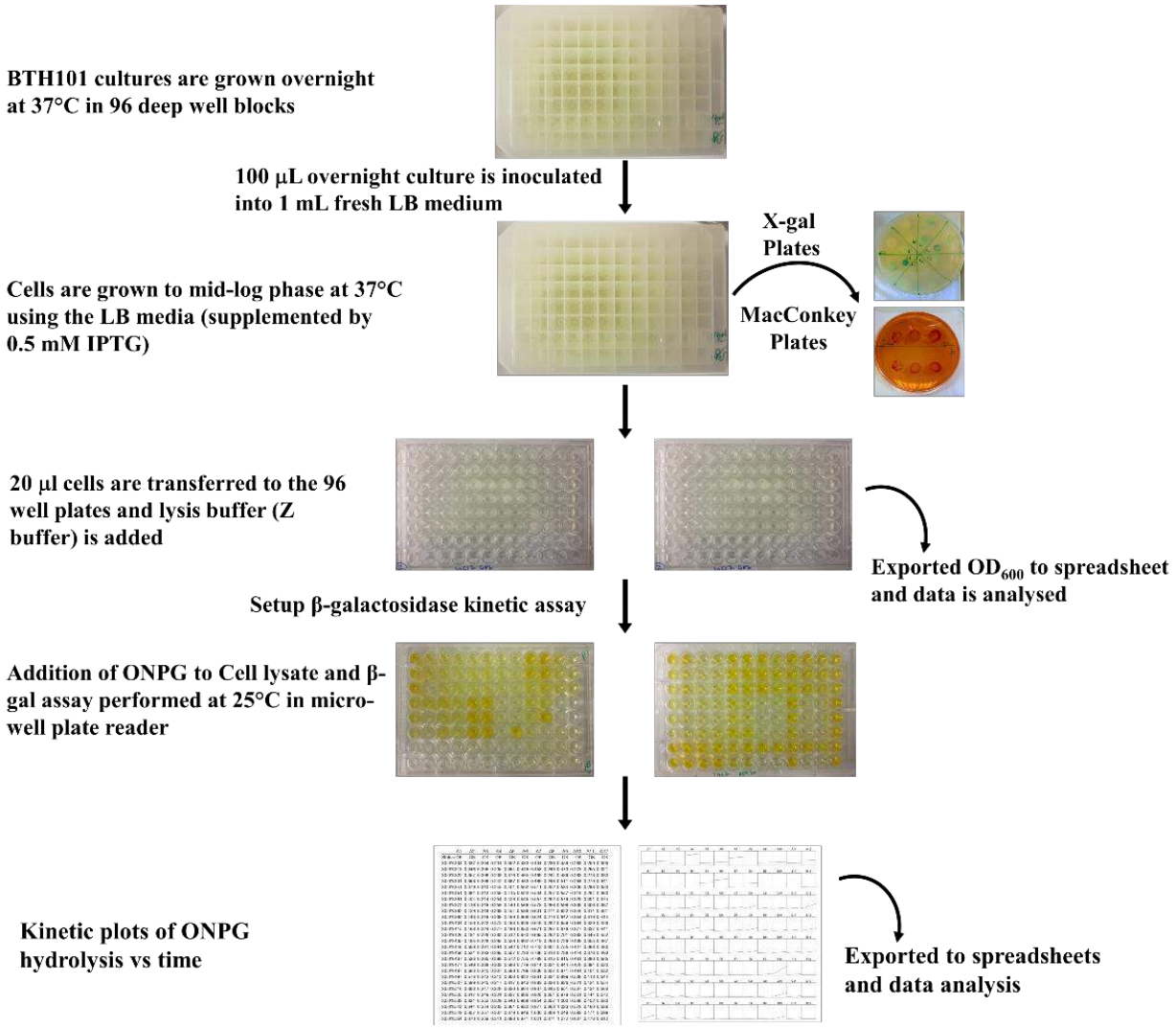


Figure 6.7 A schematic of BACTH assay. BTH101 transformants were grown at 37°C, overnight (in LB medium, supplemented with 1% glucose) in duplicate deep well blocks. From overnight cultures, 100 µL inoculum used to grow cells to mid-log phase at 37°C, in the presence of 0.5 mM IPTG. Each overnight culture was grown in triplicates. At OD_{660} ~0.32, 2 µL drops, from each culture, were spotted on the X-gal and the MacConkey agar plates. Same cultures were used for the β-galactosidase assay. 96-well plates were used for the assay. Cells were lysed using lysis buffer mentioned in methods and substrate (ONPG) was added to monitor the activity of β-galactosidase. BioTek micro-well plate reader was used to measure the OD_{600} and OD_{420}. Data were exported and tabulated for analysis in the Microsoft Excel.

The crosslinking reaction products were analysed using 4-16% gradient SDS PAGE (BioRad). SDS PAGE was performed as described in Chapter 2 (Section 2.13). For IncC1 and IncC2 crosslinking analysis, Western blots using an anti-His_6 HRP-conjugated antibody (a primary IgG HRP-conjugated, anti-His_6 mouse antibody, R & D Systems) was performed after SDS PAGE, as described by (Lal et al., 2005).

The protein bands were blotted onto PVDF membrane (Trans-Blot Turbo Mini PVDF Transfer, BioRad) using a Trans-Blot® Turbo™ Transfer System (BioRad). The PVDF membrane was incubated with skimmed (1%) milk (in TBST buffer, 20 mM Tris, 150 m NaCl, 0.05 %Tween 20) for 30 min, followed by incubation for 18 hr with an HRP -conjugated His_6 antibody. The PVDF membranes were then washed twice with Tris-borate buffer and the blotted proteins were visualized using the Amersham ECL Western blotting detection reagent in accordance with the manufacturer's instructions (GE Healthcare). The blots are exposed to the X-ray films (Kodak) overnight. The X-ray films were developed and the images were scanned for analysis.

6.7.1 Hetero-protein chemical crosslinking in the presence of DNA and nucleotides

The IncC1 and IncC2 proteins were cross-linked in the presence of each other and in the presence of KorB, KorB mutants, nucleotides and DNA. Equal amounts of proteins (IncC1, IncC2 and KorB/KorB mutants) (5 μM) were used in the hetero-protein crosslinking. To study the effect of nucleotides on protein chemical crosslinking 5 mM ATP, ADP and ATPγS were added in individual reactions. 2 μM DNA was also added in some chemical crosslinking reaction.

6.8. Crosslinking combined with Mass spectrometry (CL-MS)

Cross-linked protein bands from SDS PAGE were detected by colloidal coomassie staining, excised from the gel and sent to the Proteomic Services (Advanced Mass Spectrometry Facility, University of Birmingham) for Mass spectrometric analysis. Sample preparation in the Proteomics Services involved *in-gel* trypsin digestion. Different methods including Collision-induced dissociation (CID), Higher-energy collisional dissociation (HCD), CID coupled with HCD were employed for Mass spectrometry, using a Thermo Scientific Orbitrap Elite and Thermo Scientific Q-Exactive HF Mass spectrometer. Sample injection time was 50 ns and

resolution for MS1 was fixed from 80,000 to 120, 000 while for MS2 resolution was 7,500 to 15, 000.

6.9. CL-MS data analysis

The Mass spectrometry data was obtained as raw Thermo file format and converted to the *.mgf and *.mxXML file formats. Crosslink identification was performed using two different software programmes, *i.e.* KOJAK ver. 1.61 and Spectrum Identification Machine for Cross-linked Peptides (SIM-XL) ver. 1.4.1.5.

KOJAK is a C++ based software programme that uses Cross-Correlation and Comet Algorithms for MS/MS spectrum processing (Hoopmann et al., 2015). A list of target peptide sequences is predicted and scored using the Comet Algorithm. Cross-linked peptides are further identified using the Two-Pass Algorithm and results are exported in the position specific matrices (PSM) output files that are further analysed by Percolator. Decoy PSMs (matched to random amino acid sequences) are ignored as incorrect. Cross-linked peptides are filtered (using Percolator software) for validation and exported as separate files for inter-link, intra-link, loop-link and single cross-linked peptides. KOJAK takes the *.mzXML Mass spectrometry files and FASTA database files as input files. Database files were prepared using the IncC1, IncC2, KorA, KorB protein sequences as well as using decoy sequences (random amino acid sequence). The configuration file for KOJAK contains different parameters that are required to be modified (file format is given in Appendix 8). Spectra related to crosslinks were manually analysed using KojakSpectrumViewer, while Percolator ver 2.08 was used for PSM analysis.

SIM-XL (Lima et al., 2015) is a very sensitive crosslink identification software package that takes Thermo *.raw, *.mgf and *.mzML files as input, while results are exported as in *.simxlr file format. SIM-XL has many options to select specific crosslinks, digestive enzymes and fixed modifications in GUI of software. SIM-XL uses a Comet Algorithm to identify cross-linked peptides. Results can be analysed as 2D maps, heats maps and as peptide annotation. Cross-linked peptide spectrum can be manually analysed using a spectrum viewer. Crosslinks identified from the mass spectrum data can be mapped onto protein structure and exported as a *.pml script. The *.pml script allows crosslink visualization using PyMOL where distances between the cross-linked peptides can be mapped. Crosslinked peptides with a distance <34 Å are considered as trusted crosslinks.

6.10. Homology modelling

IncC2 homology modelling was performed using Modeller ver. 9.19. A command-based version, as well as a Chimera-based GUI of the Modeller software, was used for homology modelling. The scripts used in homology modelling are given in Appendix 9. Initially, the IncC2 sequence was submitted to EMBL BLASTP ver.2.2.29+ to search for similar structures in the Protein Data Bank (PDB). Homology modelling was performed in a sequential way as illustrated in Figure 6.8. Three structure templates, PDB ID 2BEJ, 1WCV and 5IHP, with lowest E-values and greatest sequence coverage, were selected for homology modelling (Figure 6.19). A sequence alignment was performed against the template structures and the alignments were manually checked for gaps or any realignment. A total of 50 model structures were generated against each template. Each model structure was evaluated using the DOPE, zDOPE, GA341 and RMSD scores. The best model was selected and loop modelling/structure refinement was performed using an internal algorithm in Modeller. A total of 20 structures was obtained after loop refinement and evaluated using DOPE, zDOPE, GA341 and RMSD scores.

The best model was selected based on DEPE, zDOPE and RMSD, and evaluated using ProSA (Wiederstein and Sippl, 2007), SaliLab Model Evaluation Server (Shen and Sali, 2006) and Rampage Server for Ramachandran Plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). The model was further refined using the PREFMD (Protein structure REFinement via Molecular Dynamics) (Heo and Feig, 2017) and ReFOLD web Servers (Shuid et al., 2017). Refined output models were again evaluated for zDOPE (Shen and Sali, 2006), GA341 (Melo et al., 2002) and RMSD scores using the ModEval Model Evaluation Server (https://modbase.compbio.ucsf.edu/evaluation/). The refined structures were also evaluated using ProSA, Verify 3D, and Ramachandran Plot.

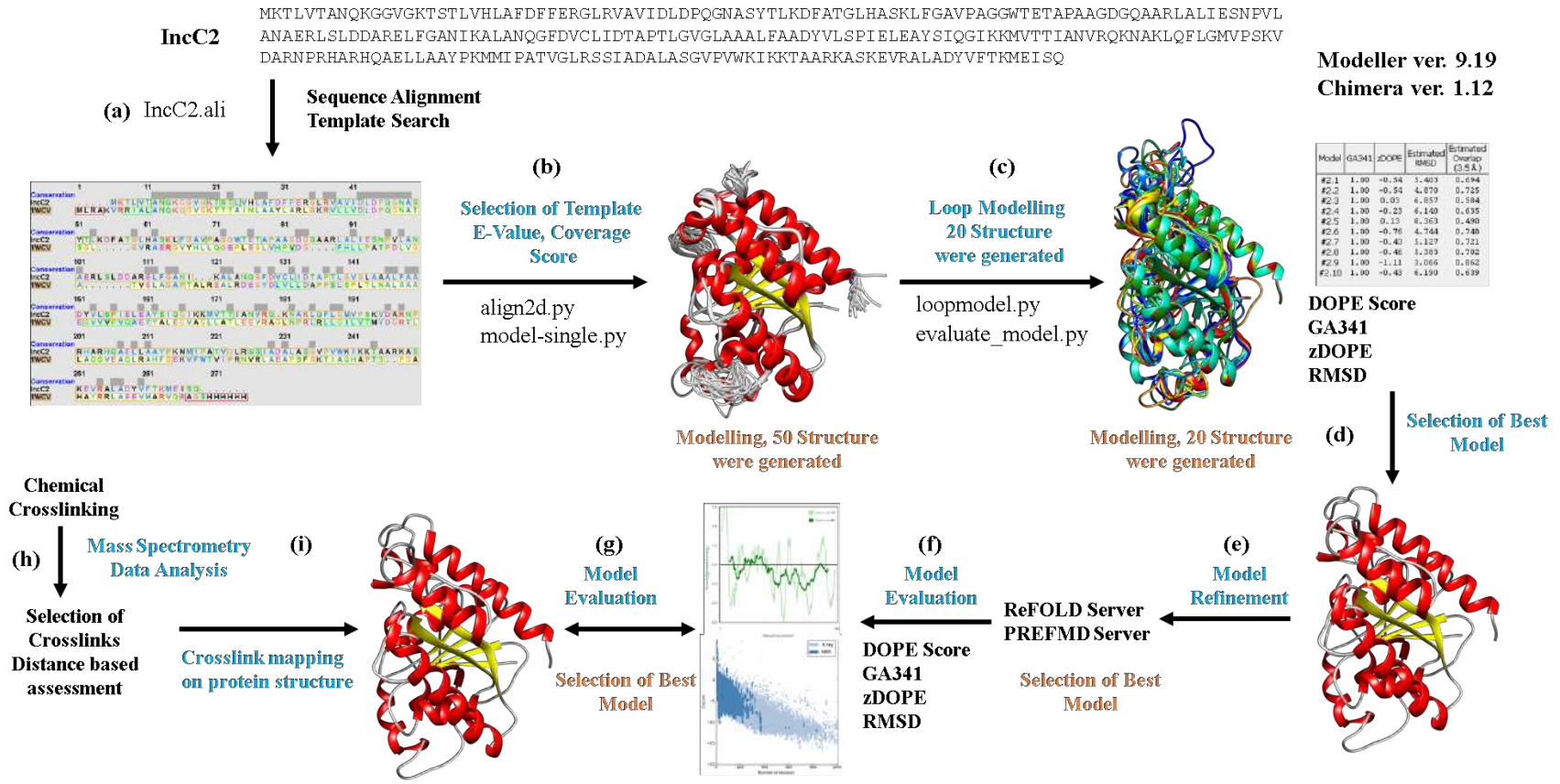


Figure 6.8 A schematic of homology modelling of IncC2 using Modeller ver 9.19. (a) The IncC2 protein sequence was aligned against selected templates from a BLAST search, using the python script align2d.py. The Alignment was manually checked for gaps and for any realignment. **(b)** Then the model-single.py script was used to build 50 different models that were further evaluated by running evaluate_model.py script. **(c)** One model out of 50 was selected on the basis of zDOPE, GA341 and RMSD scores. The selected model was further subjected to loop modelling and refinement using the loop model.py script. Twenty loop models were obtained and ranked against zDOPE, GA341 and RMSD scores **(d, e)** The best loop model was subjected to structure refinement using ReFOLD and PREFMD servers **(f, g)** The refined output models were again evaluated for zDOPE, GA341 and RMSD scores using ModEval Model Evaluation Server, ProSA, and Ramachandran Plot. The IncC2 Model with the best scores was selected **(h, i)** Peptide crosslinks, obtained from the Mass spectrometry data, were mapped onto the final structure and Cα-Cα distances for Lys-Lys residues were analysed. All scripts are given in Appendix 9.

6.10.1. IncC2 homodimer model

A model of the IncC2 homodimer was also obtained using the Swiss-Model server (Biasini et al., 2014). In the Swiss-Model server, automated homodimer modelling was performed selecting the PDB structure ID 1WCV (Leonard et al., 2005) as a modelling template. The homodimer model was evaluated using ModEval, ProSA, Verify 3D, and Ramachandran Plot. The modelled homodimer interface was verified by mapping the identified crosslinks.

6.11 Homology model mapped by crosslinking data

Crosslinks obtained from the Mass spectrometry data analysis were mapped on homology models and distance restraints (Cα-Cα) were measured. The crystal structures of two KorB domains, the C-terminal domain (PDB ID 1IGU, 1IGQ) (Delbruc et al., 2002) and the DNA binding domain (PDB ID 1R71) (Khare et al., 2004), were obtained from PDB. Crosslinks were mapped on the DNA binding domain (dimer) of KorB to evaluate the quality of crosslinking data. Crosslinks were also mapped onto the IncC1, IncC2, KorA, KorB protein sequences to get possible protein-protein interaction sites.

Results

6.12. BACTH and β-galactosidase Assay

incC1, incC2, korB, and korA were cloned to give fusion proteins with the two halves of adenylate cyclase, and protein-protein interaction were measured *in vivo* by the resulting β-galactosidase activity. The results from the β-gal Assays in lysed cells were tabulated and analysed quantitatively using MS Excel. Results were compared to the colours of the corresponding colonies on X-gal/LB and MacConkey plates. The IncC1 and IncC2 protein homodimerization and their interactions with the KorB are observed.

6.12.1. IncC1 and IncC2 protein interactions

Figure 6.9a shows the results of the β-galactosidase assays in lysed cells for IncC1 fused to T25 expressed from the low copy number vector, with the partner protein fused to the T18 fragment, expressed from the higher copy number plasmid. The pZIP controls, containing Leucine zipper fragments cloned into the vectors, show very high activity, >8000 Miller units, whereas the negative controls, containing only one of these vectors or no vector show < 100

units of activity. Most of the samples assayed show low activity, ~ 600 Miller units, apart from the pairs C1T25-C1T18, C1NT25-C1T18c, C1T25-C2T18, and C1NT25-KorBT18c with ~ 2000 units activity. These BACTH assays for IncC1 show that the protein forms homo- dimers only when one protein has a free N-terminus and the other has a free C-terminus, suggesting that it forms head to tail dimers. Similarly, with IncC2, activity is found with the pair C1T25-C2T18, (Figure 6.10a). IncC1 is also found to interact more with KorB when the C-terminus of IncC and the N-terminus of KorB is free (C1NT25-KorBT18c), but the protein does not show any significant interaction with KorA (Figure 6.9a). The X-gal and MacConkey agar plates show similar results to the β-galactosidase assays in lysed cells (Figure 6.9b), with blue colonies on X-gal plates and red colonies on MacConkey maltose plates, indicating protein-protein interactions *in vivo*.

The BACTH assays for IncC2 fused to T25 in the lower copy number plasmids (Figure 6.10a), and the partner protein fused to T18, in general, show higher activities with C2T25 than with C2NT25, suggesting that the fusion at the N-terminal may reduce activity. IncC2, like IncC1, forms homodimers however the highest activities are with both partners in the same orientation (C2T25-C2T18c, C2NT25-C2T18). In contrast, when IncC2 interacts with IncC1, strong interactions are observed when the N-terminus of IncC2 is free and the C-terminus of C1 is free (C2T25-C1T18) (Figure 6.10a). IncC2 protein also interacts with KorB, with some activity in all combinations, but the highest with C2T25-KT18c, *i.e.* when the N-termini of both proteins is free (Figure 6.10a). However, IncC2, like IncC1, does not show any significant interaction with KorA. (Figure 6.10a). The X-gal and MacConkey agar plates, in general, show similar results to the β-gal assays (Figure 6.10b). Table 6.5 summarises the IncC protein interactions with their partners.

6.12.2. KorB and KorA protein interactions

The BACTH assays for the KorB protein show that it forms homodimers when the fusion from the high copy number has a free C-terminus (KBT25-KorBT18, KBNT25-KorBT18). KorB interacts with both IncC1 and IncC2 when the N-terminus of KorB is free. For IncC1 this is with C1T18 *i.e.* the C-terminus of the IncC1 protein is free (KBT25-C1T18), but for IncC2 it is with C2T18c) *i.e.* with the N-terminus of IncC2 free (Figure 6.11a). Spots on the X-gal and

MacConkey agar plates show similar results for KorB interactions with other proteins (Figure 6.11b).

The KorA protein forms homodimers; the dimerisation domain being at the C-terminal domain (Bingle et al., 2008). The KorA-KorA interactions are strongest when the N-termini of both partners are free (KAT25-KorAT18c), but interactions are still observed when N-terminus of at least one KorA partner protein is free (Figure 6.12a).

KorA and KorB proteins are expected to interact with each other (Bingle et al., 2008), but no significant interactions are observed by BACTH. This might be due to the presence of a fusion fragment that covers the protein-protein interaction site (Figure 6.11a, 6.12a). Table 6.5 summarises the KorA and KorB protein interactions with their partners.

The spots on the X-gal and MacConkey agar plates show similar results to the BACTH assays in liquid culture.

Table 6.3 Summary of protein-protein interactions measured by BACTH Assays

		IncC1		**IncC2**		**KorB**		**KorA**	
		T18	**T18c**	**T18**	**T18c**	**T18**	**T18c**	**T18**	**T18c**
IncC1	**T25**	+++	-	+++	+	+	+	-	-
	NT25	+	++	++	++	+	+++	-	+
IncC2	**T25**	+++	++	+	+++	++	+++	-	-
	NT25	-	-	+++	++	++	++	-	-
KorB	**T25**	+++	+	+	+++	+++	+++	+	-
	NT25	-	+	+	-	-	+	-	-
KorA	**T25**	-	+	-	-	-	-	++	+++
	NT25	-	-	-	-	-	-	++	+

+++ Relatively strong Interactions (>2000 Miller Units)
++ Weak Interactions (1500-2000 Miller Units)
+ Very Weak Interactions (600-1500 Miller Units)
- Negligible Interactions (<600 Miller Units)

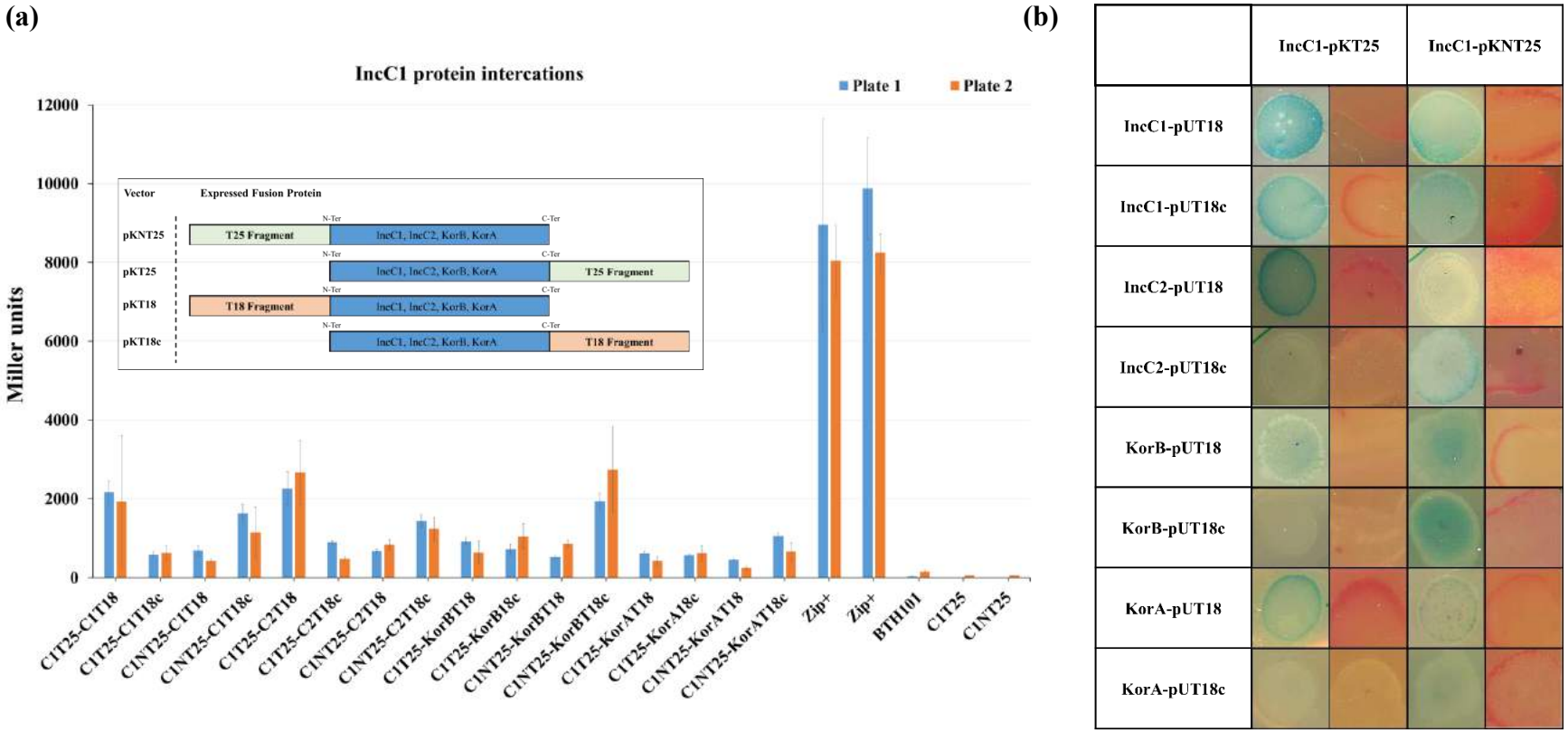


Figure 6.9 BACTH analysis for the IncC1 protein. (a) The β-galactosidase assay was used for indirect measurement of the protein-protein interaction in a semi-quantitative way. Two experiment plates were used as biological repeats (orange and blue bars) and both plates contained each sample in triplicate (error bars show the standard deviation within the triplicates on each plate). The Miller units were calculated using the Miller equation; Miller Units = (1000 × OD_{420}/min)/(volume×OD_{600}). Where 'v' is the sample volume. The higher the Miller units, the stronger the protein-protein interactions. Negative and positive controls were also analysed. The negative controls contained at least one empty vector (BTH101, C1T25 and C1NT25) and the positive controls contained T18 and T25 fragments attached to Leucine Zipper Domains (Zip+ Vectors). The Zipper domains show strong interactions. The inset figure shows the position of the fused T18 and T25 fragments in each protein. C1T25-C1T18 and C1NT25-C1T18c show IncC1-IncC1 interactions are more promising when C-terminus of one partner and N-terminus of other is free. C1T25-C2T18 shows IncC1-IncC2 interactions when N-terminus of IncC1 and C-terminus of IncC2 is free. IncC1 also interacts with KorB when its C-terminus is free (C1NT25-KorBT18c). IncC1 does not show any significant interaction with KorA. **(b)** The Co-transformed BTH101 cells were spotted on the X-Gal and MacConkey agar plates with the appropriate antibiotics (0.5 mM IPTG). In the case of MacConkey agar plates, 1 % Maltose is also added. The plates were monitored after 48 h incubation at 30°C. The greenish colour colonies in X-gal plates and red colonies in the MacConkey agar plates indicate protein-protein interactions. The results from agar plates are in agreement with the β-galactosidase assay results from the lysed cells.

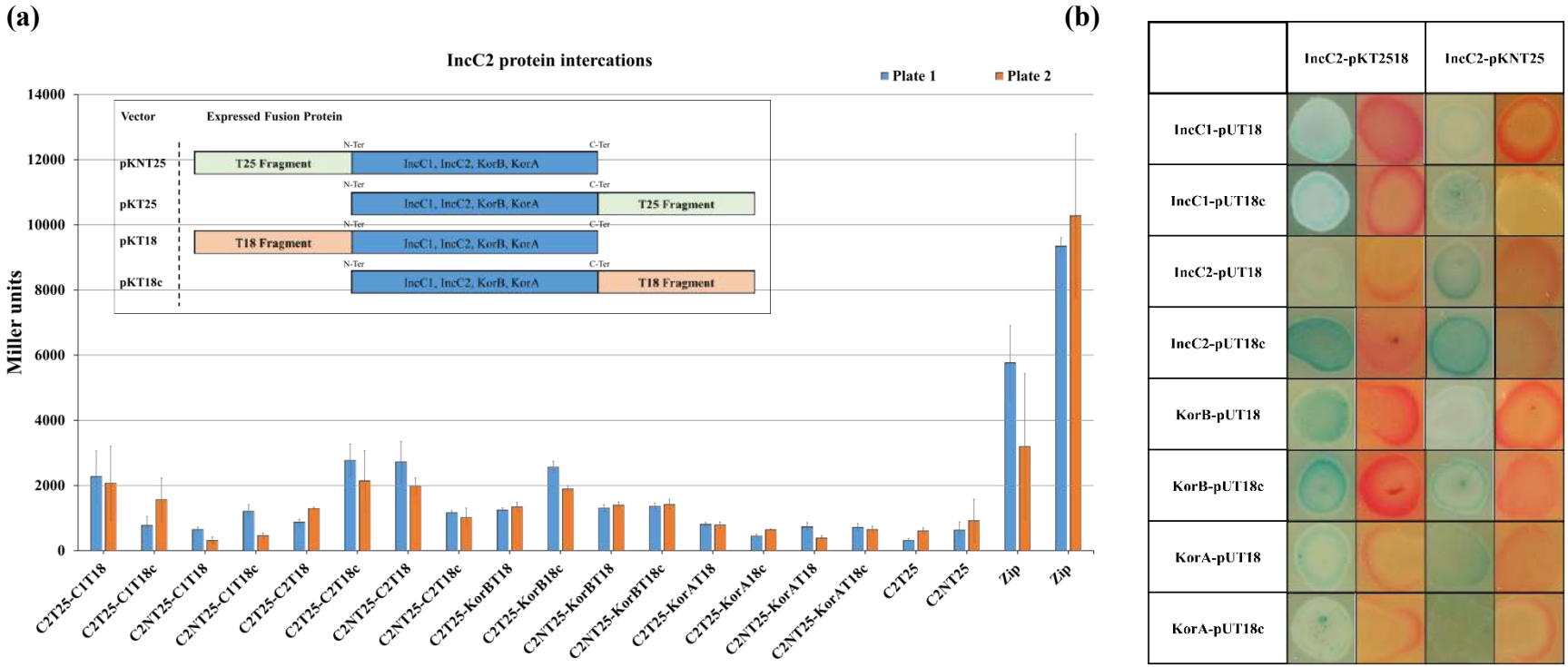
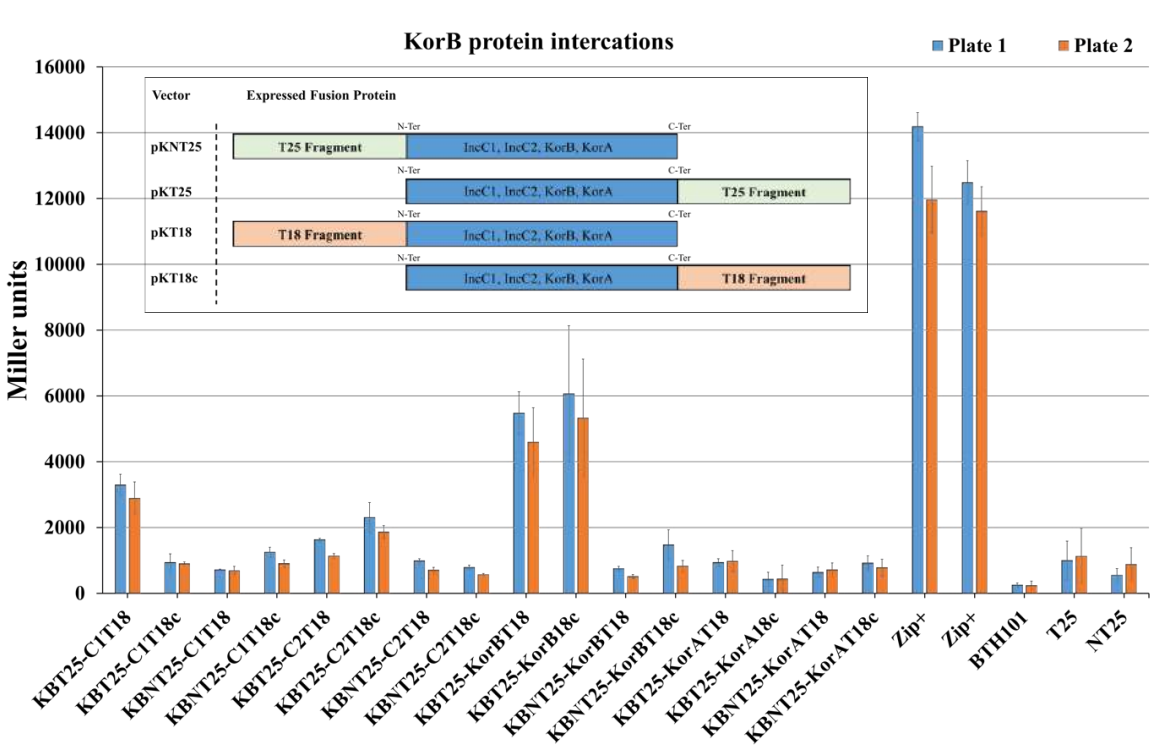


Figure 6.10 IncC2 protein interactions. (a) The β-galactosidase assay was used for indirect measurement of the protein-protein interaction in a semi-quantitative way. Two experiment plates were used as biological repeats (orange and blue bars) and both plates contained each sample in triplicate (error bars show the standard deviation within the triplicates on each plate). The Miller units were calculated using the Miller equation; Miller Units = (1000 × OD_{420}/min)/(volume×OD_{600}). Where 'v' is the sample volume. The higher the Miller units, the stronger the protein-protein interactions. Negative and positive controls were also analysed. The negative controls contained at least one empty vector (BTH101, C2T25 and C2NT25) and the positive controls contained T18 and T25 fragments attached to Leucine Zipper Domains (Zip+ Vectors). The Zipper domains show strong interactions. The inset figure shows the position of the fused T18 and T25 fragments in each protein. The IncC2 protein forms homo-dimers as shown by C2T25-C2T18c and C2NT25-C2T18. Interactions are stronger when the C-terminus of the one partner protein and N-terminus of the other partner monomer is free. C2T25-C1T18 shows IncC2 interacts with IncC1 strongly when N-terminus of IncC2 and C-terminus of IncC1 is free. IncC2 also shows stronger interaction with KorB when the N-terminus of KorB is free (C1NT25-KorBT18c). IncC2-KorA interactions are not observed. The inset figure shows the position of fused T18 and T25 fragments in proteins **(b)** The Co-transformed BTH101 cells were spotted on the X-Gal and MacConkey agar plates with the appropriate antibiotics (0.5 mM IPTG). In the case of MacConkey agar plates, 1 % Maltose is also added. The plates were monitored after 48 h incubation at 30°C. The greenish colour colonies in X-gal plates and red colonies in the MacConkey agar plates indicate protein-protein interactions. The results from agar plates are in agreement with the β-galactosidase assay results from the lysed cells.



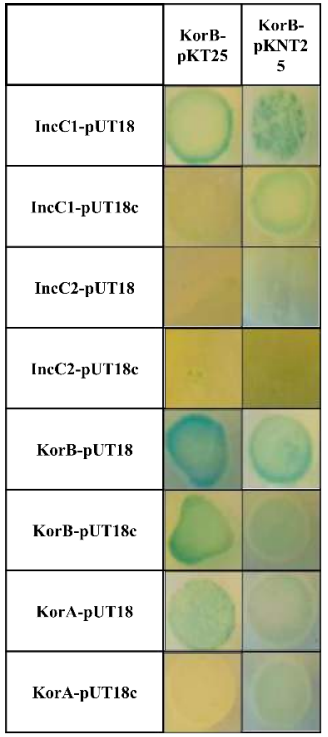
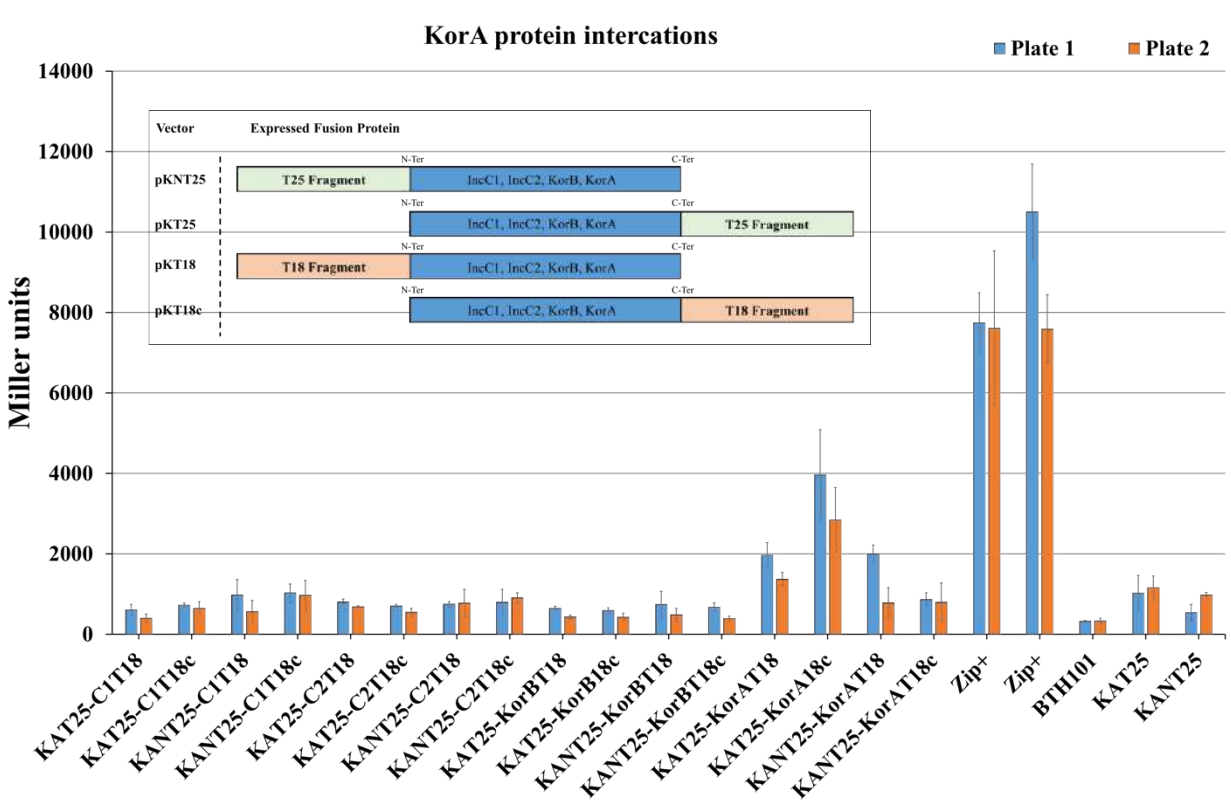


Figure 6.11 KorB protein interactions. (a) The β-galactosidase assay was used for indirect measurement of the protein-protein interaction in a semi-quantitative way. Two experiment plates were used as biological repeats (orange and blue bars) and both plates contained each sample in triplicate (error bars show the standard deviation within the triplicates on each plate). The Miller units were calculated using the Miller equation; Miller Units = (1000 × OD_{420}/min)/(volume×OD_{600}). Where 'v' is the sample volume. The higher the Miller units, the stronger the protein-protein interactions. Negative and positive controls were also analysed. The negative controls contained at least one empty vector (BTH101, T25 and NT25) and the positive controls contained T18 and T25 fragments attached to Leucine Zipper Domains (Zip+ Vectors). The Zipper domains show strong interactions. The inset figure shows the position of the fused T18 and T25 fragments in each protein. KB2T25-KorBT18 and KBT25-KorBT18c show KorB-KorB interactions. The interactions are stronger when C-terminus of one KorB monomer and N-terminus of other partner monomer is free (KBT25-KorBT18, KBT25-KorBT18c). KorB also shows stronger interactions with IncC1 when C-terminus of IncC is free (KBT25-C1T18). KorB interacts with IncC2 when the N-terminus of IncC2 is free. Strong interactions for KorB-KorA are expected but no any significant interactions are observed. Inset figure shows the position of fused T18 and T25 fragments. **(b)** The Co-transformed BTH101 cells were spotted on the X-Gal and MacConkey agar plates with the appropriate antibiotics (0.5 mM IPTG). In the case of MacConkey agar plates, 1 % Maltose is also added. The plates were monitored after 48 h incubation at 30°C. The greenish colour colonies in X-gal plates and red colonies in the MacConkey agar plates indicate protein-protein interactions. The results from agar plates are in agreement with the β-galactosidase assay results from the lysed cells.



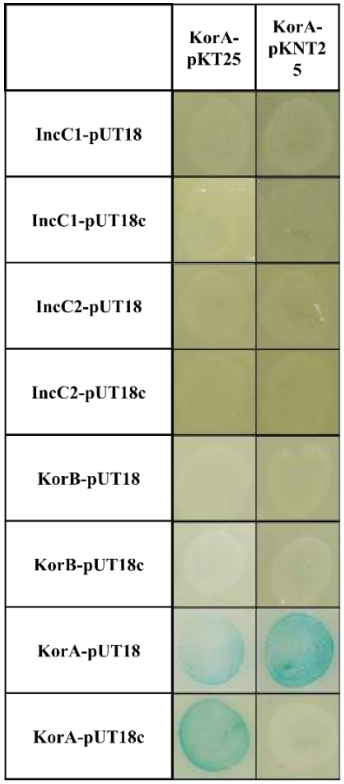


Figure 6.12 KorA protein interactions. (a) The β-galactosidase assay was used for indirect measurement of the protein-protein interaction in a semi-quantitative way. Two experiment plates were used as biological repeats (orange and blue bars) and both plates contained each sample in triplicate (error bars show the standard deviation within the triplicates on each plate). The Miller units were calculated using the Miller equation; Miller Units = (1000 × OD_{420}/min)/(volume×OD_{600}). Where 'v' is the sample volume. The higher the Miller units, the stronger the protein-protein interactions. Negative and positive controls were also analysed. The negative controls contained at least one empty vector (BTH101, KAT25 and KANT25) and the positive controls contained T18 and T25 fragments attached to Leucine Zipper Domains (Zip+ Vectors). The Zipper domains show strong interactions. The inset figure shows the position of the fused T18 and T25 fragments in each protein. KAT25-KorAT18c shows KorA forms homo-dimers when N-terminus in both monomeric partners is free. Though strong KorA-KorA interactions are also observed in KAT25-KorAT18 and KANT25-KorAT18, KorA does not show any significant interaction with IncC1, IncC2 and KorB. **(b)** The Co-transformed BTH101 cells were spotted on the X-Gal and MacConkey agar plates with the appropriate antibiotics (0.5 mM IPTG). In the case of MacConkey agar plates, 1 % Maltose is also added. The plates were monitored after 48 h incubation at 30°C. The greenish colour colonies in X-gal plates and red colonies in the MacConkey agar plates indicate protein-protein interactions. The results from agar plates are in agreement with the β-galactosidase assay results from the lysed cells.

6.13. Protein chemical crosslinking

The BS3 and DSS cross-linkers were used for the protein crosslinking. IncC1, IncC2, KorB, and KorA were cross-linked to give homo- and hetero- protein complexes.

6.13.1. IncC1 and IncC2 chemical crosslinking

Initially, the BS3 concentration was optimized for the IncC1 and IncC2 protein crosslinking. Proteins (5-10 μM) were mixed with different ratios of BS3 and the reaction mixtures were incubated at room temperature for 10 minutes and analysed by SDS PAGE (Figure 6.13a) The gel of the IncC2 crosslinking experiment shows that it forms homodimers. A cross-linked, monomeric IncC2 band was also observed. The band shows faster mobility on SDS PAGE than non-cross-linked protein monomer (Figure 6.13a). This might show that there is some conformational change in the IncC2 structure after crosslinking, or that the protein becomes less flexible/dynamic. Higher molecular weight IncC2 bands, above the dimeric band, were also observed on the SDS PAGE with molecular weights approximately equal to IncC2 trimers and tetramers. Very high molecular weight species were also observed and may have arisen due to protein over-crosslinking (Figure 6.13a). The ratio of IncC2 to BS3 of 1:1600, was selected for crosslinking. The crosslinking time was optimized using the IncC2 to BS3 ratio of 1:1600. The crosslinking mixtures were incubated at room temperature for 10-90 min and reactions were quenched using 1 M Tris. In 10-15 min, sufficient protein crosslinking was obtained (Figure 6.13b).

The IncC1 protein crosslinking was optimized in a similar way as for IncC2 and the IncC1 to BS3 ratio, 1:1600, was selected for the subsequent crosslinking experiments (Figure 6.14a). In contrast to IncC2, cross-linked IncC1 does not show a distinct cross-linked monomeric band on SDS PAGE, but protein bands with molecular weights approximate equal to IncC1 homodimer and trimer species are observed (Figure 6.14a). The crosslinking time optimization was also performed for IncC1, as with IncC2 and, again sufficient crosslinking was achieved, after a 10 min reaction (Figure 6.14b). The two proteins, IncC1 and IncC2, were each also cross-linked using different concentrations of another chemical cross-linker, DSS. The DSS crosslinking pattern for the IncC1 and IncC2 proteins is similar to that caused by BS3. Cross-linked IncC2 forms a distinct cross-linked monomer, dimers, trimers and tetramers, while IncC1 forms dimers, trimers and higher molecular weight species (Figure 6.15a and 6.15b).

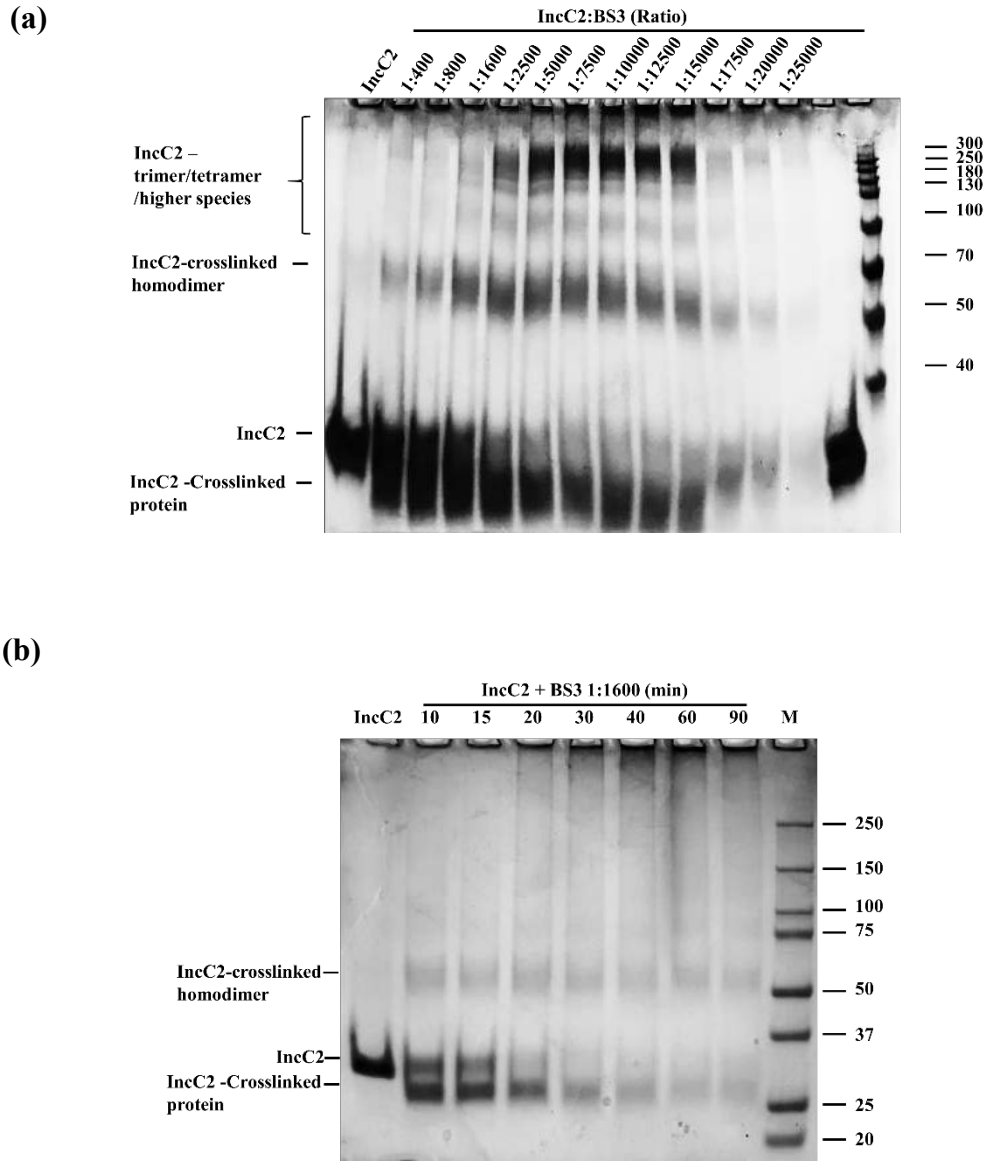


Figure 6.13 Analysis of IncC2 chemical crosslinking using SDS PAGE. The BS3 cross-linker was used for IncC2 chemical crosslinking. **(a) BS3 crosslinking optimization.** Different protein: BS3 ratios were used to optimize protein crosslinking. The IncC2: BS3 ratio, 1:1600, was found to be sufficient for IncC2 crosslinking **(b) Crosslinking reaction time optimization.** The IncC2-BS3 reaction mixtures were incubated for different times (10-90 min) and the reaction was quenched using 1 M Tris. Crosslinking products were analysed using SDS PAGE (4-20 % gradient gel). A 10 min IncC2 protein crosslinking reaction was found to be sufficient to get cross-linked products.

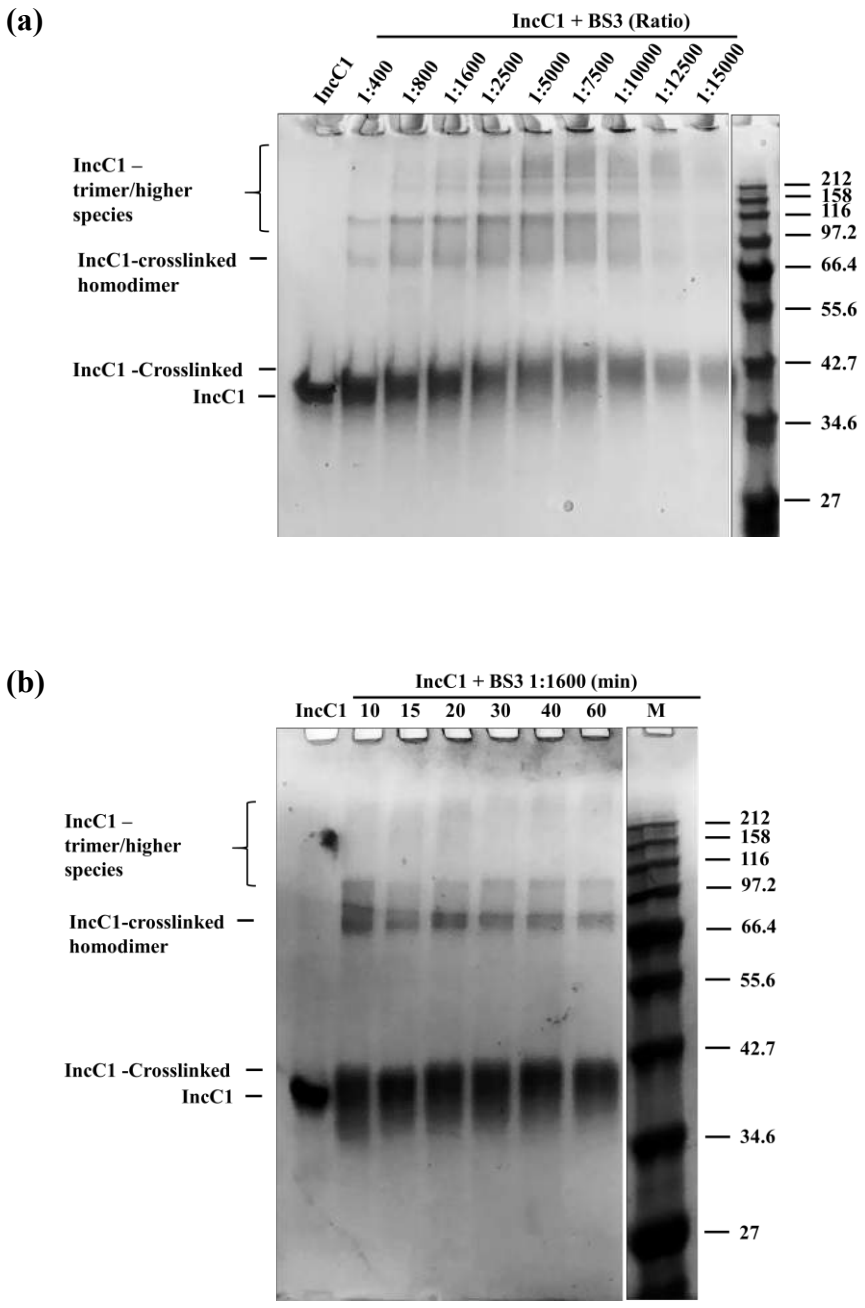


Figure 6.14 Analysis of IncC1 crosslinking using SDS PAGE. The BS3 cross-linker was used for IncC1 chemical crosslinking. **(a) BS3 crosslinking optimization.** Different protein: BS3 ratios were used to optimize protein crosslinking. The IncC1: BS3 ratio, 1:1600, is sufficient for IncC1 crosslinking **(b) Crosslinking reaction time optimization.** The IncC1-BS3 reaction mixtures were incubated for different times (10-60 min) and the reaction was quenched using 1 M Tris. Crosslinking products were analysed using SDS PAGE (4-20 % gradient gel). A 10 min IncC1 crosslinking reaction is sufficient to get cross-linked products.

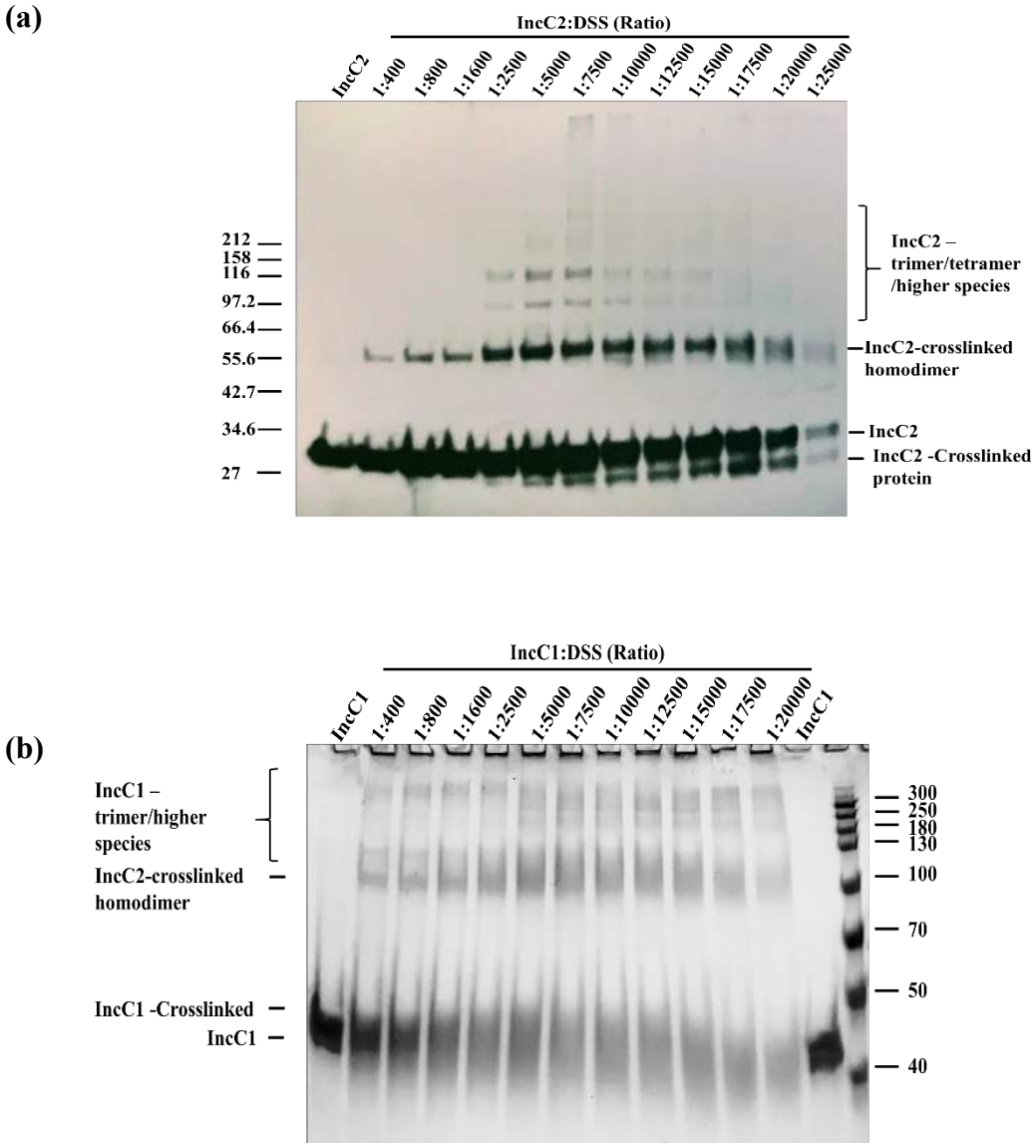


Figure 6.15 IncC1 and IncC2 chemical crosslinking using DSS. The IncC1 and IncC2 proteins were cross-linked with DSS in a similar way to that with BS3. Crosslinking reactions were performed at room temperature for 10 min before quenching with 1 M Tris **(a) The IncC2-DSS crosslinking.** Different protein: DSS ratios were used for crosslinking reaction. Western blot analysis was performed after SDS PAGE (4-20% gradient gel). The observed cross-linked protein species show a similar pattern to that observed with BS3 crosslinking but the distinct monomeric IncC2 cross-linked band is only observed at higher concentrations of DSS. **(b) The IncC1-DSS Crosslinking.** The reaction mixture for the IncC1-DSS crosslinking was analysed in a similar way to that for IncC2, using SDS PAGE (4-20 % gradient gel). The IncC1-DSS crosslinking products are similar to those obtained using the BS3 cross-linker. The IncC1 cross-linked protein bands, with molecular weights approximately those of dimers, trimers and tetramers are observed.

6.13.2. KorB and KorA chemical crosslinking

The KorB and KorA proteins (10 μM each) were cross-linked using the BS3 cross-linker. The two proteins were cross-linked separately, as well as in the presence of each other. The KorB-BS3 crosslinking experiment shows that KorB forms predominantly homodimers, while some of the KorB protein was also observed as a tetramer (Figure 6.16a). In case of KorA, the cross-linked protein showed homodimer formation (Figure 6.16b). When two proteins are cross-linked together, a diffused band was observed with a molecular weight more than a KorB tetramer and Mass spectrometry analysis of individual band showed the presence of KorA-KorB cross-linked peptides (Figure 6.16c).

6.13.3. IncC and KorB chemical crosslinking

The IncC1, IncC2 and KorB proteins were cross-linked individually as well as in the presence of each other. The crosslinking products were analysed using SDS PAGE (6-15 % gradient gels). In IncC2-KorB crosslinking experiment, KorB was predominantly dimeric, but when DNA was added to the IncC2-KorB crosslinking reaction mixture, a stronger KorB tetramer band was observed (Figure 6.17a). In case of IncC1 crosslinking to KorB, similar results were observed, the KorB dimer was prominent when KorB-IncC1 were cross-linked, but the crosslinking pattern changed in the presence of DNA to give more tetrameric KorB, and some higher species (Figure 6.17b). For IncC1-IncC2 crosslinking, SDS PAGE (4-20 % gradient gel) showed the appearance of some high molecular weight species (>212 kDa) (6.17c).

6.13.4. IncC protein crosslinking in presence of ATP, ADP and DNA

The IncC proteins were also crosslinked in the presence of ATP, ADP and DNA (57 bp, O_AO_B), and the crosslinking products were examined by SDS PAGE and Western blot analysis. Addition of ATPγS or ADP did not change the IncC2 crosslinking pattern (Figure 6.18a); however, the IncC2 crosslinking pattern was changed in the presence of ATP. In that case as most of the cross-linked protein was present as a homodimer; the higher bands disappeared and the cross-linked monomeric IncC2 band was not visible. The IncC2 protein was also cross-linked in the presence of the KorB mutant NΔ31CΔ100 (±ATP). The crosslinked products were analysed by SDS PAGE and Western blotting, using anti-His antibodies. As the KorB NΔ31CΔ100 protein is not His tagged, the protein bands from this species were not visible. The

apparent IncC2 crosslinking pattern was not changed significantly on addition of KorB NΔ31CΔ100 (Figure 6.18a), but again showed more InC2 dimer, and no higher bands, in the presence of ATP.

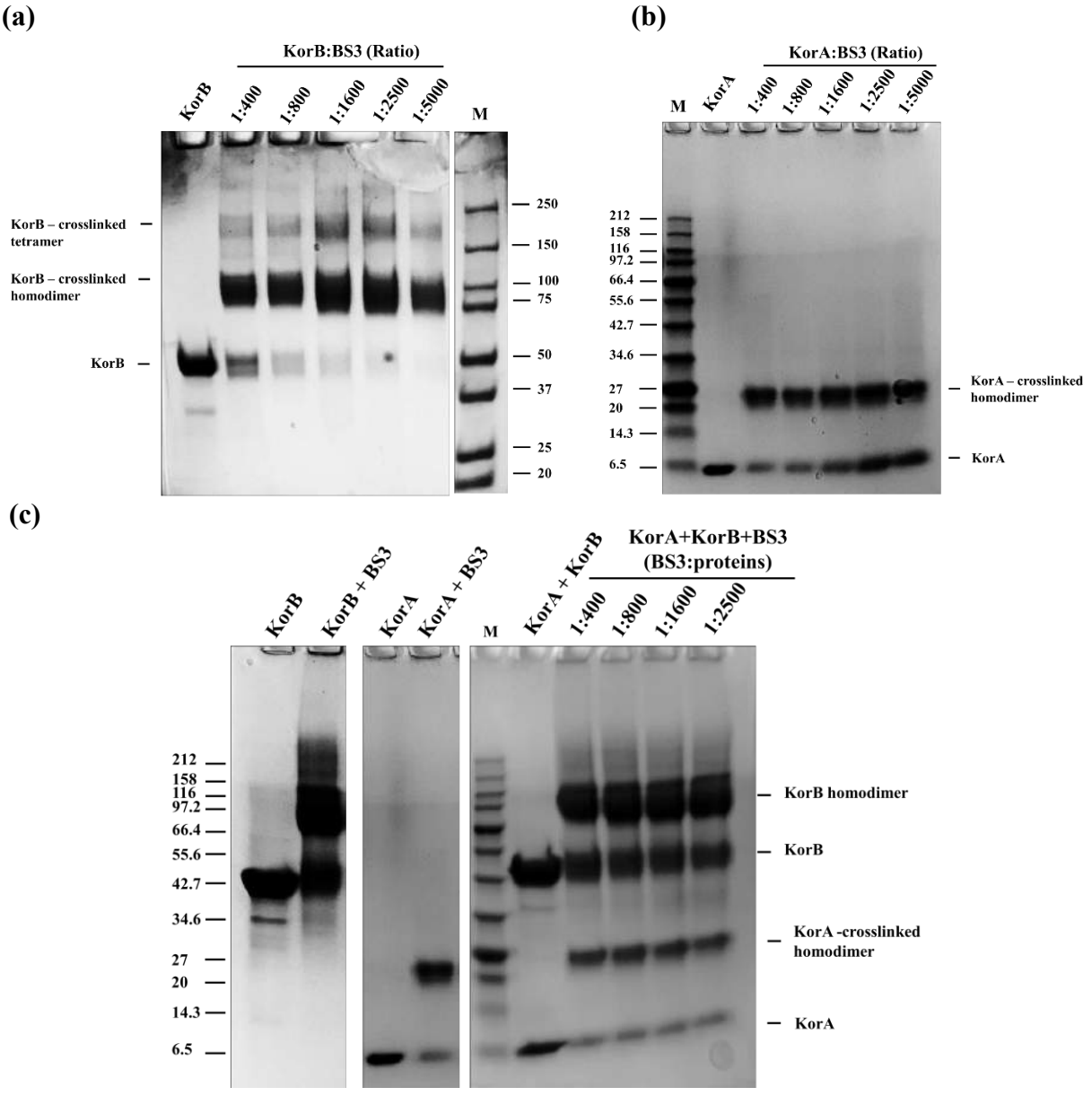


Figure 6.16 KorB and KorA chemical crosslinking analysis. The BS3 cross-linker was used for KorB and KorA (10 μM each) chemical crosslinking. Crosslinking reactions were performed at room temperature for 10 min before quenching with 1 M Tris (5 μL) **(a) KorB-BS3 crosslinking.** Different protein: BS3 ratios (1:400 to 1:2400) were used for KorB crosslinking and crosslinking products were analysed by SDS PAGE (4-20% gradient gel). KorB forms predominantly homodimers and some tetramers. The cross-linked monomer is not observed. **(b) KorA-BS3 Crosslinking.** The KorA-BS3 crosslinking products were analysed using SDS PAGE (12 % gel). KorA is found to form homodimers. **(c)** When KorA and KorB proteins are cross-linked in the presence of each other, the KorB protein is more visible as a dimer band. KorA protein is still predominantly dimeric in the presence of KorB.

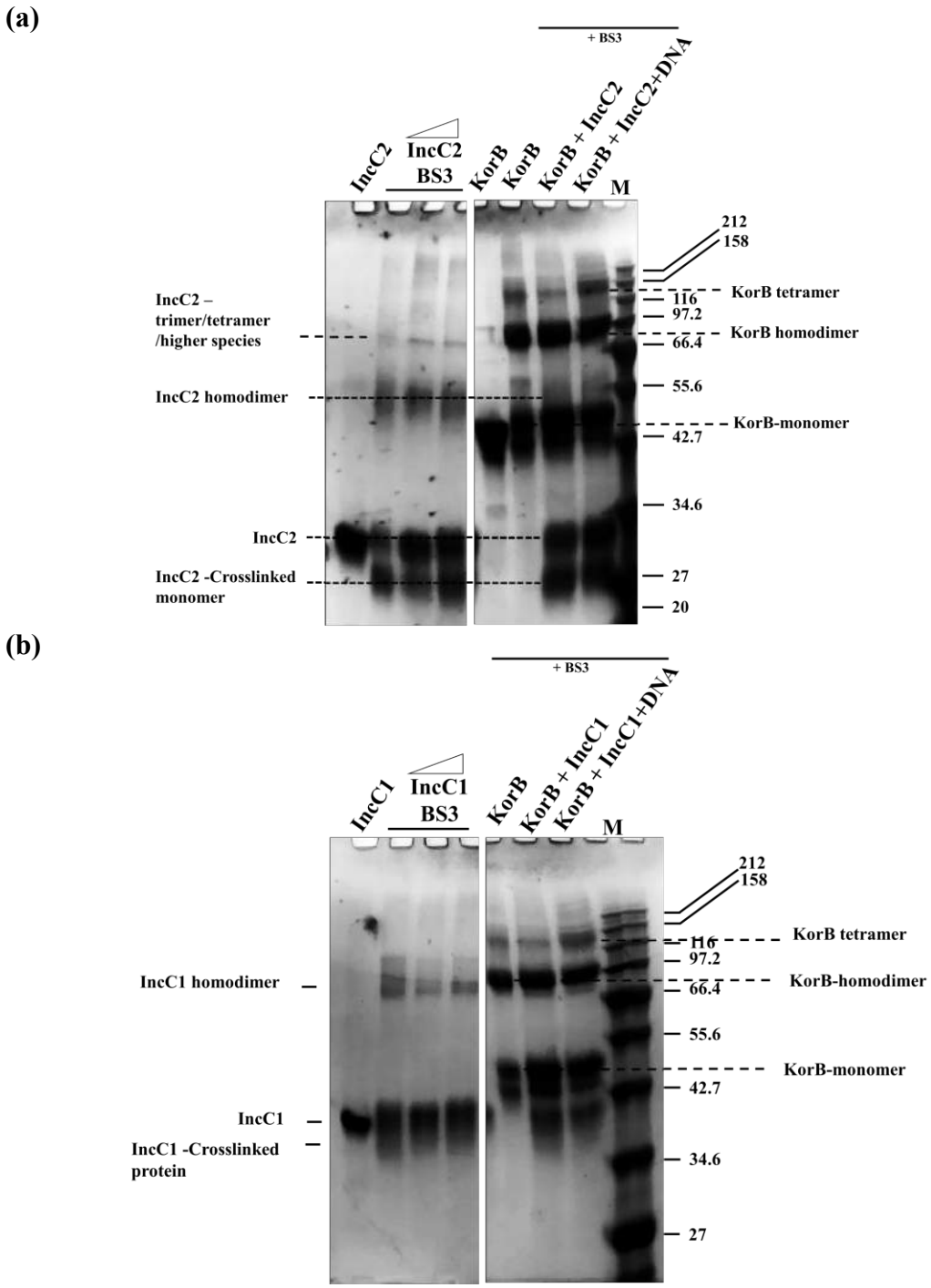


Figure 6.17 KorB-IncC and IncC-IncC chemical crosslinking. The BS3 cross-linker was used for protein crosslinking and reactions were performed at room temperature for 10 min before quenching with 1 M Tris (5 μL). The cross-linked products were analysed using SDS PAGE (6-15 % gradient gels) **(a) IncC2-KorB crosslinking.** For the IncC2-KorB chemical crosslinking equal amounts of protein were used (5 μM). In the presence of IncC2, less KorB was found to be tetramer and more protein was seen as a dimer. When DNA was added to the IncC2-KorB crosslinking reaction, the tetramer band was more prominent **(b)** For IncC1-KorB crosslinking, similar results are observed and KorB is more dimeric when cross-linked in the presence of IncC1 but upon adding DNA the KorB tetramer band is more prominent.

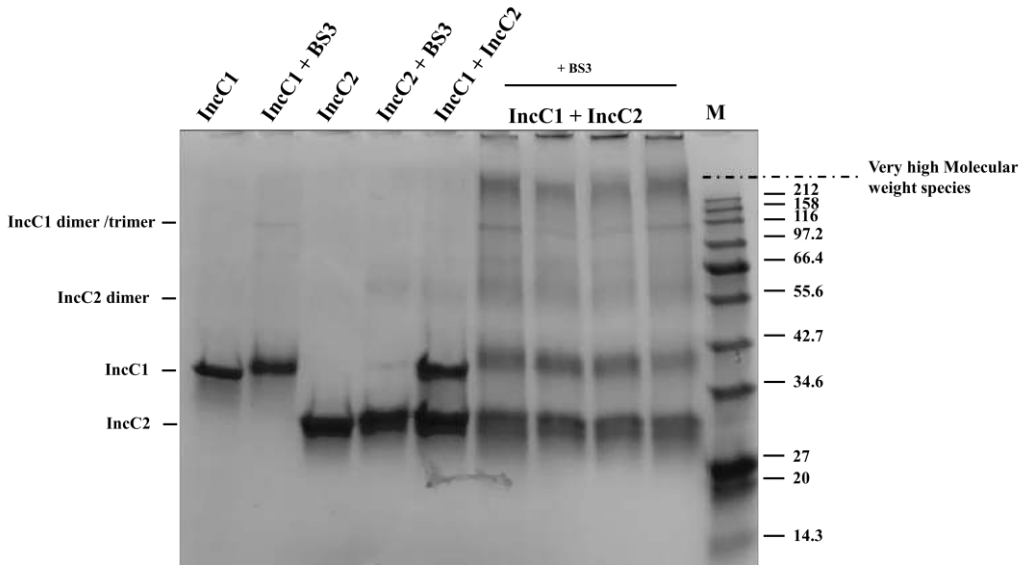


Figure 6.17 (c) IncC1-IncC2 Crosslinking. IncC1 and IncC2 protein (5 μM each) were cross-linked using BS3 chemical cross-linker. Reactions were performed at room temperature for 10 min before quenching with 1 M Tris (5 μL). Crosslinking products for IncC1-IncC2 were analysed using SDS PAGE (4-20 % gradient gel). Some high molecular weight bands (>212 kDa) were observed on the SDS PAGE gel.

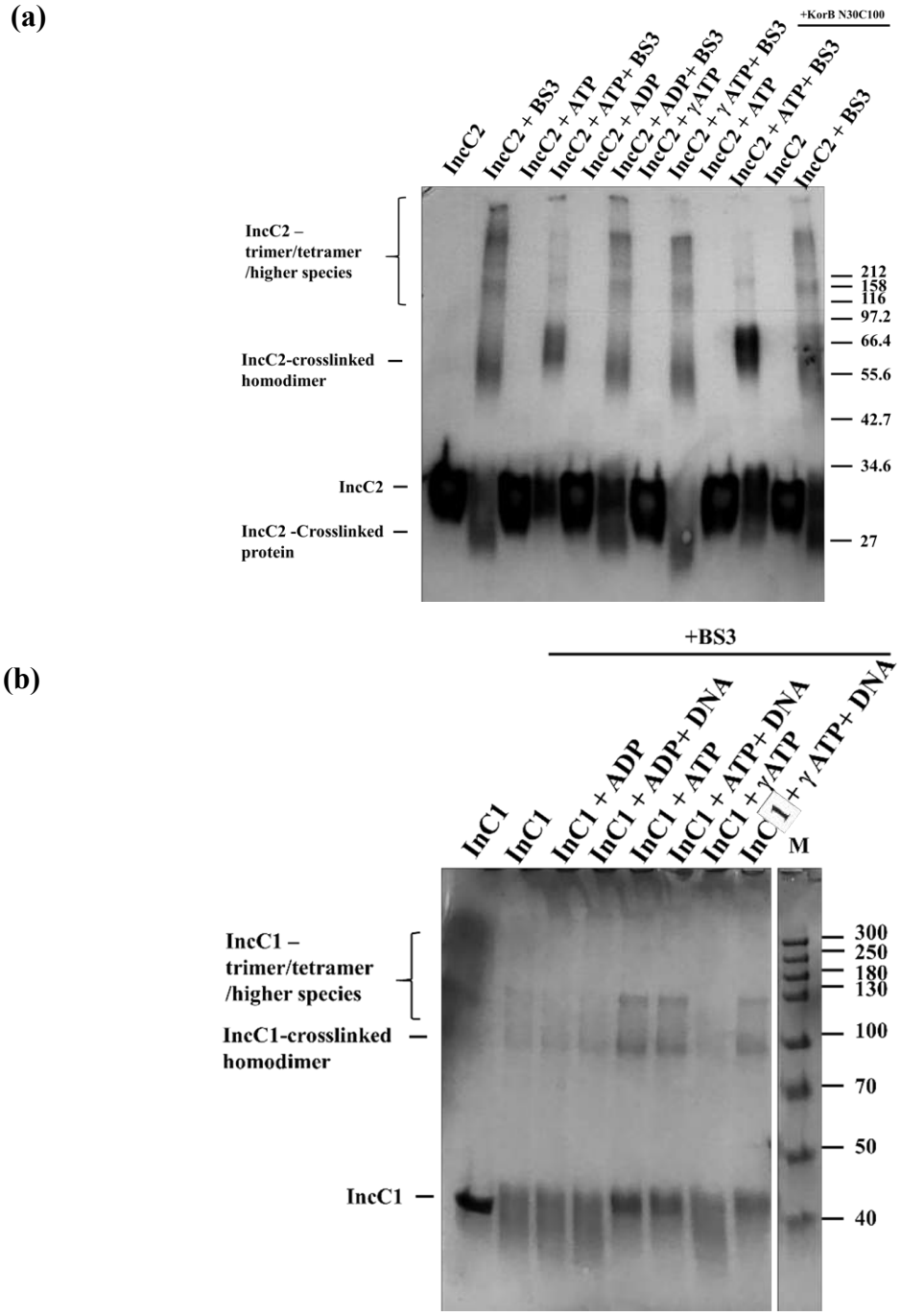


Figure 6.18 IncC1 and IncC2 chemical crosslinking in the presence of ATP, ATPγS, ADP and DNA (O_AO_B). The BS3 cross-linker was used for the IncC1 and IncC2 crosslinking. Reactions were performed at room temperature for 10 min before quenching with 1 M Tris (5 μL). Crosslinking products were analysed using SDS PAGE (4-20 % gradient gels) **(a) IncC2 crosslinking.** The IncC2 crosslinking products are analysed by Western blot after SDS PAGE. In the presence of ATP, the IncC2 protein crosslinking pattern is changed. Most of the cross-linked IncC2 protein is found as a homodimer and as a distinct monomeric cross-linked band, that is not visible in the presence of ATP. In comparison, ATPγS or ADP do not change the IncC2 crosslinking pattern. The IncC2 protein is also cross-linked in the presence of KorB mutant N31C100 and ATP. Crosslinking products were analysed by SDS PAGE and Western blot (His antibodies are used in Western blot, while N30C100 bands should not be visible after blotting as it is not a His-tagged protein). A stronger IncC2 dimer band appears in the presence of

N30C100 (±ATP). **(b) IncC1 crosslinking.** IncC1 is cross-linked in a similar way as IncC2 and cross-linked products are analysed by SDS PAGE (4-20% gradient gel). In the presence of ATP, the IncC1 protein is more dimeric and trimeric/tetrameric. The presence of ATPγS or ADP alone did not change the crosslinking pattern but addition of ATPγS in the presence of DNA gave a similar crosslinking pattern as that observed in the presence of ATP.

IncC1 was cross-linked, in the presence of ATP, ADP and DNA, and the products were analysed by the SDS PAGE (Figure 6.18b). IncC1 protein gave more dimeric and trimeric/tetrameric forms in the presence of ATP than in its absence, while there was little change in the crosslinking pattern in the presence of ATPγS or ADP alone, or with ADP and DNA. In contrast, addition of ATPγS, in the presence of DNA to IncC1, gave a crosslinking pattern similar to one observed for ATP (Figure 6.18b).

6.14. IncC2 homology modelling

The IncC1 and IncC2 proteins were highly unstable and sufficient concentration of proteins was never achieved to obtain good crystals or to do NMR. To determine the IncC2 protein structure, homology modelling was performed using Modeller ver 9.19. Figure 6.8 shows an illustration of the homology modelling method. Three structure templates (PDB IDs 2BEJ, 1WCV and 5IHP) were selected on the basis of E-Value, sequence coverage and sequence identity, and were used for homology modelling (Figure 6.19). 2BEJ (2.1 Å) and 1WCV (1.6 Å) are structures of the same ParA protein from *Thermus thermophiles*, while 5IHP (1.85 Å) is the structure of a ligase from *Mycobacterium smegmatis*.

The PDB structure templates were used individually to obtain three IncC2 homology models. The top 20 initial models for each template are shown in (Figures 6.20a, 6.21a and 6.22a) and their zDOPE and RMSD from the template in (Figures 6.20f, 6.21f and 6.22f). The model, with the lowest zDOPE and RMSD, for each template is shown in (Figures 6.20c, 6.21c and 6.22c overlaid on the template structure. Figures 6.20e, 6.21e and 6.22e, show the 20 structures generated from each selected model after loop modelling and refinement, with their zDOPE and RMSD scores in Figures 6.20g, 6.21g and 6.22g. The best structure for each is shown in Figure 6.20e, 6.21e and 6.22e, together with their evaluation from the ProSA server and Ramachandran plots (Figure 6.20 h, i, j, 6.21h, i, j and 6.22h, i, j). The refined structures obtained using the templates 1WCV 2BEJ and 5IHP had zDOPE scores -1.11, 0.46, and -0.75 respectively, with RMSD values 3.06, 4.61, and 5.74 respectively.

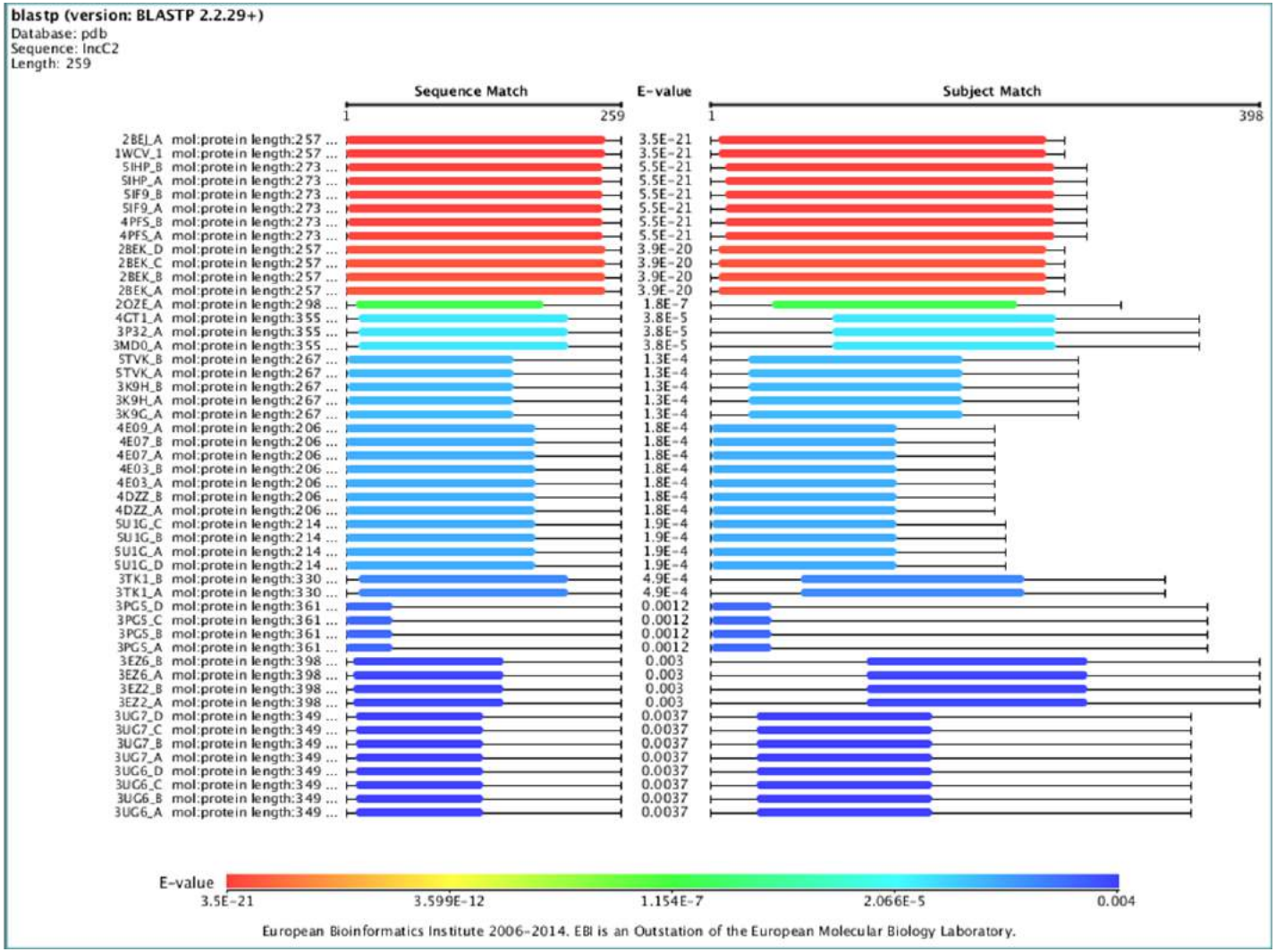
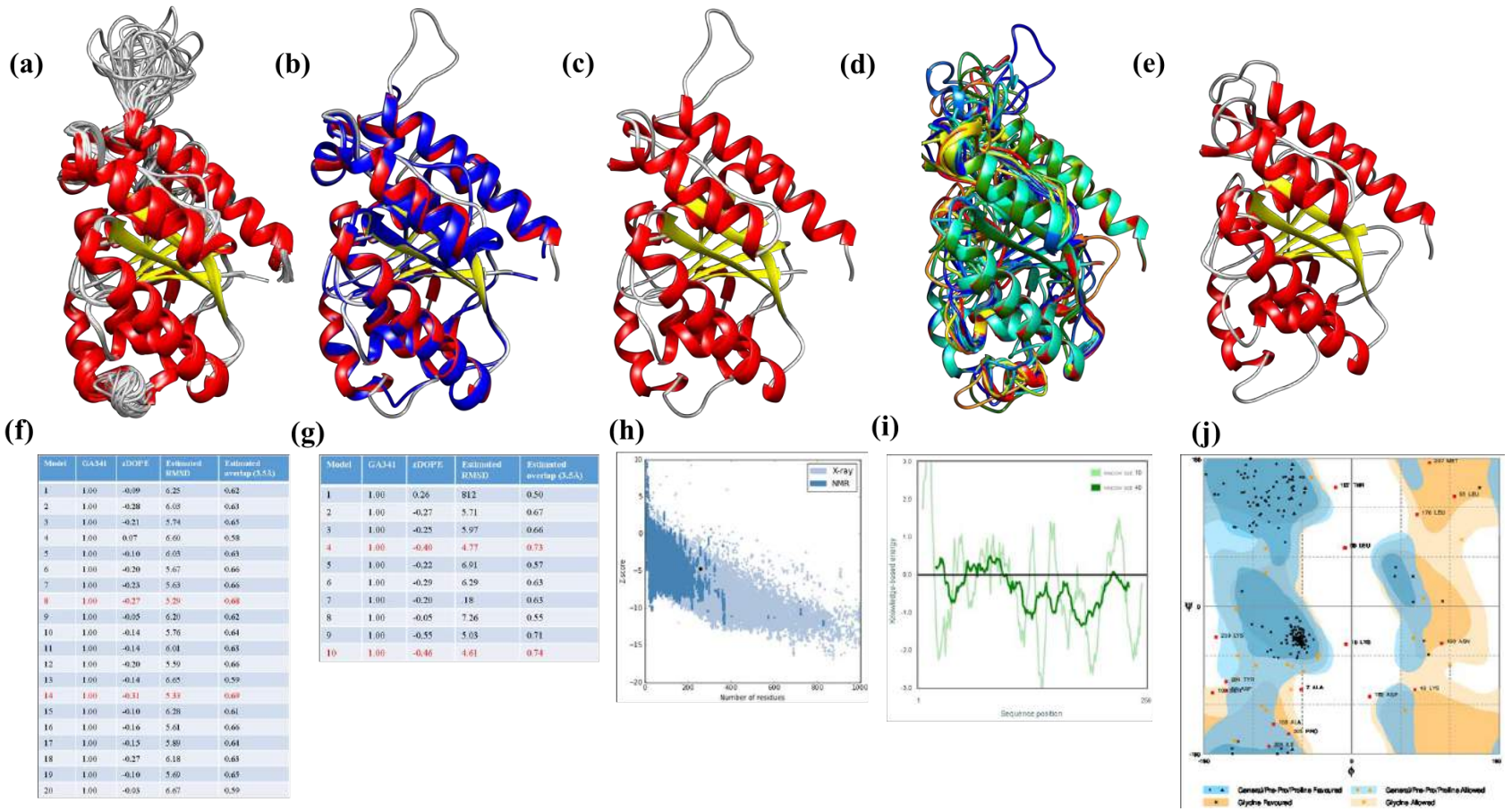


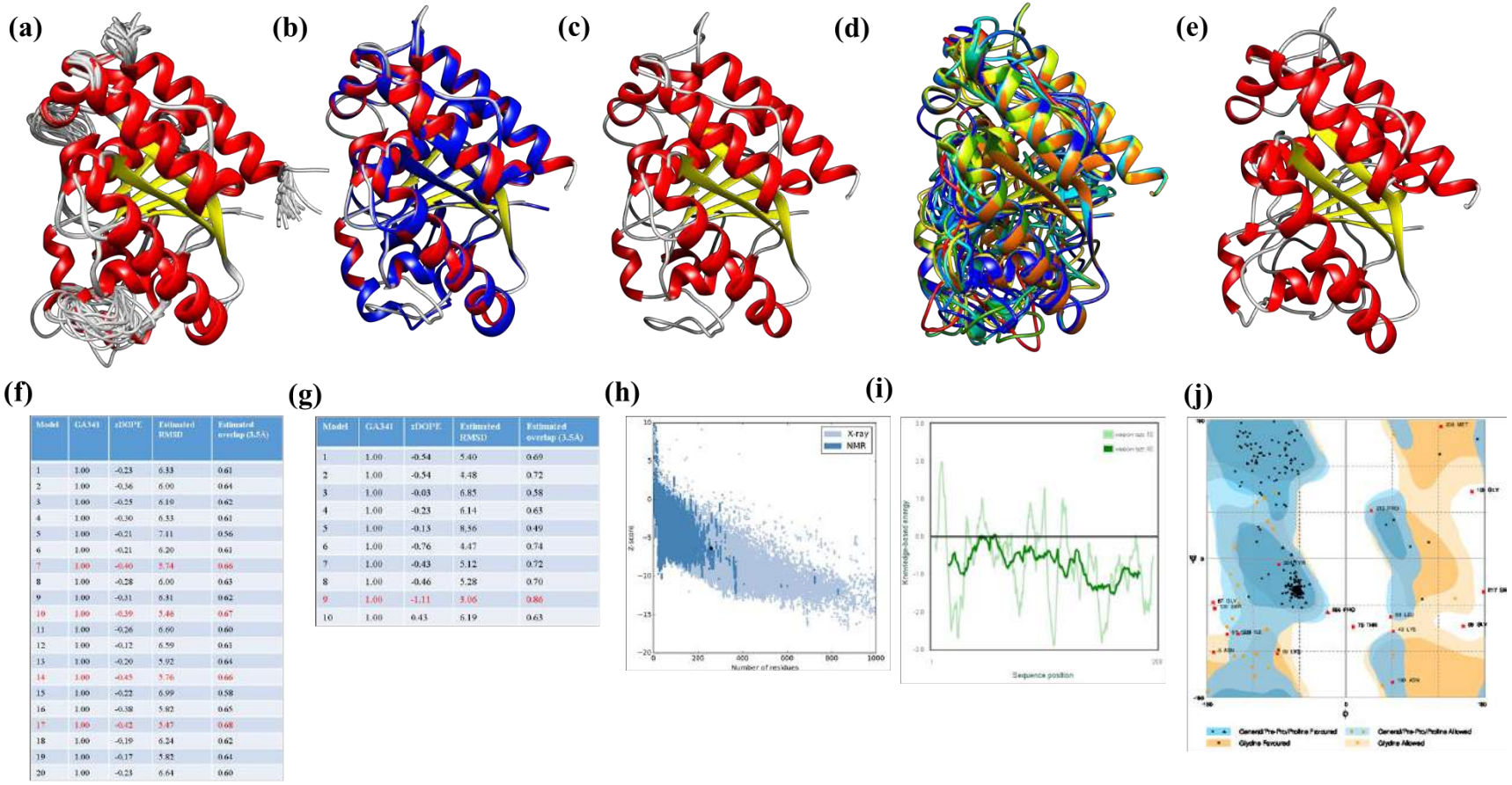
Figure 6.19 The IncC2 protein sequence BLAST against PDB. BLASTP was used to search for proteins in the PDB that are homologous to IncC2. The results are arranged with respect to their E-values and show the sequence match as well as the subject match to the complete target protein. The red colours show E-values around 3.5 E^{-21}, the green colours show 1.15 E^{-7}, light blue shows 2.06 E^{-5} and the dark blue colour shows E-value of 0.004. Three structures templates with best E-Values and sequence coverage (2BEJ, 1WCV, 5IHP) were selected for the IncC2 homology modelling.



Model	GA341	zDOPE	Estimated RMSD	Estimated overlap (5.5Å)
1	1.00	-0.09	6.25	0.62
2	1.00	-0.28	6.03	0.63
3	1.00	-0.21	5.74	0.65
4	1.00	0.07	6.60	0.58
5	1.00	-0.10	6.05	0.63
6	1.00	-0.20	5.67	0.66
7	1.00	-0.23	5.63	0.66
8	1.00	-0.27	5.29	0.68
9	1.00	-0.05	6.20	0.62
10	1.00	-0.14	5.76	0.64
11	1.00	-0.14	6.01	0.63
12	1.00	-0.20	5.59	0.66
13	1.00	-0.14	6.65	0.59
14	1.00	-0.31	5.33	0.69
15	1.00	-0.10	6.28	0.61
16	1.00	-0.16	5.61	0.66
17	1.00	-0.15	5.89	0.61
18	1.00	-0.27	6.18	0.63
19	1.00	-0.10	5.69	0.65
20	1.00	-0.03	6.67	0.59

Model	GA341	zDOPE	Estimated RMSD	Estimated overlap (5.5Å)
1	1.00	0.26	812	0.50
2	1.00	-0.27	5.71	0.67
3	1.00	-0.25	5.97	0.66
4	1.00	-0.40	4.77	0.73
5	1.00	-0.22	6.91	0.57
6	1.00	-0.29	6.29	0.63
7	1.00	-0.20	18	0.63
8	1.00	-0.05	7.26	0.55
9	1.00	-0.55	5.03	0.71
10	1.00	-0.46	4.61	0.74

Figure 6.20 IncC2 Homology modelling using the 2BEJ template. (a) Homology modelling was performed using Modeller ver 9.19. A total of 20 homology models have been overlaid each other. The core protein structure is similar in all models. **(b)** The best model (in secondary structure colours; red, yellow, grey) is superimposed on the template 2BEJ (Blue). **(c)** The model selected on the basis of zDOPE (-0.31) and RMSD (5.33) **(f)** The zDope and RMSD scores used to rank the models. **(d)** The overlay of ten structures, obtained after loop modelling of the model in (c). **(g)** the zDOPE and RMSD scores of the models in (c). **(e)** The final protein modelled from this template. **(h-j)** a structure evaluation of the model in (g) using the ProSA web server and the Ramachandran plots. For final model, the Ramachandran Plot shows 15 outliers, while zDOPE, zScore and RMSD are found to be -0.46, -5.36 and 4.61 respectively.



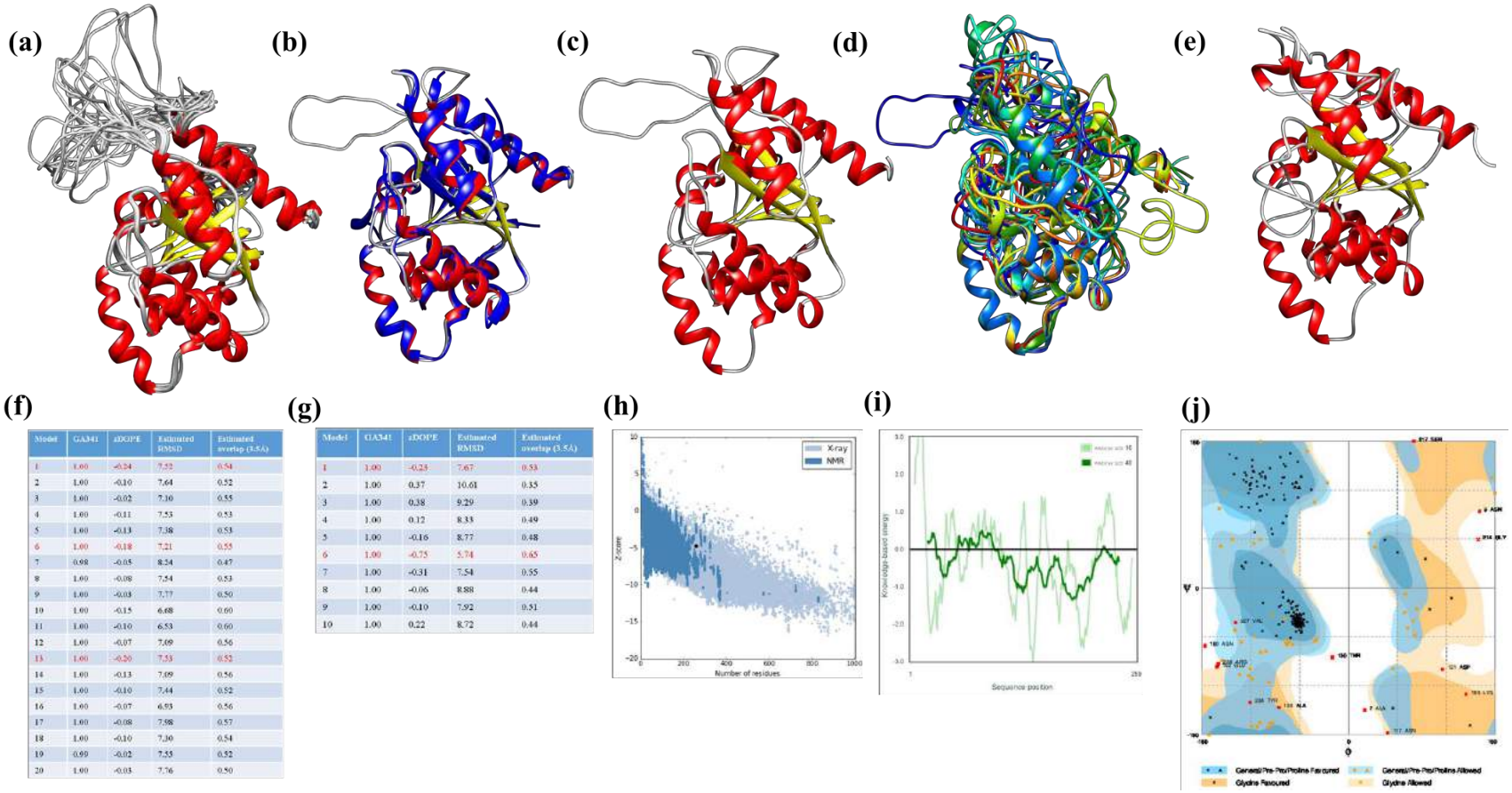
(f)

Model	GA341	zDOPE	Estimated RMSD	Estimated overlap (3.5Å)
1	1.00	-0.23	6.33	0.61
2	1.00	-0.36	6.00	0.64
3	1.00	-0.25	6.10	0.62
4	1.00	-0.30	6.33	0.61
5	1.00	-0.21	7.11	0.56
6	1.00	-0.21	6.20	0.61
7	1.00	-0.40	5.74	0.66
8	1.00	-0.28	6.00	0.63
9	1.00	-0.31	6.31	0.62
10	1.00	-0.39	5.46	0.67
11	1.00	-0.26	6.60	0.60
12	1.00	-0.12	6.59	0.61
13	1.00	-0.20	5.92	0.64
14	1.00	-0.45	5.76	0.66
15	1.00	-0.22	6.99	0.58
16	1.00	-0.38	5.82	0.65
17	1.00	-0.42	5.47	0.68
18	1.00	-0.19	6.24	0.62
19	1.00	-0.17	5.82	0.64
20	1.00	-0.23	6.64	0.60

(g)

Model	GA341	zDOPE	Estimated RMSD	Estimated overlap (3.5Å)
1	1.00	-0.54	5.40	0.69
2	1.00	-0.54	4.48	0.72
3	1.00	-0.03	6.85	0.58
4	1.00	-0.23	6.14	0.63
5	1.00	-0.13	8.56	0.49
6	1.00	-0.76	4.47	0.74
7	1.00	-0.43	5.12	0.72
8	1.00	-0.46	5.28	0.70
9	1.00	-1.11	3.06	0.86
10	1.00	0.43	6.19	0.63

Figure 6.21 IncC2 Homology modelling using the 1WCV template. (a) Homology modelling was performed using Modeller ver 9.19. A total of 20 homology models have been overlaid each other. The core protein structure is similar in all models. **(b)** Best model, selected on the basis of zDOPE (-0.45) and RMSD (5.76), (shown in secondary structure colours; red, yellow, grey) is superimposed on the template 1WCV (Blue). **(f)** The zDope and RMSD scores are used to rank models. **(d)** The overlay of ten structures, obtained after loop modelling of the model in (c). **(g)** the zDOPE and RMSD scores of the models in (c). **(e)** Final protein model is selected on the basis of zDOPE and RMSD and **(h-j)** a structure evaluation is performed using the ProSA web server and the Ramachandran plots. For final model, the Ramachandran Plot shows 14 outliers, while zDOPE, zScore and RMSD are found to be -1.11, -6.76 and 3.06 respectively.



Model	GA341	zDOPE	Estimated RMSD	Estimated overlap (3.5Å)
1	1.00	-0.24	7.52	0.54
2	1.00	-0.10	7.64	0.52
3	1.00	-0.02	7.10	0.55
4	1.00	-0.11	7.53	0.53
5	1.00	-0.13	7.38	0.53
6	1.00	-0.18	7.21	0.55
7	0.98	-0.05	8.54	0.47
8	1.00	-0.08	7.54	0.53
9	1.00	-0.03	7.77	0.50
10	1.00	-0.15	6.68	0.60
11	1.00	-0.10	6.53	0.60
12	1.00	-0.07	7.09	0.56
13	1.00	-0.20	7.53	0.52
14	1.00	-0.13	7.09	0.56
15	1.00	-0.10	7.44	0.52
16	1.00	-0.07	6.93	0.56
17	1.00	-0.08	7.98	0.57
18	1.00	-0.10	7.30	0.54
19	0.99	-0.02	7.53	0.52
20	1.00	-0.03	7.76	0.50

Model	GA341	zDOPE	Estimated RMSD	Estimated overlap (3.5Å)
1	1.00	-0.23	7.67	0.53
2	1.00	0.37	10.61	0.35
3	1.00	0.38	9.29	0.39
4	1.00	0.12	8.33	0.49
5	1.00	-0.16	8.77	0.48
6	1.00	-0.75	5.74	0.65
7	1.00	-0.31	7.54	0.55
8	1.00	-0.06	8.88	0.44
9	1.00	-0.10	7.92	0.51
10	1.00	0.22	8.72	0.44

Figure 6.22 IncC2 Homology modelling using the 5IHP template. (a) Homology modelling was performed using Modeller ver 9.19. A total of 20 homology models have been overlaid each other. Core protein structure is similar in all models **(b)** Best model selected on the basis of (-0.24) and RMSD (7.52) (in secondary structure colours; red, yellow, grey) have been superimposed on the template 5IHP (Blue). **(f)** The zDope and RMSD scores are used to rank models. **(d)** The overlay of ten structures, obtained after loop modelling of the model in (c). **(g)** the zDOPE and RMSD scores of the models in (c). **(e)** Final protein model is selected on the basis of zDOPE and RMSD and **(h-j)** a structure evaluation is performed using the ProSA web server and the Ramachandran plots. For final model, the Ramachandran Plot shows 14 outliers, while zDOPE, zScore and RMSD are found to be -0.75, -5.36 and 5.74 respectively.

6.14.1. Refinement of the IncC2 homology models

The three best IncC2 homology models, obtained using the three different templates (1WVC, 2BEJ and 5IHP), were subjected to further structure refinement using the PREFMD (Protein structure REFinement via Molecular Dynamics) (Heo and Feig, 2017) (http://feig.bch.msu.edu/prefmd) and ReFOLD servers (Shuid et al., 2017) (http://www.reading.ac.uk/bioinf/ReFOLD/). The refined structures were evaluated using the DOPE scores, and the Ramachandran Plots. The PreFMD-refined models have zDOPE and RMSD -1.23 and 2.37 for the model based on PDB ID 1WCV; -1.00 and 3.77 respectively for that based on the template 2BEJ and -0.75 and 4.34 for that based on the template PDB ID 5IHP. The 1WCV-based model is best, in terms of zDOPE (-1.23) and RMSD (2.37), in comparison to other structures (Figure 6.23 and Figure 6.24). The ReFOLD-refined structures, have zDOPE and RMSD score as -0.76 and 3.60, when modelled using the template PDB ID 1WCV; -0.51 and 4.45, when modelled using the template 2BEJ, and -0.76 and 5.03 when modelled using the template PDB ID 5IHP (Figure 6.25 and Figure 6.26). The models were assessed on the basis of their DOPE scores per residue (from the Sali server), their Ramachandran plots, and the knowledge-based energy plots from the PROSA server (Wiederstein and Sippl, 2007) (Figure 6.24 and 6.26). All the structures show fewer outliers in the Ramachandran plots after refinement Overall the PREFMD server provided better structure refinement for the IncC2 homology models (Figure 6.23-6.26) than the ReFOLD server, based on these scores. The IncC2 refined model obtained using the 1WCV template, with PREFMD, shows better zDOPE (-1.235) and RMSD (2.37Å) scores in comparison to other structures (Figure 6.23).

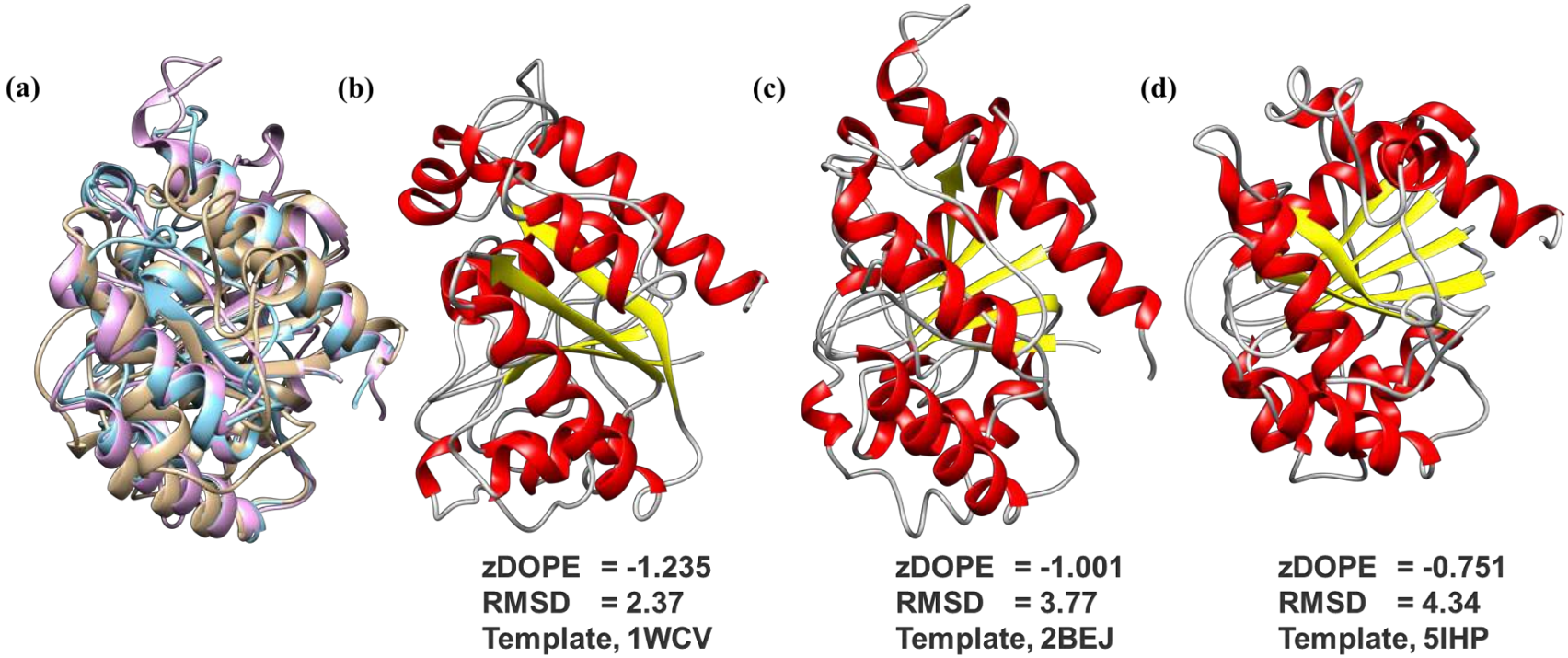


Figure 6.23 The IncC2 Homology models, based on the different templates, after structure refinement using the PREFMD server. (a) Comparison of three structures refined using the PREFMD server. The homology models show differences in their core structures as well as in the loop regions **(b, c, and d)** The PREFMD-refined structure, modelled using the 3 different templates PDB ID 1WCV, 2BEJ and 5IHP and their zDOPE and RMSD scores. The 1WCV-based model is better in terms of zDOPE (-1.23) and RMSD (2.37) in comparison to other structures.

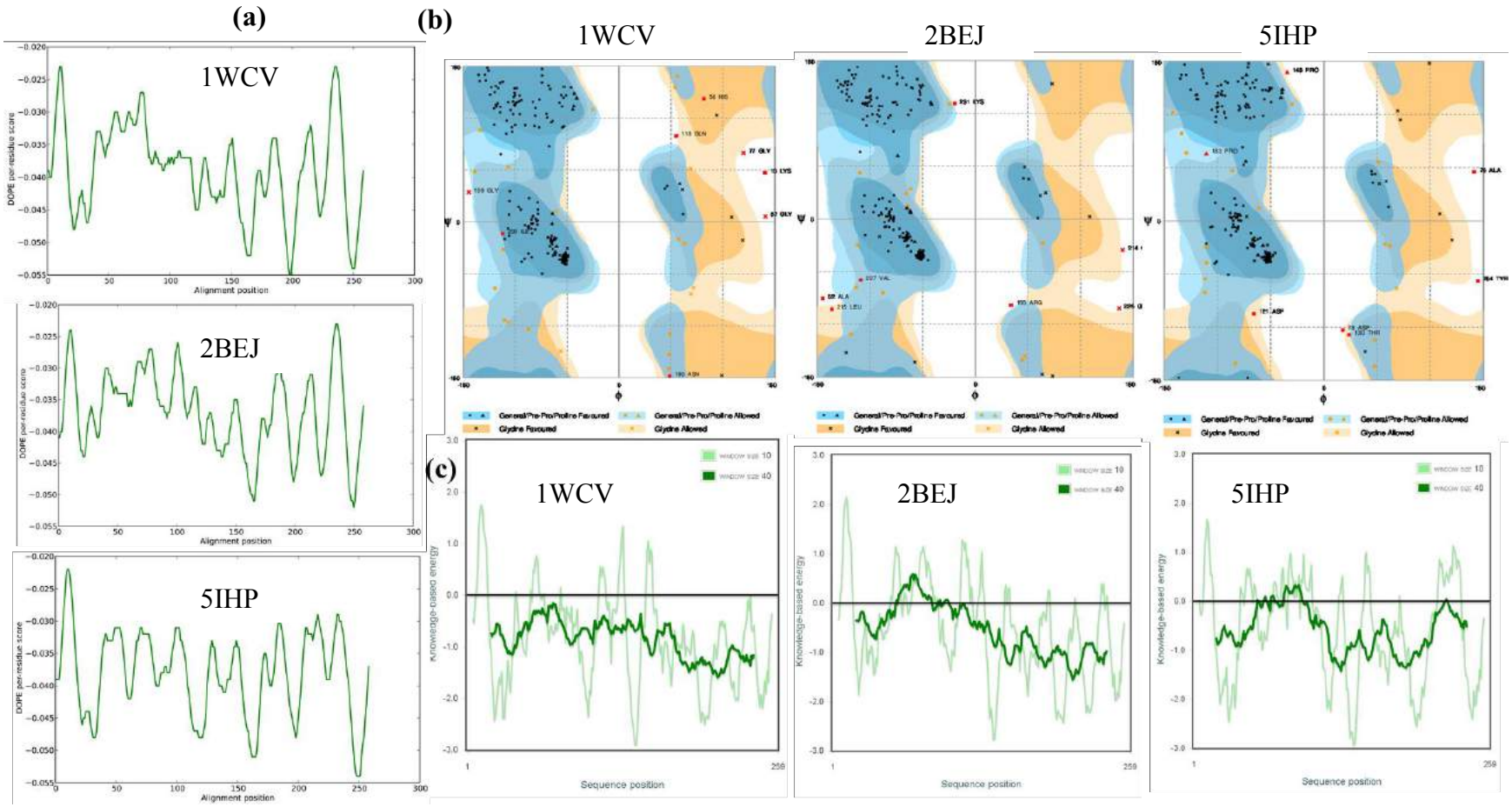


Figure 6.24 An assessment of the IncC2 homology models refined using the PREFMD server. (a) The DOPE score per residue, obtained using the SaliLab Model Evaluation Server (Shen et al., 2006). All residues in refined structures show DOPE scores < -0.20. More than 90% residues have a DOPE < -0.30. **(b)** The Ramachandran plot, for each structure, obtained using RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage2.php). Structures show fewer outliers after refinement. **(c)** The knowledge-based energy score for each residue, obtained from the ProSA server (Wiederstein and Sippl, 2007). At least 80 % of residues in refined structures showed energy a score < 0. The model structure obtained using the template 1WCV shows better scores than other models.

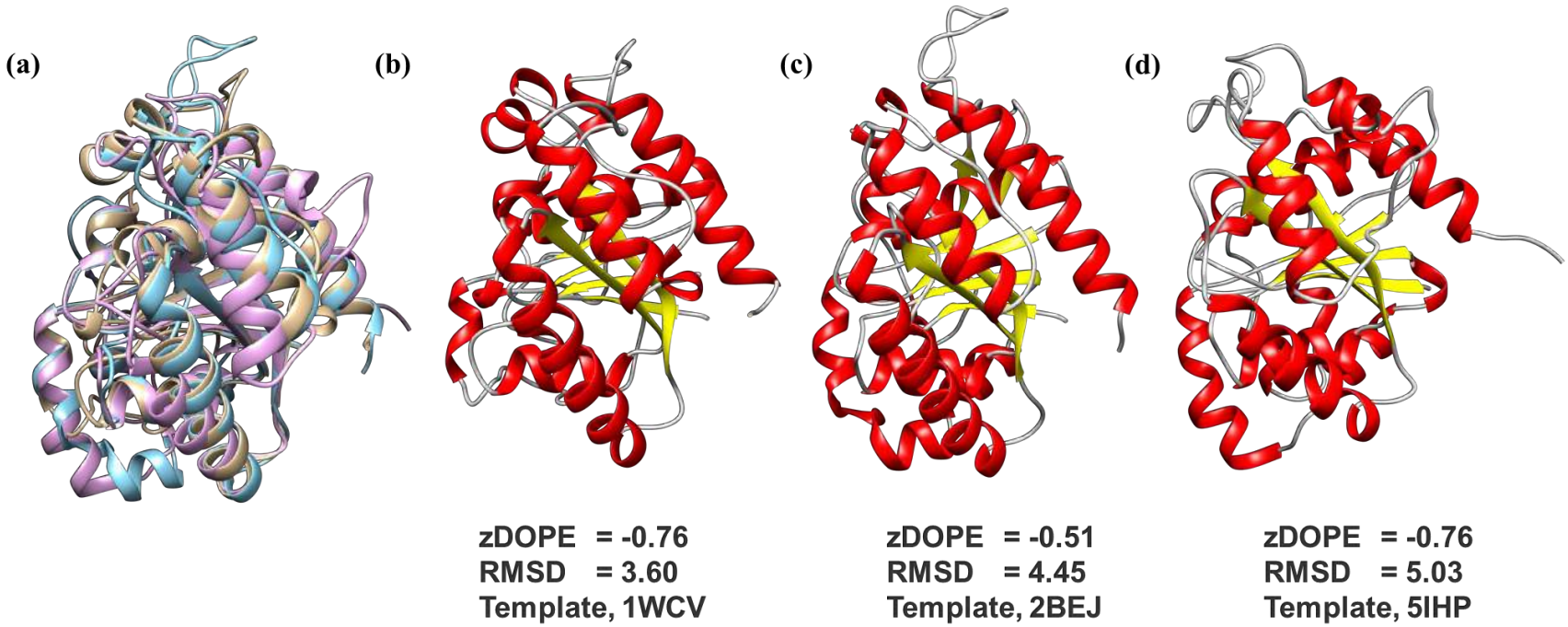


Figure 6.25 The IncC2 Homology models after structure refinement using the ReFOLD server. (a) Comparison of three structures refined using the ReFOLD server. the homology models show differences in their core structures as well as in the loop regions **(b, c and d)** The ReFOLD-refined structures, and their zDOPE and RMSDs.

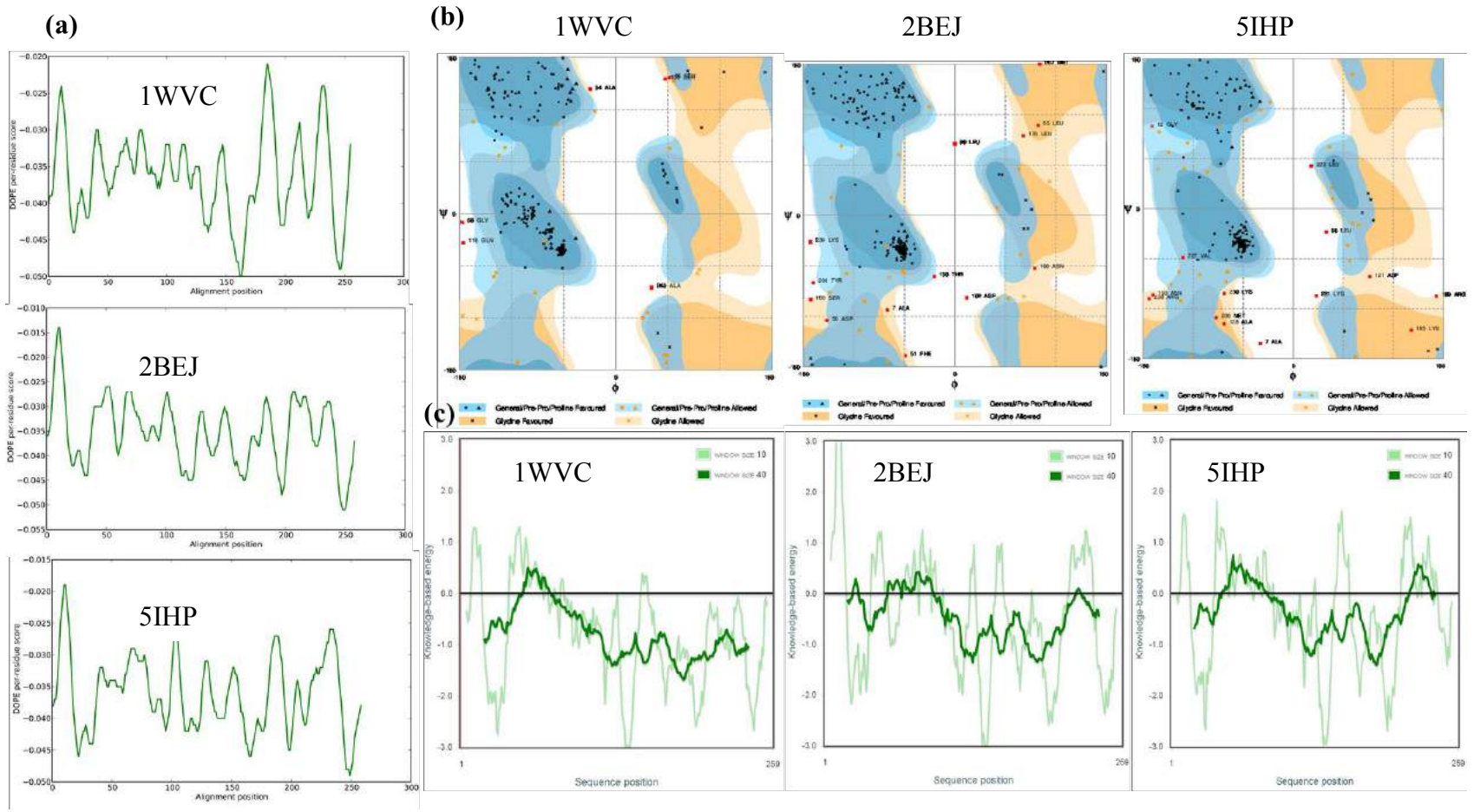


Figure 6.26 An assessment of the IncC2 homology models refined using the ReFOLD server. (a) The DOPE score per residue was obtained using the SaliLab Model Evaluation Server (Shen et al., 2006). All residues in the refined structures show a DOPE score < -0.10. More than 90% residues have a DOPE < -0.20. **(b)** The Ramachandran plot for each structure was obtained using RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage2.php). The refined structures show fewer outliers than before refinement. **(c)** The knowledge-based energy score for each residue was obtained from the ProSA server (Wiederstein and Sippl, 2007). At least 80 % of residues in refined structures show a score < 0. The model structure obtained using the template 2BEJ shows better scores than the other models.

6.15. Protein-protein interactions analysed by chemical crosslinking Mass spectrometry

The IncC2, KorB proteins were cross-linked, individually, and to each other, using the BS3 cross-linker and the cross-linked products were subjected to Mass spectrometry, after digestion with Trypsin. Cross-linked peptides were identified using KOJAK (Hoopmann et al., 2015) and SIM-XL (Lima et al., 2015) Software. A few Mass spectra of cross-linked peptides are given in Appendix 10, as examples of the process. IncC2 and KorB were also cross-linked to KorA, and treated in the same way.

Initially, identified crosslinks were mapped on the protein sequences. The inter- and intra- crosslinks identified for the IncC2, KorB and KorA proteins are shown schematically in Figure 6.27. Initially,>30 crosslinks for IncC2 and >80 crosslinks for KorB were identified. The Mapping of the cross links to the individual proteins shows that the C- and N- terminals of the IncC2 protein may come close to each other in protein structure, while for KorB, the linkers regions interact with other domains and so may bring different domains together in a close proximity Some crosslinks with >34 Å lengths were also identified. Crosslinks for the IncC2-KorB and KorB-KorA interactions were also identified and mapped on the protein sequences (Figure 6.27c and Figure 6.27d). For the IncC2-KorB pair, 20 crosslinks were identified, while for the KorB-KorA pair of proteins, 25 crosslinks were identified (Figure 6.27c and Figure 6.27d). In case of the IncC2-KorB interactions (Figure 6.27c), the C-terminals of both proteins was found to be involved in interactions, while in case of KorB-KorA, the C-terminal domain of the KorA protein is involved in interactions with KorB (Figure 6.27d).

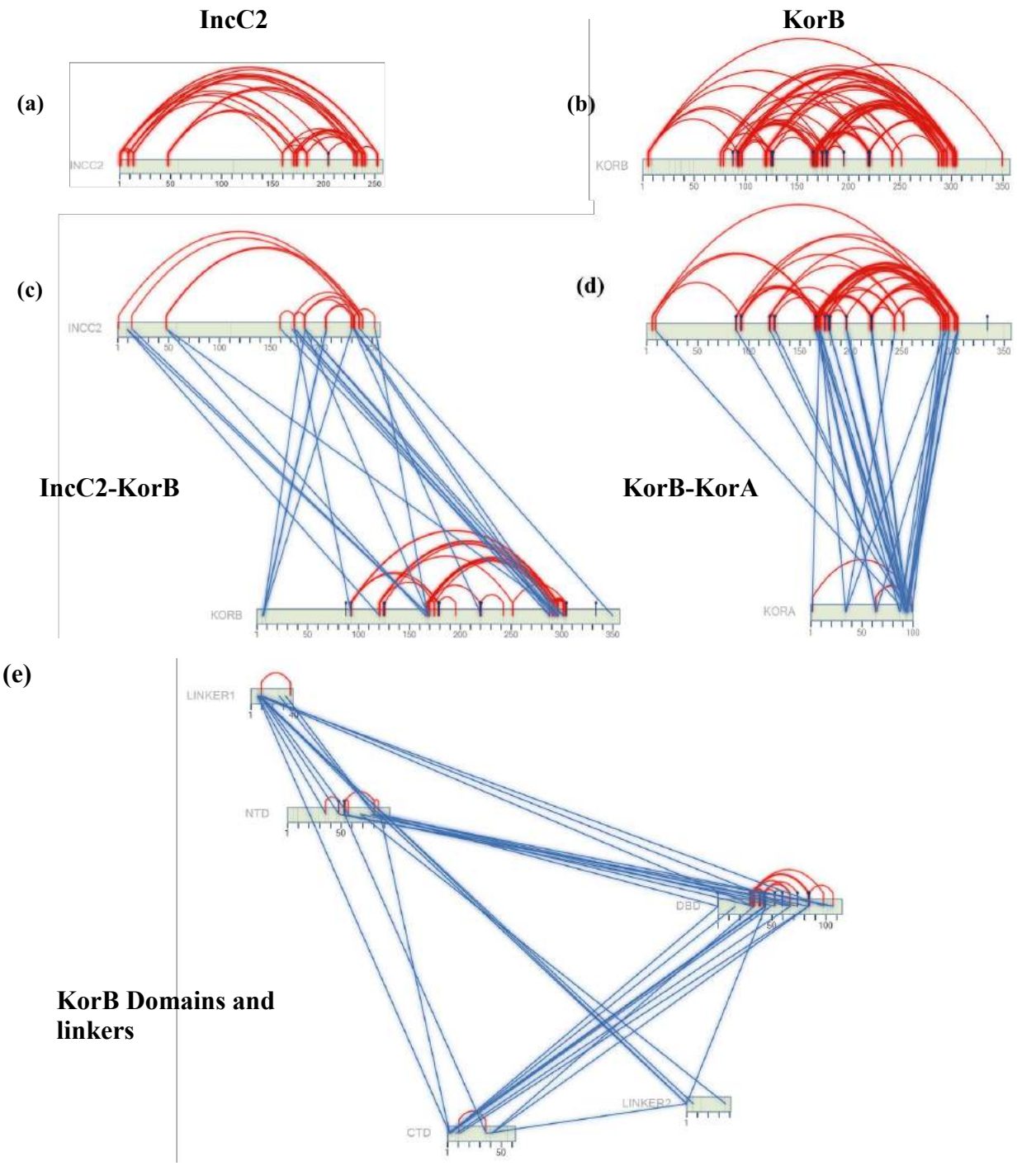


Figure 6.27 The Crosslinks observed in Mass spectra data mapped on the IncC2, KorB and KorA protein sequence. The horizontal lines show the linear sequence of the proteins. The red lines show crosslinking within a single protein, while blue lines show inter- crosslinking of 2 proteins. **(a)** For the IncC2 protein, >30 crosslinks are identified and mapped on to protein sequence. Mapping shows, that the C- and N- terminals in the IncC2 protein may come close to each other in the protein structure **(b, e)** KorB and KorB domain crosslink mapping (>80 identified crosslinks) shows that the linkers regions in the KorB structure interact with other domains and may bring different domains together in a close proximity **(c)** IncC2-KorB crosslinks show that the C- terminals of two proteins may have a major role in protein-protein interactions. **(d)** In the KorB-KorA crosslinks map, it is observed that the C-terminal of the KorB and KorA proteins are involved in protein-protein interactions.

6.15.1. Crosslink mapping of the KorB DBD (DNA binding domain)

Crystal structures for two of the KorB protein domains, the C-terminal, dimerisation, domain (CTD) and the DNA-binding domain (DBD), bound as a dimer to DNA, are available in Protein Data Bank (PDB ID 1IGQ, 1IGU, 1R71). A total of 32 crosslinks (Distance information is given in Appendix 11) were mapped on KorB DBD dimer structure (Figure 6.28). 10 were found within one monomer, and extra 2 were seen across one dimer (or is this only 1 because it is symmetrical). 3 unique crosslinks are present at the dimer interface. The Successful crosslink mapping on the KorB DBD structure shows that a reliable method was used to obtain, analyse and to map crosslinks on protein structures. The Solvent accessible surface (SAS) distances were measured using Xwalk (http://www.xwalk.org/cgi-bin/home.cgi) and more than 90 % of the crosslinks were found to be <34 Å in length.

6.15.2. Crosslink mapping of the IncC2 homology models

Crosslinks were mapped on the IncC2 homology models after their refinement from the PREFMD server, for all Lysine-Lysine (Cα- Cα) distances <34Å (distances information are given in Appendix 11). A total of 18 crosslinks were mapped on the IncC2 refined structure modelled using PDB template 1CVW, with only 16 crosslinks on the IncC2 structure modelled using template PDB ID 2BEJ and 15 crosslinks are mapped on the structure modelled using the 5IHP template (Figure 6.29). When solvent accessible surface distances in the models were mapped using Xwalk, the lengths of 5 crosslinks were found to be more than 34 Å. This may be either due to flexibility of the Lysine side chains, present in loops, or these crosslinks may map to the the dimeric or tetrameric IncC2 structures. The homology model constructed using the template PDB ID 1WCV is better than other structures as the Lysine-Lysine (Cα-Cα) distances, as it has more mapped crosslinks than the other models as well as because the zDOPE and RMSD scores of the model is better than those of the other two models.

6.15.3. Crosslink mapping on the IncC2 Dimer

The IncC2 dimer structure was modelled to determine the IncC2-IncC2 protein interface and the identified crosslinks were mapped on the dimer structure. An automated homology modelling was performed using the SWISS-MODEL server (https://swissmodel.expasy.org) (Biasini et al., 2014). The PDB structure 1WCV was selected as a template and final homodimer

structure was evaluated using zDOPE and RMSD. The zDOPE and RMSD scores for the dimer structure were calculated to be -1.01 and 2.04 respectively. Crosslinks mapped on the IncC2 homodimer model are shown in Figure 6.30. It is observed that amino acids 187-212 (including an α-helix in the 191-203 a.a region) and an α-helix consisting of amino acids 231-256 are likely to form the dimer interface of the IncC2 protein.

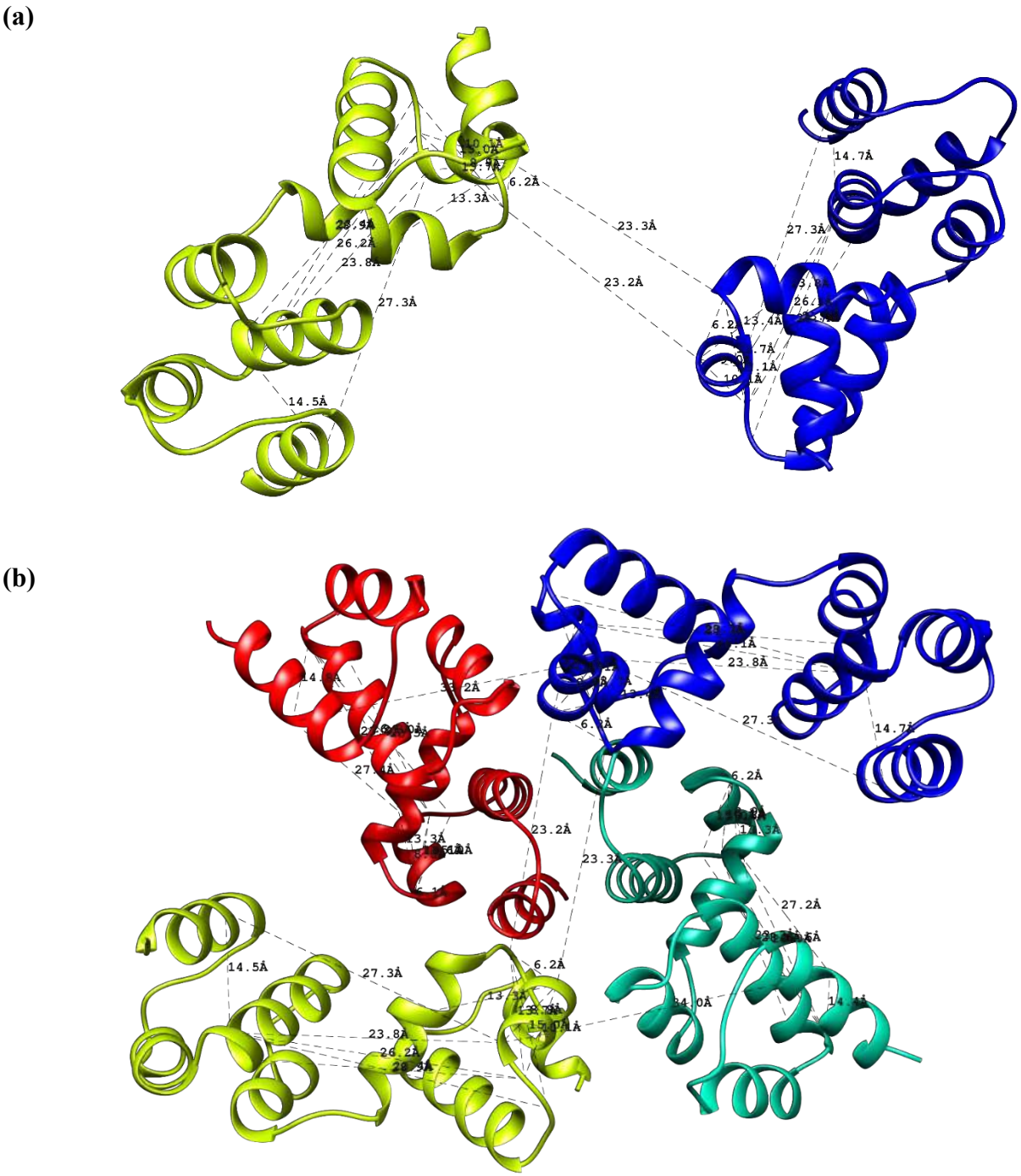


Figure 6.28 Crosslink mapping of the KorB DBD dimer and dimer of dimers (a) A total of 12 crosslinks were mapped on the KorB DBD dimer structure (1R71, Khare et al., 2004). Two out of ten crosslinks were mapped across the dimer interface. **(b)** Crosslinks were also mapped on the KorB DBD dimer of dimers structure. Mapping was performed based on Lysine-Lysine (Cα-Cα) distances. The different colours show different chains of the DBD. Dashed lines show crosslinks and numbers show the length of the crosslink. Crosslink mapping and distance calculations were performed using Chimera ver 1.12 (Pettersen et al., 2004).

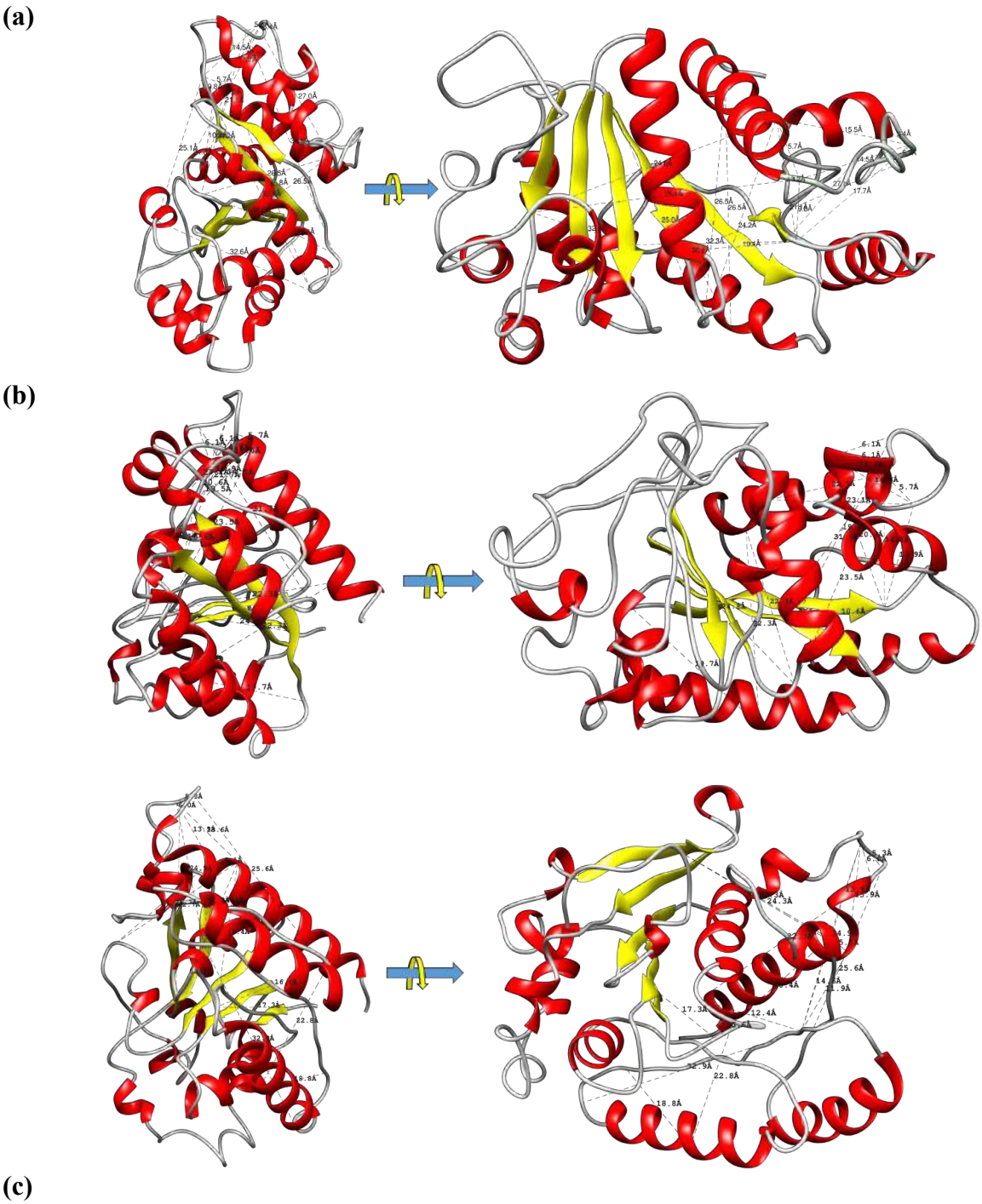


Figure 6.29 Crosslink mapping of the IncC2 monomeric homology models. Identified crosslinks mapped on the three IncC2 protein models., only Crosslinks with Lysine-Lysine (Cα-Cα) distances <34 Å were mapped on models. Red shows α-helix while yellow shows β-sheets, while loop regions are shown in grey. Dashed lines show crosslinks and numbers show lengths of respective crosslinks. **(a)** A total of 18 crosslinks mapped on the IncC2 refined structure modelled using PDB template 1WCV. **(b)** 16 crosslinks mapped to the IncC2 structure modelled using template PDB ID 2BEJ and **(c)** 15 crosslinks mapped on the structure modelled using the 5IHP template. The crosslink mapping and distance calculations were performed using Chimera ver 1.12 (Pettersen et al., 2004).

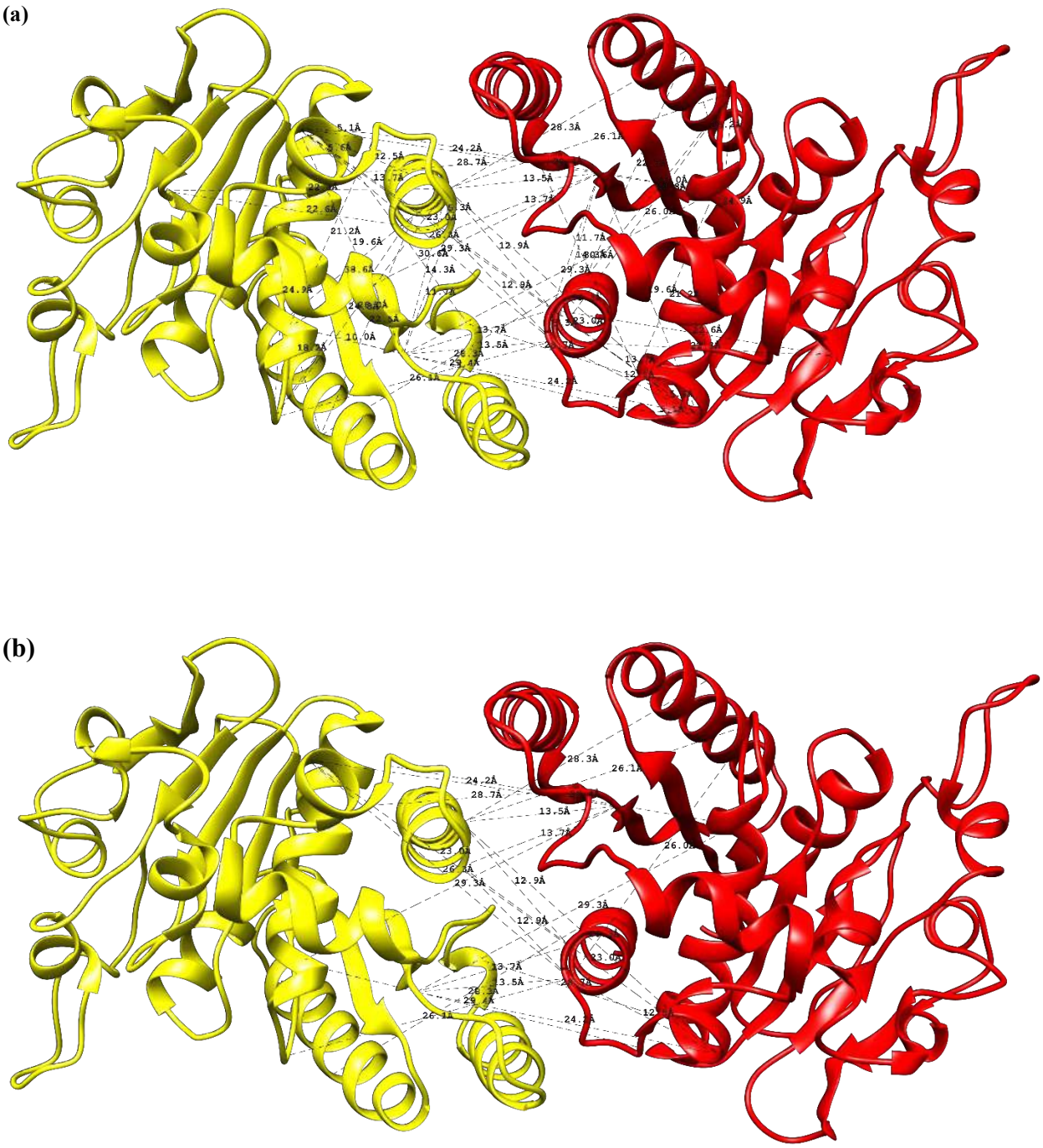


Figure 6.30 Crosslink mapping of the IncC2 dimer structure. The IncC2 homodimer model obtained using the SWISS-MODEL server and using PDB ID 1WCV as the template structure. The red and yellow colours show two different protein chains. Dashed lines show crosslinks and numbers show the lengths of respective crosslinks. The crosslinks are mapped to homodimer model. **(a)** The IncC2 homodimer model mapped with all inter- and intra-crosslinks possible. Two crosslinks were found to be >34 Å in length and not shown in the model. **(b)** Only inter-chain crosslinks on the IncC2 homodimer structure. Crosslink mapping and distance calculations were performed using Chimera ver 1.12 (Pettersen et al., 2004).

6.16. Discussion

The IncC and KorB proteins play a very important role in RK2 plasmid partitioning. These proteins interact with each other and DNA during partitioning process (Bingle et al., 2008). It has been shown previously that IncC and KorB help each other to bind to both specific and non-specific DNA (Rosche et al., 2000). In this chapter, both *in vivo* and *in vitro* experiments have been performed to examine the IncC and KorB interactions. The BACTH assay shows that IncC1 and IncC2 interact with each other as well as with KorB (Figures 6.9 and Figure 6.10). IncC1 and IncC2 interactions are strongest when the C-terminus of at least one partner protein is free. Looking at homology models of the IncC2 protein, the C- and N- termini are found close to each other and the C-terminus is involved in IncC2 dimer interface (Figure 6.30). Previously, the Glutaraldehyde chemical cross-linking of IncC1 and IncC2 show that two proteins interact with each other and that this interaction promotes IncC2 polymerization (Batt, 2008).

From the BACTH assays, The IncC1 and IncC2 proteins are found to interact with KorB via their C-terminal ends and similar results are observed when the KorB-IncC1 and KorB-IncC2 crosslinks are mapped (Figure 6.27). The N-terminus of IncC2 also shows crosslinking with KorB, that suggests that the two IncC2 terminals may form an interface for the IncC2-KorB interactions (Figure 6.27). Previously, it was observed with KorB Glutaraldehyde crosslinking that the C- terminal domain of KorB interact with the IncC proteins (Lukaszewicz et al., 2002, Batt, 2008). Initially, it was planned to probe IncC2-KorB interactions using different IncC2 deletion mutants, but the homology model of IncC2 shows that the protein has a single domain structure that may lose proper protein folding upon deletions.

The KorB-KorB interactions are stronger than all other interactions in the BACTH assay, in terms of Miller units. Weak interactions in BACTH may show transitory associations among the IncC-IncC and IncC-KorB proteins. In previous studies, a yeast two-hybrid assay also showed that the IncC1, IncC2 and KorB proteins interact with each other (Lukaszewicz et al., 2002). KorB-KorA interactions are expected as the two proteins have been reported to interact with each other (Bingle et al., 2008), but using BACTH, the KorA-KorB or KorB-KorA interactions are not observed (Figures 6.11, 6.12). There are many possibilities that can explain the absence of the KorA-KorB interactions in BACTH. Firstly, two proteins may require the presence of other partners for their interaction *i.e.* DNA or some other proteins. Secondly, the

size of the chimeric fusions may be a reason for the absence of interactions. KorA (10.5 kDa) is smaller than the fusion fragments used in BACTH (T18, T25, the catalytic domains of cyaA, are ~40 kDa). The T18 and T25 fragments may be covering the KorA-KorB interaction interface. Crosslink mapping for the KorB-KorA interactions shows that it is C-terminal of KorA that is predominantly involved in the KorB-KorA interactions (Figure 6.27), so fusions in which the large T18 or T25 fragments are at the C-terminal of KorA are not likely to interact with KorB. It has also been shown in molecular dynamics of KorA (Rajasekar et al., 2016) that the C- and N-terminals of KorA proteins come close to each other; that suggests that even N-terminal T18 or T25 fragment fusions of KorA may cover the KorB-KorA binding interface. Another reason may be a very small size and less flexible of the linker between KorA and fusion fragments. KorA-KorA interactions are observed as expected, as KorA forms predominantly homodimers (Figure 6.12) via C-terminal dimerization (Rajasekar et al., 2016). KorA is also found as a homodimer when cross-linked using BS3 cross-linker (Figure 6.16). When KorA and KorB are cross-linked together, the two proteins are found to interact with each other. Crosslink mapping shows that the C-terminal domain of KorA protein is involved in the KorA-KorB interactions (Figure 6.27), in agreement with Bingle et al., (2008). The IncC1 and IncC2 proteins predominantly showed dimer, trimer/tetramer protein bands upon crosslinking, while some higher molecular weight complexes are also observed (Figure 6.13 and Figure 6.14). Interestingly, while crosslinking IncC2 protein, a monomeric cross-linked protein band is observed that has higher SDS PAGE mobility than non-cross-linked monomeric IncC2 protein (Figure 6.13). That may suggest that, the IncC2 protein may adopt various (dynamic) conformations while in solution and it has been locked to one conformation after crosslinking. In the case of IncC1, no distinct monomeric protein band is observed (Figure 6.14) on cross-linking, which may be due to the presence of a flexible N-terminal region.

The cross-linked KorB protein is found to be predominantly dimeric, while some of the protein is a tetramer (Figure 6.16). KorB protein has been reported to exist as a dimer in solution (Leonard et al., 2004; Hyde et al., 2017) and also, upon glutaraldehyde crosslinking, the protein forms dimers and higher molecular weight complexes (Lukaszewicz et al., 2002, Batt, 2008). The P1 ParB proteins also form dimers in solution (Surtees and Funnell, 1999). During crosslinking the IncC1 and IncC2 proteins in presence of KorB, it is found that most of the KorB protein remained in dimeric form, but when DNA is added to crosslinking reaction, the protein

crosslinking pattern changes. The band showing KorB tetramer became stronger in presence of DNA. This may show that KorB is tetrameric when it interacts with DNA in presence of the IncC1 and IncC2 proteins (Figure 6.17a). In IncC1-IncC2 crosslinking, higher molecular weight complexes are observed on the SDS PAGE (6.17c). This may suggest IncC1 and IncC2 polymerization in the presence of each other. Interestingly, when ATP added to the IncC1 and IncC2 crosslinking reaction mixtures, the protein band pattern changed. The cross-linked protein remained predominantly dimeric, and higher molecular weight protein bands disappeared. This suggests that IncC1/IncC2 -ATP interactions may cause protein de-polymerization. In the presence of ATPγS, the change in the pattern of cross linking is not observed (Figure 6.18). This also suggests that ATP binding and hydrolysis may cause conformational changes in the proteins' structures. Interestingly, the IncC1 protein cross-linked in the presence of ATPγS and DNA, shows a similar crosslinking pattern as when cross linked with ATP alone. This shows that DNA may also play an important role in changing the IncC1 conformation in the presence of ATPγS (Figure 6.18), and, hence, possibly ATP.

The IncC2 homology models were obtained using Modeller ver 9.19 and refined by the PREFMD and ReFOLD server. The homology model, obtained using the template PDB ID 1WCV, is considered as the best model, as RMSD value for the model is 2.37 (Figure 6.23). The RMSD <3 is considered 1WCV is a good template to build a homology model for IncC2 (Rayan, 2009). The Ramachandran plot, as well as the residue energy plot for the IncC2 model, shows a reason structure prediction (Figures 6.23, 6.24), while Lysine-Lysine crosslinks are in agreement to the Ramachandran plot that shows a reliable IncC2 model prediction.

CHAPTER 7

Final Discussion

Chapter 7

Final Discussion

DNA partitioning involves two major proteins ParA and ParB. The two proteins interact with each other to make faithful plasmid partitioning possible. RK2 is a very low copy number plasmid and requires a partitioning system. In the RK2 plasmid, the two IncC proteins belong to the ParA family and act as ATPases, while KorB acts as a DNA-binding protein and interacts with RK2 DNA at, at least, 12 O_B/DNA binding sites (Williams et al., 1993). There are two ParA or IncC proteins that are expressed simultaneously in the cell by the RK2 plasmid; a longer IncC1 protein (364 a.a) and a shorter IncC2 protein (259 a.a). The two IncC proteins are expressed from alternative start codons from same operon, so that IncC2 is identical to the C-terminal 259 amino acids of IncC1 (Thomas and Smith, 1986; Batt et al., 2009). The difference between two proteins is the N-terminal region of 105 a.a (IncC NTD) that is only present in IncC1 (Batt et al., 2009). The IncC2 protein is similar to chromosomal ParA proteins that do not contain N-terminal extensions (Batt et al., 2009). To elucidate the role of IncC NTD in plasmid partitioning, the structure of the protein was determined using carbon-detected NMR in this thesis and the two IncC proteins were compared with respect to their structural and biophysical properties. It was found that the presence of the IncC NTD makes the two proteins quite distinct from each other in terms of DNA binding and biophysical properties. IncC1 can bind to DNA in absence of any nucleotide (ATP, ADP, ATPγ S) as in the case of IncC NTD, but DNA binding to IncC2 requires the presence of at least one of these nucleotides (ATP, ADP, ATPγ S). IncC1 was found to be a little more soluble in solution in comparison to IncC2, showed 1.5 folds higher ATPase activity in comparison with IncC2 and also showed a different crosslinking pattern, indicating differences in oligomerization between the two forms. The N-terminal region of IncC1 (IncC NTD) may play an important role in the RK2 plasmid partitioning.

IncC NTD was found to be largely disordered from the computational predictions based on its amino acid sequence, apart from a few short stretches of amino acids that were predicted to form α-helices or β strands respectively (Figure 3.6-Figure 3.8). The sequence of IncC NTD, like that of other IDRs present in large proteins, is not homologous or conserved in the other ParA proteins. The CD and Kratky plot data for IncC NTD confirmed the computational result, showing the protein structure is a random coil (Figure 3.18 and Figure 3.20). In contrast, the

computational secondary structure analysis for IncC2 protein, as well as CD studies and Kratky plots showed that it is a natively folded and predominantly helical protein (Figure 5.1, Figure 5.23 and Figure 5.28).

NMR studies also showed that the IncC NTD protein structure is a random coil. The TALOS+ and CS-Rosetta server predictions from the assigned NMR chemical shifts for the backbone resonances showed that IncC NTD may contain some small stretches of secondary structure. IncC NTD-DNA binding was observed by EMSA and NMR. The NMR studies of the DNA-IncC NTD binding show that the interaction interface consists of the regions predicted by CS-Rosetta to contain some secondary structure. The IncC NTD loop structure might undergo some transitory changes and adopt some secondary structure in the presence of a partner protein or DNA. There are some proteins which were predicted and found to be disordered in the absence of partners, but in the presence of a partner assume some secondary structure, like the p53 oligomerization region (Joerger et al., 2014) or the c-fos/c-jun/DNA junction complex (PDB ID 1FOS) (Vamosi et al., 2009). The HSQC experiment showed that the IncC NTD is disordered in the full length IncC1 protein.

Even as an intrinsically disorder region (IDR) in the IncC1 protein, the IncC NTD can play an important role by providing a highly positive flexible region interacting with DNA in different ways like sliding, hopping, and jumping along the non-specific DNA, while low DNA affinity allows the protein to associate, engage and dissociate easily. The N-terminal region of *H. Pylori* ParA was found to be involved in DNA binding as well as in interaction with ParB (Yamaoka, 2008).

Previously, it was found that the overexpression of IncC protein in the presence of KorB was toxic to the cell and that different host cells, *e.g. E. coli* and *P. putida*, express different numbers of the two IncC molecules during plasmid partitioning (Batt, 2008). The number of molecules of IncC and KorB proteins also varies between log and stationary phase (Batt, 2008). The different ratio of the two IncC proteins and the different numbers of IncC proteins present in different cells suggest that either IncC protein can allow partitioning alone. Different growth conditions and optimization of lysis buffers allowed the protein in enough soluble quantity to do biophysical characterization, but X-ray crystallography and NMR studies were not feasible for IncC1 and IncC2 proteins.

ParA proteins are ATPases, bind to DNA and its partner ParB. Studies of different ParA proteins show that nucleotides, especially ATP, play an important role in determining the function of ParA proteins. ParA proteins, being weak ATPases, do go through some conformational change (P-loop) while binding and hydrolysing ATP. ATP hydrolysis and its conversion to ADP might be involved in the regulation of ParA protein functions. ParA proteins are also found to form filaments in the presence of nucleotide and DNA (Ringgaard et al., 2009, Vecchiarelli et al., 2015). In the ParA proteins, nucleotide binding leads to dimerization (Leonard et al., 2005) and the dimers bind to DNA non-specifically (Lutkenhaus, 2007). The Type 1a ParA proteins have been reported to bind DNA non-specifically via the N-terminal region, while the Type Ib ParA proteins do not contain this region (Roberts et al., 2012). In this regards, IncC2 can be considered as a Type Ib, ParA ATPase containing a Walker-like motif, while IncC1 can be considered to be a Type Ia ParA

In the case of IncC proteins, the nucleotides seem to play important role in their function (Table 7.1). ATP, especially, was found to have an important effect on IncC activity. Kratky Plots from SAXS analysis show that the IncC proteins are more compact in the presence of ATP than in its absence or in the presence of ADP (Figure 5.23 and Figure 5.25). Even in the presence of ATPγ-S the effect was not seen, suggesting the ATP hydrolysis may provide the protein a stable transient conformation that allows protein to interact with DNA with better binding affinity. The IncC2 protein was found to interact with DNA only in the presence of nucleotides (Figure 7.1). The IncC protein crosslinking experiment, showed the IncC2 protein has dynamic conformation as the crosslinked monomer of IncC2 protein appeared to have different mobility in SDS PAGE from the unlinked protein (Figure 6.13). The experiment also suggests that in the presence of ATP, IncC2 may adopt a stable conformation that allows it to interact with DNA mostly in dimeric form. Although IncC1 did not show a different cross-linked monomer band from the unlinked monomer, in the presence of ATP the dimer and tetramer species were dominant (Figure 6.14).

SEC-MALLS analysis for IncC1 and IncC2 shows that the major population of the two proteins shifted from monomer form to dimers (Figure 5.19) in the presence of ATP. Previously, SEC analysis for a chromosomal partitioning protein, Soj, from *Bacillus subtilis* (Lee and Grossman, 2006) shows that it elutes as monomer in the absence of any nucleotide and in the presence of ADP; while a monomer-dimer equilibrium was observed in the presence of ATP

(Leonard et al., 2005). In contrast, the sedimentation results for Soj did not agree SEC MALLS results. The sedimentation results for the IncC proteins are in agreement with those for Soj protein. Both Soj and IncC showed predominantly monomeric species either in the presence or the absence of nucleotides (ATP, ADP) in sedimentation experiments (Leonard et al., 2005). It can be proposed that Soj and both IncC proteins form homodimers in the presence of ATP, but during the long sedimentation run (18 hr), the majority of added ATP is hydrolysed, to show only the ADP-bound proteins.

IncC NTD appears to make dimer more prevalent in IncC1. The compact cross-linked monomer band on SDS PAGE was only observed in case of IncC2 and not in IncC1. That shows the compactness or conformational change in the protein might be effected by the presence of IncC NTD. The IncC NTD protein binds weekly to DNA but IncC1 protein bind to the DNA with a comparable K_d as in the absence or presence of nucleotides (ATP, ADP). This also shows presence of IncC NTD in full length protein enhances the protein affinity for DNA. The presence of IncC NTD effects the IncC1 oligomerization, so as its disordered structure might affect the ATP hydrolysis and protein polymerization. Though both IncC1 and IncC2 proteins interact with KorB and DNA, ATP hydrolysis is two folds higher for IncC1 in the presence of both KorB and DNA (Table 5.6).

Table 7.1 Effect of nucleotide on IncC1 and IncC2 structure and function

Proteins	Experiments	No Nucleotides	ATP	ADP	ATPγS
IncC1	**DNA binding**	Shows DNA binding	Not required for binding, but enhanced binding affinity	Not required for binding, less binding affinity in comparison to in the presence of ATP	Not required for binding, almost the same binding affinity as in the presence of ATP
	AUC	Negligible dimerization or higher species	Negligible dimerization or higher species	N/D	N/D
	SAXS	Kratky Plot showed partially folded protein	Kratky Plot showed the protein was more folded than in the absence of any nucleotide	Similar conformation as observed in absence of any nucleotide	N/D
	Crosslinking	Dimers, tetramers and higher molecular weight species	More dimers than higher species	Similar pattern as in the absence of nucleotide	Similar pattern as in the absence of nucleotide
	SEC MALLS	Predominantly monomer, some dimer and higher molecular weight species	Conformation change and predominantly dimer	Similar conformation as observed in absence of any nucleotide	N/D
IncC2	**DNA binding**	Do not bind to DNA	Required for binding, high affinity binding	Required for binding, low affinity binding than ATP	Required for binding, allows DNA binding similar to ATP
	AUC	Predominantly monomer and some dimer	Predominantly monomer and some dimer	N/D	N/D
	SAXS	Kratky Plot showed partially folded protein	Kratky Plot showed more folded protein	Similar conformation as observed in absence of ADP	N/D
	Crosslinking	Compact monomer, dimer, tetramer and higher molecular weight species	More dimer and tetramer protein was observed than higher species	Similar pattern as in the absence of any nucleotide	Similar pattern as in the absence of nucleotide
	SEC MALLS	Largely monomer and some dimer	Conformation change and more dimerization	Same as in absence of any nucleotide	N/D

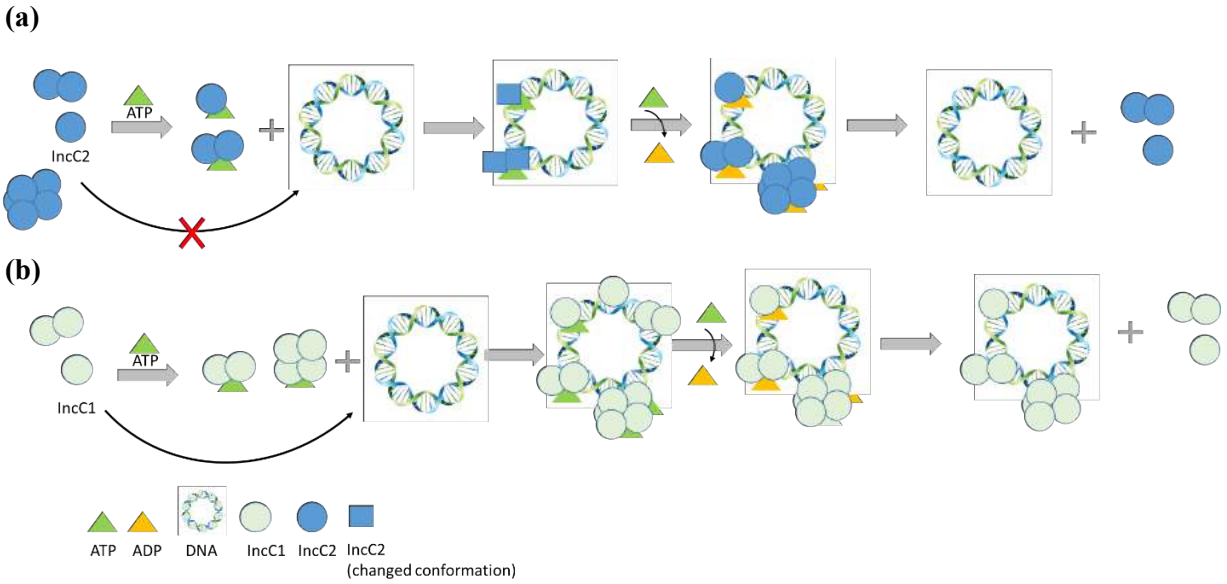


Figure 7.1 Illustrated models for effect of ATP/ADP on the DNA binding of the IncC proteins. (a) The IncC2 protein does not interact with DNA directly until it binds to ATP or ADP. Upon binding with ATP/ADP IncC2 adopts a different conformation that allows it to interact with DNA. IncC2-ADP still remains bound to DNA but the apo form of IncC2 does not bind to DNA **(b)** IncC1 protein interacts with DNA both in the presence and absence of ATP/ADP. The apo form of IncC1 still remains bound to DNA. In the presence of ATP, the IncC proteins form more dimers and tetramers than in the absence of ATP.

Homodimers for a protein MinD, from *Escherichia coli,* involved in regulation of the cytokinetic Z ring, have been previously observed in the presence of ATP (Hu and Lutkenhaus, 2003, Lutkenhaus and Sundaramoorthy, 2003). When MinD binds to ATP, it self-associates and the C-terminal amphipathic helix then interacts with the lipid bilayer of the cell (Lutkenhaus and Sucaramoorthy, 2003). The Soj protein does not contain such a hydrophobic helix that would let Soj to associate with membrane (Leonard et al., 2005). Instead, Soj binds to DNA and is thought to oscillate in the cell via nucleoid binding, with the involvement of ATP and Spo0J. The homology model for IncC2 protein shows that its core structure shares similarities with Soj and, the Soj-like proteins, but the C-terminal α-helix of IncC2 does not contain a hydrophobic face (Figure 7.2). Another significant different between Soj and IncC2 is that Soj does not bind to DNA in the presence of ADP and its DNA binding is regulated by ATP-ADP switch.

This suggests that the IncC1 and IncC2 proteins may behave like Soj and oscillate in the cell, while bound to the nucleoid. IncC1 protein does not require an interaction with ATP for this purpose. Thus, the regulatory ATP-ADP switch has less of an effect on IncC1 binding to DNA, that shows the protein may have some unusual function in the plasmid partitioning and follow a

slightly different partitioning model that is required to be elucidated by microscopic techniques. The DNA binding pattern of the IncC proteins is a little different from Soj protein, as in EMSA experiments Soj showed clear DNA shifts in the presence of ATP, while only very weak binding at high protein concentrations was observed when either ADP or no nucleotides were present. IncC2 binds to DNA in the presence of all nucleotides, ATP, ATPγS and ADP, but not in their absence, while IncC1 showed DNA binding even in the absence of nucleotides and this binding is due to presence of IncC NTD that also shows independent DNA binding. Even though the dissociation constant is very high for IncC NTD, the concentration of DNA in the cell non-specific site is very high hence proteins will be bound. A feature of the model is then it all occurs on the nucleoid or can occur on plasmid DNA but it is always bound for InCC1 but it probably will be bound because of nucleotide concentration in the cell.

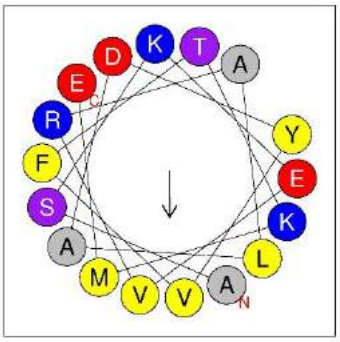


Figure 7.2 Prediction of amphipathic properties of IncC2 C-terminal helix. The C-terminal helix from IncC2 shows no hydrophobic face. An amphipathic helix prediction server (http://heliquest.ipmc.cnrs.fr/cgi-bin/ComputParams.py) was used to predict helical interface (amino acids are shown in different colours; yellow, grey-non-polar, pink-polar, red and blue-charged).

On the basis of these studies a partitioning model for IncC1 and IncC2 can be proposed (Figure 7.3). The IncC proteins (IncC1 and IncC2) hop slide, move and remain bound along the nucleotide as IncC-ATP, till find their partner KorB bound to plasmid DNA. IncC-KorB-DNA interactions stimulates the ATP hydrolysis activity of the IncC proteins. ATP hydrolysis leads to release of KorB-DNA complex that moves to the next available IncC-ATP by a directed movement guided by IncC-ATP depletion zone. Diffusion of KorB-DNA-partitioning complex along the IncC-ATP concentration gradient moves the complex towards the cell poles (Figure 7.3).

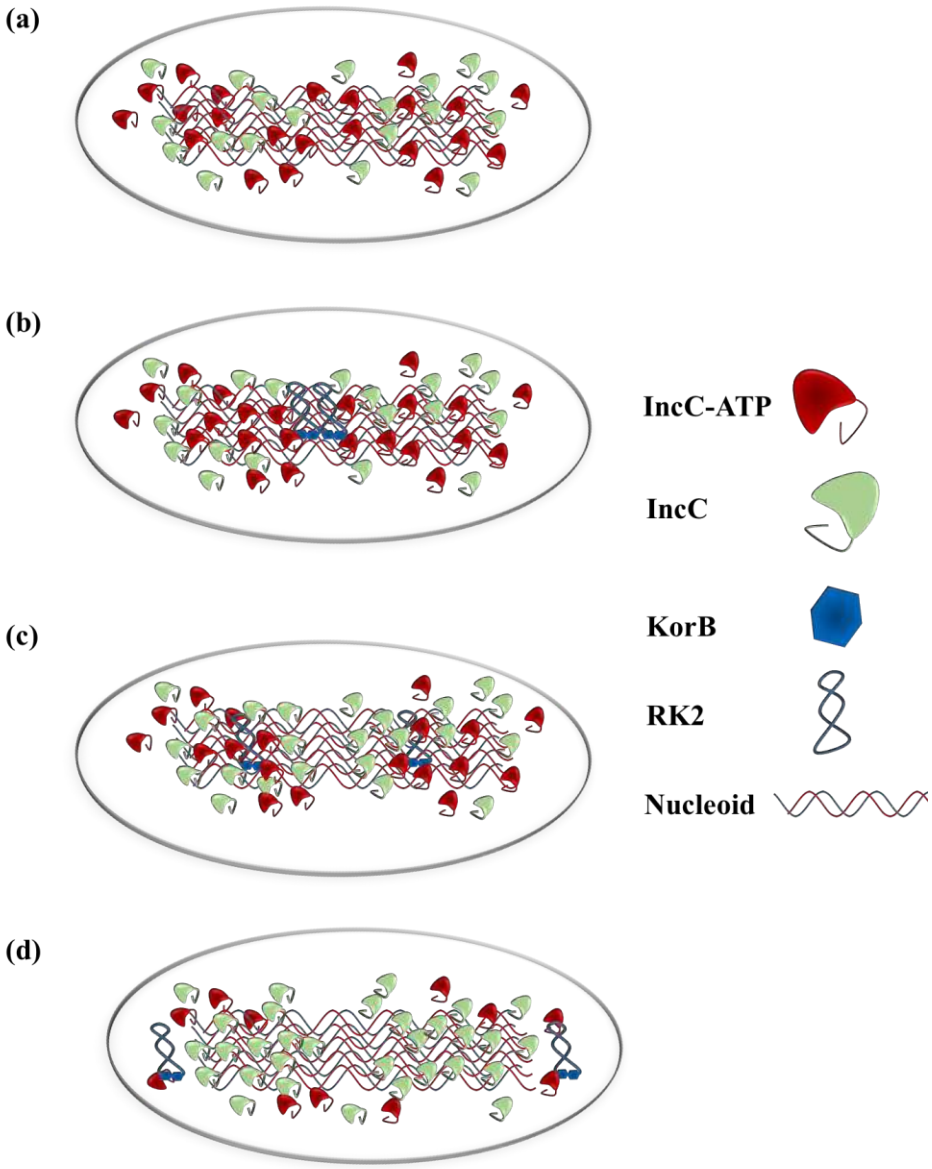


Figure 7.3 Model for plasmid RK2 partitioning based in properties of the IncC (a) IncC-ATP (shown in red) and IncC1-ADP forms clouds over the nucleoid (shown by blue and red lines) where both forms can bind to DNA. **(b)** KorB interacts with plasmid RK2 (plasmid shown as blue coil, KorB shown as blue diamonds) to form a partitioning complex. This KorB-plasmid partitioning complex interacts with the IncC and IncC-ATP proteins and stimulates ATP hydrolysis. After ATP hydrolysis, IncC-ADP (shown in green) is formed and released from the partitioning complex, while the KorB-DNA complex moves to the next available IncC-ATP by a directed movement guided by IncC-ATP depletion zone. **(c)** The KorB-DNA-partitioning complex diffuses towards the cell poles using the IncC-ATP concentration gradient as a driving force. IncC-ADP remain bound to nucleoid **(d)** By directional movement and diffusion, the RK2 plasmid reaches the cell poles.

It has been suggested that the IncC proteins are involved in interactions with the RK2 plasmid before segregation and allow clustering of the RK2 plasmid copies in the cell (Rosche et al., 2000). The clustering of plasmids may be enhanced by having two regions of IncC1 that can bind to DNA. The binding of the IncC NTD to the RK2 plasmid does not require ATP. The presence of an independent DNA binding domain in IncC1 which does not require nucleotides, may help in aggregating the RK2 plasmid copies in one cluster prior to segregation. IncC2 binding to DNA involves only one domain of the protein, may be used primarily in segregation *i.e.* in moving the plasmids apart.

Homology modelling of the IncC2 protein, using various templates, showed the presence of eight helical segments with five beta strands. The IncC and KorB proteins crosslinked individually and or in same reaction mixture were subjected to Mass spectrometry analysis. Identification and mapping of the cross-linked peptides showed the IncC-KorB interact with each other, confirming the results of the BACTH experiment. The crosslinking experiment further showed that the C-terminal of the IncC protein is involved in interactions with the C-terminal of the KorB protein. Some cross-links were also found from the C-terminal of IncC to the N-terminal of KorB. The KorB-KorB self-crosslinks show that the N-terminal of KorB is not a region hanging away from the main core, but that it may interact with the central region of KorB. In that way, the N-terminal of KorB, as well as the C-terminal of KorB, might be involved in interaction with IncC protein. This agrees with experiments on other systems, in which it has been shown that the N-terminal region of ParB proteins stimulate the activity of the ParA ATPase, resulting in de-polymerization of ParA (Leonard et al., 2005, Ringgaard et al., 2009, Barilla et al., 2007).

ParA proteins have been found to interact with N-terminal region of ParB proteins (Surtees and Funnell, 1999, Figge et al., 2003, Ravin et al., 2003) and the ATPase activity of some ParA proteins has been reported to be enhanced by interacting with this region (Leonard et al., 2005). Previously, it was reported that the KorB protein interface (174 Ile to 218 Thr) is involved in interactions with IncC (Lukaszewicz et al., 2002). The observations made in this study using chemical crosslinking show that not only this region but also the C-terminal and extreme N-terminal region of KorB is involved in interaction with IncC protein (Figure 6.27). This might show that these three regions form a binding interface and the flexible linkers present in KorB allow these regions to come in close proximity to each other. In case of IncC2, it was

observed that C-terminal deletions of KorB enhanced the ATPase activity of IncC2 to the same extent as the wild type KorB protein, while the linker deletion mutants of KorB protein showed less effect on the IncC2 ATPase activity than wildtype KorB protein (Table 5.6). This suggests that the flexible linker region may bring different domains of the KorB protein close to each other to form an interaction interface for the IncC2 protein.

The key finding of this research is the determination of the structure of N-terminal domain of IncC1 and its plausible role in the plasmid partitioning. This region provides IncC1 with different properties from IncC2 protein. As said before, only one IncC protein was found to be sufficient for plasmid partitioning, and different hosts have different IncC1:IncC2 expression levels. This work does begin to explain why these two separate proteins with same function are present in the plasmid RK2. Either protein may perform better in a particular host depending upon its interactions with other partners in the particular cell environment.

Future work should look at ways to confirm the IncC NTD stretches involved in DNA binding by making deletion mutants and point mutants in the protein. Both IncC1 and IncC2 are strongly affected by ATP hydrolysis. Point mutations in the ATP-binding site along with studies of changes in the protein conformation on ATP binding may help in understanding how the ATP switch works and causes the IncC proteins to be in a more folded and dimeric form. The homology model of the IncC2 protein also provides information about the dimer interaction interface. Point mutations in these specific regions may help in understanding the dimer interactions.

In this thesis, the cross-linked domains of KorB have been shown and crosslinks have been proposed, that may lead to modelling of complete KorB structure. The crystal structure for the KorB DNA binding domain (DBD) and the C-terminal dimerization regions have been reported already, while the NMR structure of the N-terminal of KorB has been proposed recently by Dr Hyde's Lab. An integrative model for the KorB protein can be develop using the distance restraints from the cross-linked peptides. The integrated KorB model can be docked against KorA and IncC proteins individually and as a complex in the presence or absence of DNA to get information about the interactions of these proteins. It would be beneficial to know that whether the IncC2 and KorA protein interact with KorB in the close proximity or not.

While this study contributes to the structure and function of IncC (ParA) protein, there is still a lot more to be discovered about active plasmid partitioning as well as bacterial

chromosomal partitioning systems. The ParA protein along with ParB, may interact with many other partners making the partitioning system more complicated to understand. Future work along the same lines would be beneficial to identify new targets for antibiotics and, development of drugs against segregation proteins that would help in blocking bacterial proliferation. Stopping plasmid segregation would result in the loss of antibiotic resistance plasmids, while stopping chromosomal segregation in pathogenic bacteria.

CHAPTER 8

References

REFERENCES

ABELES, A. L., FRIEDMAN, S. A. & AUSTIN, S. J. 1985. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. J Mol Biol, 185, 261-72.

ABRAHAM, S. J., HOHEISEL, S. & GAPONENKO, V. 2008. Detection of protein-ligand interactions by NMR using reductive methylation of lysine residues. J Biomol NMR, 42, 143-48.

ADACHI, S., HORI, K. & HIRAGA, S. 2006. Subcellular positioning of F plasmid mediated by dynamic localization of SopA and SopB. J Mol Biol, 356, 850-63.

ADAMCZAK, R., POROLLO, A. & MELLER, J. 2004. Accurate prediction of solvent accessibility using neural networks-based regression. Proteins, 56, 753-67.

ADAMCZYK, M. & JAGURA-BURDZY, G. 2003. Spread and survival of promiscuous IncP-1 plasmids. Acta Biochim Pol, 50, 425-53.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. J Mol Biol, 215, 403-10

AMOUTZIAS, G. D., HE, Y., LILLEY, K. S., VAN DE PEER, Y. & OLIVER, S. G. 2012. Evaluation and properties of the budding yeast phosphoproteome. Mol Cell Proteomics, 11, M111 009555.

ANAND, S. P., AKHTAR, P., TINSLEY, E., WATKINS, S. C. & KHAN, S. A. 2008. GTP-dependent polymerization of the tubulin-like RepX replication protein encoded by the pXO1 plasmid of Bacillus anthracis. Mol Microbiol, 67, 881-90.

ANDRADE, M. A., CHACON, P., MERELO, J. J. & MORAN, F. 1993. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng, 6, 383-90.

ARAI, H., WATANABE, S., KIGAWA, T. & YAMAMURA, M. 2009. A new modeling method in feature construction for the HSQC spectra screening problem. Bioinformatics, 25, 948-53.

ARLT, C., GÖTZE, M., IHLING, C. H., HAGE, C., SCHÄFER, M. & SINZ, A. 2016. Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker. Ana Chem, 88, 7930-37.

AUSTIN, S. & ABELES, A. 1983. Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. J Mol Biol, 169, 353-72.

AUTRET, S., NAIR, R. & ERRINGTON, J. 2001. Genetic analysis of the chromosome segregation protein Spo0J of *Bacillus subtilis*: Evidence for separate domains involved in DNA binding and interactions with Soj protein. Mol Microbiol, 41, 743-55.

AYLETT, C. H. & LOWE, J. 2012. Superstructure of the centromeric complex of TubZRC plasmid partitioning systems. Proc Natl Acad Sci U S A, 109, 16522-27.

BABINI, E., BERTINI, I., CAPOZZI, F., DEL BIANCO, C., HOLLENDER, D., KISS, T., LUCHINAT, C. & QUATTRONE, A. 2004. Solution structure of human beta-parvalbumin and structural comparison with its paralog alpha-parvalbumin and with their rat orthologs. Biochemistry, 43, 16076-85.

BABU, M. M. 2016. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans, 44, 1185-1200.

BABU, M. M., VAN DER LEE, R., DE GROOT, N. S. & GSPONER, J. 2011. Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol, 21, 432-40.

BAKER, J. M., HUDSON, R. P., KANELIS, V., CHOY, W. Y., THIBODEAU, P. H., THOMAS, P. J. & FORMAN-KAY, J. D. 2007. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol, 14, 738-45.

BARILLA, D. & HAYES, F. 2003. Architecture of the ParF*ParG protein complex involved in prokaryotic DNA segregation. Mol Microbiol, 49, 487-99.

BARILLA, D., CARMELO, E. & HAYES, F. 2007. The tail of the ParG DNA segregation protein remodels ParF polymers and enhances ATP hydrolysis via an arginine finger-like motif. Proc Natl Acad Sci U S A, 104, 1811-16.

BARILLA, D., ROSENBERG, M. F., NOBBMANN, U. & HAYES, F. 2005. Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF. EMBO J, 24, 1453-64.

BARTOSIK, A. A., MIERZEJEWSKA, J., THOMAS, C. M. & JAGURA-BURDZY, G. 2009. ParB deficiency in *Pseudomonas aeruginosa* destabilizes the partner protein ParA and affects a variety of physiological parameters. Microbiology, 155, 1080-92.

BATT, S. M., BINGLE, L. E., DAFFORN, T. R. & THOMAS, C. M. 2009. Bacterial genome partitioning: N-terminal domain of IncC protein encoded by broad-host-range plasmid RK2 modulates oligomerisation and DNA binding. J Mol Biol, 385, 1361-74.

BATT, S. M., 2008. The active partitioning system of plasmid RK2, PhD Thesis, University of Birmingham, UK

BATTESTI, A. & BOUVERET, E. 2012. The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli.* Methods, 58**,** 325-34.

BAXTER, J. C. & FUNNELL, B. E. 2014. Plasmid partition mechanisms. Microbiol Spectr, 2.

BECHHOFER, D. H. & FIGURSKI, D. H. 1983. Map location and nucleotide sequence of korA, a key regulatory gene of promiscuous plasmid RK2. Nucleic Acids Res, 11, 7453-69.

BERJANSKII, M., LIANG, Y., ZHOU, J., TANG, P., STOTHARD, P., ZHOU, Y., CRUZ, J., MACDONELL, C., LIN, G., LU, P. & WISHART, D. S. 2010. PROSESS: a protein structure evaluation suite and server. Nucleic Acids Res, 38, W633-40.

BERMEL, W., BERTINI, I., DUMA, L., FELLI, I. C., EMSLEY, L., PIERATTELLI, R. & VASOS, P. R. 2005. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew Chem Int Ed Engl, 44, 3089-92.

BERMEL, W., BERTINI, I., FELLI, I. C. & PIERATTELLI, R. 2009. Speeding up (13)C direct detection biomolecular NMR spectroscopy. J Am Chem Soc, 131, 15339-45.

BERMEL, W., BERTINI, I., FELLI, I. C., LEE, Y. M., LUCHINAT, C. & PIERATTELLI, R. 2006. Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc, 128, 3918-19.

BERMEL, W., BERTINI, I., FELLI, I. C., MATZAPETAKIS, M., PIERATTELLI, R., THEIL, E. C. & TURANO, P. 2007. A method for C(alpha) direct-detection in proton less NMR. J Magn Reson, 188, 301-10.

BERMEL, W., BRUIX, M., FELLI, I. C., KUMAR, M. V. V., PIERATTELLI, R. & SERRANO, S. 2013. Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR, 55, 231-37.

BERMEL, W., FELLI, I. C., KÜMMERLE, R. & PIERATTELLI, R. 2008. ^{13}C Direct-detection biomolecular NMR. Concepts in Magnetic Resonance Part A, 32A, 183-200.

BERNADO, P. & SVERGUN, D. I. 2012. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol Biosyst, 8, 151-67.

BERTINI, I., DUMA, L., FELLI, I. C., FEY, M., LUCHINAT, C., PIERATTELLI, R. & VASOS, P. R. 2004a. A Heteronuclear Direct-Detection NMR Spectroscopy Experiment for Protein-Backbone Assignment. Angewandte Chemie International Edition, 43, 2257-59.

BERTINI, I., FELLI, I. C., KUMMERLE, R., MOSKAU, D. & PIERATTELLI, R. 2004b. ^{13}C-^{13}C NOESY: an attractive alternative for studying large macromolecules. J Am Chem Soc, 126, 464-5.

BHARAT, T. A., MURSHUDOV, G. N., SACHSE, C. & LOWE, J. 2015. Structures of actin-like ParM filaments show architecture of plasmid-segregating spindles. Nature, 523, 106-10.

BIASINI, M., BIENERT, S., WATERHOUSE, A., ARNOLD, K., STUDER, G., SCHMIDT, T., KIEFER, F., GALLO CASSARINO, T., BERTONI, M., BORDOLI, L. & SCHWEDE, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res, 42, W252-58.

BIGNELL, C. & THOMAS, C. M. 2001. The bacterial ParA-ParB partitioning proteins. J Biotechnol, 91, 1-34.

BIGNELL, C. R., HAINES, A. S., KHARE, D. & THOMAS, C. M. 1999. Effect of growth rate and incC mutation on symmetric plasmid distribution by the IncP-1 partitioning apparatus. Mol Microbiol, 34, 205-16.

BINGLE, L. E. H., RAJASEKAR, K. V., MUNTAHA, S. T., NADELLA, V., HYDE, E. I. & THOMAS, C. M. 2008. A single aromatic residue in transcriptional repressor protein KorA is critical for cooperativity with its co-regulator KorB. Mol Microbiol, 70, 1502-14.

BOIVIN, S., KOZAK, S. & MEIJERS, R. 2013. Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif, 91**,** 192-206

BIRNBOIM, H. C. & DOLY, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res, 7, 1513-23.

BORG, M., MITTAG, T., PAWSON, T., TYERS, M., FORMAN-KAY, J. D. & CHAN, H. S. 2007. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. Proc Natl Acad Sci U S A, 104, 9650-55.

BORK, P., SANDER, C. & VALENCIA, A. 1992. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. Proc Natl Acad Sci U S A, 89, 7290-94.

BOUET, J. Y. & FUNNELL, B. E. 1999. P1 ParA interacts with the P1 partition complex at parS and an ATP-ADP switch controls ParA activities. EMBO J, 18, 1415-24.

BOUET, J. Y., AH-SENG, Y., BENMERADI, N. & LANE, D. 2007. Polymerization of SopA partition ATPase: regulation by DNA binding and SopB. Mol Microbiol, 63, 468-81.

BRACKEN, C., PALMER, A. G., 3RD & CAVANAGH, J. 1997. (H)N(COCA)NH and HN(COCA)NH experiments for 1H-15N backbone assignments in 13C/15N-labeled proteins. J Biomol NMR, 9, 94-100.

BROOKS, A. C. & HWANG, L. C. 2017. Reconstitutions of plasmid partition systems and their mechanisms. Plasmid, 91, 37-41.

BROWN, C. J., TAKAYAMA, S., CAMPEN, A. M., VISE, P., MARSHALL, T. W., OLDFIELD, C. J., WILLIAMS, C. J. & DUNKER, A. K. 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol, 55, 104-10.

BURKARDT, H. J., RIESS, G. & PUHLER, A. 1979. Relationship of group P1 plasmids revealed by hetero duplex experiments: RP1, RP4, R68 and RK2 are identical. J Gen Microbiol, 114, 341-48.

BUSKE, P. J. & LEVIN, P. A. 2013. A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation *in vivo*. Mol Microbiol, 89, 249-63.

BUTT, T. R., EDAVETTAL, S. C., HALL, J. P. & MATTERN, M. R. 2005. SUMO fusion technology for difficult-to-express proteins. Protein Expr Purif, 43, 1-9.

BYSTROFF, C. & SHAO, Y. 2002. Fully automated ab initio protein structure prediction using I-SITES, HMMSTR and ROSETTA. Bioinformatics, 18 Suppl 1, S54-61.

CAMPBELL, C. S. & MULLINS, R. D. 2007. In vivo visualization of type II plasmid segregation: bacterial actin filaments pushing plasmids. J Cell Biol, 179, 1059-66.

CARMELO, E., BARILLA, D., GOLOVANOV, A. P., LIAN, L. Y., DEROME, A. & HAYES, F. 2005. The unstructured N-terminal tail of ParG modulates assembly of a quaternary nucleoprotein complex in transcription repression. J Biol Chem, 280, 28683-91.

CHANG, S., CHANG, S. Y. & GRAY, O. 1987. Structural and genetic analyses of a par locus that regulates plasmid partition in *Bacillus subtilis*. J Bacteriol, 169, 3952-62.

CHEN, B. W., LIN, M. H., CHU, C. H., HSU, C. E. & SUN, Y. J. 2015. Insights into ParB spreading from the complex structure of Spo0J and parS. Proc Natl Acad Sci U S A, 112, 6613-18.

CHEN, J. W., ROMERO, P., UVERSKY, V. N. & DUNKER, A. K. 2006. Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. J Proteome Res, 5, 888-98.

CHOURA, M. & REBAI, A. 2017. Exploring disorder in the human charged biased proteins. J Recept Signal Transduct Res, 37, 386-390.

CHURION, K. A. & BONDOS, S. E. 2012. Identifying solubility-promoting buffers for intrinsically disordered proteins prior to purification. Methods Mol Biol, 896, 415-27.

COLLINS, M. O., YU, L., CAMPUZANO, I., GRANT, S. G. & CHOUDHARY, J. S. 2008. Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. Mol Cell Proteomics, 7, 1331-48.

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

COSTA, S., ALMEIDA, A., CASTRO, A. & DOMINGUES, L. 2014. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: The novel Fh8 system. Front Microbiol, 5, 63.

COUSIN, C., DEROUICHE, A., SHI, L., PAGOT, Y., PONCET, S. & MIJAKOVIC, I. 2013. Protein-serine/threonine/tyrosine kinases in bacterial signaling and regulation. FEMS Microbiol Lett, 346, 11-9.

DAM, M. & GERDES, K. 1994. Partitioning of plasmid R1. Ten direct repeats flanking the parA promoter constitute a centromere-like partition site parC, that expresses incompatibility. J Mol Biol, 236, 1289-98.

DAVEY, M. J. & FUNNELL, B. E. 1994. The P1 plasmid partition protein ParA. A role for ATP in site-specific DNA binding. J Biol Chem, 269, 29908-13.

DAVIS, G. D., ELISEE, C., NEWHAM, D. M. & HARRISON, R. G. 1999. New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnol Bioeng, 65, 382-88.

DAVIS, M. A., MARTIN, K. A. & AUSTIN, S. J. 1992. Biochemical activities of the parA partition protein of the P1 plasmid. Mol Microbiol, 6, 1141-47.

DE MARCO, A., VIGH, L., DIAMANT, S. & GOLOUBINOFF, P. 2005. Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. Cell Stress Chaperones, 10, 329-39.

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

DELBRUCK, H., ZIEGELIN, G., LANKA, E. & HEINEMANN, U. 2002. An Src homology 3-like domain is responsible for dimerization of the repressor protein KorB encoded by the promiscuous IncP plasmid RP4. J Biol Chem, 277, 4191-98.

DELEAGE, G. & GEOURJON, C. 1993. An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum. Comput Appl Biosci, 9, 197-9.

DENG, X., EICKHOLT, J. & CHENG, J. 2012. A comprehensive overview of computational protein disorder prediction methods. Mol Biosyst, 8, 114-21.

DODGE, C., SCHNEIDER, R. & SANDER, C. 1998. The HSSP database of protein structure-sequence alignments and family profiles. Nucleic Acids Res*,* 26**,** 313-15.

DOGAN, J., MU, X., ENGSTROM, A. & JEMTH, P. 2013. The transition state structure for coupled binding and folding of disordered protein domains. Sci Rep, 3, 2076.

DOSZTANYI, Z., CSIZMOK, V., TOMPA, P. & SIMON, I. 2005. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics, 21, 3433-34.

DOSZTANYI, Z., MESZAROS, B. & SIMON, I. 2009. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics, 25, 2745-46.

DRAPER, G. C. & GOBER, J. W. 2002. Bacterial chromosome segregation. Annu Rev Microbiol, 56, 567-97.

DROZDETSKIY, A., COLE, C., PROCTER, J. & BARTON, G. J. 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res, 43, W389-94.

DUMA, L., HEDIGER, S., LESAGE, A. & EMSLEY, L. 2003. Spin-state selection in solid-state NMR. J Magn Reson, 164, 187-95.

DUNKER, A. K., LAWSON, J. D., BROWN, C. J., WILLIAMS, R. M., ROMERO, P., OH, J. S., OLDFIELD, C. J., CAMPEN, A. M., RATLIFF, C. M., HIPPS, K. W., AUSIO, J., NISSEN, M. S., REEVES, R., KANG, C., KISSINGER, C. R., BAILEY, R. W., GRISWOLD, M. D., CHIU, W., GARNER, E. C. & OBRADOVIC, Z. 2001. Intrinsically disordered protein. J Mol Graph Model, 19, 26-59.

DUNKER, A. K., OLDFIELD, C. J., MENG, J., ROMERO, P., YANG, J. Y., CHEN, J. W., VACIC, V., OBRADOVIC, Z. & UVERSKY, V. N. 2008. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics, 9 Suppl 2**,** S1.

DYSON, H. J. & WRIGHT, P. E. 2002. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol, 12, 54-60.

DYSON, H. J. & WRIGHT, P. E. 2005. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol, 6, 197-208.

DZIEKANSKI, P., GRUDZIAZ, K., JARVOLL, P., KOZMINSKI, W. & ZAWADZKA-KAZIMIERCZUK, A. 2015. (13) C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing. J Biomol NMR, 62**,** 179-90.

EASTER, J., JR. & GOBER, J. W. 2002. ParB-stimulated nucleotide exchange regulates a switch in functionally distinct ParA activities. Mol Cell, 10, 427-34.

EBERSBACH, G. & GERDES, K. 2005. Plasmid segregation mechanisms. Annu Rev Genet, 39, 453-79.

EISENBERG, D., LUETHY, R. & U. BOWIE, J. 1997. VERIFY3D: Assessment of Protein Models with Three-Dimensional Profiles. Meth Enzymol, 277, 396-404.

ERDMANN, N., PETROFF, T. & FUNNELL, B. E. 1999. Intracellular localization of P1 ParB protein depends on ParA and parS. Proc Natl Acad Sci U S A, 96, 14905-10.

ERICKSON, H. P. 2009. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biol Proced Online, 11, 32-51.

ESPOSITO, D. & CHATTERJEE, D. K. 2006. Enhancement of soluble protein expression through the use of fusion tags. Curr Opinion Biotechnol, 17, 353-58.

ESWAR, N., WEBB, B., MARTI-RENOM, M. A., MADHUSUDHAN, M. S., ERAMIAN, D., SHEN, M.-Y., PIEPER, U. & SALI, A. 2006. Comparative Protein Structure Modeling Using Modeller. Current protocols in bioinformatics/editorial board, Andreas D. Baxevanis ... [et al.], 05, Unit-5.6.

FELLI, I. C. & PIERATTELLI, R. 2012. Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. IUBMB Life, 64**,** 473-81.

FERNANDEZ-LOPEZ, R., GARCILLAN-BARCIA, M. P., REVILLA, C., LAZARO, M., VIELVA, L. & DE LA CRUZ, F. 2006. Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. FEMS Microbiol Rev, 30, 942-66.

FIGGE, R. M., EASTER, J. & GOBER, J. W. 2003. Productive interaction between the chromosome partitioning proteins, ParA and ParB, is required for the progression of the cell cycle in *Caulobacter crescentus*. Mol Microbiol, 47, 1225-37.

FIGURSKI, D. H. & HELINSKI, D. R. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc Natl Acad Sci U S A, 76, 1648-52.

FINK, G. & LOWE, J. 2015. Reconstitution of a prokaryotic minus end-tracking system using TubRC centromeric complexes and tubulin-like protein TubZ filaments. Proc Natl Acad Sci U S A, 112, E1845-50.

FISHER, G. L., PASTRANA, C. L., HIGMAN, V. A., KOH, A., TAYLOR, J. A., BUTTERER, A., CRAGGS, T., SOBOTT, F., MURRAY, H., CRUMP, M. P., MORENO-HERRERO, F. & DILLINGHAM, M. S. 2017. The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. Elife, 6.

FOGEL, M. A. & WALDOR, M. K. 2006. A dynamic, mitotic-like mechanism for bacterial chromosome segregation. Genes Dev, 20, 3269-82.

FORMAN-KAY, J. D. & MITTAG, T. 2013. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. Structure, 21, 1492-9.

FRASER, C. M., CASJENS, S., HUANG, W. M., SUTTON, G. G., CLAYTON, R., LATHIGRA, R., WHITE, O., KETCHUM, K. A., DODSON, R., HICKEY, E. K., GWINN, M., DOUGHERTY, B., TOMB, J. F., FLEISCHMANN, R. D., RICHARDSON, D., PETERSON, J., KERLAVAGE, A. R., QUACKENBUSH, J., SALZBERG, S., HANSON, M., VAN VUGT, R., PALMER, N., ADAMS, M. D., GOCAYNE, J., WEIDMAN, J., UTTERBACK, T., WATTHEY, L., MCDONALD, L., ARTIACH, P., BOWMAN, C., GARLAND, S., FUJI, C., COTTON, M. D., HORST, K., ROBERTS, K., HATCH, B., SMITH, H. O. & VENTER, J. C. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature, 390, 580-86.

FRIEDMAN, S. A. & AUSTIN, S. J. 1988. The P1 plasmid-partition system synthesizes two essential proteins from an autoregulated operon. Plasmid, 19, 103-12.

FRITZSCHE, R., IHLING, C. H., GOTZE, M. & SINZ, A. 2012. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. Rapid Commun Mass Spectrom, 26, 653-58

FUNNELL, B. E. 1988. Mini-P1 plasmid partitioning: excess ParB protein destabilizes plasmids containing the centromere parS. J Bacteriol, 170, 954-60.

FUNNELL, B. E. 1991. The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology. J Biol Chem, 266, 14328-37.

FUNNELL, B. E. 2016. ParB Partition Proteins: Complex Formation and Spreading at Bacterial and Plasmid Centromeres. Front Mol Biosci, 3, 44.

FUNNELL, B. E. & GAGNIER, L. 1994. P1 plasmid partition: binding of P1 ParB protein and *Escherichia coli* integration host factor to altered parS sites. Biochimie, 76, 924-32.

FUXREITER, M. & TOMPA, P. 2012. Fuzzy complexes: a more stochastic view of protein function. Adv Exp Med Biol, 725, 1-14.

FUXREITER, M., SIMON, I. & BONDOS, S. 2011. Dynamic protein-DNA recognition: beyond what can be seen. Trends Biochem Sci, 36, 415-23.

FUXREITER, M., SIMON, I., FRIEDRICH, P. & TOMPA, P. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol, 338, 1015-26.

GALKIN, V. E., ORLOVA, A., RIVERA, C., MULLINS, R. D. & EGELMAN, E. H. 2009. Structural Polymorphism of the ParM Filament and Dynamic Instability. Structure, 17, 1253-64.

GAO, J. & XU, D. 2012. Correlation between posttranslational modification and intrinsic disorder in protein. Pac Symp Biocomput, 94-103.

GARNER, E. C., CAMPBELL, C. S. & MULLINS, R. D. 2004. Dynamic instability in a DNA-segregating prokaryotic actin homolog. Science, 306, 1021-25.

GARNER, E. C., CAMPBELL, C. S., WEIBEL, D. B. & MULLINS, R. D. 2007. Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. Science, 315, 1270-74.

GAYATHRI, P., FUJII, T., MOLLER-JENSEN, J., VAN DEN ENT, F., NAMBA, K. & LOWE, J. 2012. A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. Science, 338, 1334-37.

GAYATHRI, P., FUJII, T., NAMBA, K. & LOWE, J. 2013. Structure of the ParM filament at 8.5A resolution. J Struct Biol, 184, 33-42.

GEOURJON, C. & DELEAGE, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci, 11, 681-84.

GERDES, K., HOWARD, M. & SZARDENINGS, F. 2010. Pushing and pulling in prokaryotic DNA segregation. Cell, 141, 927-42.

GERDES, K., LARSEN, J. E. & MOLIN, S. 1985. Stable inheritance of plasmid R1 requires two different loci. J Bacteriol, 161, 292-98.

GERDES, K., MOLLER-JENSEN, J. & BUGGE JENSEN, R. 2000. Plasmid and chromosome partitioning: surprises from phylogeny. Mol Microbiol, 37, 455-66.

GLASER, P., SHARPE, M. E., RAETHER, B., PEREGO, M., OHLSEN, K. & ERRINGTON, J. 1997. Dynamic, mitotic-like behavior of a bacterial protein required for accurate chromosome partitioning. Genes Dev, 11, 1160-68.

GNAD, F., DE GODOY, L. M., COX, J., NEUHAUSER, N., REN, S., OLSEN, J. V. & MANN, M. 2009. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. Proteomics, 9, 4642-52.

GOLOVANOV, A. P., BARILLA, D., GOLOVANOVA, M., HAYES, F. & LIAN, L. Y. 2003. ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. Mol Microbiol, 50, 1141-53.

GORDON, G. S., SITNIKOV, D., WEBB, C. D., TELEMAN, A., STRAIGHT, A., LOSICK, R., MURRAY, A. W. & WRIGHT, A. 1997. Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. Cell, 90, 1113-21.

GORDON, G. S. & WRIGHT, A. 2000. DNA segregation in bacteria. Annu Rev Microbiol, 54, 681-708.

GORDON, S., RECH, J., LANE, D. & WRIGHT, A. 2004. Kinetics of plasmid segregation in *Escherichia coli.* Mol Microbiol, 51, 461-69.

GÖTZE, M., PETTELKAU, J., SCHAKS, S., BOSSE, K., IHLING, C. H., KRAUTH, F., FRITZSCHE, R., KÜHN, U. & SINZ, A. 2012. StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. J American Soc Mass Spectrom, 23, 76-87.

GRANT, S. G., JESSEE, J., BLOOM, F. R. & HANAHAN, D. 1990. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. Proc Natl Acad Sci U S A, 87, 4645-9.

GRAY, F. L., MURAI, M. J., GREMBECKA, J. & CIERPICKI, T. 2012. Detection of disordered regions in globular proteins using (1)(3)C-detected NMR. Protein Sci, 21**,** 1954-60.

GREENFIELD, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc, 1**,** 2876-90.

GRIFFITH, K. L. & WOLF, R. E., JR. 2002. Measuring beta-galactosidase activity in bacteria: cell growth, permeabilization, and enzyme assays in 96-well arrays. Biochem Biophys Res Commun, 290, 397-402.

GRZESIEK, S. & BAX, A. 1992. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. J Magnetic Resonance (1969), 96, 432-40.

GSPONER, J., FUTSCHIK, M. E., TEICHMANN, S. A. & BABU, M. M. 2008. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science, 322, 1365-68.

GUYNET, C. & DE LA CRUZ, F. 2011. Plasmid segregation without partition. Mob Genet Elements, 1, 236-41.

GUZMAN, L. M., BELIN, D., CARSON, M. J. & BECKWITH, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol, 177, 4121-30.

HABCHI, J., TOMPA, P., LONGHI, S. & UVERSKY, V. N. 2014. Introducing protein intrinsic disorder. Chem Rev, 114, 6561-88.

HAO, J. J. & YARMOLINSKY, M. 2002. Effects of the P1 plasmid centromere on expression of P1 partition genes. J Bacteriol, 184, 4857-67.

HASHIMOTO, Y., NISHIYAMA, M., YU, F., WATANABE, I., HORINOUCHI, S. & BEPPU, T. 1992. Development of a host-vector system in a *Rhodococcus* strain and its use for expression of the cloned nitrile hydratase gene cluster. J Gen Microbiol, 138, 1003-10.

HATANO, T. & NIKI, H. 2010. Partitioning of P1 plasmids by gradual distribution of the ATPase ParA. Mol Microbiol, 78, 1182-98.

HAYASHI, H. & KURUSU, Y. 2014. Analysis of a DNA region from low-copy-number plasmid pYAN-1 of *Sphingobium yanoikuyae* responsible for plasmid stability. Biosci Biotechnol Biochem, 78, 510-15.

HAYASHI, I., OYAMA, T. & MORIKAWA, K. 2001. Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus. EMBO J, 20, 1819-28.

HAYES, F. & BARILLA, D. 2006. The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. Nat Rev Microbiol, 4, 133-43.

HAYES, F. 2000. The partition system of multidrug resistance plasmid TP228 includes a novel protein that epitomizes an evolutionarily distinct subgroup of the ParA superfamily. Mol Microbiol, 37, 528-41.

HEO, L. & FEIG, M. 2017. PREFMD: a web server for protein structure refinement via molecular dynamics simulations. Bioinformatics. 34(6), 1063-65

HEIDELBERG, J. F., EISEN, J. A., NELSON, W. C., CLAYTON, R. A., GWINN, M. L., DODSON, R. J., HAFT, D. H., HICKEY, E. K., PETERSON, J. D., UMAYAM, L., GILL, S. R., NELSON, K. E., READ, T. D., TETTELIN, H., RICHARDSON, D., ERMOLAEVA, M. D., VAMATHEVAN, J., BASS, S., QIN, H., DRAGOI, I., SELLERS, P., MCDONALD, L., UTTERBACK, T., FLEISHMANN, R. D., NIERMAN, W. C., WHITE, O., SALZBERG, S. L., SMITH, H. O., COLWELL, R. R., MEKALANOS, J. J., VENTER, J. C. & FRASER, C. M. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature, 406, 477-83.

HESTER, C. M. & LUTKENHAUS, J. 2007. Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. Proc Natl Acad Sci U S A, 104, 20326-31.

HEUER, H. & SMALLA, K. 2012. Plasmids foster diversification and adaptation of bacterial populations in soil. FEMS Microbiol Rev, 36, 1083-1104.

HINNERWISCH, J., FENTON, W. A., FURTAK, K. J., FARR, G. W. & HORWICH, A. L. 2005. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. Cell, 121, 1029-41.

HISHIDA, T., IWASAKI, H., YAGI, T. & SHINAGAWA, H. 1999. Role of walker motif A of RuvB protein in promoting branch migration of holliday junctions. Walker motif A mutations affect ATP binding, ATP hydrolyzing, and DNA binding activities of Ruvb. J Biol Chem, 274, 25335-42.

HOLMES, J. A., FOLLETT, S. E., WANG, H., MEADOWS, C. P., VARGA, K. & BOWMAN, G. R. 2016. Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. Proc Natl Acad Sci U S A, 113, 12490-95.

HOOPMANN, M. R., ZELTER, A., JOHNSON, R. S., RIFFLE, M., MACCOSS, M. J., DAVIS, T. N. & MORITZ, R. L. 2015. Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes. J Proteome Res, 14, 2190-98.

HU, L., VECCHIARELLI, A. G., MIZUUCHI, K., NEUMAN, K. C. & LIU, J. 2017. Brownian Ratchet Mechanism for Faithful Segregation of Low-Copy-Number Plasmids. Biophys J, 112, 1489-1502.

HU, Z. & LUTKENHAUS, J. 2003. A conserved sequence at the C-terminus of MinD is required for binding to the membrane and targeting MinC to the septum. Mol Microbiol, 47, 345-55.

HUANG, F., OLDFIELD, C. J., XUE, B., HSU, W. L., MENG, J., LIU, X., SHEN, L., ROMERO, P., UVERSKY, V. N. & DUNKER, A. 2014. Improving protein order-disorder classification using charge-hydropathy plots. BMC Bioinformatics, 15 Suppl 17, S4.

HWANG, L. C., VECCHIARELLI, A. G., HAN, Y. W., MIZUUCHI, M., HARADA, Y., FUNNELL, B. E. & MIZUUCHI, K. 2013. ParA-mediated plasmid partition driven by protein pattern self-organization. EMBO J, 32, 1238-49.

HYDE, E. I., CALLOW, P., RAJASEKAR, K. V., TIMMINS, P., PATEL, T. R., SILIGARDI, G., HUSSAIN, R., WHITE, S. A., THOMAS, C. M. & SCOTT, D. J. 2017. Intrinsic disorder in the partitioning protein KorB persists after co-operative complex formation with operator DNA and KorA. Biochem J, 474**,** 3121-35.

IAKOUCHEVA, L. M., RADIVOJAC, P., BROWN, C. J., O'CONNOR, T. R., SIKES, J. G., OBRADOVIC, Z. & DUNKER, A. K. 2004. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res, 32, 1037-49.

INGRAM, L. C., RICHMOND, M. H. & SYKES, R. B. 1973. Molecular characterization of the R factors implicated in the carbenicillin resistance of a sequence of *Pseudomonas aeruginosa* strains isolated from burns. Antimicrob Agents Chemother, 3, 279-88.

IRETON, K., GUNTHER, N. W. T. & GROSSMAN, A. D. 1994. spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. J Bacteriol, 176, 5320-29.

ROUWENHORST J., R., PRONK, J. & P VAN DIJKEN, J. 1989. The discovery of beta-galactosidase. Trends Biochem Sci, 14, 416-18.

JACOB, F. & BRENNER, S. 1963. [On the regulation of DNA synthesis in bacteria: The hypothesis of the replicon]. C R Hebd Seances Acad Sci, 256, 298-300.

JAGURA-BURDZY, G. & THOMAS, C. M. 1995. Purification of KorA protein from broad host range plasmid RK2: definition of a hierarchy of KorA operators. J Mol Biol, 253**,** 39-50.

JAGURA-BURDZY, G., KOSTELIDOU, K., POLE, J., KHARE, D., JONES, A., WILLIAMS, D. R. & THOMAS, C. M. 1999. IncC of broad-host-range plasmid RK2 modulates KorB transcriptional repressor activity In vivo and operator binding in vitro. J Bacteriol, 181, 2807-15.

JAIN, A. & SRIVASTAVA, P. 2013. Broad host range plasmids. FEMS Microbiol Lett, 348, 87-96.

JANG, S. B., SEEFELDT, L. C. & PETERS, J. W. 2000. Insights into nucleotide signal transduction in nitrogenase: structure of an iron protein with MgADP bound. Biochemistry, 39, 14745-52.

JEEVES, M. & KNOWLES, T. J. 2015. A novel pathway for outer membrane protein biogenesis in Gram-negative bacteria. Mol Microbiol, 97, 607-11.

JENNEBACH, S., HERZOG, F., AEBERSOLD, R. & CRAMER, P. 2012. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. Nucleic Acids Res, 40, 5591-601.

JENSEN, R. B. & GERDES, K. 1997. Partitioning of plasmid R1. The ParM protein exhibits ATPase activity and interacts with the centromere-like ParR-parC complex11Edited by M. Gottesman. J Mol Biol , 269, 505-513.

JENSEN, R. B., DAM, M. & GERDES, K. 1994. Partitioning of plasmid R1. The parA operon is autoregulated by ParR and its transcription is highly stimulated by a downstream activating element. J Mol Biol, 236, 1299-309.

JENSEN, R. B., LURZ, R. & GERDES, K. 1998. Mechanism of DNA segregation in prokaryotes: Replicon pairing by parC of plasmid R1. Proc Natl Acad Sci U S A, 95, 8550-55.

JIANG, S., NARITA, A., POPP, D., GHOSHDASTIDER, U., LEE, L. J., SRINIVASAN, R., BALASUBRAMANIAN, M. K., ODA, T., KOH, F., LARSSON, M. & ROBINSON, R. C. 2016. Novel actin filaments from *Bacillus thuringiensis* form nanotubules for plasmid DNA segregation. Proc Natl Acad Sci U S A, 113, E1200-E1205

JOERGER, A. C., WILCKEN, R., & ANDREEVA, A. (2014). Tracing the Evolution of the p53 Tetramerization Domain. Structure (London, England:1993), 22(9), 1301-10.

JONES, D. T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol, 292, 195-202.

KAHRAMAN, A., MALMSTRÖM, L. & AEBERSOLD, R. 2011. Xwalk: Computing and visualizing distances in cross-linking experiments. Bioinformatics, 27, 2163-64.

KALLBERG, M., MARGARYAN, G., WANG, S., MA, J. & XU, J. 2014. RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol, 1137, 17-27.

KALLIOMAA-SANFORD, A. K., RODRIGUEZ-CASTANEDA, F. A., MCLEOD, B. N., LATORRE-ROSELLO, V., SMITH, J. H., REIMANN, J., ALBERS, S. V. & BARILLA, D. 2012. Chromosome segregation in Archaea mediated by a hybrid DNA partition machine. Proc Natl Acad Sci U S A, 109, 3754-59.

KALNIN, K., STEGALKINA, S. & YARMOLINSKY, M. 2000. pTAR-encoded proteins in plasmid partitioning. J Bacteriol, 182, 1889-94.

KARADZIC, I., MAUPIN-FURLOW, J., HUMBARD, M., PRUNETTI, L., SINGH, P. & GOODLETT, D. R. 2012. Chemical cross-linking, mass spectrometry, and in silico modeling of proteasomal 20S core particles of the haloarchaeon *Haloferax volcanii*. Proteomics, 12, 1806-14.

KARIMOVA, G., PIDOUX, J., ULLMANN, A. & LADANT, D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A, 95, 5752-6.

KAY, L. E., IKURA, M., TSCHUDIN, R. & BAX, A. 2011. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. J Mag Resonance, 213, 423-41.

KAY, L., KEIFER, P. & SAARINEN, T. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc, 114, 10663-65.

KEARSEY, S., CAMPBELL, J. L., ALMOUZNI, G., POULTON, J., YASUKAWA, T., BELL, S., LASKEY, R., KUNKEL, T., CHONG, J. & GE, X. 2009. Molecular Themes in DNA Replication, Royal Society of Chemistry, doi.org/10.1039/9781847559852

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modelling, prediction and analysis. Nature prot, 10, 845-58.

KHARE, D., ZIEGELIN, G., LANKA, E. & HEINEMANN, U. 2004. Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. Nat Struct Mol Biol, 11, 656-63.

KIM, D. E., CHIVIAN, D. & BAKER, D. 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res, 32, W526-W531.

KLECKNER, I. R. & FOSTER, M. P. 2011. An introduction to NMR-based approaches for measuring protein dynamics. Biochim Biophys Acta, 1814, 942-68.

KOONIN, E. V. 1993. A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. J Mol Biol, 229, 1165-74.
KORNACKI, J. A., BALDERES, P. J. & FIGURSKI, D. H. 1987. Nucleotide sequence of korB, a replication control gene of broad host-range plasmid RK2. J Mol Biol, 198**,** 211-22.
KOSOL, S., CONTRERAS-MARTOS, S., CEDENO, C. & TOMPA, P. 2013b. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. Molecules*,* 18**,** 10802-28.
KOSTELIDOU, K., JAGURA-BURDZY, G. & THOMAS, C. M. 1998. Mutational analysis of the global regulator KorA of broad-host-range plasmid RK2. J Mol Biol, 281**,** 453-63.
KOZLOWSKI, L. P. & BUJNICKI, J. M. 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinfo, 13, 111.
KRIEGER, E., NABUURS, S. B. & VRIEND, G. 2003. Homology modeling. Meth Biochem Ana, 44, 509-24.
KRIEGER, E., NABUURS, S. B. & VRIEND, G. 2005. Homology Modeling. Struc Bioinfo. John Wiley & Sons, Inc.
KUMAR, D., SHARMA, N. & GIRI, R. 2017. Therapeutic Interventions of Cancers Using Intrinsically Disordered Proteins as Drug Targets: c-Myc as Model System. Cancer Inform, 16, 1176935-94.
KWAN, A. H., MOBLI, M., GOOLEY, P. R., KING, G. F. & MACKAY, J. P. 2011. Macromolecular NMR spectroscopy for the non-spectroscopist. FEBS J, 278, 687-703.
LADANT, D. & ULLMANN, A. 1999. *Bordatella pertussis* adenylate cyclase: A toxin with multiple talents. Trends Microbiol, 7, 172-76.
LAL, A., HAYNES, S. R. & GOROSPE, M. 2005. Clean Western blot signals from immunoprecipitated samples. Mol Cell Probes, 19**,** 385-88.
LARSEN, R. A., CUSUMANO, C., FUJIOKA, A., LIM-FONG, G., PATTERSON, P. & POGLIANO, J. 2007. Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis. Genes Dev, 21, 1340-52.
LA VERDE, V., DOMINICI, P. AND ASTEGNO, A. 2017. Determination of Hydrodynamic Radius of Proteins by Size Exclusion Chromatography. Bio-protocol, 7(8), e2230.
LASKOWSKI, R., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst, 26, 283-91
LAUE, T., SHAH, B., RIDGEWAY, T. & PELLETIER, S. 1992. Computer-aided Interpretation of Sedimentation Data for Proteins. In Harding, SE, Rowe, AJ, Horton, JC (Eds.). Anal Ultracentrifugation Biochem Pol Sci. Cambridge Royal Society of Chemistry.
LAVALLIE, E. R., LU, Z., DIBLASIO-SMITH, E. A., COLLINS-RACIE, L. A. & MCCOY, J. M. 2000. Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. Methods Enzymol, 326, 322-40.
LEBENDIKER, M. & DANIELI, T. 2014. Production of prone-to-aggregate proteins. FEBS Lett, 588, 236-46.
LEBOWITZ, J., LEWIS, M. S. & SCHUCK, P. 2002. Modern analytical ultracentrifugation in protein science: A tutorial review. Protein Sci, 11, 2067-79.
LEDERBERG, J. 1950. The beta-d-galactosidase of *Escherichia coli*, strain K-12. J Bacteriol, 60, 381-92.
LEE, P. S. & GROSSMAN, A. D. 2006. The chromosome partitioning proteins Soj (ParA) and Spo0J (ParB) contribute to accurate chromosome partitioning, separation of replicated sister origins, and regulation of replication initiation in *Bacillus subtilis*. Mol Microbiol, 60, 853-69.
LEIPE, D. D., WOLF, Y. I., KOONIN, E. V. & ARAVIND, L. 2002. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol, 317, 41-72.
LEITNER, A., WALZTHOENI, T. & AEBERSOLD, R. 2013. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. Nature Prot, 9, 120.
LEONARD, T. A., BUTLER, P. J. & LOWE, J. 2004. Structural analysis of the chromosome segregation protein Spo0J from *Thermus thermophilus*. Mol Microbiol, 53, 419-32.
LEONARD, T. A., BUTLER, P. J. & LOWE, J. 2005. Bacterial chromosome segregation: structure and DNA binding of the Soj dimer--a conserved biological switch. EMBO J, 24, 270-82.

LEPLAE, R., LIMA-MENDEZ, G. & TOUSSAINT, A. 2006. A first global analysis of plasmid encoded proteins in the ACLAME database. FEMS Microbiol Rev, 30, 980-94.

LI, J., WHITE, J. T., SAAVEDRA, H., WRABL, J. O., MOTLAGH, H. N., LIU, K., SOWERS, J., SCHROER, T. A., THOMPSON, E. B. & HILSER, V. J. 2017. Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. Elife, 6.

LI, Y., DABRAZHYNETSKAYA, A., YOUNGREN, B. & AUSTIN, S. 2004. The role of Par proteins in the active segregation of the P1 plasmid. Mol Microbiol, 53, 93-102.

LIBANTE, V., THION, L. & LANE, D. 2001. Role of the ATP-binding site of SopA protein in partition of the F plasmid. J Mol Biol, 314, 387-99.

LIM, G. E., DERMAN, A. I. & POGLIANO, J. 2005. Bacterial DNA segregation by dynamic SopA polymers. Proc Natl Acad Sci U S A, 102, 17658-63.

LIM, H. C., SUROVTSEV, I. V., BELTRAN, B. G., HUANG, F., BEWERSDORF, J. & JACOBS-WAGNER, C. 2014. Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. Elife, 3, e02758.

LIMA, D., B. DE LIMA, T., S. BALBUENA, T., C. NEVES-FERREIRA, A. G., C. BARBOSA, V., C. GOZZO, F. & C. CARVALHO, P. 2015. Using SIM-XL to identify and annotate cross-linked peptides analyzed by mass spectrometry, J Proteomics, 129, 51-55.

LIN, D. C. & GROSSMAN, A. D. 1998. Identification and characterization of a bacterial chromosome partitioning site. Cell, 92, 675-85.

LIN, K., SIMOSSIS, V. A., TAYLOR, W. R. & HERINGA, J. 2005. A simple and fast secondary structure prediction method using hidden neural networks. Bioinformatics, 21, 152-59.

LINDING, R., RUSSELL, R. B., NEDUVA, V. & GIBSON, T. J. 2003. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res, 31, 3701-708.

LIU, M. A., KWONG, S. M., JENSEN, S. O., BRZOSKA, A. J. & FIRTH, N. 2013. Biology of the staphylococcal conjugative multiresistance plasmid pSK41. Plasmid, 70, 42-51.

LOBSTEIN, J., EMRICH, C. A., JEANS, C., FAULKNER, M., RIGGS, P. & BERKMEN, M. 2012. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact, 11, 56.

LOPEZ-MONTERO, I., LOPEZ-NAVAJAS, P., MINGORANCE, J., VELEZ, M., VICENTE, M. & MONROY, F. 2013. Membrane reconstitution of FtsZ-ZipA complex inside giant spherical vesicles made of *E. coli* lipids: large membrane dilation and analysis of membrane plasticity. Biochim Biophys Acta, 1828, 687-98.

LUKASZEWICZ, M., KOSTELIDOU, K., BARTOSIK, A. A., COOKE, G. D., THOMAS, C. M. & JAGURA-BURDZY, G. 2002. Functional dissection of the ParB homologue (KorB) from IncP-1 plasmid RK2. Nucleic Acids Res, 30, 1046-55.

LUTKENHAUS, J. 2012. The ParA/MinD family puts things in their place. Trends Microbiol, 20, 411-18.

LUTKENHAUS, J. & SUNDARAMOORTHY, M. 2003. MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. Mol Microbiol, 48, 295-303.

LUTKENHAUS, J. 2007. Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. Annu Rev Biochem, 76, 539-62.

MACKENZIE, H. W. & HANSEN, D. F. 2017. A (13) C-detected (15) N double-quantum NMR experiment to probe arginine side-chain guanidinium (15)N(eta) chemical shifts. J Biomol NMR, 69, 123-132.

MAGNAN, C. N. & BALDI, P. 2014. SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. Bioinformatics, 30, 2592-97.

MALAKHOV, M. P., MATTERN, M. R., MALAKHOVA, O. A., DRINKER, M., WEEKS, S. D. & BUTT, T. R. 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Funct Genomics, 5, 75-86.

MAO, A. H., LYLE, N. & PAPPU, R. V. 2013. Describing sequence-ensemble relationships for intrinsically disordered proteins. Biochem J, 449, 307-18.

MARCHINI, F. K., DE GODOY, L. M., RAMPAZZO, R. C., PAVONI, D. P., PROBST, C. M., GNAD, F., MANN, M. & KRIEGER, M. A. 2011. Profiling the *Trypanosoma cruzi* phosphoproteome. PLoS One, 6, e25381.

MARKLEY, J. L., ULRICH, E. L., BERMAN, H. M., HENRICK, K., NAKAMURA, H. & AKUTSU, H. 2008. BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): New policies affecting biomolecular NMR depositions. J Biomol NMR, 40, 153-5.

MARSTON, A. L. & ERRINGTON, J. 1999. Dynamic movement of the ParA-like Soj protein of B. subtilis and its dual role in nucleoid organization and developmental regulation. Mol Cell, 4, 673-82.

MEACOCK, P. A. & COHEN, S. N. 1980. Partitioning of bacterial plasmids during cell division: A cis-acting locus that accomplishes stable plasmid inheritance. Cell, 20, 529-42.

MELAMUD, E. & MOULT, J. 2003. Evaluation of disorder predictions in CASP5. Proteins, 53 Suppl 6, 561-5.

MELO, F., SANCHEZ, R. & SALI, A. 2002. Statistical potentials for fold assessment. Protein Sci, 11, 430-48.

MENDOZA-ESPINOSA, P., GARCIA-GONZALEZ, V., MORENO, A., CASTILLO, R. & MAS-OLIVA, J. 2009. Disorder-to-order conformational transitions in protein structure and its relationship to disease. *Mol Cell Biochem,* 330**,** 105-20.

MEYER, R. & HINDS, M. 1982. Multiple mechanisms for expression of incompatibility by broad-host-range plasmid RK2. J Bacteriol, 152, 1078-90.

MEYER, R. J. & HELINSKI, D. R. 1977. Unidirectional replication of the P-group plasmid RK2. Biochim Biophys Acta, 478, 109-13.

Miller J.H. Cold Spring Harbor; NY: 1972. Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press.

MILLION-WEAVER, S. & CAMPS, M. 2014. Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? Plasmid, 75, 27-36.

MITCHELL, C. & OLIVER, D. 1993. Two distinct ATP-binding domains are needed to promote protein export by *Escherichia coli* SecA ATPase. Mol Microbiol, 10, 483-97.

MOFFATT, B. A. & STUDIER, F. W. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell, 49, 221-7.

MOHL, D. A., EASTER, J., JR. & GOBER, J. W. 2001. The chromosome partitioning protein, ParB, is required for cytokinesis in *Caulobacter crescentus*. Mol Microbiol, 42, 741-55.

MOHL, D. A. & GOBER, J. W. 1997. Cell cycle-dependent polar localization of chromosome partitioning proteins in *Caulobacter crescentus*. Cell, 88, 675-84.

MOLLER-JENSEN, J., BORCH, J., DAM, M., JENSEN, R. B., ROEPSTORFF, P. & GERDES, K. 2003. Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an Actin-like insertional polymerization mechanism. Mol Cell, 12, 1477-87.

MOLLER-JENSEN, J., JENSEN, R. B., LOWE, J. & GERDES, K. 2002. Prokaryotic DNA segregation by an Actin-like filament. EMBO J, 21, 3119-27.

MOLLER-JENSEN, J., RINGGAARD, S., MERCOGLIANO, C. P., GERDES, K. & LOWE, J. 2007. Structural analysis of the ParR/parC plasmid partition complex. EMBO J, 26, 4413-22.

MORI, H., MORI, Y., ICHINOSE, C., NIKI, H., OGURA, T., KATO, A. & HIRAGA, S. 1989. Purification and characterization of SopA and SopB proteins essential for F plasmid partitioning. J Biol Chem, 264, 15535-41.

MOTLAGH, H. N. & HILSER, V. J. 2012. Agonism/antagonism switching in allosteric ensembles. Proc Natl Acad Sci U S A, 109, 4134-39.

MULLER-SPATH, S., SORANNO, A., HIRSCHFELD, V., HOFMANN, H., RUEGGER, S., REYMOND, L., NETTELS, D. & SCHULER, B. 2010. From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc Natl Acad Sci U S A, 107, 14609-14.

MURAYAMA, K., ORTH, P., DE LA HOZ, A. B., ALONSO, J. C. & SAENGER, W. 2001. Crystal structure of omega transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 A resolution. J Mol Biol, 314, 789-96.

NALLAMSETTY, S. & WAUGH, D. S. 2006. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expr Purif, 45, 175-82.

NI, L., XU, W., KUMARASWAMI, M. & SCHUMACHER, M. A. 2010. Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. Proc Natl Acad Sci U S A, 107, 11763-68.

NIELSEN, M., LUNDEGAARD, C., LUND, O. & PETERSEN, T. N. 2010. CPHmodels-3.0-Remote homology modeling using structure-guided sequence profiles. Nucleic Acids Res, 38, W576-81.

NIESEN, F. H., BERGLUND, H. & VEDADI, M. 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc, 2, 2212-21.

NIKI, H. & HIRAGA, S. 1997. Subcellular distribution of actively partitioning F plasmid during the cell division cycle in *E. coli*. Cell, 90, 951-57.

NOVACEK, J., HABA, N. Y., CHILL, J. H., ZIDEK, L. & SKLENAR, V. 2012. 4D non-uniformly sampled HCBCACON and (1)J(NCalpha)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR, 53, 139-48.

OGURA, T. & HIRAGA, S. 1983. Partition mechanism of F plasmid: Two plasmid gene-encoded products and a cis-acting region are involved in partition. Cell, 32, 351-60.

OHANA, R. F., ENCELL, L. P., ZHAO, K., SIMPSON, D., SLATER, M. R., URH, M. & WOOD, K. V. 2009. HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. Protein Expr Purif, 68, 110-20.

OLDFIELD, C. J. & DUNKER, A. K. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem, 83, 553-84.

OLDFIELD, C. J., CHENG, Y., CORTESE, M. S., BROWN, C. J., UVERSKY, V. N. & DUNKER, A. K. 2005. Comparing and combining predictors of mostly disordered proteins. Biochemistry, 44, 1989-2000.

ORLOVA, A., GARNER, E. C., GALKIN, V. E., HEUSER, J., MULLINS, R. D. & EGELMAN, E. H. 2007. The structure of bacterial ParM filaments. Nat Struct Mol Biol, 14, 921-26.

PALMER, A. G., CAVANAGH, J., WRIGHT, P. E. & RANCE, M. 1991. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. J Mag Resonance (1969), 93, 151-70.

PANAGIOTIDIS, C. H., REYES, M., SIEVERTSEN, A., BOOS, W. & SHUMAN, H. A. 1993. Characterization of the structural requirements for assembly and nucleotide binding of an ATP-binding cassette transporter. The maltose transport system of *Escherichia coli*. J Biol Chem, 268, 23685-96.

PANCHAL, S. C., BHAVESH, N. S. & HOSUR, R. V. 2001. Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. J Biomol NMR, 20, 135-47.

PANTOJA-UCEDA, D. & SANTORO, J. 2013. Direct correlation of consecutive C'-N groups in proteins: a method for the assignment of intrinsically disordered proteins. J Biomol NMR, 57, 57-63.

PARK, K. T., WU, W., LOVELL, S. & LUTKENHAUS, J. 2012. Mechanism of the asymmetric activation of the MinD ATPase by MinE. Mol Microbiol, 85, 271-81.

PARRISH, J. R., YU, J., LIU, G., HINES, J. A., CHAN, J. E., MANGIOLA, B. A., ZHANG, H., PACIFICO, S., FOTOUHI, F. & DIRITA, V. J. 2007. A proteome-wide protein interaction map for *Campylobacter jejuni*. Genome biology, 8, R130.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera-A visualization system for exploratory research and analysis. J Comput Chem, 25**,** 1605-12.

PAZOS, F., PIETROSEMOLI, N., GARCIA-MARTIN, J. A. & SOLANO, R. 2013. Protein intrinsic disorder in plants. Front Plant Sci, 4, 363.

PEREIRA, S. F., GOSS, L. & DWORKIN, J. 2011. Eukaryote-like serine/threonine kinases and phosphatases in bacteria. Microbiol Mol Biol Rev, 75, 192-212.

PEREZ, V., CORPA, J. M., GARCIA MARIN, J. F., ADURIZ, J. J. & JENSEN, H. E. 1999. Generalized aspergillosis in dairy sheep. Zentralbl Veterinarmed B, 46, 613-21.

POPP, D., NARITA, A., ODA, T., FUJISAWA, T., MATSUO, H., NITANAI, Y., IWASA, M., MAEDA, K., ONISHI, H. & MAEDA, Y. 2008. Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability. EMBO J, 27, 570-79.

POPP, D., XU, W., NARITA, A., BRZOSKA, A. J., SKURRAY, R. A., FIRTH, N., GHOSHDASTIDER, U., MAEDA, Y., ROBINSON, R. C. & SCHUMACHER, M. A. 2010. Structure and filament dynamics of the pSK41 actin-like ParM protein: Implications for plasmid DNA segregation. J Biol Chem, 285, 10130-40.

PORTERFIELD, J. Z. & ZLOTNICK, A. 2010. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. Virology, 407**,** 281-88.

PRAKASH, S., TIAN, L., RATLIFF, K. S., LEHOTZKY, R. E. & MATOUSCHEK, A. 2004. An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat Struct Mol Biol, 11, 830-37.

PRATTO, F., SUZUKI, Y., TAKEYASU, K. & ALONSO, J. C. 2009. Single-molecule analysis of proteinxDNA complexes formed during partition of newly replicated plasmid molecules in *Streptococcus pyogenes*. J Biol Chem, 284, 30298-306.

PROVENCHER, S. W. & GLOCKNER, J. 1981. Estimation of globular protein secondary structure from circular dichroism. Biochemistry, 20, 33-7.

PTACIN, J. L., LEE, S. F., GARNER, E. C., TORO, E., ECKART, M., COMOLLI, L. R., MOERNER, W. E. & SHAPIRO, L. 2010. A spindle-like apparatus guides bacterial chromosome segregation. Nat Cell Biol, 12, 791-98.

QING, G., MA, L. C., KHORCHID, A., SWAPNA, G. V., MAL, T. K., TAKAYAMA, M. M., XIA, B., PHADTARE, S., KE, H., ACTON, T., MONTELIONE, G. T., IKURA, M. & INOUYE, M. 2004. Cold-shock induced high-yield protein production in *Escherichia coli*. Nat Biotechnol, 22, 877-82.

QUISEL, J. D., LIN, D. C. & GROSSMAN, A. D. 1999. Control of development by altered localization of a transcription factor in B. subtilis. Mol Cell, 4, 665-72.

RAJASEKAR, K., MUNTAHA, S. T., TAME, J. R., KOMMAREDDY, S., MORRIS, G., WHARTON, C. W., THOMAS, C. M., WHITE, S. A., HYDE, E. I. & SCOTT, D. J. 2010. Order and disorder in the domain organization of the plasmid partition protein KorB. J Biol Chem, 285, 15440-49.

RAJASEKAR, K. V., LOVERING, A. L., DANCEA, F., SCOTT, D. J., HARRIS, S. A., BINGLE, L. E., ROESSLE, M., THOMAS, C. M., HYDE, E. I. & WHITE, S. A. 2016. Flexibility of KorA, a plasmid-encoded, global transcription regulator, in the presence and the absence of its operator. Nucleic Acids Res*,* 44**,** 4947-56.

RAMAKRISHNAN, C., DANI, V. S. & RAMASARMA, T. 2002. A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. Protein Eng, 15, 783-98.

RAVIN, N. V., RECH, J. & LANE, D. 2003. Mapping of functional domains in F plasmid partition proteins reveals a bipartite SopB-recognition domain in SopA. J Mol Biol, 329, 875-89.

RAYAN, A. 2009. New tips for structure prediction by comparative modeling. Bioinformation*,* 3**,** 263-7.

REHM, T., HUBER, R. & HOLAK, T. A. 2002. Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. Structure, 10, 1613-18.

RINGGAARD, S., VAN ZON, J., HOWARD, M. & GERDES, K. 2009. Movement and equipositioning of plasmids by ParA filament disassembly. Proc Natl Acad Sci U S A, 106, 19369-74.

ROBERTS, M. A., WADHAMS, G. H., HADFIELD, K. A., TICKNER, S. & ARMITAGE, J. P. 2012. ParA-like protein uses nonspecific chromosomal DNA binding to partition protein complexes. Proc Natl Acad Sci U S A, 109, 6698-703.

RODIONOV, O., LOBOCKA, M. & YARMOLINSKY, M. 1999. Silencing of genes flanking the P1 plasmid centromere. Science, 283, 546-49.

ROMERO, P. R., ZAIDI, S., FANG, Y. Y., UVERSKY, V. N., RADIVOJAC, P., OLDFIELD, C. J., CORTESE, M. S., SICKMEIER, M., LEGALL, T., OBRADOVIC, Z. & DUNKER, A. K. 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci U S A, 103, 8390-95.

ROMERO, P., OBRADOVIC, Z. & DUNKER, A. K. 1997. Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Informatics, 8, 110-24.

ROMERO, P., OBRADOVIC, Z., LI, X., GARNER, E. C., BROWN, C. J. & DUNKER, A. K. 2001. Sequence complexity of disordered protein. Proteins, 42, 38-48.

RONDAHL, H., NILSSON, B. & HOLMGREN, E. 1992. Fusions to the 5' end of a gene encoding a two-domain analogue of staphylococcal protein A. J Biotechnol, 25, 269-87.

ROSCHE, T. M., SIDDIQUE, A., LARSEN, M. H. & FIGURSKI, D. H. 2000. Incompatibility protein IncC and global regulator KorB interact in active partition of promiscuous plasmid RK2. J Bacteriol, 182, 6014-26.

ROST, B., FARISELLI, P. & CASADIO, R. 1996. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci, 5, 1704-18.

ROST, B., YACHDAV, G. & LIU, J. 2004. The PredictProtein server. Nucleic Acids Res, 32, W321-6.

ROUWENHORST, R. J., PRONK, J. T. & VAN DIJKEN, J. P. 1989. The discovery of beta-galactosidase. Trends Biochem Sci, 14**,** 416-8.

ROWLETT, V. W. & MARGOLIN, W. 2013. The bacterial Min system. Curr Biol, 23, R553-56.

RYAN, K. R. & SHAPIRO, L. 2003. Temporal and spatial regulation in prokaryotic cell cycle progression and development. Annu Rev Biochem, 72, 367-94.

SAHU, D., BASTIDAS, M. & SHOWALTER, S. A. 2014. Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. Anal Biochem*,* 449**,** 17-25.

SALJE, J., GAYATHRI, P. & LOWE, J. 2010. The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. Nat Rev Microbiol, 8, 683-92.

SAMBROOK, J. & RUSSELL, D. W. 2001. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

SANCHEZ, A., RECH, J., GASC, C. & BOUET, J. Y. 2013. Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance. Nucleic Acids Res, 41, 3094-103.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, 74, 5463-67.

SCHEUERMANN, T. H., PADRICK, S. B., GARDNER, K. H. & BRAUTIGAM, C. A. 2016. On the acquisition and analysis of Microscale Thermophoresis data. Ana Biochem, 496, 79-93.

SCHILLING, B., ROW, R. H., GIBSON, B. W., GUO, X. & YOUNG, M. M. 2003. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. J Am Soc Mass Spectrom, 14, 834-50.

SCHMIDT, A. & HALL, M. N. 1998. Signaling to the actin cytoskeleton. Annu Rev Cell Dev Biol, 14, 305-38

SCHLEPER, C., HOLZ, I., JANEKOVIC, D., MURPHY, J. & ZILLIG, W. 1995. A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. J Bacteriol, 177, 4417-26.

SCHUCK, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J, 78, 1606-19.

SCHUMACHER, M. A. & FUNNELL, B. E. 2005. Structures of ParB bound to DNA reveal mechanism of partition complex formation. Nature, 438, 516-19.

SCHUMACHER, M. A. 2008. Structural biology of plasmid partition: Uncovering the molecular mechanisms of DNA segregation. Biochem J, 412, 1-18.

SCHUMACHER, M. A. 2012. Bacterial plasmid partition machinery: a minimalist approach to survival. Curr Opin Struct Biol, 22, 72-79.

SCHUMACHER, M. A., GLOVER, T. C., BRZOSKA, A. J., JENSEN, S. O., DUNHAM, T. D., SKURRAY, R. A. & FIRTH, N. 2007. Segrosome structure revealed by a complex of ParR with centromere DNA. Nature, 450, 1268-71.

SCHWEDE, T., KOPP, J., GUEX, N. & PEITSCH, M. C. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res, 31, 3381-85.

SEEFELDT, L. C., MORGAN, T. V., DEAN, D. R. & MORTENSON, L. E. 1992. Mapping the site(s) of MgATP and MgADP interaction with the nitrogenase of *Azotobacter vinelandii*. Lysine 15 of the iron protein plays a major role in MgATP interaction. J Biol Chem, 267, 6680-88.

SENGUPTA, M. & AUSTIN, S. 2011. Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. Infect Immun, 79, 2502-09.

SHARMA, R., RADULY, Z., MISKEI, M. & FUXREITER, M. 2015. Fuzzy complexes: Specific binding without complete folding. FEBS Letters, 589, 2533-42.

SHEN, M.-Y. & SALI, A. 2006. Statistical potential for assessment and prediction of protein structures. Protein Science: A Publication of the Protein Society, 15, 2507-24.

SHEN, Y. & BAX, A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR, 56, 227-41.

SHEN, Y., DELAGLIO, F., CORNILESCU, G. & BAX, A. 2009a. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR, 44, 213-23.

SHEN, Y., VERNON, R., BAKER, D. & BAX, A. 2009b. De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR, 43, 63-78.

SHIMBA, N., KOVACS, H., STERN, A. S., NOMURA, A. M., SHIMADA, I., HOCH, J. C., CRAIK, C. S. & DOTSCH, V. 2004. Optimization of 13C direct detection NMR methods. J Biomol NMR, 30, 175-79.

SHOEMAKER, B. A., PORTMAN, J. J. & WOLYNES, P. G. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A, 97, 8868-73.

SHUID, A. N., KEMPSTER, R. & MCGUFFIN, L. J. 2017. ReFOLD: a server for the refinement of 3D protein models guided by accurate quality estimates. Nucleic Acids Res. 3, 45(W1), W422-W428.

SIDDIQUE, A. & FIGURSKI, D. H. 2002. The active partition gene incC of IncP plasmids is required for stable maintenance in a broad range of hosts. J Bacteriol, 184, 1788-93.

SIEGEL, L. M. & MONTY, K. J. 1966. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. Biochim Biophys Acta, 112, 346-62.

SIMPSON, A. E., SKURRAY, R. A. & FIRTH, N. 2003. A single gene on the staphylococcal multi resistance plasmid pSK1 encodes a novel partitioning system. J Bacteriol, 185, 2143-52.

SMITH, D. B. & JOHNSON, K. S. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene, 67, 31-40.

SOBERON, N. E., LIOY, V. S., PRATTO, F., VOLANTE, A. & ALONSO, J. C. 2011. Molecular anatomy of the *Streptococcus pyogenes* pSM19035 partition and segrosome complexes. Nucleic Acids Res, 39, 2624-37.

SÖDING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res, 33, W244-W248.

SPOLAR, R. S. & RECORD, M. T., JR. 1994. Coupling of local folding to site-specific binding of proteins to DNA. Science, 263, 777-84.

SREERAMA, N. & WOODY, R. W. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem, 287, 252-60.

SREERAMA, N. & WOODY, R. W. 2004. On the analysis of membrane protein circular dichroism spectra. Protein Sci, 13, 100-12.

SREERAMA, N., VENYAMINOV, S. Y. & WOODY, R. W. 1999. Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. Protein Sci, 8, 370-80.

STABY, L., O'SHEA, C., WILLEMOES, M., THEISEN, F., KRAGELUND, B. B. & SKRIVER, K. 2017. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. Biochem J, 474, 2509-32.

STORY, R. M. & STEITZ, T. A. 1992. Structure of the recA protein-ADP complex. Nature, 355, 374-76.

STUDIER, F. W. & DUNN, J. J. 1983. Organization and expression of bacteriophage T7 DNA. Cold Spring Harb Symp Quant Biol, 47 Pt 2**,** 999-1007.

STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol, 189, 113-30.

SUN, X., JONES, W. T. & RIKKERINK, E. H. 2012. GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. Biochem J, 442, 1-12.

SURTEES, J. A. & FUNNELL, B. E. 1999. P1 ParB Domain Structure Includes Two Independent Multimerization Domains. J Bacteriol, 181, 5898-5908.

SURTEES, J. A. & FUNNELL, B. E. 2001. The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex. J Biol Chem, 276, 12385-94.

SURTEES, J. A. & FUNNELL, B. E. 2003. Plasmid and chromosome traffic control: how ParA and ParB drive partition. Curr Top Dev Biol, 56, 145-80.

SUZUKI, H., YANO, H., BROWN, C. J. & TOP, E. M. 2010. Predicting plasmid promiscuity based on genomic signature. J Bacteriol, 192, 6045-55.

THEOPHILUS, B. D. & THOMAS, C. M. 1987. Nucleotide sequence of the transcriptional repressor gene korB which plays a key role in regulation of the copy number of broad host range plasmid RK2. Nucleic Acids Res, 15, 7443-50.

THOMAIDES, H. B., FREEMAN, M., KAROUI, M. E. & ERRINGTON, J. 2001. Division site selection protein DivIVA of *Bacillus subtilis* has a second distinct function in chromosome segregation during sporulation. Gen Dev, 15, 1662-73.

THOMAS, C. M. & HUSSAIN, A. A. 1984. The korB gene of broad host range plasmid RK2 is a major copy number control element which may act together with *trfB* by limiting *trfA* expression. EMBO J, 3, 1513-19.

THOMAS, C. M. & SMITH, C. A. 1986. The *trfB* region of broad host range plasmid RK2: the nucleotide sequence reveals *incC* and key regulatory gene *trfB/korA/korD* as overlapping genes. Nucleic Acids Res, 14, 4453-69.

THOMAS, C. M., MEYER, R. & HELINSKI, D. R. 1980. Regions of broad-host-range plasmid RK2 which are essential for replication and maintenance. J Bacteriol, 141, 213-22.

THOMAS, J. G. & BANEYX, F. 1996. Protein folding in the cytoplasm of *Escherichia coli*: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. Mol Microbiol, 21, 1185-96.

TILLY, K., CHECROUN, C. & ROSA, P. A. 2012. Requirements for *Borrelia burgdorferi* plasmid maintenance. Plasmid, 68, 1-12.

TINSLEY, E. & KHAN, S. A. 2006. A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. J Bacteriol, 188, 2829-35.

TOMPA, P. & FUXREITER, M. 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci, 33, 2-8.

TOMPA, P. 2002. Intrinsically unstructured proteins. Trends Biochem Sci, 27, 527-33.

TOMPA, P. 2005. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett, 579, 3346-54.

TOMPA, P., DOSZTANYI, Z. & SIMON, I. 2006. Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. J Proteome Res, 5, 1996-2000.

TOTO, A., CAMILLONI, C., GIRI, R., BRUNORI, M., VENDRUSCOLO, M. & GIANNI, S. 2016. Molecular recognition by templated folding of an intrinsically disordered protein. Sci Rep, 6, 21994.

TRNKA, M. J., BAKER, P. R., ROBINSON, P. J. J., BURLINGAME, A. L. & CHALKLEY, R. J. 2014. Matching cross-linked peptide spectra: Only as good as the Wworse identification. Mol Cell Proteom: MCP, 13, 420-34.

TYPAS, A. & HENGGE, R. 2006. Role of the spacer between the -35 and -10 regions in sigmas promoter selectivity in *Escherichia coli*. Mol Microbiol, 59, 1037-51.

UETZ, P., GIOT, L., CAGNEY, G., MANSFIELD, T. A., JUDSON, R. S., KNIGHT, J. R., LOCKSHON, D., NARAYAN, V., SRINIVASAN, M., POCHART, P., QURESHI-EMILI, A., LI, Y., GODWIN, B., CONOVER, D., KALBFLEISCH, T., VIJAYADAMODAR, G., YANG, M., JOHNSTON, M., FIELDS, S. & ROTHBERG, J. M. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature, 403, 623-27.

UVERSKY, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. Protein Sci, 11, 739-56.

UVERSKY, V. N. 2013. Intrinsic disorder-based protein interactions and their modulators. Curr Pharm Des, 19, 4191-213.
UVERSKY, V. N. 2015. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. Front Aging Neurosci, 7, 18.
UVERSKY, V. N., GILLESPIE, J. R. & FINK, A. L. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins, 41, 415-27.

VÁMOSI, G., BAUDENDISTEL, N., VON DER LIETH, C.-W., SZALÓKI, N., MOCSÁR, G., MÜLLER, G., TÓTH, K. (2008). Conformation of the c-Fos/c-Jun complex *in Vivo*: A combined FRET, FCCS, and MD-Modeling study. Biophy J, 94(7), 2859-68.
VAN DEN ENT, F., MOLLER-JENSEN, J., AMOS, L. A., GERDES, K. & LOWE, J. 2002. F-actin-like filaments formed by plasmid segregation protein ParM. EMBO J, 21, 6935-43
VAN DER LEE, R., LANG, B., KRUSE, K., GSPONER, J., SANCHEZ DE GROOT, N., HUYNEN, M. A., MATOUSCHEK, A., FUXREITER, M. & BABU, M. M. 2014. Intrinsically disordered segments affect protein half-life in the cell and during evolution. Cell Rep, 8, 1832-44.
VECCHIARELLI, A. G., HAN, Y. W., TAN, X., MIZUUCHI, M., GHIRLANDO, R., BIERTUMPFEL, C., FUNNELL, B. E. & MIZUUCHI, K. 2010. ATP control of dynamic P1 ParA-DNA interactions: A key role for the nucleoid in plasmid partition. Mol Microbiol, 78, 78-91.
VECCHIARELLI, A. G., HWANG, L. C. & MIZUUCHI, K. 2013. Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. Proc Natl Acad Sci U S A, 110, E1390-97.
VECCHIARELLI, A. G., MIZUUCHI, K. & FUNNELL, B. E. 2012. Surfing biological surfaces: Exploiting the nucleoid for partition and transport in bacteria. Mol Microbiol, 86, 513-23.
VECCHIARELLI, A. G., NEUMAN, K. C. & MIZUUCHI, K. 2014. A propagating ATPase gradient drives transport of surface-confined cellular cargo. Proc Natl Acad Sci U S A, 111, 4880-5.
VECCHIARELLI, A. G., TAYLOR, J. A. & MIZUUCHI, K. 2015. Reconstituting ParA/ParB-mediated transport of DNA cargo. Methods Cell Biol, 128, 243-69.
VUZMAN, D., AZIA, A. & LEVY, Y. 2010. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. J Mol Biol, 396, 674-84.
WAKE, R. G. & ERRINGTON, J. 1995. Chromosome Partitioning in Bacteria. Ann Rev Gen, 29, 41-67.
WALKER, J. E., SARASTE, M., RUNSWICK, M. J. & GAY, N. J. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO Journal, 1, 945-51.
WALSH, I., MARTIN, A. J. M., DI DOMENICO, T. & TOSATTO, S. C. E. 2012. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics, 28, 503-09.
WARD, J. J., MCGUFFIN, L. J., BRYSON, K., BUXTON, B. F. & JONES, D. T. 2004. The DISOPRED server for the prediction of protein disorder. Bioinformatics, 20, 2138-39.
WEISEMANN, R., RUTERJANS, H. & BERMEL, W. 1993. 3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. J Biomol NMR, 3, 113-20.
WELLS, J. A. & MCCLENDON, C. L. 2007. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature, 450, 1001-9.
WEN, J., ARAKAWA, T. & PHILO, J. S. 1996. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Anal Biochem, 240, 155-66.
WHITMORE, L. & WALLACE, B. A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res, 32, W668-73.
WIEDERSTEIN, M. & SIPPL, M. J. 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res, 35, W407-W410.
WILKINSON, B. M., ESNAULT, Y., CRAVEN, R. A., SKIBA, F., FIESCHI, J., K'EPES, F. & STIRLING, C. J. 1997. Molecular architecture of the ER translocase probed by chemical crosslinking of Sss1p to complementary fragments of Sec61p. EMBO J, 16, 4549-59.

WILLIAMS, D. R., MACARTNEY, D. P. & THOMAS, C. M. 1998. The partitioning activity of the RK2 central control region requires only *incC*, *korB* and KorB-binding site O(B)3 but other KorB-binding sites form destabilizing complexes in the absence of O(B)3. Microbiology, 144 (Pt 12), 3369-78.

WILLIAMS, D. R., MOTALLEBI-VESHAREH, M. & THOMAS, C. M. 1993. Multifunctional repressor KorB can block transcription by preventing isomerization of RNA polymerase-promoter complexes. Nucleic Acids Res, 21, 1141-48.

WRIGHT, P. E. & DYSON, H. J. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol, 293, 321-31.

WU, S. & ZHANG, Y. 2007. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res, 35, 3375-82.

WU, S. & ZHANG, Y. 2008. MUSTER: Improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins, 72, 547-56.

XU, D., XU, Y. & UBERBACHER, E. C. 2000. Computational tools for protein modeling. Curr Protein Pept Sci, 1, 1-21.

XU, H., HSU, P. H., ZHANG, L., TSAI, M. D. & FREITAS, M. A. 2010. Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry. J Proteome Res, 9, 3384-93.

XU, J., LI, M., LIN G., KIM, D., & XU, Y. 2003. Protein structure prediction by linear programming. Pac Symp Biocomput. 2003, 264-75.

XUE, B., DUNBRACK, R. L., WILLIAMS, R. W., DUNKER, A. K. & UVERSKY, V. N. 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta, 1804, 996-1010.

YAMAOKA, Y. 2008. *Helicobacter pylori*: molecular genetics and cellular biology, Wymondham, Caister Academic.

YAN, R., XU, D., YANG, J., WALKER, S. & ZHANG, Y. 2013. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. Sci Rep, 3.

YANG, J. & ZHANG, Y. 2015. Protein Structure and Function Prediction Using I-TASSER. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.], 52, 5.8.1-5.815.

YANG, Z. R., THOMSON, R., MCNEIL, P. & ESNOUF, R. M. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics, 21, 3369-76.

ZAMPINI, M., DEROME, A., BAILEY, S. E., BARILLA, D. & HAYES, F. 2009. Recruitment of the ParG segregation protein to different affinity DNA sites. J Bacteriol, 191, 3832-41.

ZEEB, M. & BALBACH, J. 2003. Single-stranded DNA binding of the cold-shock protein CspB from *Bacillus subtilis*: NMR mapping and mutational characterization. Protein Sci, 12**,** 112-23.

ZELTER, A., HOOPMANN, M. R., VERNON, R., BAKER, D., MACCOSS, M. J. & DAVIS, T. N. 2010. Isotope signatures allow identification of chemically cross-linked peptides by Mass Spectrometry: A novel method to determine interresidue distances in protein structures through cross-linking. J Proteome Res, 9, 3583-89.

ZEYMER, C., WERBECK, N. D., ZIMMERMANN, S., REINSTEIN, J. & HANSEN, D. F. 2016. Characterizing active site conformational heterogeneity along the trajectory of an enzymatic phosphoryl transfer reaction. Angew Chem Int Ed Engl, 55, 11533-7.

ZHANG, H. & SCHUMACHER, M. A. 2017. Structures of partition protein ParA with nonspecific DNA and ParB effector reveal molecular insights into principles governing Walker-box DNA segregation. Genes Dev, 31, 481-92.

ZHANG, W., YANG, J., HE, B., WALKER, S. E., ZHANG, H., GOVINDARAJOO, B., VIRTANEN, J., XUE, Z., SHEN, H. B. & ZHANG, Y. 2016. Integration of QUARK and I-TASSER for Ab Initio protein structure prediction in CASP11. Proteins, 84 Suppl 1, 76-86.

ZHAO, H., BRAUTIGAM, C. A., GHIRLANDO, R. & SCHUCK, P. 2013. Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. Curr Protoc Protein Sci, Chapter 20, Unit 20, 12.

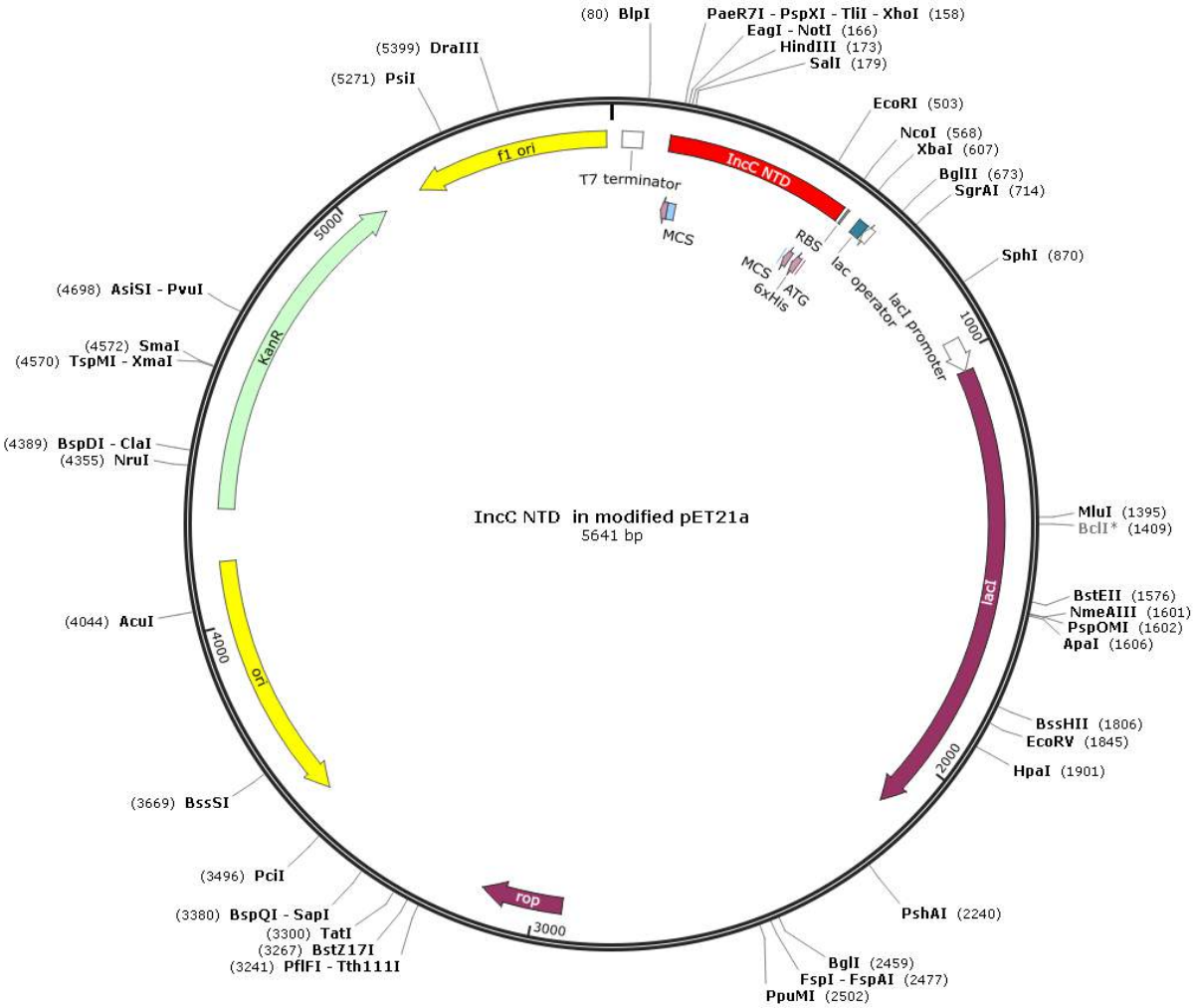
ZHOU, P. & WAGNER, G. 2010. Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. J Biomol NMR, 46, 23-31.
ZHOU, P., LUGOVSKOY, A. A. & WAGNER, G. 2001. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR, 20, 11-4.
ZIELENKIEWICZ, U. & CEGLOWSKI, P. 2001. Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. Acta Biochim Pol, 48, 1003-23.

Appendix 1

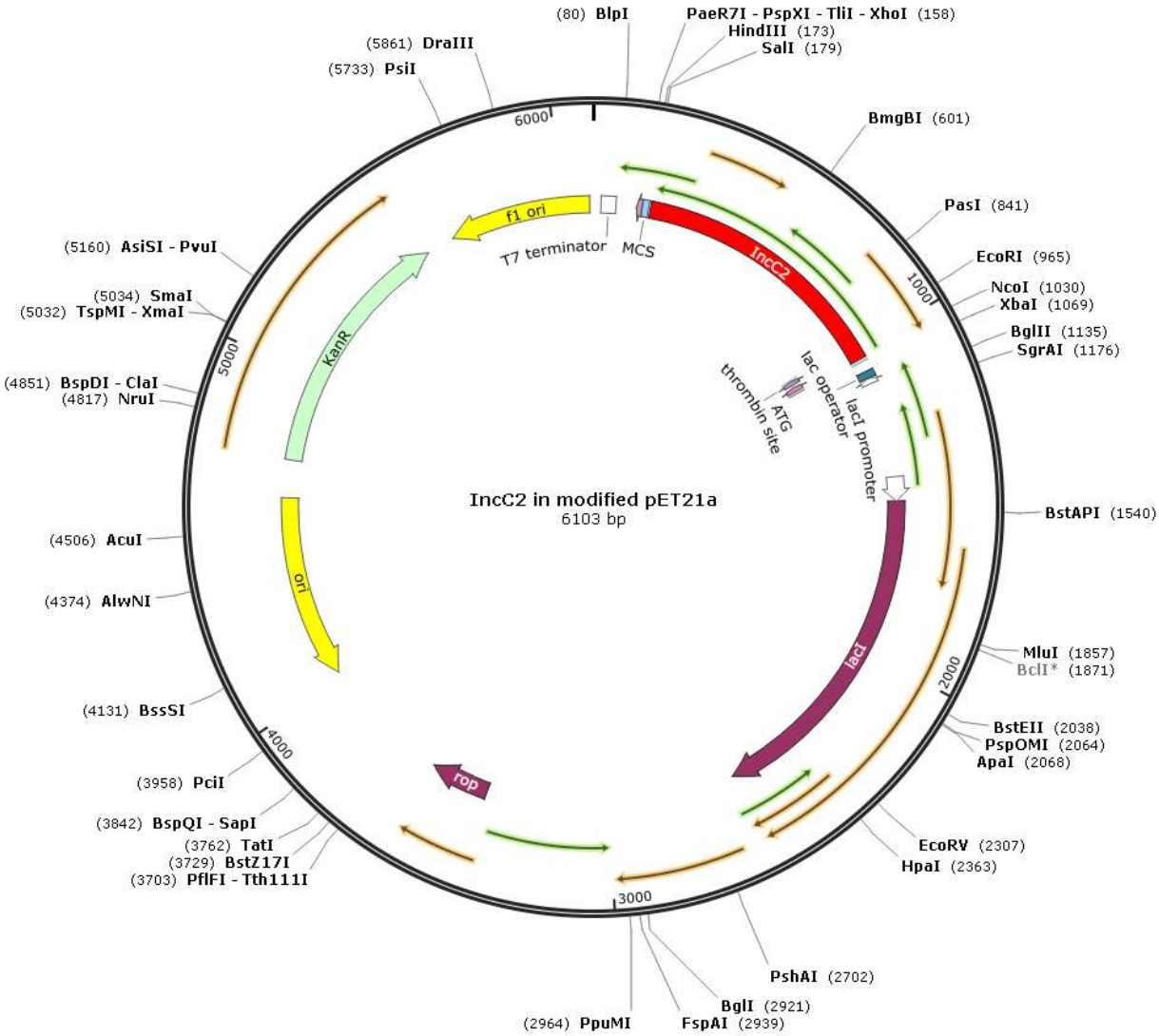
Plasmid Maps

The maps for the plasmids used in the thesis are given below: All maps have been created by SnapGene Ver. 2.3.2

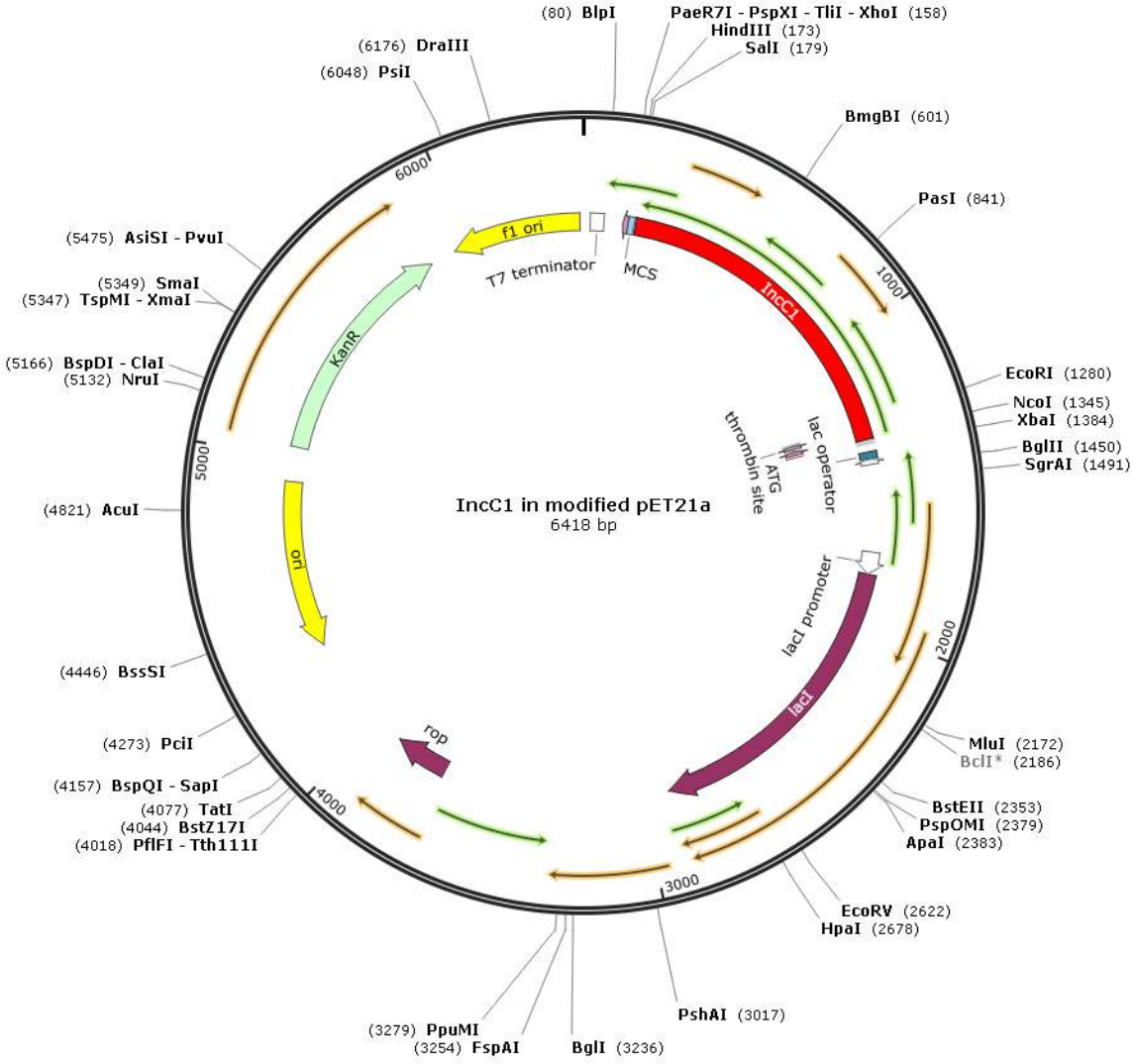
pET21a-incC-ntd



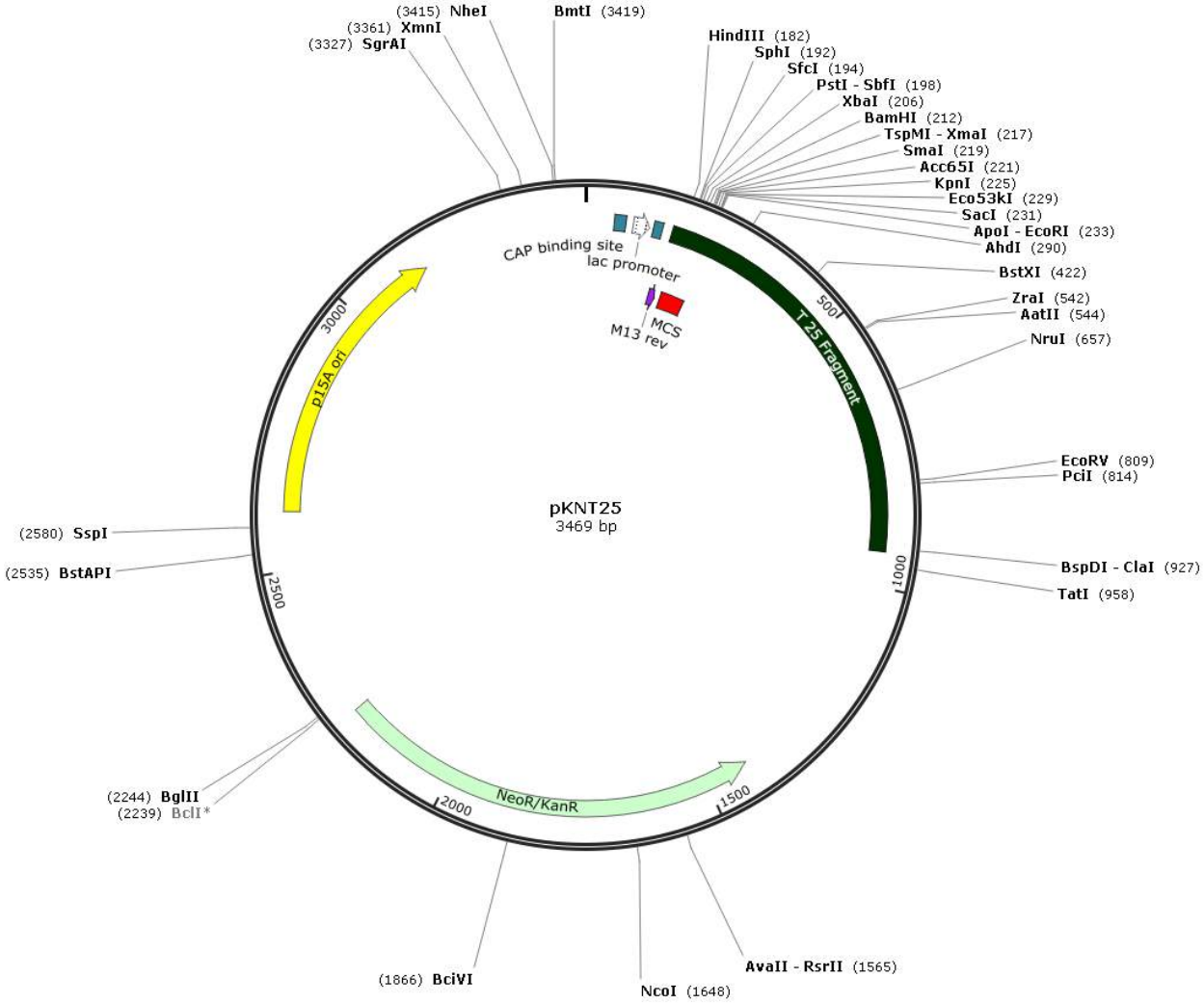
pET21a-incC2



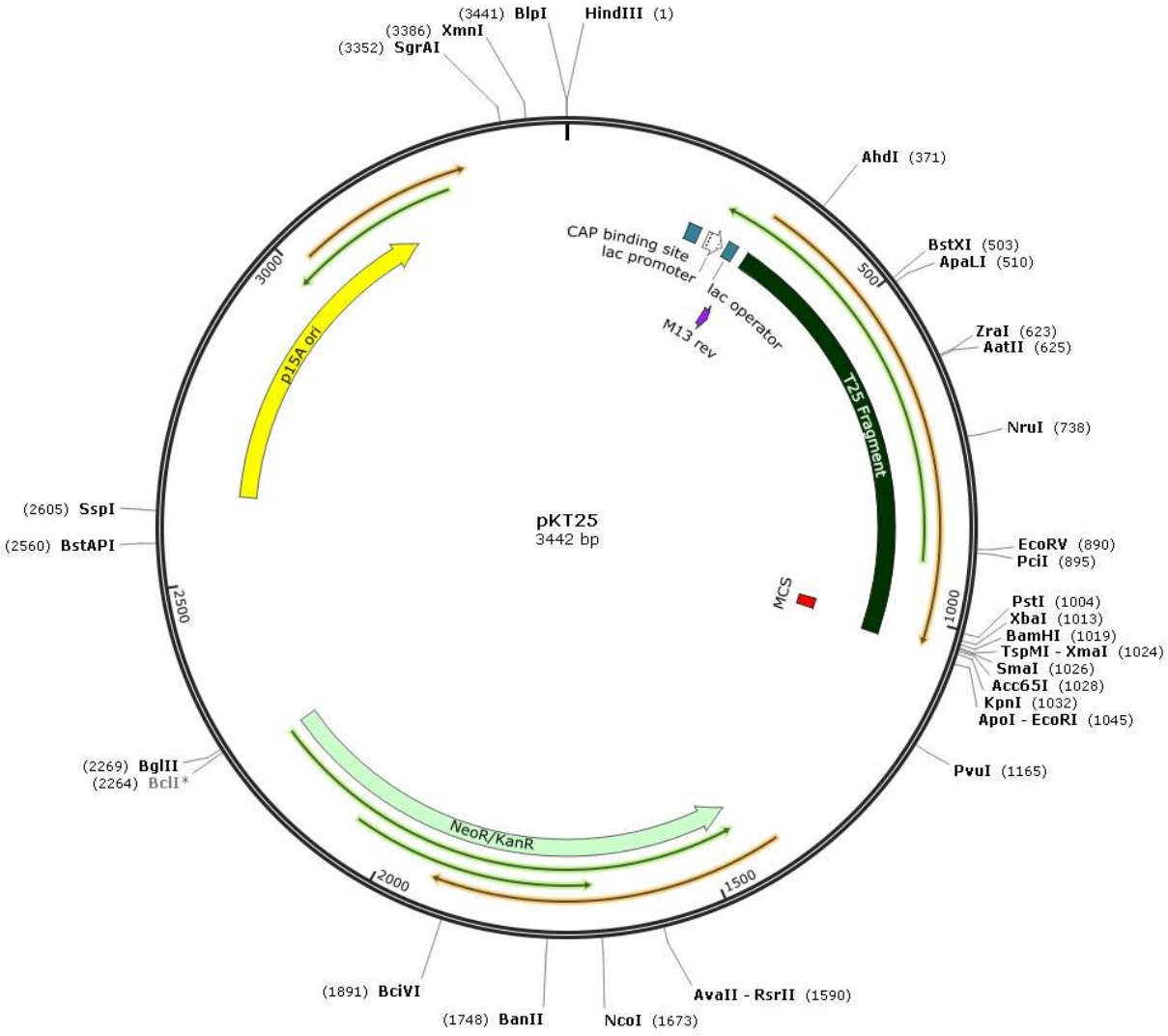
pET21a-inc1


(80) BlpI
PaeR7I - PspXI - TliI - XhoI (158)
HindIII (173)
SalI (179)
(6176) DraIII
(6048) PsiI
BmgBI (601)
PasI (841)
(5475) AsiSI - PvuI
(5349) SmaI
(5347) TspMI - XmaI
(5166) BspDI - ClaI
(5132) NruI
EcoRI (1280)
NcoI (1345)
XbaI (1384)
BglII (1450)
SgrAI (1491)
(4821) AcuI
(4446) BssSI
(4273) PciI
(4157) BspQI - SapI
(4077) TatI
(4044) BstZ17I
(4018) PflFI - Tth111I
MluI (2172)
BclI* (2186)
BstEII (2353)
PspOMI (2379)
ApaI (2383)
EcoRV (2622)
HpaI (2678)
PshAI (3017)
(3279) PpuMI
(3254) FspAI
BglI (3236)
6000
1000
2000
3000
4000
5000
f1 ori
T7 terminator
MCS
IncC1
KanR
thrombin site
ATG
lac operator
lacI promoter
lacI
rop
ori
IncC1 in modified pET21a
6418 bp

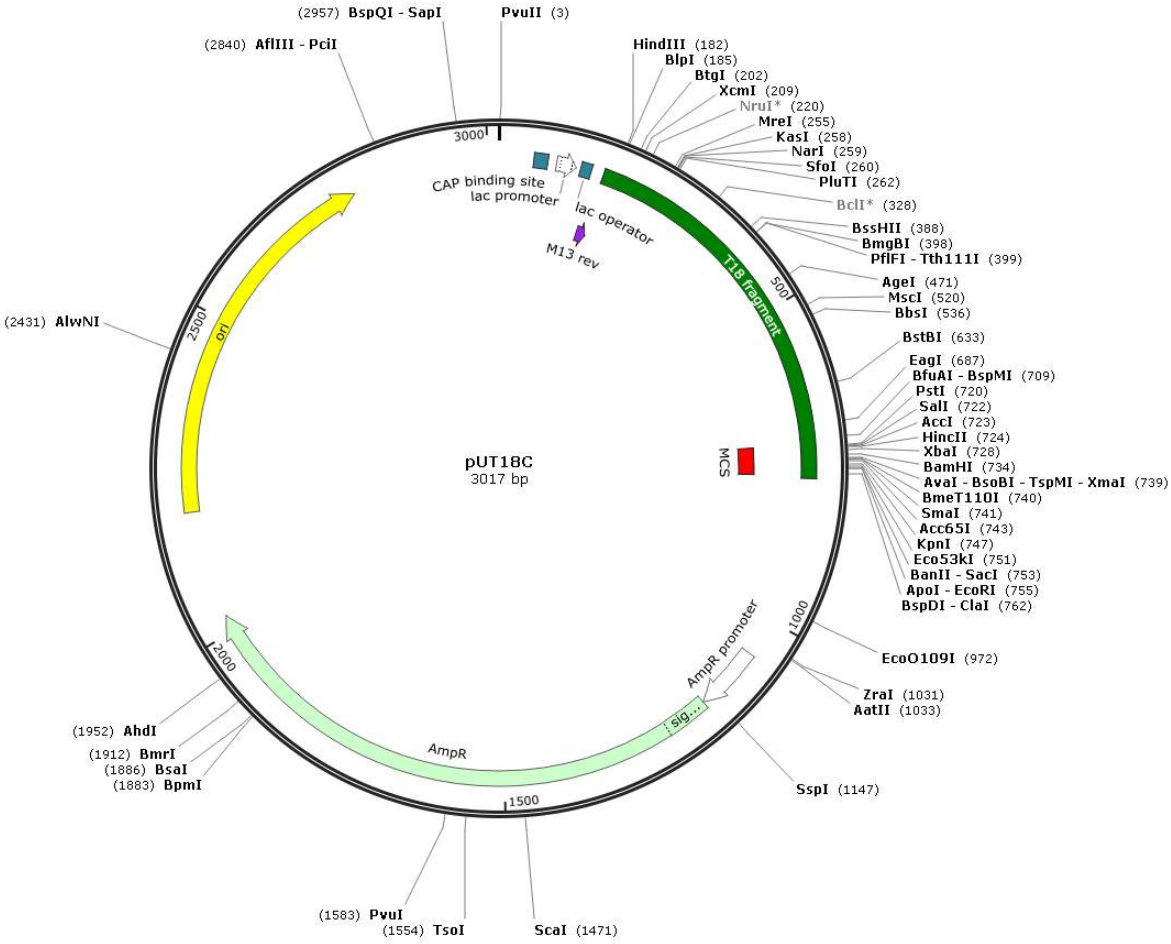
pKNT25


pKNT25
3469 bp
(3415) NheI
BmtI (3419)
(3361) XmnI
(3327) SgrAI
HindIII (182)
SphI (192)
SfcI (194)
PstI - SbfI (198)
XbaI (206)
BamHI (212)
TspMI - XmaI (217)
SmaI (219)
Acc65I (221)
KpnI (225)
Eco53kI (229)
SacI (231)
ApoI - EcoRI (233)
AhdI (290)
BstXI (422)
ZraI (542)
AatII (544)
NruI (657)
EcoRV (809)
PciI (814)
BspDI - ClaI (927)
TatI (958)
(2580) SspI
(2535) BstAPI
(2244) BglII
(2239) BclI*
(1866) BciVI
NcoI (1648)
AvaII - RsrII (1565)
CAP binding site
lac promoter
MCS
M13 rev
T25 Fragment
p15A ori
NeoR/KanR
500
1000
1500
2000
2500
3000

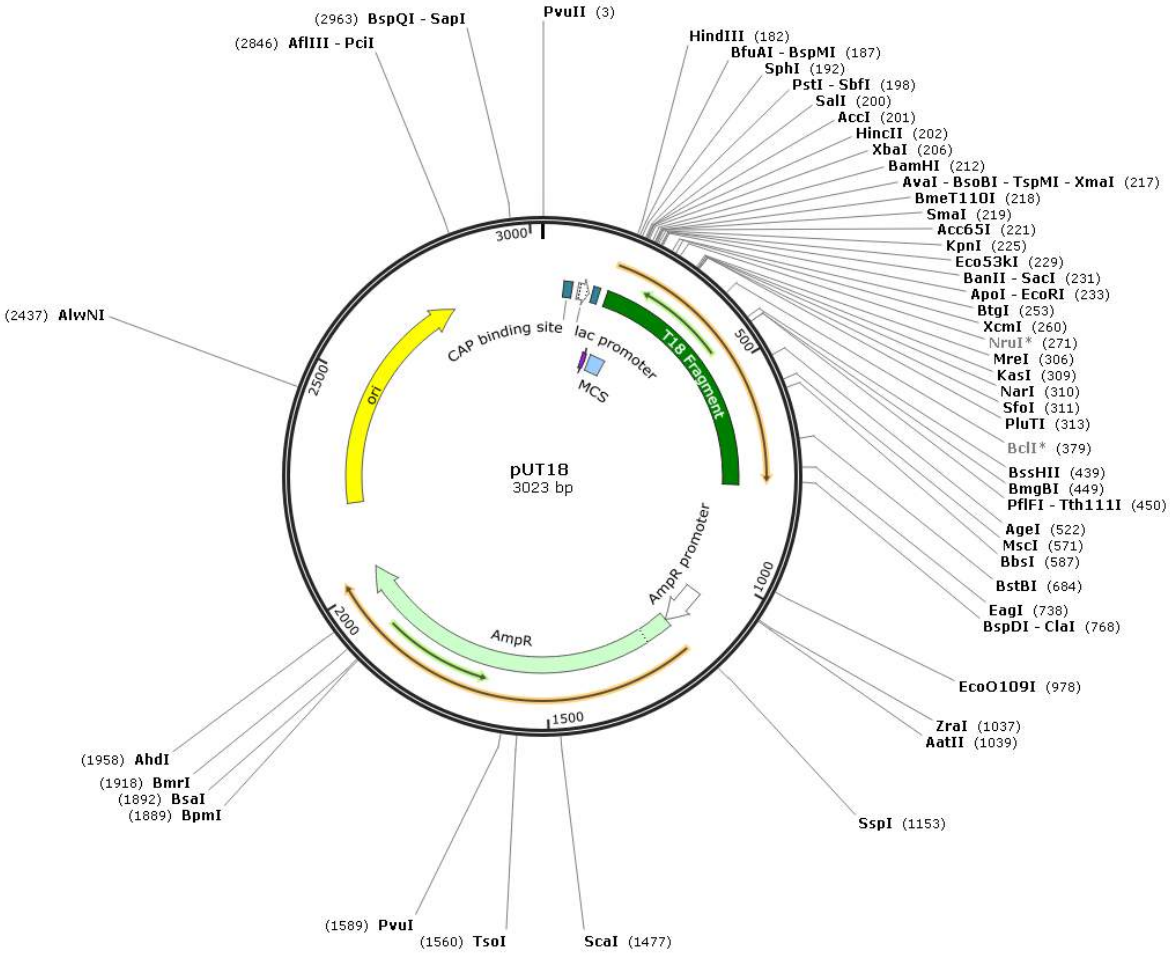
pKT25


pKT25
3442 bp
(3441) BlpI
HindIII (1)
(3386) XmnI
(3352) SgrAI
AhdI (371)
BstXI (503)
ApaLI (510)
ZraI (623)
AatII (625)
NruI (738)
EcoRV (890)
PciI (895)
PstI (1004)
XbaI (1013)
BamHI (1019)
TspMI - XmaI (1024)
SmaI (1026)
Acc65I (1028)
KpnI (1032)
ApoI - EcoRI (1045)
PvuI (1165)
AvaII - RsrII (1590)
NcoI (1673)
(1748) BanII
(1891) BciVI
(2264) BclI*
(2269) BglII
(2560) BstAPI
(2605) SspI
CAP binding site
lac promoter
lac operator
M13 rev
T25 Fragment
MCS
p15A ori
NeoR/KanR
500
1000
1500
2000
2500
3000

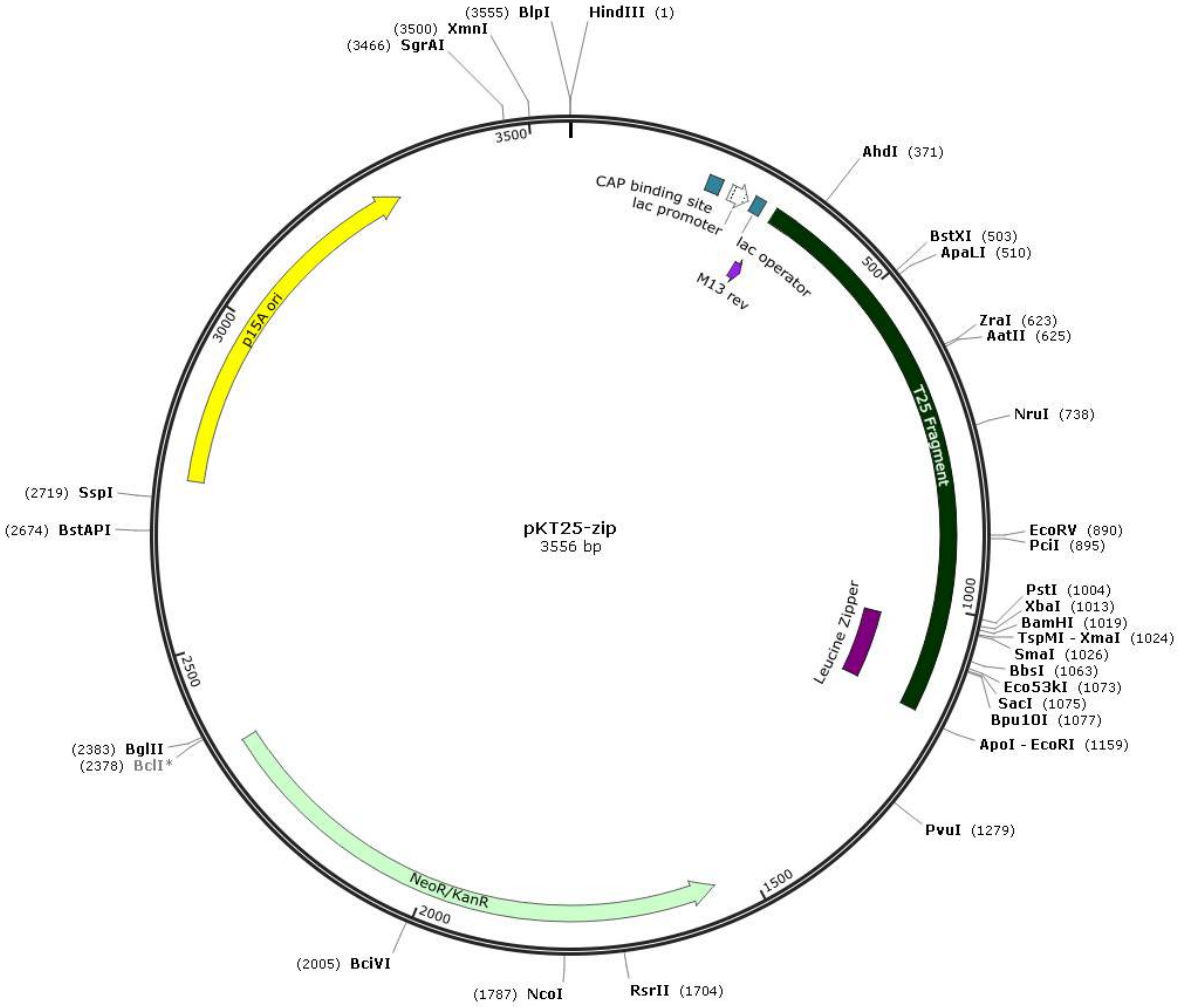
pUT18c


pUT18C
3017 bp
(2957) BspQI - SapI
PvuII (3)
(2840) AflIII - PciI
HindIII (182)
BlpI (185)
BtgI (202)
XcmI (209)
NruI* (220)
MreI (255)
KasI (258)
NarI (259)
SfoI (260)
PluTI (262)
BclI* (328)
BssHII (388)
BmgBI (398)
PflFI - Tth111I (399)
AgeI (471)
MscI (520)
BbsI (536)
BstBI (633)
EagI (687)
BfuAI - BspMI (709)
PstI (720)
SalI (722)
AccI (723)
HincII (724)
XbaI (728)
BamHI (734)
AvaI - BsoBI - TspMI - XmaI (739)
BmeT110I (740)
SmaI (741)
Acc65I (743)
KpnI (747)
Eco53kI (751)
BanII - SacI (753)
ApoI - EcoRI (755)
BspDI - ClaI (762)
EcoO109I (972)
ZraI (1031)
AatII (1033)
SspI (1147)
(1583) PvuI
(1554) TsoI
ScaI (1471)
(1952) AhdI
(1912) BmrI
(1886) BsaI
(1883) BpmI
(2431) AlwNI
CAP binding site
lac promoter
lac operator
M13 rev
T18 fragment
MCS
ori
AmpR promoter
sig...
AmpR
500
1000
1500
2000
2500
3000

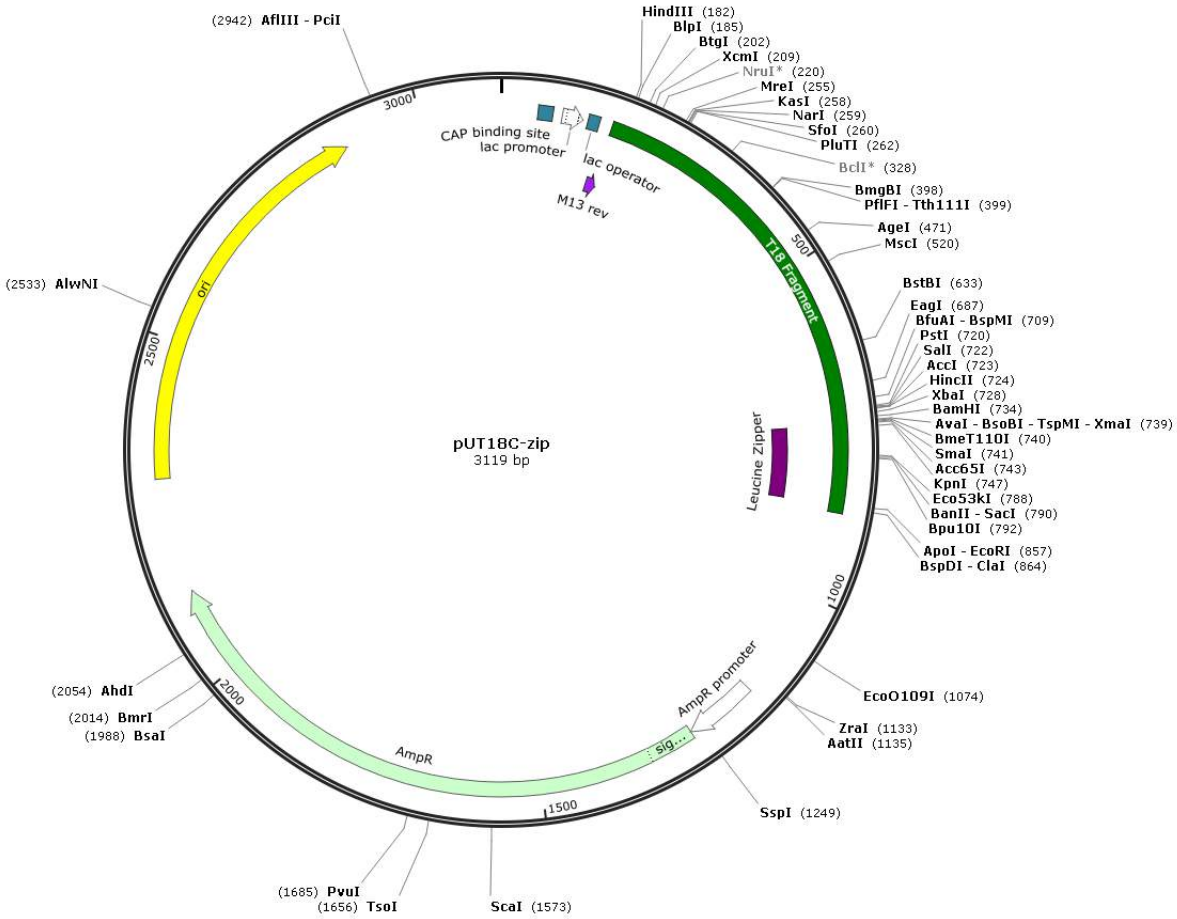
pUT18


pUT18
3023 bp
CAP binding site
lac promoter
T18 Fragment
MCS
ori
AmpR promoter
AmpR
500
1000
1500
2000
2500
3000
PvuII (3)
HindIII (182)
BfuAI - BspMI (187)
SphI (192)
PstI - SbfI (198)
SalI (200)
AccI (201)
HincII (202)
XbaI (206)
BamHI (212)
AvaI - BsoBI - TspMI - XmaI (217)
BmeT110I (218)
SmaI (219)
Acc65I (221)
KpnI (225)
Eco53kI (229)
BanII - SacI (231)
ApoI - EcoRI (233)
BtgI (253)
XcmI (260)
NruI* (271)
MreI (306)
KasI (309)
NarI (310)
SfoI (311)
PluTI (313)
BclI* (379)
BssHII (439)
BmgBI (449)
PflFI - Tth111I (450)
AgeI (522)
MscI (571)
BbsI (587)
BstBI (684)
EagI (738)
BspDI - ClaI (768)
EcoO109I (978)
ZraI (1037)
AatII (1039)
SspI (1153)
ScaI (1477)
TsoI (1560)
PvuI (1589)
BpmI (1889)
BsaI (1892)
BmrI (1918)
AhdI (1958)
AlwNI (2437)
AflIII - PciI (2846)
BspQI - SapI (2963)

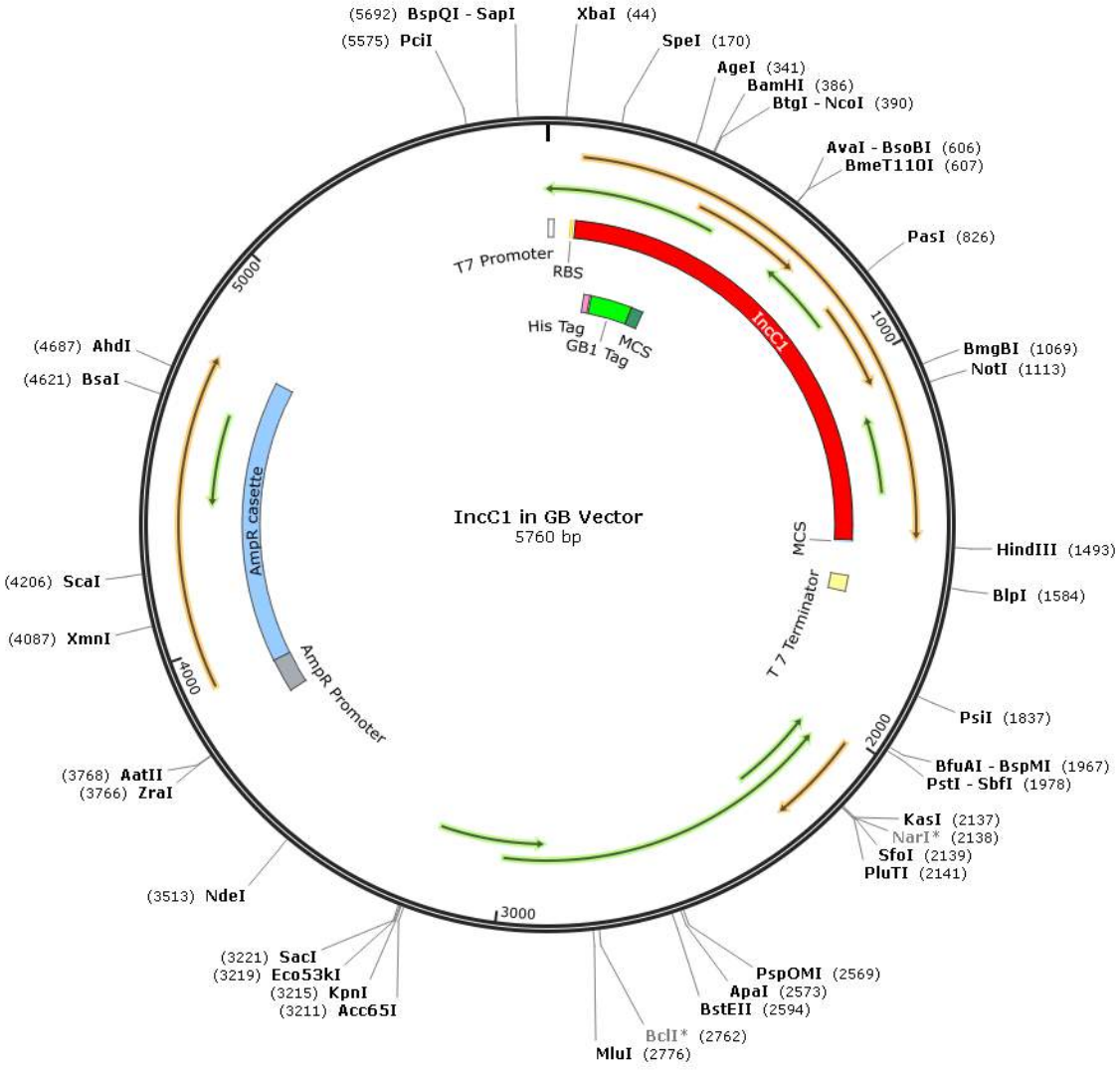
pKT25-zip


pKT25-zip
3556 bp
(3555) BlpI
HindIII (1)
(3500) XmnI
(3466) SgrAI
3500
AhdI (371)
CAP binding site
lac promoter
lac operator
M13 rev
BstXI (503)
ApaLI (510)
500
ZraI (623)
AatII (625)
T25 Fragment
NruI (738)
3000
P15A ori
(2719) SspI
(2674) BstAPI
EcoRV (890)
PciI (895)
PstI (1004)
XbaI (1013)
BamHI (1019)
TspMI - XmaI (1024)
SmaI (1026)
BbsI (1063)
Eco53kI (1073)
SacI (1075)
Bpu10I (1077)
1000
Leucine Zipper
2500
(2383) BglII
(2378) BclI*
ApoI - EcoRI (1159)
PvuI (1279)
NeoR/KanR
1500
2000
(2005) BciVI
(1787) NcoI
RsrII (1704)

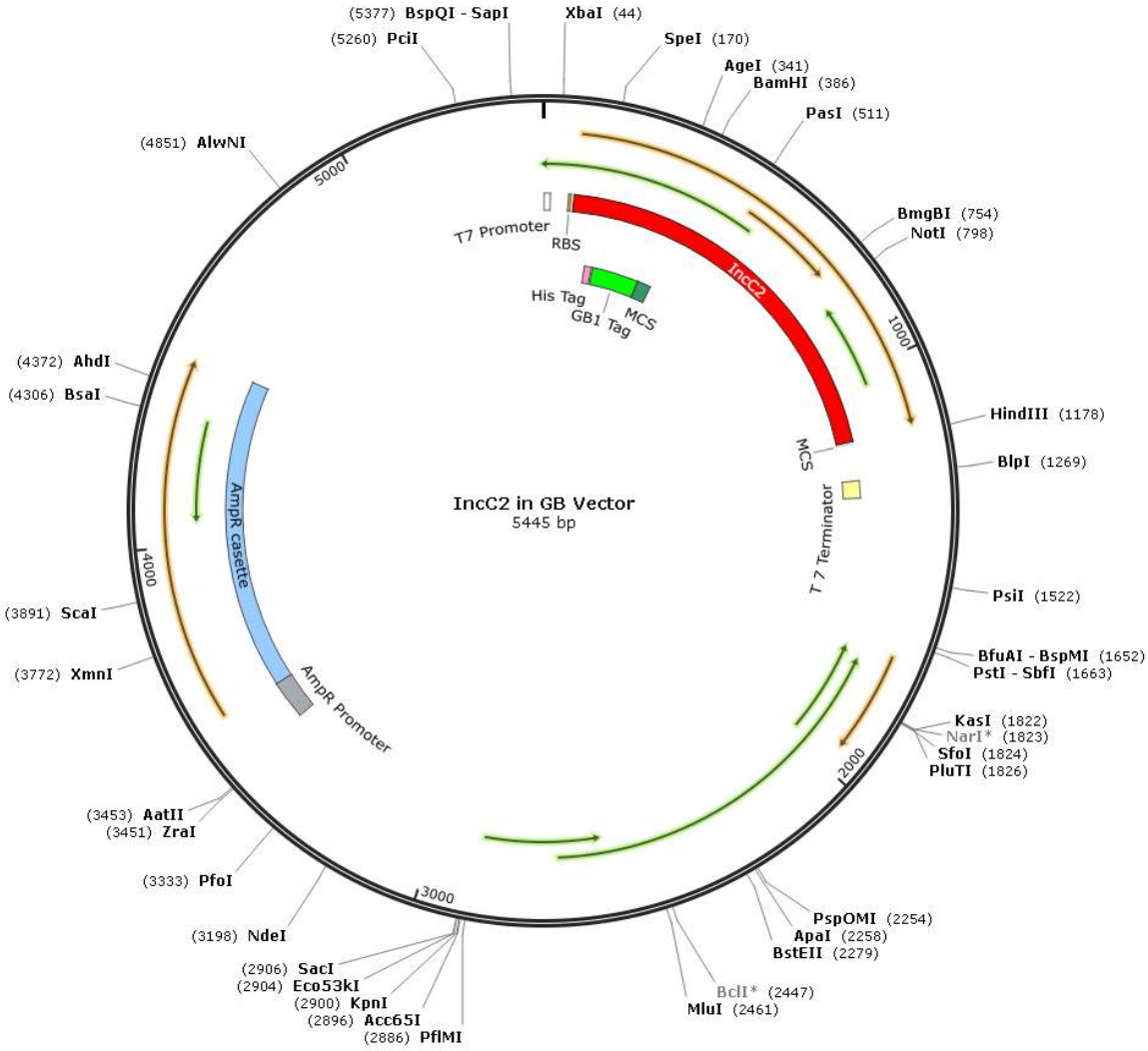
pUT18c-zip


pUT18C-zip
3119 bp
HindIII (182)
BlpI (185)
BtgI (202)
XcmI (209)
NruI* (220)
MreI (255)
KasI (258)
NarI (259)
SfoI (260)
PluTI (262)
BclI* (328)
BmgBI (398)
PflFI - Tth111I (399)
AgeI (471)
MscI (520)
BstBI (633)
EagI (687)
BfuAI - BspMI (709)
PstI (720)
SalI (722)
AccI (723)
HincII (724)
XbaI (728)
BamHI (734)
AvaI - BsoBI - TspMI - XmaI (739)
BmeT110I (740)
SmaI (741)
Acc65I (743)
KpnI (747)
Eco53kI (788)
BanII - SacI (790)
Bpu10I (792)
ApoI - EcoRI (857)
BspDI - ClaI (864)
EcoO109I (1074)
ZraI (1133)
AatII (1135)
SspI (1249)
ScaI (1573)
(1685) PvuI
(1656) TsoI
(2054) AhdI
(2014) BmrI
(1988) BsaI
(2533) AlwNI
(2942) AflIII - PciI
CAP binding site
lac promoter
lac operator
M13 rev
T18 fragment
Leucine Zipper
AmpR promoter
AmpR
ori
500
1000
1500
2000
2500
3000

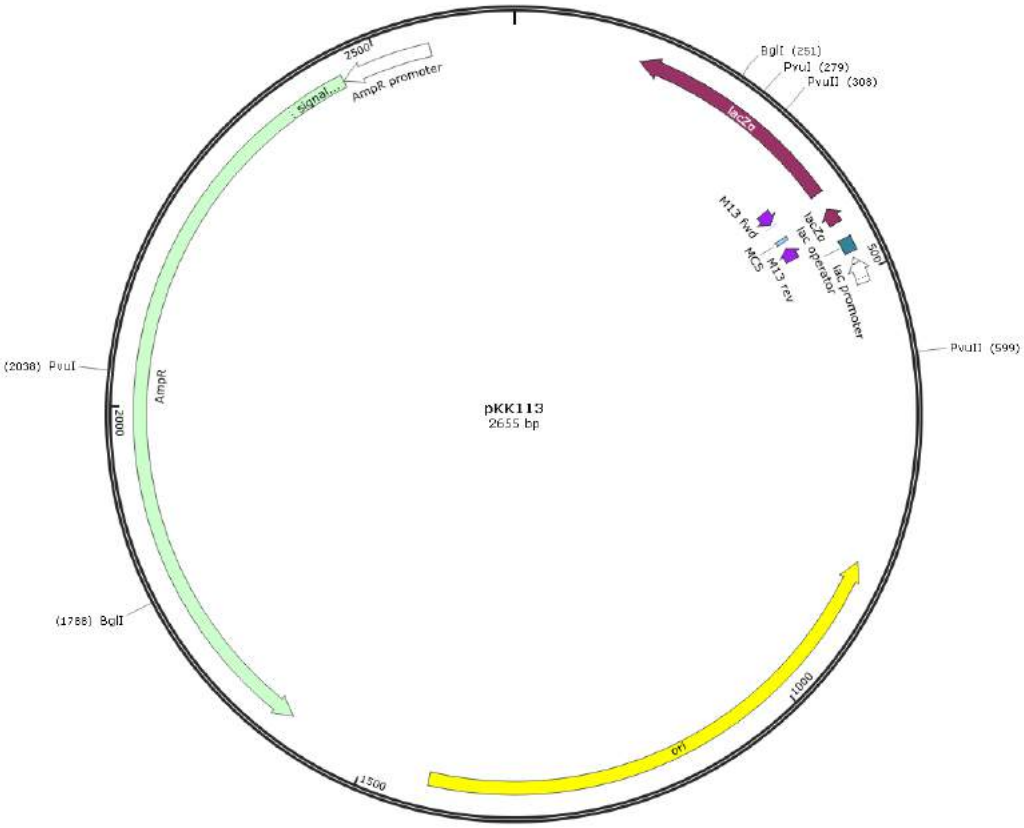
IncC1-GB


(5692) BspQI - SapI
XbaI (44)
(5575) PciI
SpeI (170)
AgeI (341)
BamHI (386)
BtgI - NcoI (390)
AvaI - BsoBI (606)
BmeT110I (607)
PasI (826)
T7 Promoter
RBS
His Tag
GB1 Tag
MCS
IncC1
5000
1000
BmgBI (1069)
NotI (1113)
(4687) AhdI
(4621) BsaI
AmpR casette
IncC1 in GB Vector
5760 bp
MCS
HindIII (1493)
(4206) ScaI
BlpI (1584)
(4087) XmnI
T 7 Terminator
AmpR Promoter
4000
2000
PsiI (1837)
BfuAI - BspMI (1967)
(3768) AatII
(3766) ZraI
PstI - SbfI (1978)
KasI (2137)
NarI* (2138)
SfoI (2139)
PluTI (2141)
(3513) NdeI
3000
(3221) SacI
(3219) Eco53kI
(3215) KpnI
(3211) Acc65I
PspOMI (2569)
ApaI (2573)
BstEII (2594)
BclI* (2762)
MluI (2776)

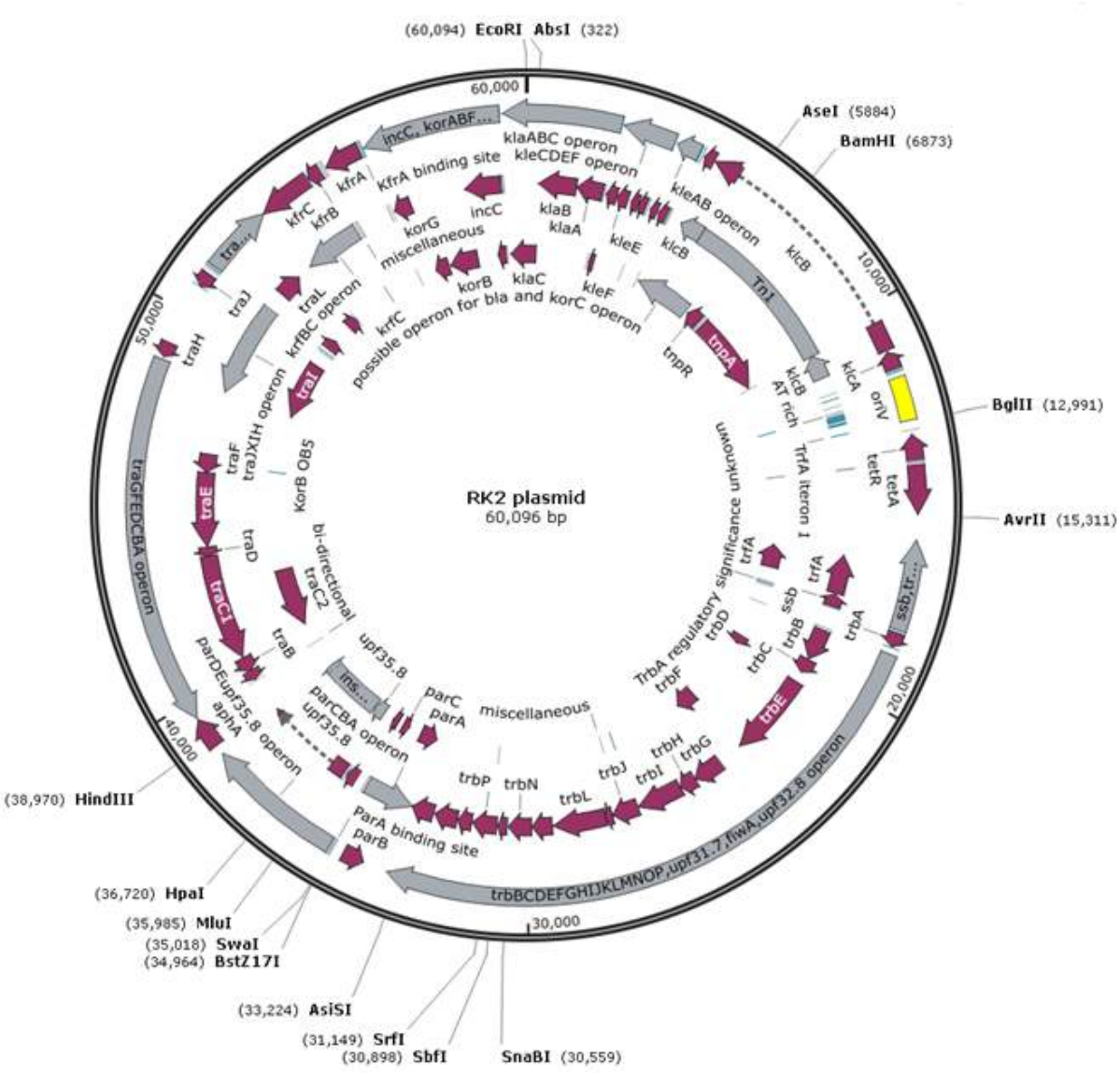
IncC2-GB


(5377) BspQI - SapI
XbaI (44)
(5260) PciI
SpeI (170)
AgeI (341)
BamHI (386)
PasI (511)
(4851) AlwNI
5000
T7 Promoter
RBS
IncC2
BmgBI (754)
NotI (798)
His Tag
GB1 Tag
MCS
1000
(4372) AhdI
(4306) BsaI
HindIII (1178)
BlpI (1269)
MCS
IncC2 in GB Vector
5445 bp
T7 Terminator
AmpR casette
4000
(3891) ScaI
PsiI (1522)
(3772) XmnI
BfuAI - BspMI (1652)
PstI - SbfI (1663)
AmpR Promoter
KasI (1822)
NarI* (1823)
SfoI (1824)
PluTI (1826)
2000
(3453) AatII
(3451) ZraI
(3333) PfoI
3000
PspOMI (2254)
ApaI (2258)
BstEII (2279)
(3198) NdeI
BclI* (2447)
MluI (2461)
(2906) SacI
(2904) Eco53kI
(2900) KpnI
(2896) Acc65I
(2886) PflMI

PKK113


pKK113
2655 bp
BglI (251)
PvuI (279)
PvuII (308)
lacZα
M13 fwd
MCS
M13 rev
lacZα
lac operator
lac promoter
500
PvuII (599)
ori
1000
1500
(1788) BglI
AmpR
2000
(2038) PvuI
signal
AmpR promoter
2500

Plasmid RK2


RK2 plasmid
60,096 bp
(60,094) EcoRI AbsI (322)
60,000
AseI (5884)
BamHI (6873)
10,000
BglII (12,991)
AvrII (15,311)
20,000
30,000
(33,224) AsiSI
(31,149) SrfI
(30,898) SbfI
SnaBI (30,559)
(35,018) SwaI
(34,964) BstZ17I
(35,985) MluI
(36,720) HpaI
(38,970) HindIII
40,000
50,000
incC, korABF...
klaABC operon
kleCDEF operon
kleAB operon
KfrA binding site
kfrA
kfrC
kfrB
korG
miscellaneous
incC
klaB
klaA
kleE
klcB
Tn1
tnpA
tnpR
klcA
oriV
AT rich
TrfA iteron 1
tetR
tetA
unknown significance
trfA
ssb
trfA
ssb;tr...
trbA
trbB
trbC
trbD
TrbA regulatory
trbF
trbE
trbH
trbG
trbI
trbJ
trbL
trbN
trbP
miscellaneous
parC
parA
ins...
upf35.8
parCBA operon
upf35.8
parDEupf35.8
aphA
operon
parA binding site
parB
trbBCDEFGHIJKLMNOP,upf31.7,fliwA,upf32.8 operon
traG
traF
traJXIH operon
traL
traI
traH
traJ
traL
krfBC operon
krfC
possible operon for bla and korC operon
KorB OBS
bi-directional
traC2
traD
traC1
traB
traE
traGFEDCBA operon

Appendix 2

Tris Glycine SDS- PAGE

For the 12 % gel following solutions were used for resolving gel.

- mL acrylamide (30 % w/v acrylamide, 0.8 % w/v bis-acrylamide stock)
- 2.5 mL lower Tris (1.5M Tris-HCl pH 8.8, 0.4 % SDS)
- 3.5 mL distilled water
- 150 µL 10 % APS (Ammonium Persulphate)
- 6 µL TEMED (N, N`, N``, N```- Tetramethylethylenediamine)

Resolving gel was poured (3.7 mL) quickly after adding TEMED between the glass plates assembled at gel polymerizing assembly. The gel was overlaid with water saturated butanol and allowed to polymerise completely.

The upper stacking gel (3 %) was prepared using the following solutions

- 0.54 mL of acrylamide (30 % w/v acrylamide, 0.8 % w/v bis-acrylamide stock)
- 1.25 mL upper Tris (0.5 M Tris-HCl pH 6.8, 0.4 % SDS)
- 3.24 mL distilled water
- 100 µL 10 % APS (Ammonium Persulphate)
- 5 µL TEMED (N, N`, N``, N```- Tetramethylethylenediamine)

Stacking gel was poured on the top of the resolving gel (once the isopropanol had been removed and washed). The comb was inserted between the spacers at the top of the plate and gel was allowed to polymerise for half an hour.

- **SDS loading buffer or urea loading buffer**

(0.4 mL of 1 % bromophenol blue, 2.5 mL 1 M Tris.HCl pH 6.8, 2 mL glycerol, 5 M urea, 5 mL distilled water and 0.2 mL β-mercaptoethanol).

- **Running Tris-Glycine buffer**

24 mM Tris base, 250 mM glycine, 0.1% w/v SDS) was used as running buffer.

Tricine SDS PAGE

Gel buffer (4X) used to cast the gel contained 3 M Tris-HCl and 0.3 % SDS pH 8.45. Anode buffer (10X) contained 1 M Tris HCl pH 8.9 while cathode buffer (10X) contained 1 M Tris, 1 M Tricine, 1 % SDS pH 8.25 (all solutions were checked for pH and kept at room temperature).

Resolving gel (15 %) was used to separate small proteins and had following composition

- 5 mL acrylamide (30% w/v acrylamide, 0.8% w/v bis-acrylamide stock)
- 2.5 mL gel buffer (4X)
- 2 mL 50 % Glycerol
- 0.5 mL distilled water
- 120 µL 10 % APS (Ammonium Persulphate)
- 6 µL TEMED (N, N`, N``, N```- Tetramethylethylenediamine)

Stacking gel was used to stack protein sample and had following composition

- 54 µL of acrylamide (30 % w/v acrylamide, 0.8 % w/v bis-acrylamide stock)
- 1.25 mL gel buffer
- 3.24 mL distilled water
- 80 µL 10 % APS (Ammonium Persulphate)
- 5 µL TEMED (N, N`, N``, N```- Tetramethylethylenediamine)

Resolving gel was first poured and polymerized followed by pouring the stacking gel and comb was inserted to make wells to load protein samples.

Reference

1. SCHAGGER, H. 2006. Tricine-SDS-PAGE. Nat Protoc, 1, 16-22.

Appendix 3

Protein Sequence and properties

His Tag – at N-terminal of IncC NTD, IncC1, IncC2 and KorB

1- MGSSHHHHHHSSGLVPRGSHSEF -23

>IncC NTD

```
         10          20          30          40          50          60
MGVIHEETAY RKPVPGGDPG AGSGAADHRD SAGRLSRWEA TGDVRNVAGT DQGRSVASGA
         70          80          90         100    105
SRVGRVRGQE LARGVRAGNG GSAGTSGVHR PEVGSGRQEK TGNQT
```

>IncC1

```
         10          20          30          40          50          60
MGVIHEETAY RKPVPGGDPG AGSGAADHRD SAGRLSRWEA TGDVRNVAGT DQGRSVASGA
         70          80          90         100         110         120
SRVGRVRGQE LARGVRAGNG GSAGTSGVHR PEVGSGRQEK TGNQTMKTLV TANQKGGVGK
        130         140         150         160         170         180
TSTLVHLAFD FFERGLRVAV IDLDPQGNAS YTLKDFATGL HASKLFGAVP AGGWTETAPA
        190         200         210         220         230         240
AGDGQAARLA LIESNPVLAN AERLSLDDAR ELFGANIKAL ANQGFDVCLI DTAPTLGVGL
        250         260         270         280         290         300
AAALFAADYV LSPIELEAYS IQGIKKMVTT IANVRQKNAK LQFLGMVPSK VDARNPRHAR
        310         320         330         340         350         360
HQAELLAAYP KMMIPATVGL RSSIADALAS GVPVWKIKKT AARKASKEVR ALADYVFTKM
   364
EISQ
```

>IncC2

```
         10          20          30          40          50          60
MKTLVTANQK GGVGKTSTLV HLAFDFFERG LRVAVIDLDP QGNASYTLKD FATGLHASKL
         70          80          90         100         110         120
FGAVPAGGWT ETAPAAGDGQ AARLALIESN PVLANAERLS LDDARELFGA NIKALANQGF
        130         140         150         160         170         180
DVCLIDTAPT LGVGLAAALF AADYVLSPIE LEAYSIQGIK KMVTTIANVR QKNAKLQFLG
        190         200         210         220         230         240
MVPSKVDARN PRHARHQAEL LAAYPKMMIP ATVGLRSSIA DALASGVPVW KIKKTAARKA
        250         259
SKEVRALADY VFTKMEISQ
```

>KorB

```
     10         20         30         40         50         60
MTAAQAKTTK KNTAAAAQEA AGAAQPSGLG LDSIGDLSSL LDAPAASQGG SGPIELDLDL
        70         80         90        100        110        120
IDEDPHQPRT ADNPGFSPES IAEIGATIKE RGVKSPISVR ENQEQPGRYI INHGARRYRG
       130        140        150        160        170        180
SKWAGKKSIP AFIDNDYNEA DQVIENLQRN ELTPREIADF IGRELAKGKK KGDIAKEIGK
       190        200        210        220        230        240
SPAFITQHVT LLDLPEKIAD AFNTGRVRDV TVVNELVTAF KKRPEEVEAW LDDDTQEITR
       250        260        270        280        290        300
GTVKLLREFL DEKGRDPNTV DAFNGQTDAE RDAEAGDGQD GEDGDQDGKD AKEKGAKEPD
       310        320        330        340        350       358
PDKLKKAIVQ VEHDERPARL ILNRRPPAEG YAWLKYEDDG QEFEANLADV KLVALIEG
```

>KorA

```
        10         20         30         40         50         60
MKKRLTESQF QEAIQGLEVG QQTIEIARGV LVDGKPQATF ATSLGLTRGA VSQAVHRVWA
        70         80         90        101
AFEDKNLPEG YARVTAVLPE HQAYIVRKWE ADAKKKQETKR
```

Table 1: Sequence-based protein properties calculated from protpram server (https://web.expasy.org/cgi-bin/protparam/protparam)

Protein	Extinction coefficient	pI	Mol wt (Da)
IncC NTD	6990	10.65	10619.47
IncC1	25440	9.73	38134.35
IncC2	18450	9.47	27532.89
KorB	23950	4.76	39011.18
KorA	13980	9.57	11305.92

Appendix 4

>IncC_NTD [Pseudomonas aeruginosa]

MGVIHEETAYRKPVPGGDPGAGSGAADHRDSAGRLSRWEATGDVRNVAGTDQGRSVA
SGASRVGRVRGQELARGVRAGNGGSAGTSGVHRPEVGSGRQE

>IncC1_[Bordetella pertussis]

MGAIHEETANRSPIPGGHQGAGDRAADYRHSTWRAGRWASPGGIRHVAGVDQGGSVA
SGQSRMDGSGGTAPRGVRAGDGGTAGASSVHRQEVGSRRQEE

>Protein_IncC_[Proteobacteria]

MGAIHEETANRSPIPDGHQGAGDRAADHRHSARRAGRWPAPGGVCHFAGADQGGGVA
GGQSRMGSSRGTAPRGLRASDGGTAGASSVHRQEVGSRRQEE

>IncC_[Plasmid pMCBF1]

MGVIHEETANRSPIPGGNQGPGSGSADHRDSAGRAGRWQAPGGLCGRAGINQGGGVAS
SQSGLGSGAAHCTRGVRANNGGAARTSSLHRQEVGRGRQEK

>IncC_[Comamonadaceae]

MGVLHEETANRSPIPGGDQGPGDRATDHRHSARRAGRWKAPGRVCGLAGADQGGGIA
SGQSRVGSGGGSAPAGLRAGDGGTAGASSVHRQALGSRRQGK

>IncC2_[Comamonas testosteroni]

MGVLHEETANRSPIPGGDQGPGDRATDHRHSARRAGRWKAPGRVCGLAGADQGGGIA
SGQSRVGSGGGSAPAGLRAGDGGTAGASSVHRQALGSRRQGK

>Inclusion_membrane_protein_[Achromobacter xylosoxidans]

MGVLHEETANRSPIPGGDQGPGDRATDHRHSARRAGRWKAPGRVCGLAGADQGGGIA
SGQSRVGSGGGSAPAGLRAGDGGTAGASSVHRQALGSRRQGK

>IncC_[Delftia acidovorans]

MGVLHEETANRSPIPGGDQGPGDRATDHRHSARRAGRGEAPGRVCGLAGAVQGGGIAS
GQSRVGSSGGSAPRGLRAGDGGTAGASSVHRQALGSRRQGK

>plasmid_partitioning_protein_ParA_[Mycobacterium abscessus]

MGVLHEETANRSPIPGGDQGPGDRATDHRHSARRAGRWKAPGRVCGLAGADQGGGIA
SGQPRVGSGGGSAPAGLRAGDGGTAGASSVHRQALGSRRQGK

>IncC_[Sphingomonas sp. A1]

MGVLHEETANRSPIPGGDQGPGDRAADHRHSARRAGRGKAPGRVCGLAGAVQGGGIA
SGQPRVGSSGGSAPRGIRAGDGGTAGASSVHRQALGSRRQGK

>IncC_[Comamonas testosteroni]

MGVLHEETANRSPIPGGDQGPGDRATDHRHSARRAGRWKAPGRVCGLAGADQGGGIA
SGQPRVGSGGGSAPAGLRAGDGGTAGASSVHRQALGSRRQGK

>Cobyrinic acid_a,c-diamide_synthase_[Acidovorax sp. JS42]

MGVLHEETANRSPISGGDQGPGDRAADHRHSARRAGRGKAPGRVCGLAGAVQGGGVA
SGQPRVGSSGGSAPRGIRAGDGGTAGASSVHRQALGSRRQGK

>IncC1_[Achromobacter denitrificans]

MGAIHEETANRSPIPRGHKGLGHWPADHRNSAGRAGFGPAASGFHHVAGTDQRGSLSC
REQGVDGVFVQERTPGVRAGFGSAAGASGVHRQEVGRGRGD

>IncC1_protein_[Photobacterium damselae]

MGAIHEEKANRVSISGCNQEFGSGTSNTTDSARGLSRGKTSNRICDVVGSDEGGGITSRQ
EGMGSSHGTSAGRIRANNGNTAGTSSVHRQEVGRGRKKE

>plasmid_partitioning_protein_ParA_[Xylella fastidiosa]

MGVFHEETANGSTVPSCHKGSGSKSTNTRDSSWRFSRGRISNKIRCITESIKGSSVANGES
SVAVARRTERPERFRASNCSASRTSSVHCQEMGSRGSE

>plasmid_partition_protein_[Yersinia pseudotuberculosis IP 31758]

MGNVHEKTALESRVPGSHKRLGCGPTNHRYSKWCLGRGDAASNVCSDVRAFPRGCISS
SESCLDSTSGEKSAGGIREGYSDTSGTKSIHSQEMGRGNSK

Appendix 5

Minimal growth media for 15N 13C labelled proteins overexpression

Solution 1 – 100 mL

concentrated HCl	8 mL
$FeCl_2•4H_2O$	5 g
$CaCl_2•2H_2O$	184 mg
H_3BO_3	64 mg
$MnCl_2•4H_2O$	40 mg
$CoCl_2•6H_2O$	18 mg
$CuCl_2•2H_2O$	4 mg
$ZnCl_2$	340 mg
$NaMoO_4•2H_2O$	605 mg

Make up to 100 mL with H_2O

Solution 2: 1 L

Vitamins

Biotin	1.1 mg
Folic acid	1.1 mg
Paba	110 mg
Riboflavin	110 mg
Pantothenic acid	220 mg
Pyridoxine HCl	220 mg
Thiamine HCl	220 mg
Niacinamide	220 mg

Add 500 mL H_2O + 500 mL high-purity ethanol, Store at 4°C

Solution 3: 500 mL

KH_2PO_4	16.5 g
K_2HPO_4	87.5 g
NaCl	18.25 g

Add H_2O to 500 mL and autoclaved

Solution 4: 500 mL

28.8 g $MgCl_2$ + 10 mL Solution 1

Added water to make 500 mL

Solution 5: 10 mL

Thiamine Solution (1 mg/mL)

1 L Minimal Media

H_2O (autoclaved)	940 mL
Solution 2 (filter sterilized)	1
Solution 3 (autoclaved)	40 mL
Solution 4(filter sterilized)	2 mL
Solution 5 (filter sterilized)	1 mL

1g $^{15}NH_4Cl$ dissolved in 5 mL H_2O (filter sterilized and added)

4 g $^{13}C^6$ glucose dissolved in 10 mL H_2O (filter sterilized and added)

Reference

Weber DJ, Gittis AG, Mullen GP, Abeygunawardana C, Lattman EE, and Mildvan AS. "NMR docking of a substrate into the X-ray structure of staphylococcal nuclease." Proteins (1992) 4, 275-287.

Table 1 Chemical Shifts obtained from Backbone assignments of IncC NTD. HA shows chemical shifts for ^{1}H, CA for Cα and CB for Cβ, CO for carbonyl carbon and NH for N^{H}. The chemical shift table have been exported from the CCPNNmr Analysis Software.

#NUM	AA	HA	CA	CB	CO	NH
1	M	55.260	32.270	176.154	122.449	0.000
2	G	43.259	0.000	169.890	109.030	0.000
3	V	62.200	32.765	175.938	119.012	8.380
4	I	60.852	38.454	175.821	124.657	8.216
5	H	55.246	29.251	174.087	123.749	8.551
6	E	56.671	30.275	176.185	123.510	8.483
7	E	0.000	30.071	176.659	122.716	8.655
8	T	62.047	69.588	174.232	115.147	8.133
9	A	52.564	19.072	177.127	126.034	8.188
10	Y	57.916	38.631	175.282	119.504	7.999
11	R	55.221	31.006	175.021	124.039	7.900
12	K	54.307	32.273	0.000	124.713	8.187
13	P	62.940	31.408	174.411	137.109	0.000
14	V	0.000	0.000	0.000	0.000	0.000
15	P	0.000	0.000	0.000	0.000	0.000
16	G	44.983	0.000	174.541	0.000	0.000
17	G	44.812	0.000	173.304	108.517	8.160
18	D	52.084	41.397	175.052	121.452	8.239
19	P	63.990	31.854	175.039	137.844	0.000
20	G	0.000	0.000	0.000	0.000	0.000
21	A	52.720	19.178	178.328	123.649	0.000
22	G	0.000	0.000	0.000	0.000	0.000
23	S	0.000	0.000	0.000	0.000	0.000
24	G	0.000	0.000	0.000	0.000	0.000
25	A	52.597	19.090	177.721	0.000	0.000
26	A	52.594	18.932	177.293	122.738	8.222
27	D	54.014	41.090	176.012	118.581	8.073
28	H	55.549	28.691	174.510	118.877	8.281
29	R	56.244	30.977	176.162	121.810	8.243
30	D	54.098	41.235	176.758	121.184	8.430
31	S	59.108	63.601	174.685	141.099	8.249
32	A	52.784	18.769	178.261	125.212	8.222
33	G	45.400	0.000	174.351	107.789	8.276
34	R	56.344	30.709	176.464	120.794	8.096
35	L	55.253	42.217	177.388	122.635	8.244
36	S	58.305	63.823	174.435	116.825	8.227
37	R	56.453	30.351	175.975	122.412	8.201

38	W	57.356	57.310	176.066	121.648	8.064
39	E	56.218	30.598	175.735	122.280	8.024
40	A	52.602	19.123	177.854	125.567	8.282
41	T	61.889	69.959	175.123	112.103	7.953
42	G	45.257	0.000	173.632	110.571	8.264
43	D	54.171	41.407	176.376	120.727	8.181
44	V	62.542	32.104	176.198	120.596	8.089
45	R	56.211	30.777	175.971	123.625	8.326
46	N	53.161	38.877	175.077	119.701	8.314
47	V	62.313	32.599	175.857	120.353	8.025
48	A	52.697	19.146	178.133	127.183	8.337
49	G	45.245	0.000	174.481	107.999	8.286
50	T	61.816	70.009	174.461	112.691	7.982
51	D	54.497	40.990	176.446	122.377	8.416
52	Q	56.399	29.060	176.636	120.656	8.337
53	G	45.546	0.000	174.284	109.154	8.418
54	R	56.232	30.800	176.400	120.158	7.985
55	S	58.297	63.793	174.617	116.913	8.300
56	V	62.287	32.672	176.031	121.562	8.091
57	A	52.556	18.962	177.775	127.269	8.301
58	S	58.498	63.844	0.000	115.141	8.216
59	G	45.382	0.000	174.065	0.000	0.000
60	A	52.668	19.161	177.890	123.551	8.070
61	S	58.386	63.706	174.628	114.601	8.199
62	R	56.188	30.587	176.209	123.004	8.285
63	V	62.564	32.665	176.528	120.507	8.013
64	G	45.171	0.000	173.791	112.480	8.395
65	R	56.080	30.928	176.300	120.763	0.000
66	V	62.354	32.790	176.110	121.818	8.190
67	R	56.461	30.593	176.770	125.362	8.463
68	G	45.459	0.000	174.257	110.434	8.492
69	Q	56.128	29.560	176.057	120.060	8.222
70	E	56.969	29.871	176.573	122.150	8.569
71	L	55.172	42.264	177.191	122.984	8.165
72	A	52.533	18.857	177.579	124.507	8.159
73	R	56.274	30.730	176.860	119.915	8.188
74	G	45.238	0.000	174.021	109.686	8.313
75	V	62.461	32.730	176.090	119.443	7.915
76	R	55.841	30.869	0.000	0.000	0.000
77	A	52.442	19.242	177.806	125.905	8.319
78	G	45.222	0.000	172.930	108.709	8.287
79	N	54.719	40.943	176.432	123.947	7.944

80	G	0.000	0.000	0.000	0.000	0.000
81	G	0.000	0.000	0.000	0.000	0.000
82	S	58.370	63.873	174.534	115.765	0.000
83	A	52.779	19.069	0.000	125.834	8.365
84	G	0.000	0.000	174.554	0.000	0.000
85	T	61.834	69.852	174.875	112.761	8.018
86	S	58.548	63.840	174.795	117.903	8.328
87	G	45.283	0.000	173.826	110.740	8.352
88	V	62.205	32.761	175.853	119.119	7.855
89	H	55.105	29.194	174.051	123.251	8.530
90	R	53.955	30.129	0.000	124.790	8.364
91	P	0.000	0.000	0.000	0.000	0.000
92	E	56.575	30.033	176.614	0.000	0.000
93	V	62.544	32.700	176.668	121.173	8.161
94	G	45.253	0.000	174.249	112.422	8.468
95	S	58.526	63.842	175.189	115.656	0.000
96	G	45.301	0.000	174.057	110.894	8.392
97	R	55.975	30.614	176.222	120.451	8.117
98	Q	55.739	29.288	175.809	121.859	8.452
99	E	56.396	30.348	176.349	122.715	8.449
100	K	56.184	32.820	176.616	122.404	8.400
101	T	61.795	69.961	175.017	114.911	8.135
102	G	45.271	0.000	173.846	110.729	8.385
103	N	53.198	38.974	175.066	118.745	8.329
104	Q	56.099	29.462	175.288	120.953	8.358
105	T	63.339	70.663	0.000	20.590	7.809

Appendix 6

Bacterial Lysis buffers

20 mM Tris pH 7.5	NaCl 100mM, 1 mM EDTA
	NaCl 100mM, 10 mM imidazole, 10 % glycerol
	NaCl 100mM, 10 mM imidazole, 50 mM Sodium glutamate, 10 % glycerol
	NaCl 500mM, 1 mM DTT
	NaCl 100mM, 10 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol
	NaCl 100mM, 10 mM imidazole, 50 mM Sodium glutamate, 10 % glycerol, 10 mM $MgCl_2$
	KCl 150 mM, 10 mM imidazole, 10 % glycerol
	KCl 150 mM, 10 mM imidazole, 10 % glycerol, 0.05% Tween 20 **(B1 Buffer)**
	KCl 300 mM, 50 mM Sodium Glutamate, 10 mM imidazole, 1 mM DTT, 10 % glycerol **(B2 Buffer)**
	KCl 100 mM, $LiCl_2$ 100 mM, 10 % glycerol
	$(NH_4)_2SO_4$ 100 mM, 10 % glycerol **(B3 Buffer)**
	NaCl 100mM, 10 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol, 1 mM DTT, 1 % NP-40
	NaCl 100mM, 10 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol, 1 mM DTT, 1% Sodium Deoxycholate
	NaCl 150 mM, Na2 EDTA 1 mM, 1% NP-40, 1% sodium deoxycholate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate **(RIPA Buffer/B4 Buffer)**
	KCl 150 mM, 10 mM imidazole, 10 % glycerol, 10 % isopropanol
	KCl 150 mM, $(NH_4)_2SO_4$ 100 mM 10 mM imidazole, 10 % glycerol
	KCl 150 mM, $(NH_4)_2SO_4$ 100 mM 10 mM imidazole, 10 % glycerol, 1mM DTT

20 mM HEPES pH 7.5	KCl 150mM, 10 % glycerol **(B5 Buffer)**
	NaCl 100mM, 10 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol
	NaCl 300 mM, 50 mM Sodium Glutamate, 10 mM imidazole, 1 mM DTT, 10 % glycerol **(B6 Buffer)**
	NaCl 100mM, 10 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol, 1 mM DTT, 1% sodium deoxycholate
	NaCl 100mM, 100 mM $CaCl_2$, 10 % glycerol
	NaCl 500mM, 1 mM DTT, 0.05% Tween20, 10 % glycerol **(B7 Buffer)**
	NaCl 500mM, 10 mM Imidazole, 1 mM DTT **(B8 Buffer 8)**

Buffer Screen for pH Optimization (homemade Buffers) use for thermal shift assay

	1	2	3	4	5	6	7	8	9	10	11	12
A	(1.0 M Sodium acetate 4.4)	(1.0 M 1.0 M Sodium acetate 4.4)	(1.0 M 1.0 M Sodium acetate 4.8)	(1.0 M 1.0 M Sodium acetate 4.8)	(1.0 M Sodium Citrate 4.8)	(1.0 M Sodium Citrate 4.8)	(1.0 M Sodium acetate 5.2)	(1.0 M Sodium acetate 5.2)	(1.0 M Citrate 5.3)	(1.0 M Citrate 5.3)	(1.0 M MES 5.5)	(1.0 M MES 5.5)
B	(1.0 M Sodium Citrate 5.8)	(1.0 M Sodium Citrate 5.8)	(1.0 M MES 6.0)	(1.0 M MES 6.0)	(1.0 M ADA 6.0)	(1.0 M ADA 6.0)	(1.0 M BIS-TRIS 6.0)	(1.0 M BIS-TRIS 6.0)	(1.0 M MES 6.5)	(1.0 M MES 6.5)	(1.0 M ADA 6.5)	(1.0 M ADA 6.5)
C	(1.0 M BIS-TRIS 6.5)	(1.0 M BIS-TRIS 6.5)	(1.0 M ADA 7.0)	(1.0 M ADA 7.0)	(1.0 M BIS-TRIS 7.0)	(1.0 M BIS-TRIS 7.0)	(1.0 M HEPES 7.0)	(1.0 M HEPES 7.0)	(1.0 M HEPES 7.5)	(1.0 M HEPES 7.5)	(1.0 M TRIS 7.5)	(1.0 M TRIS 7.5)
D	(1.0 M HEPES 8.0)	(1.0 M HEPES 8.0)	(1.0 M TRIS 8.0)	(1.0 M TRIS 8.0)	(1.0 M TRIS 8.5)	(1.0 M TRIS 8.5)	(1.0 M CHES 8.8)	(1.0 M CHES 8.8)	(1.0 M CHES 9.3)	(1.0 M CHES 9.3)	(1.0 M CHES 9.8)	(1.0 M CHES 9.8)
E	(1.0 M Sodium acetate 4.4)	(1.0 M Sodium acetate 4.4)	(1.0 M Sodium acetate 4.8)	(1.0 M Sodium acetate 4.8)	(1.0 M Sodium Citrate 4.8)	(1.0 M Sodium Citrate 4.8)	(1.0 M Sodium acetate 5.2)	(1.0 M Sodium acetate 5.2)	(1.0 M Citrate 5.3)	(1.0 M Citrate 5.3)	(1.0 M MES 5.5)	(1.0 M MES 5.5)
F	(1.0 M Sodium Citrate 5.8)	(1.0 M Sodium Citrate 5.8)	(1.0 M MES 6.0)	(1.0 M MES 6.0)	(1.0 M ADA 6.0)	(1.0 M ADA 6.0)	(1.0 M BIS-TRIS 6.0)	(1.0 M BIS-TRIS 6.0)	(1.0 M MES 6.5)	(1.0 M MES 6.5)	(1.0 M ADA 6.5)	(1.0 M ADA 6.5)
G	(1.0 M BIS-TRIS 6.5)	(1.0 M BIS-TRIS 6.5)	(1.0 M ADA 7.0)	(1.0 M ADA 7.0)	(1.0 M BIS-TRIS 7.0)	(1.0 M BIS-TRIS 7.0)	(1.0 M HEPES 7.0)	(1.0 M HEPES 7.0)	(1.0 M HEPES 7.5)	(1.0 M HEPES 7.5)	(1.0 M TRIS 7.5)	(1.0 M TRIS 7.5)
H	(1.0 M HEPES 8.0)	(1.0 M HEPES 8.0)	(1.0 M TRIS 8.0)	(1.0 M TRIS 8.0)	(1.0 M TRIS 8.5)	(1.0 M TRIS 8.5)	(1.0 M CHES 8.8)	(1.0 M CHES 8.8)	(1.0 M CHES 9.3)	(1.0 M CHES 9.3)	(1.0 M CHES 9.8)	(1.0 M CHES 9.8)

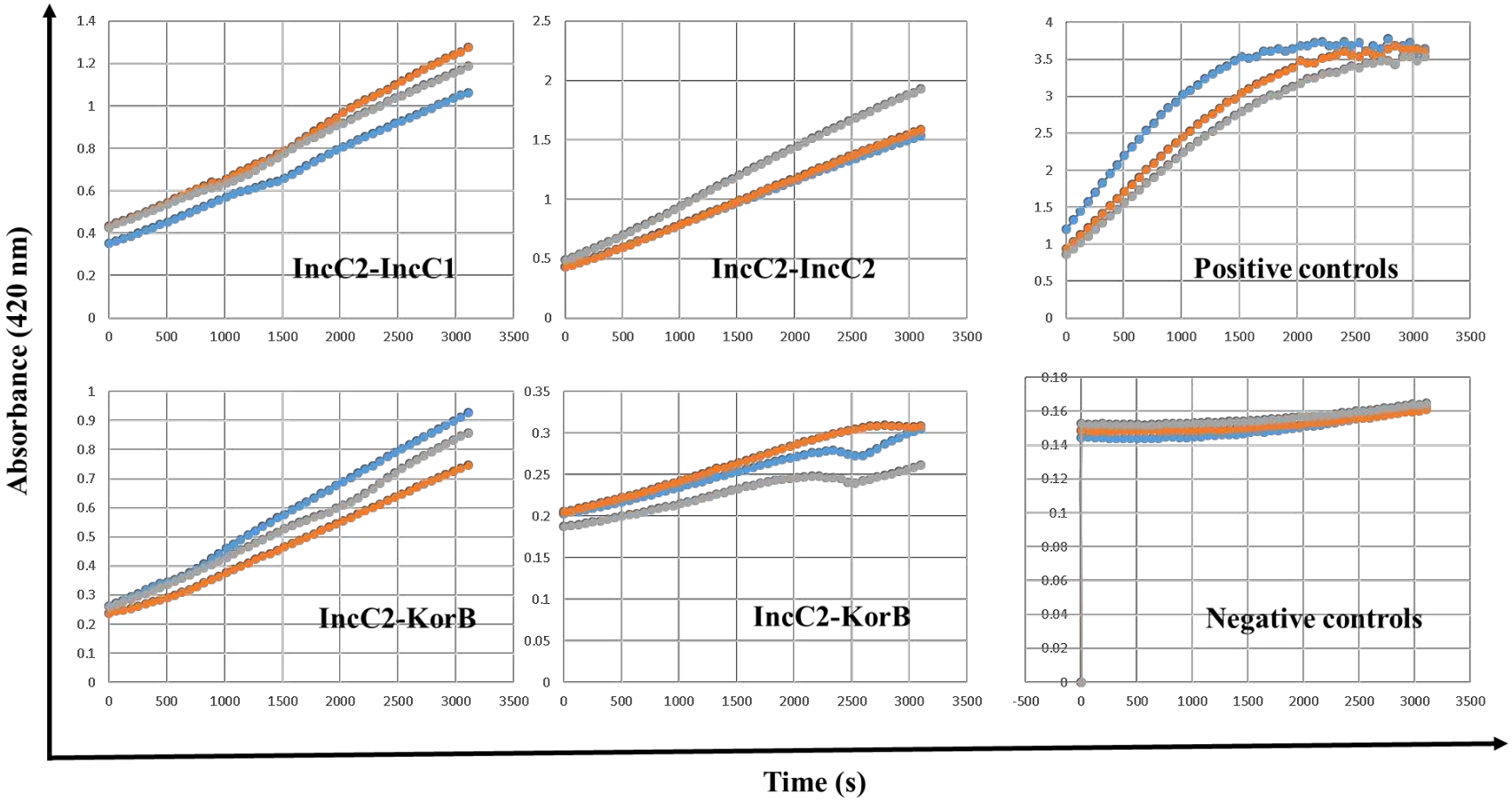
White Shading: + 150 mM NaCl
Gray shading: High Salt: + 500 mM NaCl

Buffer Screen for pH Optimization (1.0 M pH Slice Kit from Hapmton Research) used for thermal shift assay

	1	2	3	4	5	6	7	8	9	10	11	12
A	(1.0 M 1.0 M Citric acid pH 3.5)	(1.0 M 1.0 M Citric acid pH 3.8)	(1.0 M 1.0 M Citric acid pH 4.1)	(1.0 M 1.0 M Citric acid pH 4.4)	(1.0 M 1.0 M Sodium citrate tribasic dihydrate pH 3.6)	(1.0 M 1.0 M Sodium citrate tribasic dihydrate pH 3.9)	(1.0 M 1.0 M Sodium citrate tribasic dihydrate pH 4.2)	(1.0 M 1.0 M Sodium citrate tribasic dihydrate pH 4.5)	(1.0 M 1.0 M Sodium acetate trihydrate pH 3.7)	(1.0 M 1.0 M Sodium acetate trihydrate pH 4.0)	(1.0 M 1.0 M Sodium acetate trihydrate pH 4.3)	(1.0 M 1.0 M Sodium acetate trihydrate pH 4.6)
B	(1.0 M 1.0 M Sodium acetate trihydrate pH 4.9)	(1.0 M 1.0 DL-Malic pH 4.7)	(1.0 M 1.0 DL-Malic pH 5.0)	(1.0 M 1.0 DL-Malic pH 5.3)	(1.0 M 1.0 DL-Malic pH 5.6)	(1.0 M 1.0 DL-Malic pH 5.9)	(1.0 M 1.0 M Succinic acid pH 4.8)	(1.0 M 1.0 M Succinic acid pH 5.1)	(1.0 M 1.0 M Succinic acid pH 5.4)	(1.0 M 1.0 M Succinic acid pH 5.7)	(1.0 M 1.0 M Succinic acid pH 6.0)	(1.0 M 1.0 M Sodium cacodylate trihydrate pH 5.2)
C	(1.0 M 1.0 M Sodium cacodylate trihydrate pH 5.5)	(1.0 M 1.0 M Sodium cacodylate trihydrate pH 5.8)	(1.0 M 1.0 M Sodium cacodylate trihydrate pH 6.1)	(1.0 M 1.0 M Sodium cacodylate trihydrate pH 6.4)	(1.0 M 1.0 MES monohydrate pH 5.3)	(1.0 M 1.0 MES monohydrate pH 5.6)	(1.0 M 1.0 MES monohydrate pH 5.9)	(1.0 M 1.0 MES monohydrate pH 6.2)	(1.0 M 1.0 MES monohydrate pH 6.5)	(1.0 M 1.0 M BIS-TRIS pH 5.7)	(1.0 M 1.0 M BIS-TRIS pH 6.0)	(1.0 M 1.0 M BIS-TRIS pH 6.3)
D	(1.0 M 1.0 M BIS-TRIS pH 6.6)	(1.0 M 1.0 M BIS-TRIS pH 6.9)	(1.0 M 1.0 M ADA pH 5.8)	(1.0 M 1.0 M ADA pH 6.1)	(1.0 M 1.0 M ADA pH 6.4)	(1.0 M 1.0 M ADA pH 6.7)	(1.0 M 1.0 M ADA pH 7.0)	(1.0 M 1.0 M Imidazole pH 6.2)	(1.0 M 1.0 M Imidazole pH 6.5)	(1.0 M 1.0 M Imidazole pH 6.8)	(1.0 M 1.0 M Imidazole pH 7.1)	(1.0 M 1.0 M Imidazole pH 7.4)
E	(1.0 M 1.0 M BIS-TRIS propane pH 6.4)	(1.0 M 1.0 M BIS-TRIS propane pH 6.7)	(1.0 M 1.0 M BIS-TRIS propane pH 7.0)	(1.0 M 1.0 M BIS-TRIS propane pH 7.3)	(1.0 M 1.0 M MOPS pH 6.5)	(1.0 M 1.0 M MOPS pH 6.8)	(1.0 M 1.0 M MOPS pH 7.1)	(1.0 M 1.0 M MOPS pH 7.4)	(1.0 M 1.0 M MOPS pH 7.7)	(1.0 M 1.0 M HEPES sodium pH 6.6)	(1.0 M 1.0 M HEPES sodium pH 6.9)	(1.0 M 1.0 M HEPES sodium pH 7.2)
F	(1.0 M 1.0 M HEPES sodium pH 7.5)	(1.0 M 1.0 M HEPES pH 6.8)	(1.0 M 1.0 M HEPES pH 7.1)	(1.0 M 1.0 M HEPES pH 7.4)	(1.0 M 1.0 M HEPES pH 7.7)	(1.0 M 1.0 M TRIS hydrochloride pH 7.2)	(1.0 M 1.0 M TRIS hydrochloride pH 7.5)	(1.0 M 1.0 M TRIS hydrochloride pH 7.8)	(1.0 M 1.0 M TRIS hydrochloride pH 8.1)	(1.0 M 1.0 M Tris pH 7.3)	(1.0 M 1.0 M Tris pH 7.6)	(1.0 M 1.0 M Tris pH 7.9)
G	(1.0 M 1.0 M Tris pH 8.2)	(1.0 M 1.0 M Tris pH 8.5)	(1.0 M 1.0 M Tricine pH 7.4)	(1.0 M 1.0 M Tricine pH 7.7)	(1.0 M 1.0 M Tricine pH 8.0)	(1.0 M 1.0 M Tricine pH 8.3)	(1.0 M 1.0 M Tricine pH 8.6)	(1.0 M 1.0 M BICINE pH 7.5)	(1.0 M 1.0 M BICINE pH 7.8)	(1.0 M 1.0 M BICINE pH 8.1)	(1.0 M 1.0 M BICINE pH 8.4)	(1.0 M 1.0 M BICINE pH 8.7)
H	(1.0 M 1.0 M BIS-TRIS propane pH 8.5)	(1.0 M 1.0 M BIS-TRIS propane pH 8.8)	(1.0 M 1.0 M BIS-TRIS propane pH 9.1)	(1.0 M 1.0 M BIS-TRIS propane pH 9.4)	(1.0 M 1.0 M Glycine pH 8.6)	(1.0 M 1.0 M Glycine pH 8.9)	(1.0 M 1.0 M Glycine pH 9.2)	(1.0 M 1.0 M Glycine pH 9.5)	(1.0 M 1.0 M AMPD pH 8.7)	(1.0 M 1.0 M AMPD pH 9.0)	(1.0 M 1.0 M AMPD pH 9.3)	(1.0 M 1.0 M AMPD pH 9.6)

Appendix 7

Some graph used to monitor linear area for slope calculations after data tabulation (BACTH analysis)



Appendix 8

Kojak configuration file

```
#files used for data set MFR1
#Computational resources
#
threads     =   64  # Increase to use more cores/nodes
# Data input files: include full path if not in current working directory
#
database            =   all_proteins.fasta
export_percolator=          1
export_pepXML               =           1
MS_data_file                =   041216_MFR_1.mzXML
percolator_version          =   2.08
#
# Parameters used to described the data being input to Kojak
#
enrichment      =   0       #Values between 0 and 1 to describe 18O APE.
                    #For example, 0.25 equals 25 APE.
instrument      =   0       #Values are: 0=Orbitrap, 1=FTICR (such as Thermo LTQ-FT)
MS1_centroid    =   0       #0=no, 1=yes
MS2_centroid    =   1       #0=no, 1=yes
MS1_resolution  =   120000 #Resolution at 400 m/z,
                    #already centroided
MS2_resolution  =   15000   #Resolution at 400 m/z, e
                    #already centroided
# Cross-link and mono-link masses allowed. May have more than one of each parameter.
#
# Format for cross_link is [amino acids] [amino acids] [mass mod] [identifier]
# Format for mono_link is [amino acids] [mass mod]
# One or more amino acids (uppercase only!!) can be specified for each linkage moiety
# Use lowercase 'n' or 'c' to indicate protein N-terminus or C-terminus
#
cross_link   =      nK        nK        138.0680742       BS3
cross_link   =      nK        nT        138.0680742       BS3
cross_link   =      nK        nS        138.0680742       BS3
cross_link   =      nK        nY        138.0680742       BS3
cross_linl   =      nK        nK        142.09          BS304
mono_link  =      nK        156.0786

#
# Fixed modifications. Add as many as necessary.
#
fixed_modification  =   C   57.02146
fixed_modification_protC =          0
fixed_modification_protN =          0

#
# Differential modifications. Add as many as necessary. Uppercase only for amino acids!
# n = peptide N-terminus, c = peptide C-terminus
#
# If more than one modification is possible for an amino acid,
# list all modifications on separate lines
#
```

```
modification         =      M      15.9949
modification_protC                  =                  0
modification_protN                  =                  0
diff_mods_on_xl                     =                  0
max_mods_per_peptide   =      2
mono_links_on_xl                    =                  0

# Digestion enzyme rules.
#
enzyme =      [KR]|{P} Trypsin

# Scoring algorithm parameters
#
# fragment_bin_offset and fragment_bin_size influence algorithm precision and memory usage.
# They should be set appropriately for the data analyzed.
# For ion trap ms/ms:  1.0005 size, 0.4 offset
# For high res ms/ms:    0.03 size, 0.0 offset
#
fragment_bin_offset        =  0.4    #between 0.0 and 1.0
fragment_bin_size          =  1.0005  #in Thomsons
ion_series_A                        =         0
ion_series_B                        =         1
ion_series_C                        =         0
ion_series_X                        =         0
ion_series_Y                        =         1
ion_series_Z                        =         0                  #Z-dot values are used


#
# Additional parameters used in Kojak analysis
#
decoy_filter          =  Random   #identifier for all decoys in the database.
                         #Default value is "random" (without quotes)
isotope_error         =  1      #account for errors in precursor peak identification.
                         #Searches this number of isotope peak offsets.
                                                                          #Values are 0,1,or 2.
max_miscleavages       =  2      #number of missed trypsin cleavages allowed
max_peptide_mass       =  8000.0  #largest allowed peptide mass in Daltons
min_peptide_mass       =  500.0   #lowest allowed peptide mass in Daltons
max_spectrum_peaks     =  0      #top N peaks to use during analysis. 0 uses all peaks.
ppm_tolerance_pre      =  15.0    #mass tolerance on precursor when searching
prefer_precursor_pred  =  2      #prefer precursor mono mass predicted by
                         #instrument software.
                         #  0 = ignore previous predictions
                         #  1 = use only previous predictions
                         #  2 = supplement predictions with additional analysis
spectrum_processing    =  0      #0=no, 1=yes
top_count              =  20     #number of top scoring single peptides to combine
                         #in relaxed analysis
truncate_prot_names    =  0      #Max protein name character to export, 0=off
turbo_button               =  1      #Generally speeds up analysis. Special cases cause reverse
                         #effect, thus this is allowed to be disabled. 0=off
```

Appendix 9

Script files used in Modeller for IncC2 homology modelling

1. build_profile.py

```
from modeller import *
log.verbose()
env = environ()
#-- Prepare the input files
#-- Read in the sequence database
sdb = sequence_db(env)
sdb.read(seq_database_file='pdball.pir', seq_database_format='PIR',
        chains_list='ALL', minmax_db_seq_len=(30, 4000), clean_sequences=True)

#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdball.bin', seq_database_format='BINARY',
         chains_list='ALL')
#-- Now, read in the binary database
sdb.read(seq_database_file='pdball.bin', seq_database_format='BINARY',
        chains_list='ALL')

#-- Read in the target sequence/alignment
aln = alignment(env)
aln.append(file='IncC2.ali', alignment_format='PIR', align_codes='ALL')

#-- Convert the input sequence/alignment into
#   profile format
prf = aln.to_profile()

#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
         gap_penalties_1d=(-500, -50), n_prof_iterations=1,
         check_profile=False, max_aln_evalue=0.01)

#-- Write out the profile in text format
prf.write(file='build_profile.prf', profile_format='TEXT')

#-- Convert the profile back to alignment format
aln = prf.to_alignment()

#-- Write out the alignment file
aln.write(file='build_profile.ali', alignment_format='PIR')
```

2. align2d.py

```
from modeller import *

env = environ()
aln = alignment(env)
mdl = model(env, file='5ihp', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='5ihpA', atom_files='5ihp.pdb')
aln.append(file='IncC2.ali', align_codes='IncC2')
aln.align2d()
aln.write(file='IncC2-5ihp.ali', alignment_format='PIR')
```

```
aln.write(file='IncC2-5ihp.pap', alignment_format='PAP')
```

4. model-single.py

```
from modeller import *
from modeller.automodel import *
#from modeller import soap_protein_od

env = environ()
a = automodel(env, alnfile='IncC2-5ihp.ali',
              knowns='5ihpA', sequence='IncC2',
              assess_methods=(assess.DOPE,
                              #soap_protein_od.Scorer(),
                              assess.GA341))
a.starting_model = 1
a.ending_model = 100
a.make()
```

5. evaluate_model.py

```
from modeller import *
from modeller.scripts import complete_pdb

log.verbose()    # request verbose output
env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

# read model file
mdl = complete_pdb(env, 'IncC2.B99990012.pdb')

# Assess all atoms with DOPE:
s = selection(mdl)
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='IncC2.profile',
              normalize_profile=True, smoothing_window=15)
```

6. plot_profiles.py

```
import pylab
import modeller

def r_enumerate(seq):
   """Enumerate a sequence in reverse order"""
   # Note that we don't use reversed() since Python 2.3 doesn't have it
   num = len(seq) - 1
   while num >= 0:
      yield num, seq[num]
      num -= 1

def get_profile(profile_file, seq):
   """Read `profile_file` into a Python array, and add gaps corresponding to
     the alignment sequence `seq`."""
   # Read all non-comment and non-blank lines from the file:
```

```
    f = open(profile_file)
    vals = []
    for line in f:
        if not line.startswith('#') and len(line) > 10:
            spl = line.split()
            vals.append(float(spl[-1]))
    # Insert gaps into the profile corresponding to those in seq:
    for n, res in r_enumerate(seq.residues):
        for gap in range(res.get_leading_gaps()):
            vals.insert(n, None)
    # Add a gap at position '0', so that we effectively count from 1:
    vals.insert(0, None)
    return vals

e = modeller.environ()
a = modeller.alignment(e, file='IncC2-5ihp.ali')

template = get_profile('5ihp.profile', a['5ihpA'])
model = get_profile('Incc2.profile', a['Incc2'])

# Plot the template and model profiles in the same plot for comparison:
pylab.figure(1, figsize=(10,6))
pylab.xlabel('Alignment position')
pylab.ylabel('DOPE per-residue score')
pylab.plot(model, color='red', linewidth=2, label='Model')
pylab.plot(template, color='green', linewidth=2, label='Template')
pylab.legend()
pylab.savefig('dope_profile.png', dpi=65)
```

7. loopmodel.py

```
# Loop refinement of an existing model
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = './:../atom_files'

# Create a new class based on 'loopmodel' so that we can redefine
# select_loop_atoms (necessary)
class MyLoop(loopmodel):
    # This routine picks the residues to be refined by loop modeling
    def select_loop_atoms(self):
        # 10 residue insertion
        return selection(self.residue_range('67', '87'))

m = MyLoop(env,
           inimodel='IncC2-mult.pdb', # initial model of the target
           sequence='IncC2')          # code of the target

m.loop.starting_model= 1           # index of the first loop model
```

```
m.loop.ending_model  = 50         # index of the last loop model
m.loop.md_level = refine.very_fast # loop refinement method; this yields
                          # models quickly but of low quality;
                          # use refine.slow for better models

m.make()
```

8. assess_dope.py

```
from modeller import *
from modeller.scripts import complete_pdb

env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# Read a model previously generated by Modeller's automodel class
mdl = complete_pdb(env, 'IncC2-mult.pdb')

# Select all atoms in the first chain
atmsel = selection(mdl.chains[0])

score = atmsel.assess_dope()
```

9. assess_normalized_dope.py

```
# Example for: model.assess_normalized_dope()

from modeller import *
from modeller.scripts import complete_pdb

env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# Read a model previously generated by Modeller's automodel class
mdl = complete_pdb(env, 'IncC2-mult.pdb')

zscore = mdl.assess_normalized_dope()
```

10. trajectory.py

```
from modeller import *
from modeller.scripts import complete_pdb
from modeller.optimizers import molecular_dynamics, actions

env = environ()
env.io.atom_files_directory = ['.']
env.edat.dynamic_sphere = True
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')
```

```
code = 'IncC2-mult'
mdl = complete_pdb(env, code)

# Stereochemical restraints on all atoms:
atmsel = selection(mdl)
mdl.restraints.make(atmsel, restraint_type='stereo', spline_on_site=False)

# Write a PSF
mdl.write_psf(code+'.psf')

# Run 100 steps of MD, writing a CHARMM binary trajectory every 5 steps
md = molecular_dynamics(output='REPORT')
md.optimize(atmsel, temperature=300, max_iterations=10000,
        actions=actions.charmm_trajectory(50, filename=code+'.dcd'))
```

11. assess_ga341.py

```
from modeller import *
from modeller.scripts import complete_pdb

env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# Read a model previously generated by Modeller's automodel class
mdl = complete_pdb(env, 'IncC2-mult.pdb')

# Set template-model sequence identity. (Not needed in this case, since
# this is written by Modeller into the .pdb file.)
mdl.seq_id = 37.037

score = mdl.assess_ga341()
```

Appendix 10

Some of mass spectra from crosslinked peptides

IncC2-IncC2

KorB-KorB

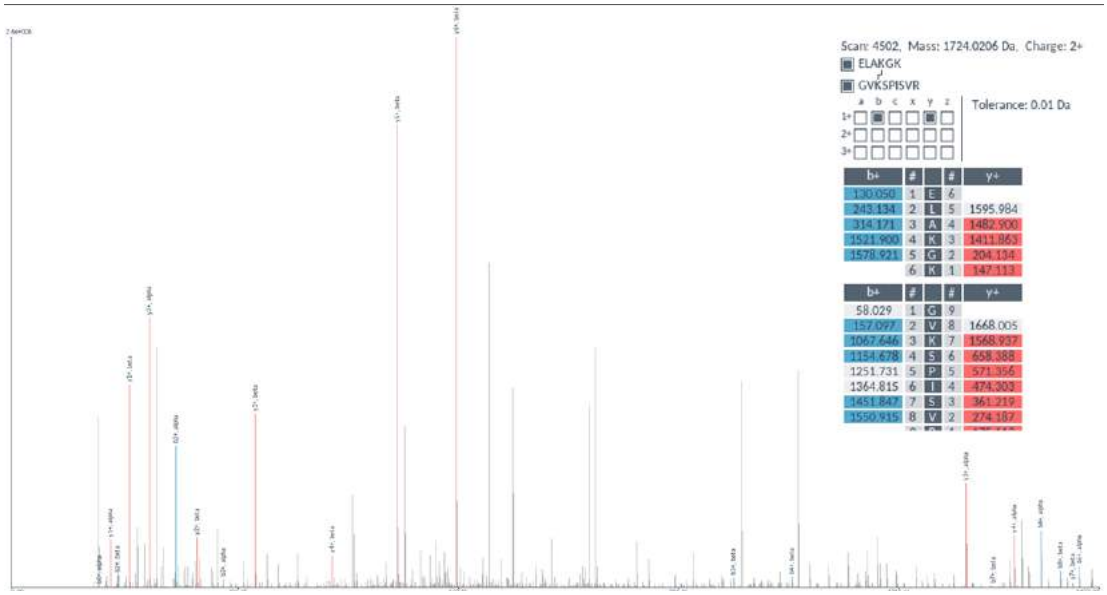


Figure 1 Mass spectra and peptide annotation of a IncC2-IncC2 and KorB-KorB crosslinked peptide

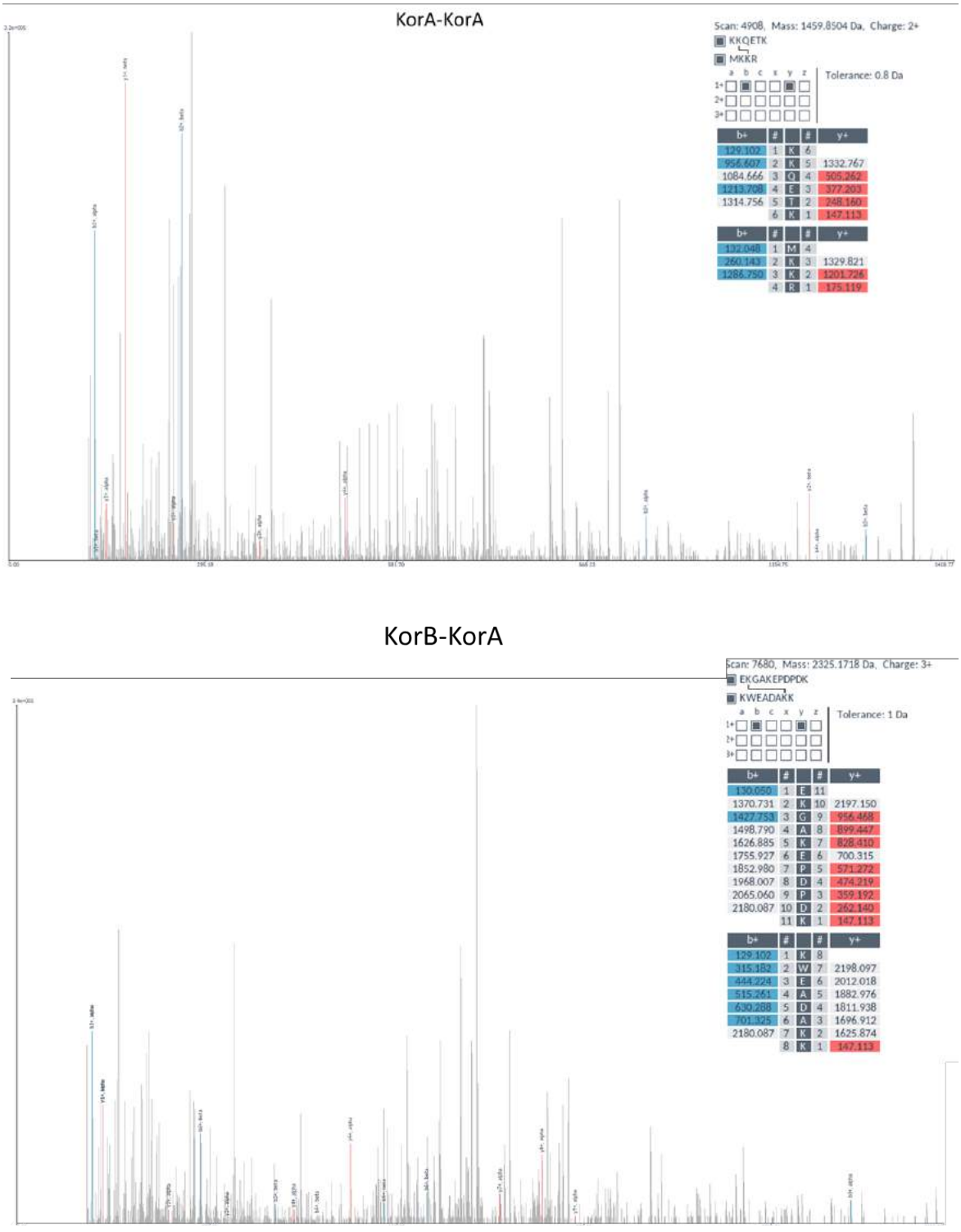


Figure 2 Mass spectra and peptide annotation of a KorA-KorA and KorA-KorB crosslinked peptide

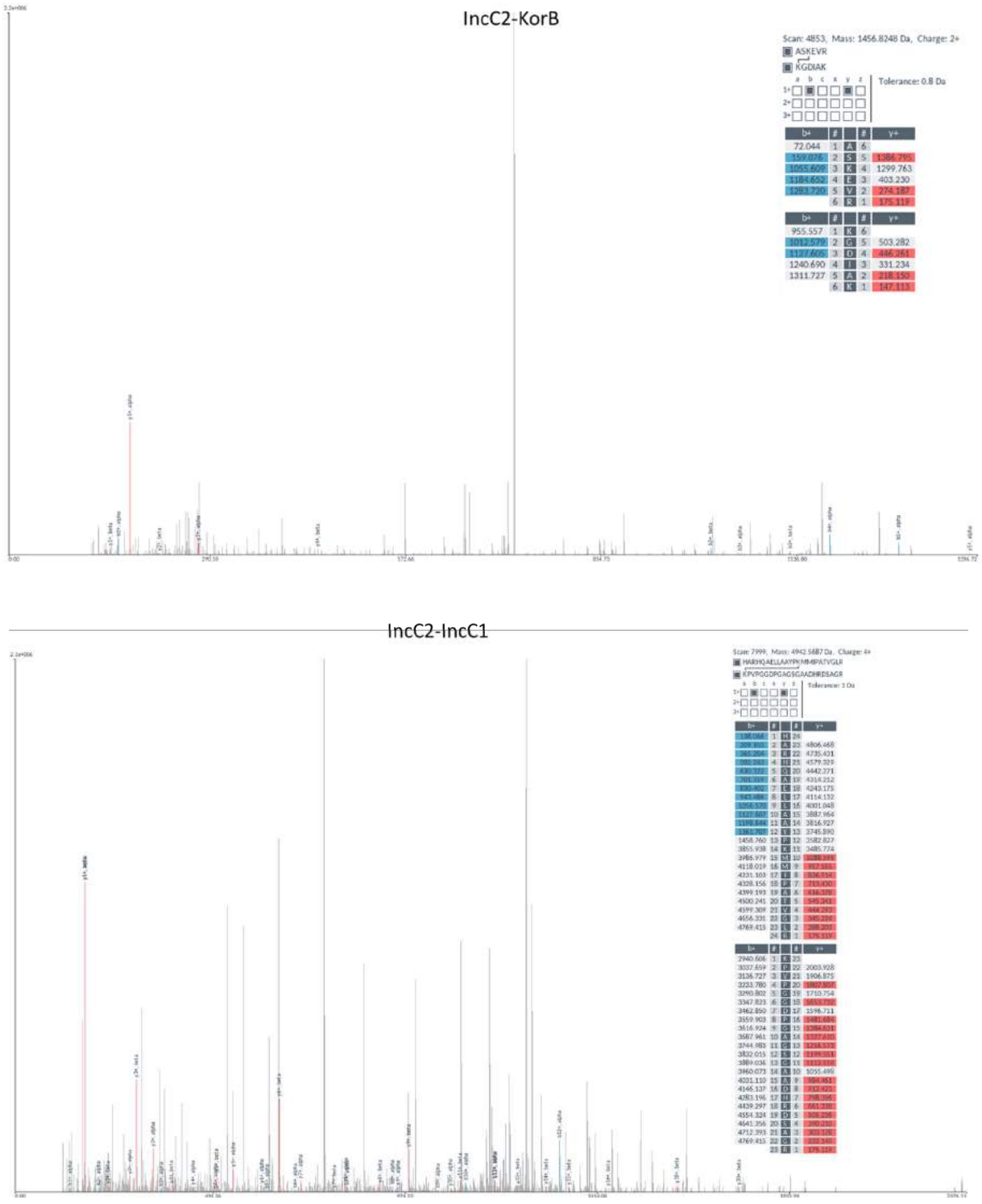


Figure 3 Mass spectra and peptide annotation of a IncC2-KorB and IncC2-IncC1 crosslinked peptide

Appendix 11

Crosslink distances mapped on DBD of KorB

Distance information for crosslinks mapped on DBD binding domain of KorB (The first amino acid in DBD was considered as residue # 1)

- LYS 33.A CA <-> LYS 38.A CA: 9.0
- LYS 33.B CA <-> LYS 38.B CA: 8.9
- LYS 33.C CA <-> LYS 38.C CA: 8.9
- LYS 33.D CA <-> LYS 38.D CA: 8.8
- LYS 33.A CA <-> LYS 106.A CA: 27.3
- LYS 33.B CA <-> LYS 106.B CA: 27.2
- LYS 33.C CA <-> LYS 106.C CA: 27.3
- LYS 33.D CA <-> LYS 106.D CA: 27.4
- LYS 33.A CA <-> LYS 106.D CA: 33.2
- LYS 33.C CA <-> LYS 106.B CA: 34.0
- LYS 42.A CA <-> LYS 31.A CA: 15.1
- LYS 42.B CA <-> LYS 31.B CA: 15.0
- LYS 42.C CA <-> LYS 31.C CA: 15.0
- LYS 42.D CA <-> LYS 31.D CA: 15.0
- LYS 42.A CA <-> LYS 38.A CA: 6.2
- LYS 42.B CA <-> LYS 38.B CA: 6.2
- LYS 42.C CA <-> LYS 38.C CA: 6.2
- LYS 42.D CA <-> LYS 38.D CA: 6.1
- LYS 42.A CA <-> LYS 38.C CA: 23.3
- LYS 42.C CA <-> LYS 38.A CA: 23.2
- LYS 84.A CA <-> LYS 32.A CA: 26.1
- LYS 84.B CA <-> LYS 32.B CA: 26.0
- LYS 84.C CA <-> LYS 32.C CA: 26.2
- LYS 84.D CA <-> LYS 32.D CA: 26.3
- LYS 42.A CA <-> LYS 32.A CA: 13.7
- LYS 42.B CA <-> LYS 32.B CA: 13.6
- LYS 42.C CA <-> LYS 32.C CA: 13.7
- LYS 42.D CA <-> LYS 32.D CA: 13.6
- LYS 38.A CA <-> LYS 32.A CA: 10.1
- LYS 38.B CA <-> LYS 32.B CA: 10.1
- LYS 38.C CA <-> LYS 32.C CA: 10.1
- LYS 38.D CA <-> LYS 32.D CA: 10.1
- THR 48.A CA <-> LYS 38.A CA: 13.4
- THR 48.B CA <-> LYS 38.B CA: 13.3
- THR 48.C CA <-> LYS 38.C CA: 13.3
- THR 48.D CA <-> LYS 38.D CA: 13.3
- LYS 83.A CA <-> LYS 32.A CA: 23.3
- LYS 83.B CA <-> LYS 32.B CA: 23.3
- LYS 83.C CA <-> LYS 32.C CA: 23.4
- LYS 83.D CA <-> LYS 32.D CA: 23.5
- LYS 84.A CA <-> LYS 33.A CA: 23.8
- LYS 84.B CA <-> LYS 33.B CA: 23.6
- LYS 84.C CA <-> LYS 33.C CA: 23.8
- LYS 84.D CA <-> LYS 33.D CA: 23.9
- LYS 84.A CA <-> LYS 31.A CA: 28.7

- LYS 84.B CA <-> LYS 31.B CA: 28.6
- LYS 84.C CA <-> LYS 31.C CA: 28.9
- LYS 84.D CA <-> LYS 31.D CA: 29.0
- LYS 84.A CA <-> LYS 106.A CA: 14.7
- LYS 84.B CA <-> LYS 106.B CA: 14.4
- LYS 84.C CA <-> LYS 106.C CA: 14.5
- LYS 84.D CA <-> LYS 106.D CA: 14.8

Crosslink distances mapped on IncC2 homology model (Dimer)

Distance information for crosslinks mapped on IncC2 dimer (The first amino acid in each dimer was considered as residue # 1)

- LYS 231.A CA <-> LYS 185.A CA: 19.6
- LYS 231.B CA <-> LYS 185.B CA: 19.6
- LYS 231.A CA <-> LYS 185.B CA: 24.2
- LYS 231.B CA <-> LYS 185.A CA: 24.2
- LYS 233.A CA <-> LYS 206.A CA: 30.6
- LYS 233.A CA <-> LYS 206.B CA: 28.7
- LYS 233.B CA <-> LYS 206.B CA: 30.6
- LYS 233.B CA <-> LYS 206.A CA: 28.7
- LYS 239.A CA <-> LYS 49.A CA: 22.6
- LYS 239.B CA <-> LYS 49.B CA: 22.6
- LYS 239.B CA <-> LYS 185.B CA: 11.7
- LYS 239.A CA <-> LYS 185.A CA: 11.7
- LYS 239.A CA <-> LYS 185.B CA: 13.7
- LYS 239.B CA <-> LYS 185.A CA: 13.7
- LYS 242.A CA <-> LYS 10.A CA: 26.0
- LYS 242.B CA <-> LYS 10.B CA: 26.0
- LYS 242.A CA <-> LYS 10.B CA: 26.1
- LYS 242.B CA <-> LYS 10.A CA: 26.1
- LYS 175.A CA <-> LYS 161.A CA: 18.2
- LYS 175.B CA <-> LYS 161.B CA: 18.2
- LYS 242.B CA <-> LYS 185.B CA: 14.3
- LYS 242.A CA <-> LYS 185.A CA: 14.3
- LYS 242.A CA <-> LYS 185.B CA: 13.5
- LYS 242.B CA <-> LYS 185.A CA: 13.5
- LYS 10.A CA <-> LYS 2.A CA: 24.9
- LYS 10.B CA <-> LYS 2.B CA: 24.9
- LYS 242.A CA <-> LYS 172.A CA: 38.6
- LYS 239.A CA <-> LYS 231.A CA: 13.7
- LYS 239.B CA <-> LYS 231.B CA: 13.7
- LYS 239.A CA <-> LYS 231.B CA: 23.0
- LYS 239.B CA <-> LYS 231.A CA: 23.0
- LYS 234.A CA <-> LYS 231.A CA: 5.1
- LYS 234.B CA <-> LYS 231.B CA: 5.1
- LYS 242.A CA <-> LYS 231.A CA: 12.5
- LYS 242.B CA <-> LYS 231.B CA: 12.5
- LYS 242.A CA <-> LYS 231.B CA: 26.3
- LYS 242.B CA <-> LYS 231.A CA: 26.3
- LYS 242.A CA <-> LYS 49.A CA: 22.3
- LYS 242.B CA <-> LYS 49.B CA: 22.3

- LYS 242.A CA <-> LYS 239.A CA: 5.3
- LYS 242.B CA <-> LYS 239.B CA: 5.3
- LYS 242.A CA <-> LYS 239.B CA: 12.9
- LYS 242.B CA <-> LYS 239.A CA: 12.9
- LYS 254.A CA <-> LYS 10.A CA: 24.8
- LYS 254.B CA <-> LYS 10.B CA: 24.8
- LYS 234.A CA <-> LYS 15.A CA: 21.2
- LYS 234.B CA <-> LYS 15.B CA: 21.2
- LYS 234.A CA <-> LYS 15.B CA: 29.4
- LYS 234.B CA <-> LYS 15.A CA: 29.4
- LYS 254.B CA <-> LYS 161.A CA: 28.3
- LYS 254.A CA <-> LYS 161.A CA: 22.5
- LYS 254.B CA <-> LYS 161.B CA: 22.5
- LYS 254.A CA <-> LYS 161.B CA: 28.3
- LYS 185.A CA <-> LYS 15.B CA: 29.3
- LYS 185.A CA <-> LYS 15.A CA: 10.0
- LYS 185.B CA <-> LYS 15.B CA: 10.0
- LYS 185.B CA <-> LYS 15.A CA: 29.3
- LYS 231.A CA <-> LYS 233.A CA: 5.6
- LYS 231.B CA <-> LYS 233.B CA: 5.6