Structural studies of the DNA partitioning protein KorB from the plasmid RK2

by

Anmol Gautam

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

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
November 2017

Abstract

The partitioning of low-copy number broad host range plasmids depend on a centromere binding protein (CBP) that binds to a centromere-like site on plasmid. For RK2 plasmid, the CBP is a 358 residue, multi-domain protein, KorB. KorB contains an N-terminal domain (NTD), a central DNA-binding domain (DBD), and a C-terminal dimerisation (CTD) domain and the protein binds to the O_B site on the plasmid as a dimer. Structures of central DBD and CTD have been elucidated whilst limited information is available about the N-terminal of the KorB.

In this study, NTD KorB protein was expressed and purified. Size exclusion chromatography and analytical ultracentrifugation data confirm that the NTD KorB behaves as a monomer in solution. Using solution-state NMR spectroscopy data, majority of the backbone and side-chain resonances for the NTD KorB are assigned and an ensemble structure of the protein is calculated. The flexibility of the NTD KorB is studied with coarse-grain Molecular Dynamics simulation package, AWSEM. The N-terminal of the NTD KorB is mostly unstructured whilst two α-helices towards the C-terminal of the protein exhibit limited motion.

Two C-terminal KorB deletion mutants capable of binding the O_B DNA (CΔ100 KorB and NΔ31CΔ100 KorB) were purified and characterised using circular dichroism (CD) and mass spectrometry. DNA-binding properties of these two deletion mutants are compared to the KorB wild-type (WT) using circular dichroism, fluorescence anisotropy and microscale thermophoresis measurements, indicating weaker binding of the deletion mutants with respect to the KorB WT to its centromere-like site (O_B). Considering the current state of structural information about the KorB and homologous proteins, a DNA partitioning model for the KorB is proposed.

Dedicated to Mom and Dad and the universe...

Acknowledgements

First, I would like to express my gratitude to Dr Eva Hyde for her invaluable scientific guidance and the opportunity to work on such a stimulating project. I would also like to issue a huge thanks to Dr Timothy Knowles for agreeing to supervise me and providing engaging NMR related Matrix analogies which made jigsaw-esque NMR relatable. I also wish to extend my deepest thanks to Dr Scott White for his scientific contribution especially his outstanding crystallographic knowledge and wizardry.

The research presented in this dissertation would simply not have been possible without the financial support from The Darwin Trust of Edinburgh. I would like to express my thanks to Prof David Finnegan and Heather Hall at the University of Edinburgh for their continued support throughout my doctoral degree. I would like to thank past and present members of Eva's lab including Emily Lythell, Thomas Penfare, and Tamara Lewis. A special thanks to Muhammad Fayyaz ur Rehman for being an excellent friend and a fantastic colleague throughout the duration of this monumental task. A big thanks to Rosemary Parslow and Pooja Sridhar for taking me under their wing and helping me rescue protein purification experiments. I would like to thank Dr Mark Jeeves and Dr Sara Whittaker of the Biomolecular NMR facility at the Henry Wellcome Building for their help during the NMR experiments.

A discussion with Johnny Lau nudged me to apply for Universitas21 scholarship, which presented me the opportunity to visit the lab of Prof Garegin Papoian at the University of Maryland. I would to thank Hao Wu and Mary Pitman in Papoian lab for helping me with Molecular dynamics simulations and introducing me to bash scripting, regular expressions and vector plotting packages. In addition, I want to extend my thanks to the super-computing facilities at the University of Birmingham (Blue Bear) and University of Maryland (Deep Thought) for granting time to run extensive and resource hungry protein simulations.

Much of the research presented in this work would simply not have been possible without the scientific expertise and invaluable contributions from several collaborators, both internally and externally, particularly Prof Christopher Thomas, Dr David Scott and Dr Peter Winn for critical scientific discussions throughout the evolution of my projects.

A special mention to one of my best friends Kamal Goswami and his supervisor Dr Alessandro Mottura who introduced me to the incredibly pragmatic and beautiful typesetting world of LaTeX. I am grateful to my colleagues and friends on the level 7 including Jack Charlton, Maria Azam, Penny, Richard Meek, Richard Logan, Charles and Haydn Little.

On a personal note, a big thanks must go to my friends and family and for their sustained support and understanding. I would like to say thanks to Nishtha Chandra, Janani Saikumar, Aditi Sharma, Muhammad Yasir, Sahara Rai, Sabarinath Krishnan, Kanika Gupta, Arhontissa Kanavaki, Tania Sakanaka, Mark Mshila, Sasha Hulsken, Beate Loesch, Jose Carlos, Bruno Theodoro, Rachana Mane for putting up with my gourmet makeshift delicacies and endless incompetence. And I express a deep sense of reverence for my parents and extended family members for their never ending affection and moral support.

Contents

List of Figures

List of Tables

1

Introduction

1 Introduction

Every organism needs to reproduce and fidelity in segregation of genetic material is a quintessential criterion for self-perpetuation. DNA partitioning (in bacteria) or DNA segregation (in eukaryotes) is the process by which genetic material is evenly distributed to newly formed daughter cells. Chromosome segregation in eukaryotic cells is accomplished by formation of spindle fibres, which helps the chromosomes to migrate to the opposite poles (Heald, 2000; Heald and Khodjakov, 2015). Accurate segregation of genetic material, chromosomes as well as plasmids, in bacteria is also crucial for their existence but so far the process has been less clear.

1.1 Plasmids

Plasmids are self-transmissible genetic elements that are extra-chromosomal DNA with their own replicative machinery (Pansegrau et al., 1994). Incompatible plasmids are said to be those plasmids that cannot coexist in the same host for several generations. This can be due to a shared replication mechanism or a similar partitioning system. As the host cannot distinguish between the types of plasmid, one plasmid is lost in due course of time from host population. One such group of incompatible plasmids is IncP, capable of maintaining in a few gram-positive bacteria and all gram-negative bacteria. Based on the analysis of the plasmid replication region and transfer region, the IncP plasmids are divided into IncP-α, (RK2-like plasmids) and IncP-β (R751-like plasmids) (Datta and Hedges, 1972).

1.1.1 RK2 plasmid

The RK2 plasmid is a broad-host range plasmid that belongs to the IncP-1α family and carries multiple antibiotic resistance genes. The RK2 plasmids were first identified at the Burns Research Unit, Birmingham (UK) as they conferred resistance to several an-

tibiotics in *Pseudomonas sp*. (Lowbury et al., 1969; Macartney et al., 1997). RK2 plasmid contains 60 kbp of DNA (62% G+C content), with a low-copy number of four to eight units in a single cell (Figurski and Helinski, 1979). All IncP plasmids diverged into α and β groups and evolved from a common ancestor, a conjugative P plasmid, as indicated by the sequence similarity in the replication region and transfer region (Chikami et al., 1985). IncP α and β plasmids can be distinguished on the basis of restriction enzyme analysis. Plasmids RP1, RP4, RK2, R18, R68 are different isolates, isolated at different times, belonging to IncPα subgroup and are indistinguishable from each other (Pansegrau et al., 1994).

1.2 Plasmid partitioning

Plasmid partitioning is a dynamic process where upon cell division, each daughter cell gets at least one copy of the plasmid. The high-copy number plasmids passively diffuse to daughter cells. On the other hand, most low-copy number plasmids utilise partition (par) systems for partitioning. These partition systems are found in most bacterial chromosomes and plasmids (Baxter and Funnell, 2014). Our understanding of prokaryotic partitioning has primarily resulted from cellular, genetic and biochemical data of low-copy number plasmids (Schumacher, 2008). In most cases, plasmid partitioning system consists of two proteins and one or more conserved regions on the plasmid (partition sites). The partition or par sites are centromere-like sites on the plasmid that direct the action of partitioning machinery (Baxter and Funnell, 2014). Partitioning depends on a site-specific centromere-binding protein (CBP) that binds to a centromere-like site on the plasmid and a nucleotide triphosphatase (NTPase) for mediating the plasmid partitioning. Both CBP and motor (NTPase) proteins are encoded on the plasmid and form the minimal machinery required for partitioning. Both CBP and motor proteins interact with plasmid DNA to accomplish the partitioning and the mechanism is briefly outlined in figure 1; a) the CBP binds to a centromere-like site on plasmid DNA to form

a protein-DNA complex *i.e.* partition complex, b) this complex then recruits a motor protein or an NTPase (not shown), c) further to recruitment, the NTPase mediates plasmid separation using a mechanism that has remained elusive. In this process, the CBP acts as an adaptor protein linking the NTPase and the plasmid DNA by interacting with the NTPase while simultaneously binding to the centromere-like site. The partitioning systems have been classified into multiple types, depending on the type of cytoskeletal-like NTPase (Schumacher, 2008). The mechanisms for partitioning associated with individual systems differ and are discussed in the next section (1.3).

1.3 Models of bacterial partitioning systems

Multiple partitioning mechanisms have been proposed in the literature regarding partitioning of plasmids in bacteria. These mechanisms are based primarily on the type of NTPase present in the partition system and in this section, different types of partition systems have been discussed (Gerdes et al., 2000).

The most common type of partition system is the type I system (pulling model). In type I systems, the NTPases are called ParA and CBPs are called ParB. Based on the size and sequence of the CBP and NTPase proteins, the type I partition system (section 1.3.1) can be divided into Ia and Ib. In type II system the NTPases are called ParM and CBPs are called ParR. R1 plasmid of *E. coli* is the best studied type II partition system (pushing model, section 1.3.2). The type III system, primarily studied in the pBtoxis plasmid (tramming model, section 1.3.3) utilises NTPases and CBPs called TubZ and TubR respectively as the principle proteins for partitioning. In addition, section 1.3.4 and 1.3.5 describes type IV partitioning system and diffusion ratchet model for plasmid partitioning. The partition proteins are encoded on the plasmid and the amounts of these are carefully auto-regulated at the transcriptional level (Schumacher, 2008). NTPases are responsible for auto-regulation in the type Ia systems whilst CBP proteins are responsible for auto-regulation in the type Ib, II and III systems (Schumacher, 2012).

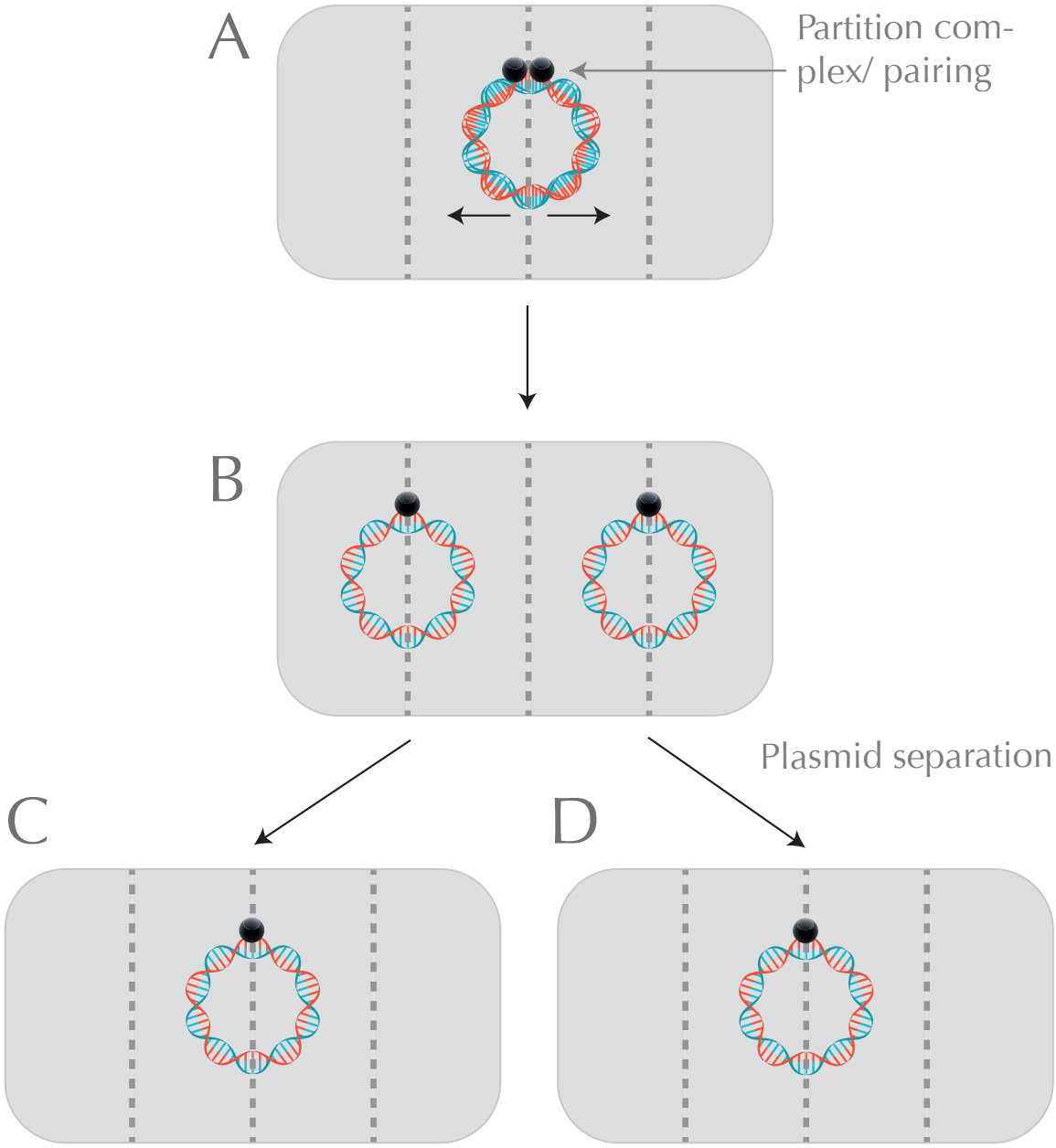

Figure 1: Illustration of general steps of plasmid partitioning. (A) The circular DNA in red and blue represents the plasmid at mid-cell position and two black dots on the plasmid refer to centromere-binding proteins (CBP). The CBP binds to the plasmid at centromere-like site and forms the partition complex. **(B)** After plasmid pairing, motor protein (not shown) gets recruited to the partition complex and both the proteins drive the two plasmids towards the opposite poles. In partitioning, the CBP acts as an adaptor protein by interacting with the motor protein while simultaneously binding to the plasmid. **(C-D)** After partitioning, the plasmids are at the mid-cell position in the daughter cells. The illustration is adapted from Schumacher, 2008.

1.3.1 Type I partition system

The type I partition systems consist of a centromere binding protein, an NTPase belonging to Walker-type ATPase and a centromere-like DNA site. As mentioned earlier, type I systems have been further divided into type Ia and Ib, depending on the genetic organisation of the par operon and on the basis of sequence and size of the Par proteins (Schumacher, 2008). In type Ia systems, the CBP gene is sandwiched between the centromere-like site and the gene encoding for the NTPase. In type Ib systems, the CBP gene is downstream of the motor protein, which in turn is downstream of the centromere site. Figure 2(A1-A3) illustrates the type I partitioning system. The components and structural biology of type I partition systems are given in table 1 and table 2 respectively. Well-studied type Ia systems include those from P1 and F plasmids from bacterium *E. coli* (Ogura and Hiraga, 1983).

For the P1 plasmid, the DNA-binding protein, ParB, binds to a cis-acting centromere-like sequence, *parS* and forms a higher order protein-DNA complex. This ParB-*parS* complex is then pulled towards the opposite poles of bacterial cell by a filament-forming ATPase protein, ParA (Bingle and Thomas, 2001). Formation of the ParA filament initiates with a nucleating core when ParA-ATP dimers binds to the bacterial chromosome (nucleoid), the filament then polymerises rapidly. Meanwhile, ParB binds to *parS* to form a complex. The ATPase activity of the ParA-ATP is triggered when it encounters a ParB bound to *parS*. ATP hydrolysis results in depolymerisation as ParA-ADP gets released from the nucleoid DNA. On each ParA-ATP hydrolysis and depolymerisation event (retraction of the ParA filament), ParB-*parS* complex is carried or 'pulled' along the length of the filament towards the cell pole (Gerdes et al., 2010; Schumacher, 2012).

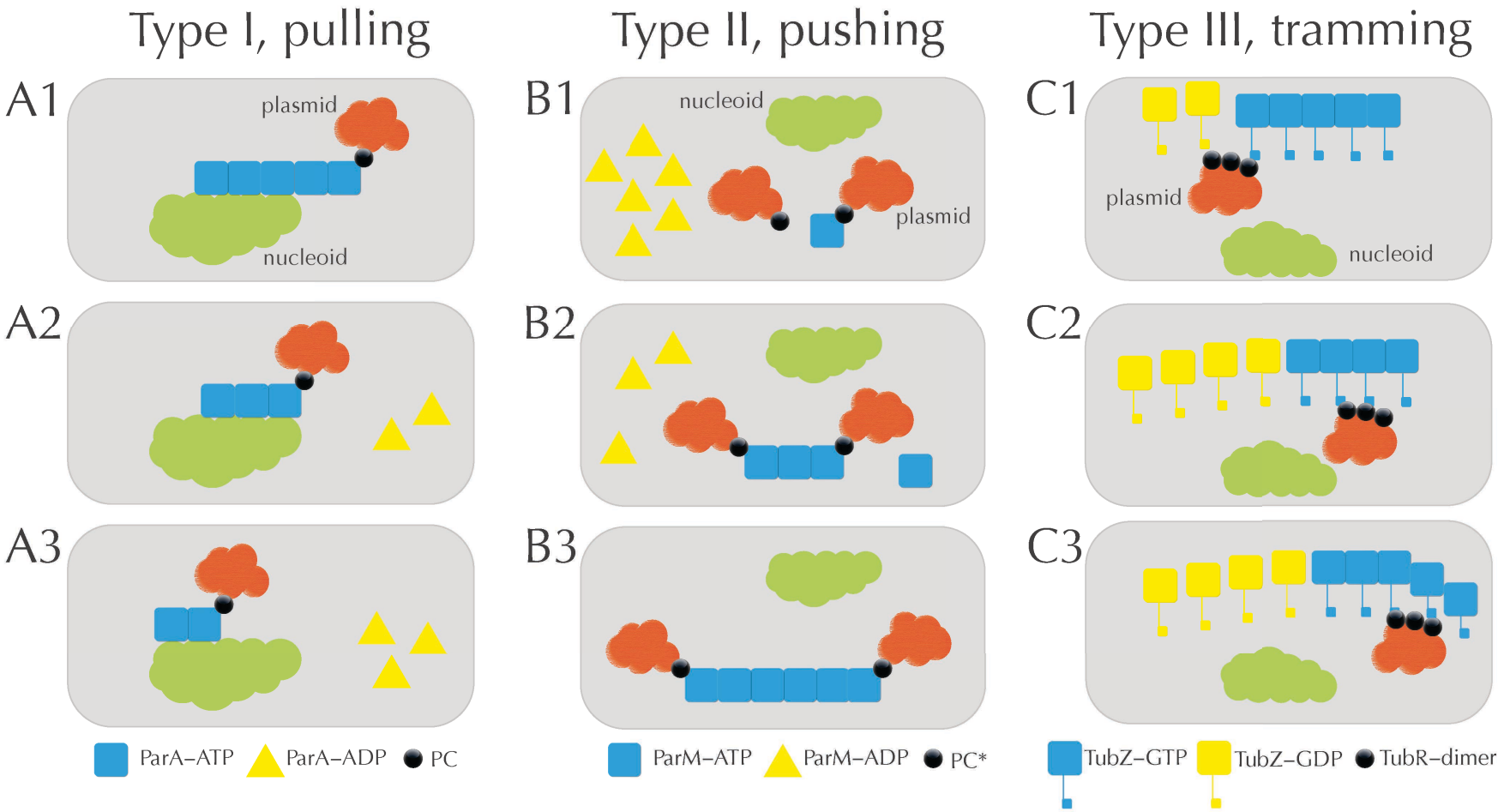

Figure 2: An overview of the partitioning systems. (A1-A3) Representation of Type I partitioning system (pulling model). ParA-ATP (blue square) binds to the nucleoid (green cloud). Upon encountering the ParB-plasmid partition complex (PC), ParA gets activated (ParA-ATP) resulting in ATP hydrolysis and depolymerisation occurs (ParA-ADP, yellow triangle). Partition complex is carried or 'pulled' in the direction of retreating ParA filament. **(B1-B3) Representation of Type II partitioning system (pushing model).** ParM filament (blue square) binds to the ParR-plasmid partition complex (PC*) on both ends and the complex is pushed to opposite poles via the elongating ParM filament (also known as insertional polymerisation model). **(C1-C3) Representation of Type III partitioning system (tramming model).** Schematic model for TubZ (GTPase) filament that exhibits tramming or treadmilling behaviour after encountering TubR-plasmid partition complex. Upon reaching the cell pole, TubZ filament bends to release the partition complex. The illustration is adapted from Schumacher et al., 2012.

Table 1: Type I ATPase partition system.

Host (bacterium)	Plasmid	*par* site	CBP	ATPase	Reference
E. coli, *Pseudomonas sp.*	RK2, RP4	O_B	KorB	IncC	(Delbruck et al., 2002; Khare et al., 2004; Williams et al., 1998)
E. coli	P1, P7	*parS*	ParB	ParA	(Radnedge et al., 1998)
E. coli	F	*sopC*	SopB	SopA	(Ravin et al., 2003)

Table 2: Structural biology of type I partition system.

Protein	Plasmid	PDB ID	Reference
CBP: KorB (dimerisation domain)	RP4	1IGQ, 1IGU	(Delbruck et al., 2002)
CBP: KorB (DNA-binding domain)	RP4	1R71	(Khare et al., 2004)
CBP: ParB	P1	1ZX4, 2NTZ	(Schumacher and Funnell, 2005; Schumacher et al., 2007b)
CBP: SopB (dimerisation domain)	F	3KZ5	(Schumacher et al., 2010)
CBP: SopB (DNA-binding domain)	F	3MKW, 3MKY, 3MKZ	(Schumacher et al., 2010)
ATPase: ParA	P1	3EZ7, 3EZ2, 3EZ6	(Dunham et al., 2009)
ATPase: ParA	P7	3EZ9, 3EZF	(Dunham et al., 2009)

1.3.2 Type II partition system

The type II partition system consists of a CBP, an ATPase belonging to the actin/hsp70 superfamily and a centromere-like site (Gerdes and Molin, 1986). The genetic organisation of type II resembles the type Ib partition system. The system includes an ATPase encoding gene sandwiched between a centromere-like site (upstream) and a gene coding for the CBP (downstream) (Schumacher, 2008). Figure 2(B1-B3) illustrates the type II partitioning system. The components and structural biology of type II partition systems are given in table 3 and table 4 respectively.

R1 plasmid of *E. coli* is the best studied type II system so far and an insertional polymerisation model is proposed for plasmid partitioning. In this system, the NTPases are called ParM and CBPs are called ParR. ParM is an ATP-dependent filament forming protein that polymerises bidirectionally and contain pockets for ATP binding (Moller-Jensen et al., 2003). ParR (from pB171 plasmid) is a dimer of dimers containing ribbon-helix-helix (RHH) motif that binds to the DNA. The C-terminal domain of ParR forms a three helix 'cap' that has been proposed to strengthen the dimer itself and stabilise the dimer of dimer (Moller-Jensen et al., 2007). Six dimers of dimers of ParR assemble over 360° turn with N-termini facing outward and C-termini facing inward towards helix centre. The ParR/*parC* superhelix once assembled, forms a clamp over the growing ParM filaments. The complex caps and stabilises the terminal of single ParM filament and binds to two ParM-ATP units using the ParR C-termini in the interior of the superhelix. Only one ParM-ATP subunit can be added at a time to the ParM filament due to steric constraints. ATP hydrolysis allows the release of ParR/*parC* superhelix on one side only and this then attaches to newly introduced ParM-ATP unit, resulting in translocation. The translocation introduces necessary space for ParM-ATP unit to be added. Iteration of the translocation process alongside continued growth of ParM filament, pushes the plasmids towards the pole and facilitates the partitioning process (Salje and Lowe, 2008; Schumacher, 2012).

Table 3: Type II ATPase partition system.

Host (bacterium)	Plasmid	*par* site	CBP	GTPase	Reference
E. coli	R1, pB171	*parC*	ParR	ParM	(Ebersbach and Gerdes, 2001; Jensen et al., 1998)

Table 4: Structural biology of type II partition system.

Protein	Plasmid	PDB ID	Reference
CBP: ParR	pB171	2JD3	(Moller-Jensen et al., 2007)
CBP: ParR	pSK41	2Q2K	(Schumacher et al., 2007a)
ATPase: ParM	R1	1MWK, 1MWM, 2ZHC, 4A61, 4A6J, 2ZGY, 2ZGZ, 4A62	(Galkin et al., 2009; Gayathri et al., 2012; Popp et al., 2008; van den Ent et al., 2002)

1.3.3 Type III partition system

The system was identified in the plasmid pBtoxis from *B. thuringiensis*. Type III partition system consists of TubZ (GTPase), TubR (DNA-binding protein) and *tubC* as the partition site (reviewed by Schumacher, 2008). Type III system has a distinct genetic organisation in comparison to type I and II partition systems. The gene for the *tubR* is upstream of the *tubZ*. In this system, the TubR protein (with putative HTH domain) binds to the *tubC*. TubZ, being the motor protein of this system exhibits GTPase activity and is distantly related to the tubulin/FtsZ GTPase superfamily (Larsen et al., 2007). The components and structural biology of type III partition systems are given in table 5 and table 6 respectively.

Table 5: Type III GTPase system.

Host (bacterium)	Plasmid	*par* site	CBP	GTPase	Reference
B. anthracis	pXO1	*tubC*	TubR	TubZ	(Ni et al., 2010)
B. thuringiensis	pBtoxis	*tubC*	TubR	TubZ	(Larsen et al., 2007)

Table 6: Structural biology of type III partition system.

Protein	Plasmid	PDB ID	Reference
GTPase: TubZ	pBtoxis	3M8K, 2XKB, 2XKA, 3M89	(Aylett and Lowe, 2012; Larsen et al., 2007)
CBP: TubR	pBtoxis	3M8E, 3M8F, 3M9A, 4ASO (with 24-mer *tubC*), 4ASS (with 26-mer *tubC*)	(Aylett et al., 2010; Ni et al., 2010)

Figure 2(C1-C3) illustrates the type III partitioning system. TubZ assembles into double-stranded helical fragments and exhibits tramming or treadmilling behaviour like free microtubules, both in presence and absence of the DNA-binding protein TubR, which might help in binding TubZ to the plasmid. In this model, the TubR binds to the pBtoxis plasmid to form TubR-pBtoxis complex. Using the solvent exposed C-terminal domain of TubZ, the TubR-pBtoxis complex interacts with the filaments of TubZ. The GTP hydrolysis within the TubZ filament assists in the treadmilling, polymerisation at the plus end, retraction at the minus end and transporting TubR-plasmid to one of the cell's pole. Bending of the TubZ filament can stimulate detachment of TubR-plasmid. Now, TubZ continues to treadmill in the opposite pole, attaching to another TubR-plasmid complex and transporting it to the opposite pole (Larsen et al., 2007; Schumacher, 2012).

1.3.4 Type IV partition system

The type IV partitioning system requires only one protein (Schumacher, 2008). The protein is labelled as Par (245 residues) and is identified in pSK1 plasmid from *Staphylococcus aureus* (Simpson et al., 2003). Structural predictions suggest Par protein to have a putative centromere-binding HTH motif in the N-terminal domain and hints towards a central coiled-coil domain that might help the Par protein to polymerise, similar to cytoskeleton proteins in eukaryotic cells (Weitao et al., 2000).

It is unlikely that a second protein is involved in the process as only one open reading frame has been identified for a protein in the coding region (Simpson et al., 2003). Conceivably, Par protein has structural attributes to perform both the functions in partitioning, but it is still unclear as to how Par protein performs both DNA-binding and polymerisation roles.

1.3.5 A recent partitioning model: the diffusion ratchet model

The diffusion ratchet model is a more recent model of plasmid partitioning. Both P1 and F plasmid encode an ATPase, ParA responsible for plasmid movement (in conjunction with the bacterial nucleoid). In this model, the ParB-*parS* complex binds ParA but ParA is no longer thought to form filaments. Instead, the movement of the plasmid is guided by following the concentration of the ParA.

According to the model ParA exists in two forms; one is the active ParA form that binds to the plasmid. The other form is the inactive ParA (diffusible form) that diffuses in to cytoplasm. The conversion of ParA between the active form and the diffusible form is a slow process and this conversion between the two forms rely on the ParA-ATP binding cycle. The slow conversion of the ParA essentially creates two pools of the protein; one that is bound to the nucleoid and the other one, which is free to diffuse and creates a zone of depletion.

Figure 3 illustrates the diffusion ratchet model. On the nucleoid (not shown in the figure), ParB-plasmid DNA complex interacts with ParA, resulting in the conversion of active form of ParA to the inactive form. The hydrolysis of the active ParA-ATP gives rise to the ParA-ADP (inactive). The inactive ParA diffuses from its location, creating a zone of depletion as it takes time to charge the ParA with ATP. ParA moves until it encounters an ATP molecule and becomes active again. In a dividing bacterial cell, when two ParB bound plasmids are present, the protein-DNA complex ratchets along the surface of the nucleoid following the gradient of ParA protein and moves towards higher concentration of active ParA (Baxter and Funnell, 2014; Hwang et al., 2013; Song and Loparo, 2015; Vecchiarelli et al., 2010, 2013, 2012, 2014).

1.4 ParB superfamily of proteins

The ParB superfamily of proteins are centromere binding proteins (CBP) that share common structural domains with similar functions: an N-terminal domain (NTD) for oligomerisation and protein-protein interactions, a central DNA-binding region for site-specific *parS* binding and a C-terminal domain (CTD) for ParB dimerisation (Schumacher, 2008). Figure 4 illustrates a simple model to understand the spatial organisation of protein-DNA interactions such as 1D spreading and 3D bridging. The interactions are applicable to ParB proteins binding to either chromosomal or plasmid DNA. In this section, emphasis is given on the ParB proteins containing a helix-turn-helix (HTH) motif and common features shared by the N-terminal of CBPs. ParB proteins include KorB and Spo0J as both the proteins contain a flexible N-terminal domain, a central DNA-binding domain and a C-terminal dimerisation domain. Centromere binding proteins with an HTH motif along and the role of N-terminal of the CBPs are discussed in sections 1.4.1 and 1.4.2 respectively.

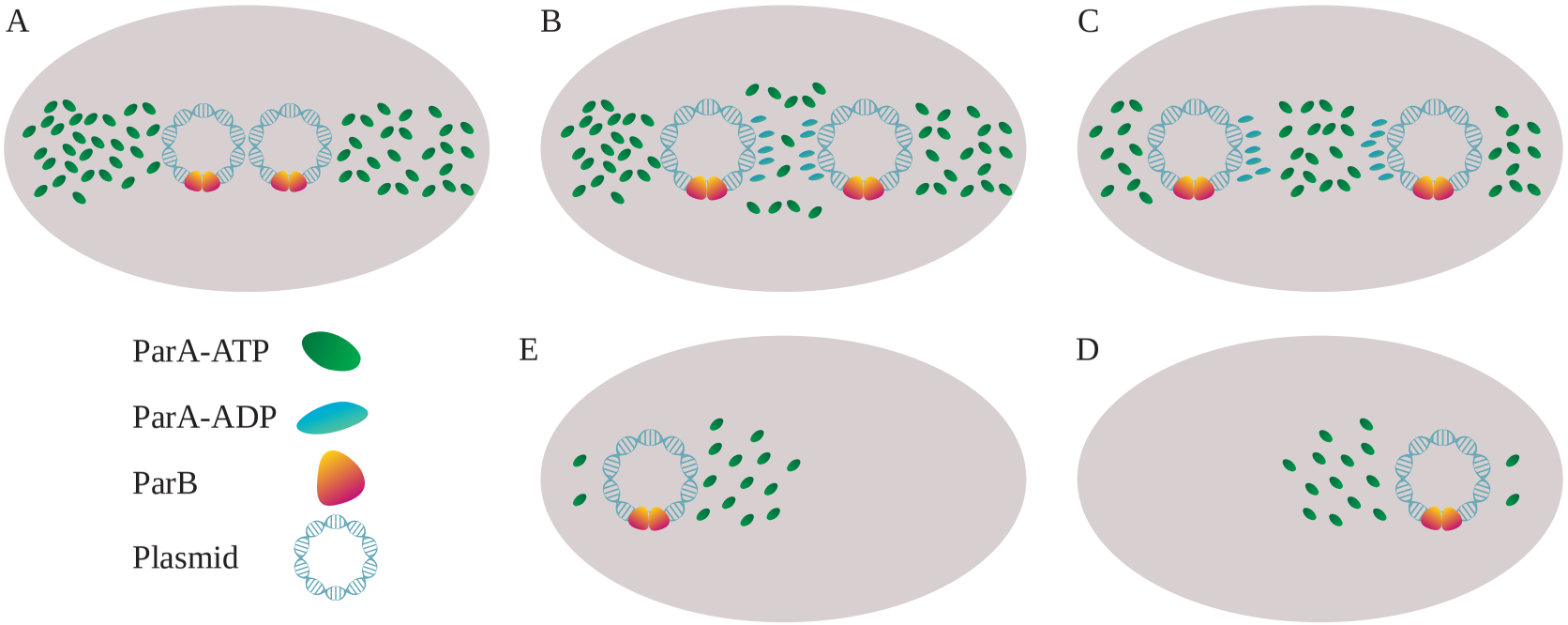

Figure 3: Diffusion ratchet model for plasmid partitioning. A simple illustration of diffusion ratchet model to understand the spatial organisation of plasmid and partitioning proteins, ParA and ParB. ParA exists in two forms, ParA-ATP and ParA-ADP, where ParA-ATP (active) is capable of binding to the nucleoid whilst ParA-ADP is diffusible in cytoplasm. The conversion between these two forms of ParA is slow as there is a delay in the formation of ParA-ATP complex. **(A)** DNA-binding protein, ParB binds as a dimer to the plasmid at *parS* site to form the partitioning complex. **(B)** The ParB-plasmid complex interacts with activated ParA-ATP bound to nucleoid (not shown), stimulating ATP hydrolysis and converting active ParA into its inactive form. The inactive ParA moves from its location, creating a zone of depletion. Following brownian motion, ParB-plasmid complex interacts with next available ParA-ATP and ratchets along. The depletion zone favours directional movement of the partitioning complex where ATP concentration gradient provides the driving force. **(C)** The adjacent activated ParA-ATP proteins diffuse towards depletion zone to refill and cause the partitioning complex to move and the cell divides into daughter cells (not shown). **(D-E)** The daughter plasmids approaching mid-cell position using the same mechanism. The illustration is adapted from Baxter and Funnell, 2014.

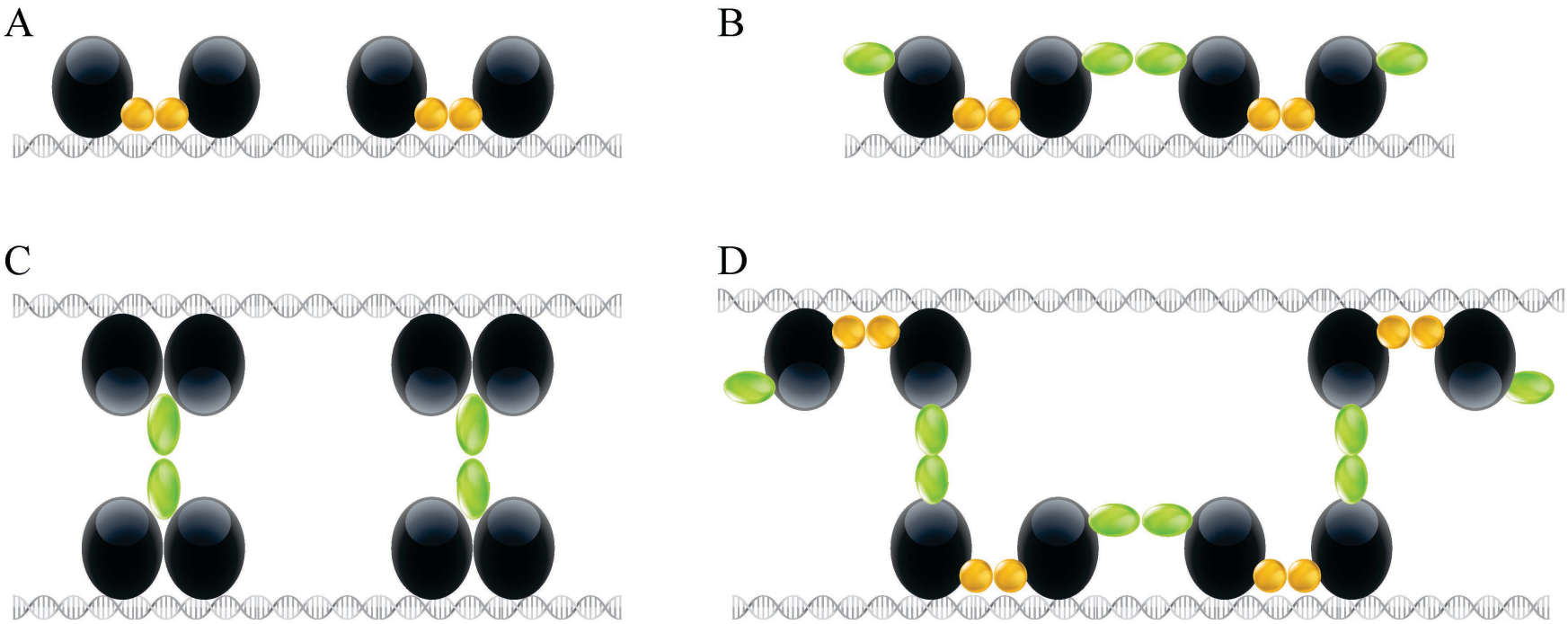

Figure 4: Model for DNA-bound protein interactions. A simple model is proposed to understand the spatial organisation of DNA-bound protein interactions. DNA is described as a self-avoiding, linear chain where proteins can interact in different ways. Here, DNA is shown in grey, protein molecules are shown in black ovals where orange circles represent protein dimer interaction and green ovals represent non-dimer protein interactions. **(A)** Non-interacting dimers: protein binds as a dimer to the DNA. **(B)** Spreading interactions: protein binds as a dimer to the DNA and the protein spreads along the DNA using protein-protein interactions. **(C)** Bridging interactions: protein binds as a dimer to the DNA and also bridges the DNA. **(D)** Spreading or bridging interactions: here each protein monomer contains two interaction domains, one of the domains can either be in spreading or bridging mode. The model is adapted from Broedersz et al., 2014.

1.4.1 Centromere binding proteins with HTH motif

The CBPs with an HTH motif in the DNA-binding domains are mainly observed in bacterial chromosome and plasmid partitioning type Ia system (Baxter and Funnell, 2014). The sequence similarity of residues is not high but the structural and functional aspect of domain organisation is conserved (Lukaszewicz et al., 2002; Ravin et al., 2003; Surtees and Funnell, 1999, 2001).

CBPs with an HTH motif typically bind to DNA sequences with inverted repeats, as is the case with the *parS* sequences (Baxter and Funnell, 2014). KorB from RP4 plasmid is a CBP and is a global transcription regulator that binds to an operator sequence called O_B (Bignell and Thomas, 2001). The CBPs of P1 and P7 plasmid in *E. coli* have bipartite recognition sites with two distinct sequence elements.

One of the recognition sites is called the A-box (an inverted repeat sequence) and the other is a hexamer sequence known as the B-box (Funnell and Slavcev, 2004). The *sopC* site of F plasmid consists of a 43 bp repeat sequence with 12 copies in the plasmid (Schumacher et al., 2010). Out of the 12 copies, only one is essential for the partitioning process. The CBP for F plasmid *i.e.* SopB binds specifically to a 16 bp inverted repeat within the 43 bp long *sopC* site and the rest of the sequence has been proposed to be required for optimal spacing (Pillet et al., 2011). The CBPs with an HTH motif have been found to oligomerise and spread along the DNA at a distance from their *parS* sites (Baxter and Funnell, 2014).

The crystal structures for domains of three CBPs with an HTH motif have been solved. These include KorB, ParB and SopB from RP4, P1 and F plasmid respectively (Delbruck et al., 2002; Khare et al., 2004; Schumacher and Funnell, 2005; Schumacher et al., 2010). The crystal structures of KorB, ParB and SopB are primarily helical in nature, are similar to each other and contain a DNA-binding HTH motif (Baxter and Funnell, 2014). The 'recognition helix' in the HTH region of the protein sits in the major-groove of the DNA and the HTH part of the CBP binds to an inverted repeat sequence.

The regions that impart specificity to the protein in relation to binding a specific DNA sequence are different in these three CBPs (KorB, ParB, SopB). Residues in the α3 helix of ParB make specific contacts with the inverted repeat sequence on the plasmid P1 (Schumacher and Funnell, 2005; Schumacher et al., 2007b). SopB utilises R219 along with α3 helix in the HTH region for making specific contacts with F plasmid (Sanchez et al., 2013; Schumacher et al., 2010). In case of KorB from RP4 plasmid, T211 (α6 helix) and R240 (α8 helix) are the residues responsible for specificity and α3-α4 helix in the HTH region are responsible for non-specific DNA-binding (Khare et al., 2004). The NTD of these three CBPs fail to crystallise because the proteins are flexible in nature (Hanai et al., 1996; Rajasekar et al., 2010; Ravin et al., 2003; Surtees and Funnell, 1999). Possibly, flexibility in the NTD permits the proteins to bind to more than one partner and carry out diverse biological roles.

1.4.2 N-terminal domain of CBPs

The NTD of the HTH CBPs are involved in forming oligomers on the DNA chain. ParB proteins encoded by both bacterial chromosome and plasmid have the capacity to bridge DNA. In addition, the NTD interacts with the NTPase involved in the plasmid partitioning (Ah-Seng et al., 2009; Radnedge et al., 1998; Ravin et al., 2003; Surtees and Funnell, 1999). Two positively charged and conserved residues, Lys3 and Lys7 in the NTD of Spo0J, a ParB superfamily protein have been shown to interact with the ATPase Soj from *Bacillus subtilis* (Scholefield et al., 2011). Spo0J regulates the ATPase activity along with the dimerisation state of Soj. A recent study by Graham et al., (2014) shows that Spo0J is involved in DNA bridging. DNA bridging capabilities were diminished with R79A, R80A, and R82A mutants *in vitro* and these three mutants could not spread *in vivo*. The stretch of residues spanning amino acids (aa) 79-82 has been defined as 'arginine patch' and the residues have been implicated in the DNA bridging interactions (Graham et al., 2014).

The Spo0J from *Helicobacter pylori* (see section 1.4.3.2) has a flexible NTD involved in protein-protein interactions and the protein folds into an elongated structure with *parS* binding capacity through the conserved central DNA-binding domain. Chen et al., (2015) proposed that adjacent and traverse protein-protein interactions led to the oligomerisation of Spo0J on the DNA. The protein-protein interactions occur via the 'arginine patch', which is highly conserved in the ParB family. Using simulation studies, a similar model was presented by Broedersz et al., (2014) suggesting that the NTD of CBP is involved in 1D spreading and 3D bridging over plasmid DNA (see figure 4); they also proposed that these interactions can be important for the assembly of high-order protein-DNA complex resulting into ParB spreading on to the DNA (Broedersz et al., 2014). The exact stoichiometry between the ParB and plasmid necessary for the process of partitioning is still unclear. How CBPs interact with each other around the *parS* sites non-specifically and coat the DNA is yet to be deciphered.

1.4.3 KorB homologues

Spo0J (stage 0 sporulation protein J) is a member of the ParB family of proteins involved in bacterial chromosomal segregation. Spo0J is an essential part of the ParABS partition system, where ParA is the motor protein, ParB is the DNA-binding protein and *parS* is the centromere-like site on the chromosomal DNA. Spo0J is a KorB homologue as it binds to the *parS* sites *in vivo* with an HTH motif to facilitate bacterial chromosomal segregation. The crystal structures of Spo0J from *Thermus thermophilus* and *Helicobacter pylori* are available in the protein data bank (PDB) and both of the structures are discussed in the sections 1.4.3.1 and 1.4.3.2 respectively.

1.4.3.1 Spo0J protein from *Thermus thermophilus*

Spo0J is a KorB homologue from *Thermus thermophilus* which binds to *parS* sites *in vivo*. The structure of the N-terminus of the Spo0J protein (1-222) from the bacterium *Thermus thermophilus* has been determined (PDB ID: 1VZ0). The Spo0J protein dimer was

crystallised in the absence of DNA but the dimer does contain appropriate distance (34 Å) to bind DNA (Leonard et al., 2004). Figure 5A shows the Spo0J dimer with the HTH motif shown in yellow. Since Spo0J has been crystallised, based on sequence alignment and secondary structure prediction, there is increased chance that KorB construct with amino acids 60-222 can form diffraction quality crystals.

1.4.3.2 Spo0J protein from *Helicobacter pylori*

The structure of the CTD-truncated Spo0J from *Helicobacter pylori* has been reported (Chen et al., 2015). The Spo0J was crystallised as a dimer (at a resolution of 3.1 Å) in the presence of 24 bp *parS* site (PDB ID: 4UMK). Two Spo0J molecules interact with the DNA and the CTD primarily functions as the dimerisation domain. The N-terminal of Spo0J being flexible in nature serves as a contact point for spreading on the DNA (Chen et al., 2015; Graham et al., 2014; Havey et al., 2012). Figure 5B illustrates the flexibility of the NTD as four Spo0J protein molecules observed in the same crystal structure (without the DNA) are superimposed for structural comparison. In the four monomers, the helices in the C-terminal share a similar DNA-binding domain (112-227 aa with RMSD of 0.3–0.6 Å for Cα atoms) whilst the N-terminal of the protein does not superimpose (35-111 aa with RMSD of 3.9–8.1 Å for Cα atoms). Various loop regions are observed in the N-terminal suggesting high flexibility of the N-terminal domains (Chen et al., 2015).

ParB binds to the chromosomal DNA specifically at *parS* sites and also non-specifically (Breier and Grossman, 2007; Murray et al., 2006). Using simulations, Broedersz et al., (2014) proposed that it is the combination of 1D spreading and 3D bridging which is responsible for ParB coating the DNA. Based on experimental data, Chen et al. (2015) have proposed a ParB spreading model on DNA (*parS* site). The model suggests that multiple ParB protein molecules bind horizontally and vertically on a long and looping DNA molecule exploiting the NTD interactions. The model proposes a high-order protein DNA complex with ParB bound to DNA (Chen et al., 2015).

ParB spreads along the chromosomal DNA and this helps to segregate the DNA. ParB utilises three domains to achieve this. The central DNA-binding domain of ParB binds to the DNA at specific *parS* sites. ParB-*parS* complex is stabilised as ParB dimerises with adjacent protein molecule using the CTD. The flexible NTD allows for multiple protein-protein interactions and adjacent ParB dimers interact using horizontal interactions, which help in spreading of the protein on the DNA. Dimer ParB on the looped DNA interacts using vertical and traverse interactions because of the protruded NTD. Considering the (non) specific interactions, the result is a multi-order protein DNA complex. It is postulated that the protein DNA complex recruits energy rich ParA protein *i.e.* ParA-ATP and hydrolysis of nucleotide imparts the driving force necessary to segregate the bacterial chromosome (Chen et al., 2015).

1.5 KorB protein from RK2 plasmid

In case of RK2 plasmid (low copy number and broad host range IncP-1 plasmid), the motor protein is an ATPase, IncC and the CBP is the KorB protein and both IncC and KorB are required for plasmid partitioning (Jagura-Burdzy et al., 1999; Motallebi-Veshareh et al., 1990). In addition to a partitioning protein, KorB has also been shown to play role in coordinating gene expression (Bingle et al., 2005; Kostelidou and Thomas, 2000). On the RK2 plasmid (RK2), KorB is encoded within the central control region (CCR). The CCR encodes KorA, IncC and KorB and is regulated by both KorA and KorB (Batt et al., 2009).

KorA is a repressor protein that acts in conjunction with KorB to regulate the expression of the CCR (Bingle et al., 2005). IncC contains an ATP-binding motif and it belongs to the ParA family of proteins (Motallebi-Veshareh et al., 1990). IncC can modulate KorB binding to its DNA operators (Jagura-Burdzy et al., 1999). The DNA-binding functionality of KorB in plasmid partitioning, with its protein partner IncC, makes KorB an important structural target to understand partitioning at molecular level.

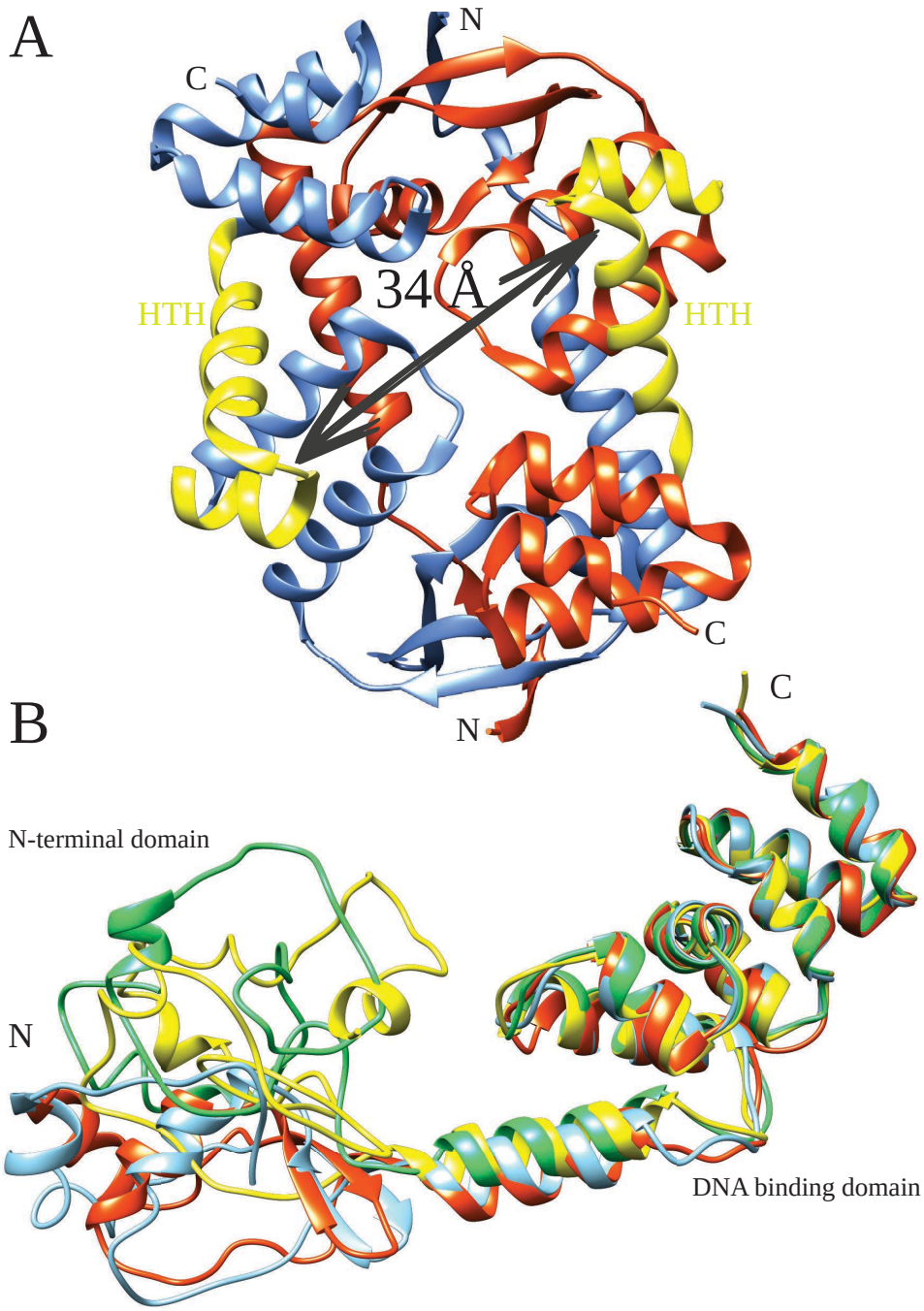

Figure 5: Structures of Spo0J protein. Spo0J protein is a KorB homologue and binds to *parS* sites *in vivo* using an HTH motif. **(A)** Ribbon representation of Spo0J dimer from *Thermus thermophilus* is shown. One of the monomers is shown in blue, the other one in orange and the HTH motif in both monomers is shown in yellow. The dimer is crystallised in the absence of the DNA but the dimer does contain appropriate distance (34 Å) to bind DNA (Leonard et al., 2004). **(B)** Ribbon representation of Spo0J from *Helicobacter pylori* is shown. Four Spo0J monomers (from the same crystal structure), chains A (55–226, red), B (35–226, light blue), C (45–227, yellow), and D (51–226, green) are superimposed for structural comparison. In the four monomers, the helices in the C-terminal (112-227 aa) share a similar DNA-binding domain whilst the N-terminal (35-111 aa) of the protein does not superimpose. Various loop regions are observed in the N-terminal suggesting high flexibility of the N-terminal domains (Chen et al., 2015). The image is rendered with POV-ray in Chimera.

KorB is a 358 aa, multidomain, DNA-binding ParB homologue. The subunit mass of KorB is 39.01 kDa and there are approximately 2000 molecules of KorB per bacterium cell. KorB binds specifically to the operator sequence (O_B site) on the RK2 plasmid. The palindromic O_B site [5'-TTTAGC(G/C)GCTAAA-3'] is present 12 times in the 60 kb plasmid (Balzer et al., 1992; Delbruck et al., 2002; Pansegrau et al., 1994). The full-length KorB behaves as a dimer in solution (Balzer et al., 1992). Intact or wild-type (WT) KorB includes an N-terminal domain (1-137 aa), a central DNA-binding domain (DBD, 138-258 aa) and a C-terminal dimerisation domain (297-358 aa). Of these, the central DNA-binding domain and the C-terminal domain have previously been crystallised and these domains are discussed in sections 1.5.1 and 1.5.2 respectively.

Two additional regions in KorB are predicted from their sequence to be flexible and intrinsically disordered. Secondary structure prediction based on amino acid sequence indicates that 1-64 aa and 258-297 aa of KorB are flexible in nature (Rajasekar et al., 2010). To date there is no structure of an intact KorB protein with or without DNA. This is possibly due to the disordered regions described above, which pose a hindrance to crystal formation of KorB and the intact protein is too big to be studied by solution state NMR spectroscopy alone.

1.5.1 Central DNA-binding domain of KorB

The crystal structure of the central DNA-binding domain of KorB with residues 138-252 (figure 6A) has been determined in the presence of O_B DNA (Khare et al., 2004). The protein was crystallised in presence of a 17 bp operator O_B DNA oligonucleotide in B-form conformation. The resultant model (KorB-O_B DNA) contains two protein dimer-DNA complexes *i.e.* two DNA oligonucleotides and four protein monomers. The central region behaves as monomer in solution, indicated by size exclusion chromatography and cross-linking study with glutaraldehyde. Helices α3 and α4 form the HTH motif, typical of other DNA-binding proteins such as prokaryotic and eukaryotic

transcription factors. However, unusually, the HTH motif was observed to bind the operator DNA using non-sequence specific water mediated H-bonds and van-der Waals contacts at the major-groove of the DNA. Using site-directed mutagenesis, two residues located on helices α2 and α4 helix, outside the HTH (T211 and R240) were shown to play crucial roles in nucleotide base recognition using direct H-bond contacts to C and G respectively (Khare et al., 2004).

1.5.2 C-terminal domain of KorB

The C-terminal domain of KorB (CTD KorB) facilitates in the dimerisation of the protein and has been crystallised (PDB ID: 1IGQ and 1IGU) as a dimer (Delbruck et al., 2002). CTD KorB consists of 62 amino acid residues, corresponding to 297-358 aa of the full length protein and this was the first crystal structure determined for KorB from RP4 plasmid (figure 6B). The CTD KorB structure is an all β-conformation with five antiparallel β-strands (β1, β2, β3, β4 and β5). These up-and-down strands are connected with loops (L1, L2, L3 and L4). The structure of the CTD KorB is strikingly similar to eukaryotic signalling proteins with Src Homology 3 (SH3)-like fold (Pawson and Schlessingert, 1993). In KorB, β5 is elongated in comparison to eukaryotic SH3 domains. Also the structure lacks the elongated loop between strand β1 and β2, which is important for recognising motifs rich in proline content in SH3 proteins (Musacchio et al., 1994). The CTD KorB is a dimer and does not form higher oligomers on its own, even in the presence of the cross-linking agent, glutaraldehyde. Also, KorB mutants without the C-terminal domain fail to dimerise (Lukaszewicz et al., 2002). C-terminal domain promotes the dimerisation ability of KorB protein and does not play a role in DNA-binding on its own. However, C-terminal domain does augment the binding capacity of N-terminal of the KorB (1-294) towards the O_B DNA (Delbruck et al., 2002).

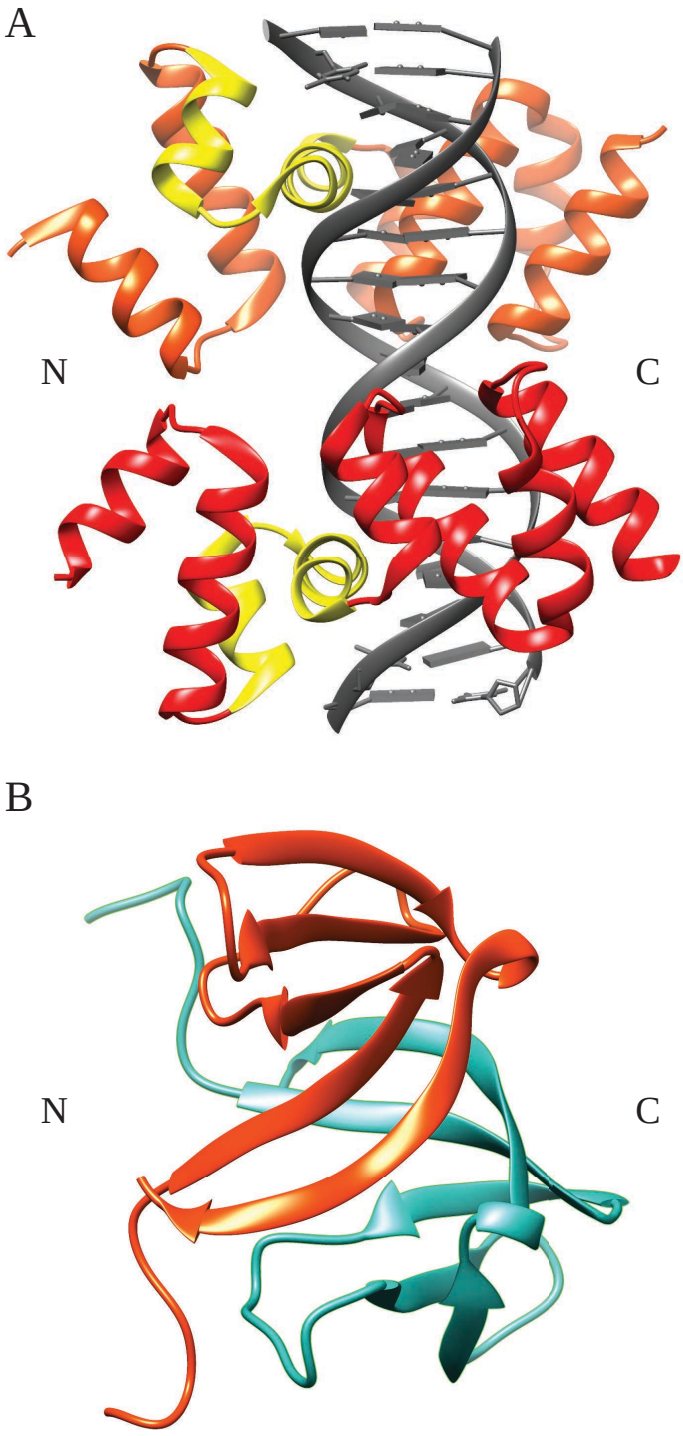

Figure 6: Structure of KorB domains. **(A)** Ribbon representation of central DNA-binding domain of KorB (138-252 aa, PDB ID: 1R71) bound to O_B DNA with an HTH motif. The DNA-binding domain is crystallised as a dimer and the structure is determined to a resolution of 2.2 Å (Khare et al., 2004). The chains A and B are shown in red and orange respectively. The HTH motif in both chains A and B is shown in yellow and the DNA molecule is shown in grey. **(B)** Ribbon representation of C-terminal domain of KorB (297-358 aa, PDB ID: 1IGQ). The C-terminal domain is crystallised as a dimer and the structure is determined to a resolution of 1.7 Å (Delbruck et al., 2002). The chains A and B are shown in orange and cyan respectively. The DNA-binding domain of KorB is primarily helical whilst the C-terminal domain is majorly strand. The structures are rendered with POV-ray in Chimera.

1.6 KorA protein from RK2 plasmid

The central control region in the IncP-1 plasmids encodes KorA, KorB and IncC (Bechhofer and Figurski, 1983; Macartney et al., 1997; Theophilus and Thomas, 1987). KorA is a repressor protein that acts in conjunction with KorB to regulate the expression of central control region and they act together as co-repressors (Bingle et al., 2005; Jagura-Burdzy and Thomas, 1995; Kostelidou et al., 1999).

KorA is a 101 aa, homodimeric DNA-binding protein with two domains, an N-terminal DNA-binding domain (DBD) containing an HTH motif and a C-terminal dimerisation domain. (Kostelidou et al., 1999). The two domains of KorA are joined by a highly flexible four residue linker (Rajasekar et al., 2016).

The crystal structures of free KorA and also of KorA bound to DNA has been determined (Rajasekar et al., 2016). In the free protein, two dimers of KorA (PDB ID: 5CKT) were contained in the asymmetric unit (ASU). Figure 7 shows the two dimers of KorA contained in the ASU. Each KorA monomer consists of an N-terminal DBD with four α-helices and a CTD containing an α-helix and a β-strand. In the ASU, difference in conformation was observed for KorA dimers; this is because of the linker connecting the DBD and CTD being highly flexible.

The structures of individual domains of KorA have also been determined using solution state NMR data. However, because of the lack of observable NOE information between the individual domains, the relative orientation of the domains could not be determined. The structures of each domain are consistent with the KorA structure elucidated using X-ray crystallography (XRC). The crystal structures of KorA bound to O_A DNA (PDB ID: 5CM3) suggest that the DBDs of KorA are symmetrically bound to the O_A DNA while the orientation of the CTD differs. Even during the 100 ns molecular dynamics (MD) simulation run, the KorA-DNA complex was observed to be flexible with the CTD exploring a wide range of orientations (Rajasekar et al., 2016).

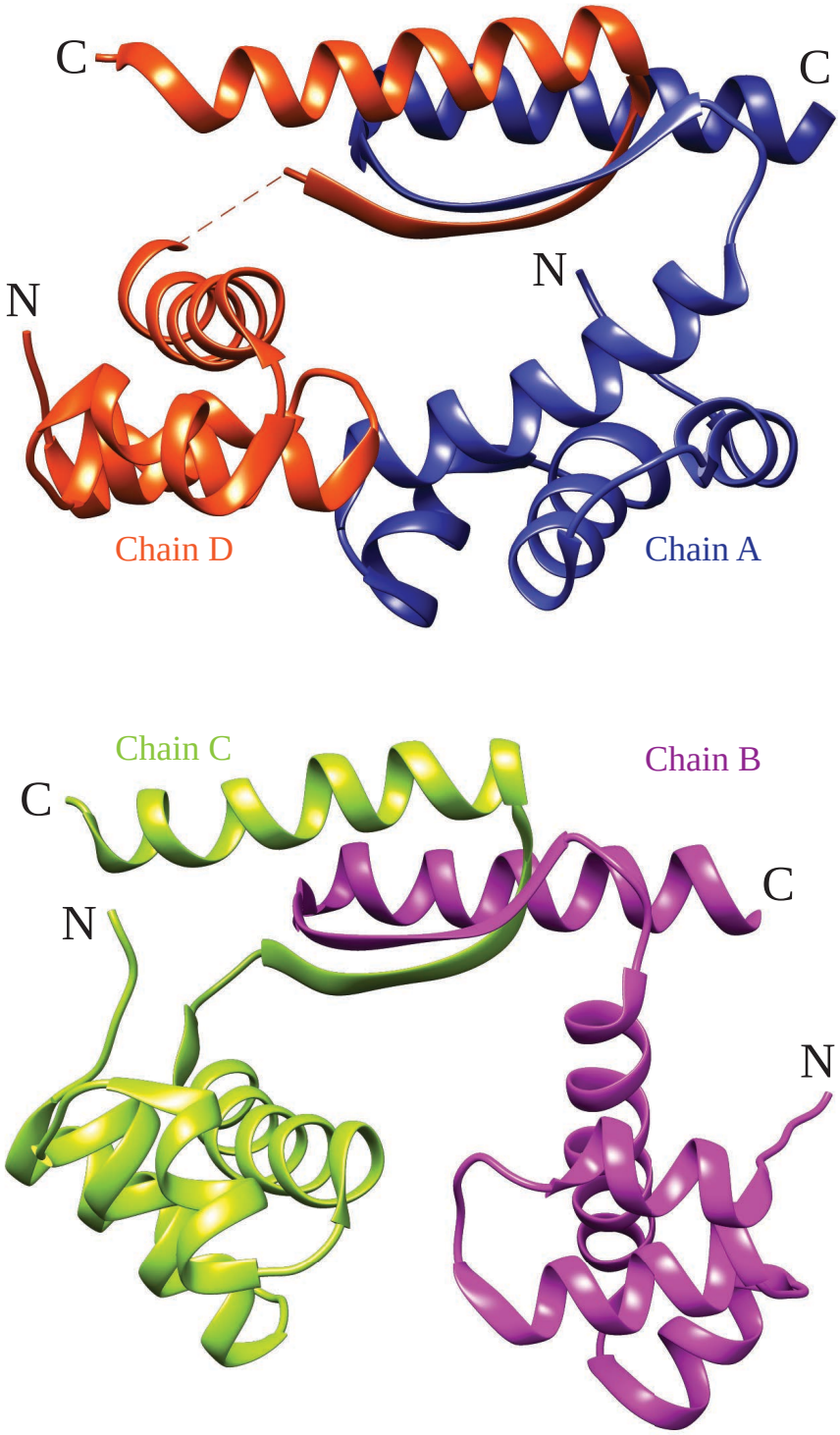

Figure 7: The crystal structure of KorA protein. Ribbon representation of crystal structure of KorA dimers. Two dimers of KorA are contained in the asymmetric unit (PDB ID: 5CKT) and each monomer chain is labelled from A–D. Each KorA monomer consists of an N-terminal DNA-binding domain (DBD) with four α-helices and a CTD containing an α-helix and a β-strand. The KorA DBD binds to the O_A DNA with an HTH motif. **(A)** Ribbon representation of KorA chain A (blue) and chain D (orange). In chain D, the dashed line represents the hinge region (residues 65–69) between the two domains and structural information about the region is not available. **(B)** Ribbon representation of KorA's chain B (magenta) and chain C (lime). The structures are rendered with POV-ray in Chimera.

Based on NMR spectroscopy and mutational analysis data, the KorA CTD is involved in interacting with KorB (Bingle et al., 2008). The mobility within the linker between the two domains of KorA while bound to DNA, allows for KorA-KorB contact at different relative orientations and contributes to KorA and KorB cooperativity (Bingle et al., 2005, 2008). Rajasekar et al. (2016), have shown that a short but flexible linker can be highly dynamic, suggesting that dynamics can be as crucial as 3D structure in understanding the function of a protein. The P1 plasmid partitioning protein, ParB also contains a highly flexible four residue linker allowing the HTH domain to rotate independently of the dimerisation domain (Schumacher and Funnell, 2005). Similarly, the extensive flexibility of linkers in the DNA-binding partitioning proteins allow the domains to contact DNA-binding sites in different arrangements, required for partitioning.

1.7 Intrinsically disordered proteins and flexible KorB

An intrinsically disordered protein (IDP) is a protein that lacks a regular secondary structure. IDPs can exist as completely unstructured to partially structured proteins. Also, IDPs can contain short stretches of amino acids that form flexible or unfolded structures called intrinsic disordered regions (IDR) in large multi-domain proteins. (Bu and Callaway, 2011; Dunker et al., 2001, 2008; Dyson and Wright, 2005). The term IDP includes proteins that contain IDRs within an otherwise well-ordered protein as well as fully disordered proteins. Essentially, IDPs are a large and functionally important class of proteins having distinct properties in terms of sequence, structure, function and interactions and can be considered different from structured proteins in many ways considering that IDPs lack regular secondary structure (van der Lee et al., 2014).

When compared to bacterial and archaeal proteins, IDRs are more prevalent in viral and eukaryal proteins (Chen et al., 2006a,b; Oldfield et al., 2005; Pancsa and Tompa, 2012). About 10-35% of prokaryotic proteins contain unstructured regions and 15-45%

of the total eukaryotic proteins exhibit intrinsic flexibility in their structure (Pancsa and Tompa, 2012; Tompa, 2012). Inside a cell, these proteins play a central role in multiple biological functions including but not limited to molecular recognition, protein folding and regulating the gene expression (Gsponer et al., 2008; Hartl, 1996; Tompa, 2005) IDPs interacting with partner(s) may involve conformational changes. IDPs can function by folding before binding to a partner, 'conformational selection' or they can fold after binding, 'induced folding' (Tompa, 2012). Both conformational selection or induced fit are usually governed by complex kinetics so it is often difficult to distinguish between the two mechanisms, however these two mechanisms can be distinguished by varying the concentration of the reactants. (Gianni et al., 2014). In addition, Vogt and Di Cera (2013) suggest that in order to better describe the kinetics data in general, conformational selection is a more versatile model.

Disorder enables a protein to have large surface area available for binding and to possess multiple conformational states. This proves helpful for binding multiple different partners to the same protein, with specific but weak binding. Disorder in proteins has been implicated in cell signalling, regulation, transcription, remodelling and modifying chromatin (Iakoucheva et al., 2002; Sandhu, 2009).

IDRs in protein often occur as highly flexible linkers connecting structured domains. The length of linker sequences vary but the flexibility of linkers allow the connecting domains to freely twist and rotate to recruit their binding partners and may induce large scale conformational changes (Bu and Callaway, 2011; Dunker et al., 2001). As mentioned in section 1.6, a short but highly flexible linker in KorA helps in binding DNA and the partner protein, KorB. Likewise, flexibility in disordered regions in KorB may help to bind multiple protein partners and/or DNA, which is important for plasmid partitioning.

KorB can be understood as an ordered protein [the central DNA-binding domain (DBD) and the C-terminal domain (CTD)] containing intrinsically disordered regions (the N-terminal domain and the linker connecting DBD and CTD). Obtaining a refined

description of KorB in DNA partitioning requires atomic-level information from structural and dynamics studies. Therefore, the overarching goal of this project is to understand the structural properties of KorB in order to study its molecular interactions with DNA and partner protein, KorA. The approach used in this thesis, involves overexpression and purification of wild-type and mutants of KorB followed by Thrombin digestion wherever necessary. The purified KorB constructs will be studied using multiple biophysical and structural techniques.

In this study, circular dichroism (CD) method will be used to gain preliminary secondary structure information about the protein constructs, whilst analytical ultracentrifugation (AUC) will provide information about the molecular mass of the protein molecules in solution. NMR spectroscopy will be used to gather structural information about the N-terminal of the KorB. Binding of KorB mutants containing the DNA-binding domain will be tested with O_B DNA using CD, fluorescence anisotropy (FA) and microscale thermophoresis (MST). The data from all these methods will complement each other and will allow us to build structural model of the KorB with DNA. This will help to understand DNA partitioning at molecular level. This understanding will assist in developing novel strategies in combating bacteria containing IncP plasmids carrying antibiotic resistance genes. This approach can also be extended to other biological processes wherein proteins with intrinsically flexible/disordered regions play a pivotal role.

1.8 Aims and outline of the thesis

The primary focus of this research is to study and further our understanding about the structure and function of DNA-binding protein, KorB from the RK2 plasmid using structural and biophysical methods. In terms of domain organisation, KorB can be divided into an N-terminal domain (1-137 aa), a central DNA-binding domain (138-258 aa) and a C-terminal dimerisation domain (297-358 aa). The current state of structural information about the KorB is shown in figure 8.

Crystal structures of central DNA-binding domain and C-terminal dimerisation domain have been previously elucidated (Delbruck et al., 2002; Khare et al., 2004) where as limited structural information is available on the N-terminal domain of the KorB. This became the motivation of studying primarily the N-terminal domain (NTD) of the KorB. Illustration of all the mutants of the KorB and the protein nomenclature used in this study is given in chapter 2 (see figure 9).

NTD KorB's purification scheme along with biophysical characterisation of the protein using analytical ultracentrifugation and circular dichroism is documented in chapter 3. Also, the partial disorder of NTD KorB is discussed in this chapter.

Through the application of solution state NMR spectroscopy, an ensemble structure of N-terminal domain of the KorB is calculated. Molecular Dynamics (MD) simulations are performed on the solution state NTD KorB structure using the coarse-grain simulation package, AWSEM to study the mobility of the protein. The details about the flexibility of NTD KorB are discussed in chapter 4 suggesting important structural insights.

In chapter 5, optimisation of purification for two constructs containing the DNA binding domain (CΔ100 KorB and NΔ31CΔ100 KorB) of the KorB is described. Both constructs are characterised using mass spectrometry and circular dichroism and their biophysical insights are discussed in this chapter.

In chapter 6, KorB-DNA interaction studies undertaken using the techniques including fluorescence anisotropy, circular dichroism and microscale thermophoresis are discussed. Various mutants of the KorB (WT, CΔ100 KorB and NΔ31CΔ100 KorB) capable of binding to the DNA are tested and binding affinities of the KorB-DNA interactions are compared and discussed in this chapter. KorB-KorA interactions studied using the 2D NMR spectroscopy (HSQCs) are also discussed.

Finally, general conclusion and outlook summarising structural characterisation of NTD KorB and KorB-DNA interactions are presented and considering the current state of information about the KorB and homologous proteins, a DNA partitioning model for

KorB is proposed in chapter 7.

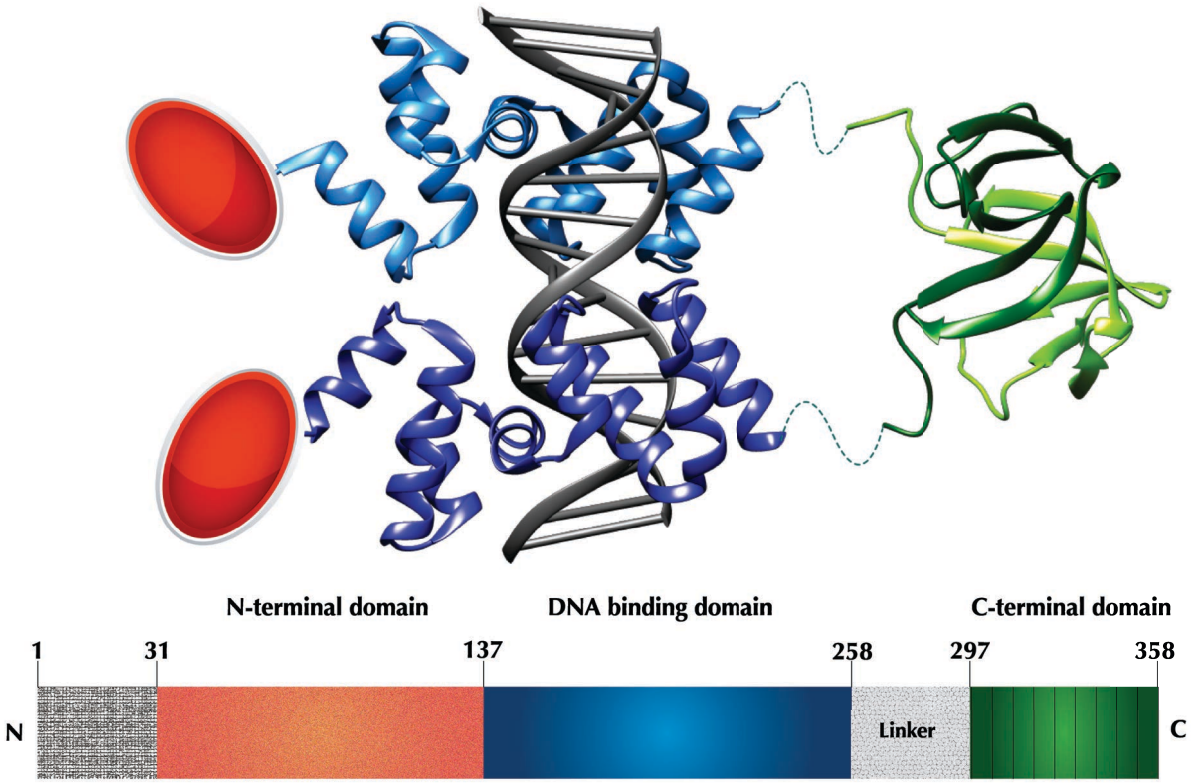

Figure 8: Current state of the KorB structure. The crystal structure of C-terminal domain (CTD) and central DNA-binding domain (DBD) are shown in green and blue respectively (top). The DBD is primarily helical whilst the CTD is majorly strand. The DBD binds to the O_B DNA shown in grey. The dashed line represents the linker connecting the two domains and structural information about the linker is not available. The N-terminal domain of the KorB is shown in red and structural information about the domain is limited. The coloured ribbon (bottom) represents the corresponding amino acids and domain organisation of the KorB. The protein structures are rendered with POV-ray in Chimera.

2

Materials and Methods

2 Materials and Methods

2.1 The pET System

The pET system (Rosenberg et al., 1987; Studier and Moffatt, 1986) is one of the most used expression systems for cloning and expression of recombinant proteins in the *E. coli*. The gene of interest is cloned into a specific pET plasmid and the gene is under the control of bacteriophage T7 transcriptional elements. The expression of protein is induced by the addition of a suitable source of T7 RNA polymerase in the *E. coli*. The pET vector is usually used in conjunction with a modified host containing a λDE3 lysogen [BL21(DE3)]. Expression of the target protein can be induced by adding IPTG as the host contains a chromosomal copy of T7 RNA polymerase gene under the control of *lacUV5*. Being selective and active, T7 RNA polymerase dictates the host's resources to express the target gene. The resulting protein can reach up to 50% of the total cell protein within a few hours after induction. The schematic representation of pET28a(+) vector (Novagen, Merck Biosciences, Darmstadt, Germany) with its multiple cloning site (MCS) is given in appendix (page 228).

2.2 Protein sample preparation

2.2.1 Protein expression

2.2.1.1 Bacterial transformation with DNA

Plasmid DNA in 100 mM Tris HCl, pH 7.5, 1 mM EDTA was used to transform competent *E. coli* BL21 cells using the heat-shock method. Competent *E. coli* BL21 (λDE3) (New England Biolabs, Ipswich, USA) cells were used to transform the plasmid DNA containing the gene for the protein of interest. Aliquoted BL21 cells stored at -80 °C were thawed on ice for $\sim$ 2 min. Plasmid DNA (1 μL of DNA at 40-80 ng/μL) was

used to transform 20 μL of competent cells in a microcentrifuge tube. Competent BL21 cell-plasmid DNA mixture was incubated on ice for 20 min immediately proceeded by a heat-shock treatment at 42 °C for exactly 90 seconds. The transformed cells were incubated for 2 min on ice followed by addition of 80 μL of Super Optimal Broth (SOC) or Lysogeny Broth (LB) without the addition of any antibiotic. Transformed cells were left for 45 min in a shaking incubator with 200 rpm at 37 °C. Transformed cells were plated onto LB-agar-kanamycin (15 g/L agar and 30 μg/mL kanamycin) plates at room temperature (RT) and were left to grow overnight at 37 °C for colonies to develop and appear on the plate. The bacterial strains used in this study are given in table 7.

2.2.1.2 Isolation and purification of DNA for KorB constructs

DNA for the KorB constructs was transformed into competent *E. coli* subcloning efficiency DH5α cells (Invitrogen) cells (see section 2.2.1.1 for transformation protocol), plated on LB agar (+ antibiotic) plates followed by overnight incubation. The KorB constructs used in this study are shown in figure 9. The protein constructs with their respective biochemical characteristics are given in table 8 and table 9. A single bacterial colony was then picked with a smear loop to inoculate 5 mL of LB, left overnight in a shaking incubator with 200 rpm at 37 °C. The overnight bacterial culture was pelleted in an Eppendorf® 5810R centrifuge with rotor S-4-104 (Eppendorf) at 4,000 rpm for 6 min at 4 °C. All of the KorB DNA constructs used in the present study were purified following the instructions supplied with the QIAprep Spin Miniprep Kit (Qiagen). The concentration of purified DNA was measured in ng/μL using NanoDrop 1000 (Thermo Fisher Scientific) and the DNA was stored at -20 °C for future use.

Strain	Genotype	Features
BL21 (λDE3) (NEB)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	Competent *E. coli* strain used for transformation and high-level expression of recombinant protein. This is a T7 expression strain deficient in two proteases: lon and outer membrane protease, ompT.
DH5α (Invitrogen)	*F^- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(r_k^-, m_k^+) phoA supE44 thi-1 gyrA96 relA1 λ^-*	Competent *E. coli* strain with a high transformation efficiency of plasmid DNA used for routine subcloning procedures. This strain contains recA1 which prevents unwanted recombination and endA1 which improves the yield of plasmid DNA.
XL10-Gold (Agilent)	*Tet^r Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB$lacI^q$ZΔM15 Tn10 (Tet^r) Amy Cam^r]*	Competent *E. coli* strain offering highest transformation efficiencies for cloning large plasmids and ligated DNA. This strain contains the high transformation phenotype (Hte). The strain is endonuclease deficient (endA) and recombination deficient (recA) which enhances the quality of miniprep DNA and provides insert stability respectively.

Table 7: Bacterial strains used in this study.

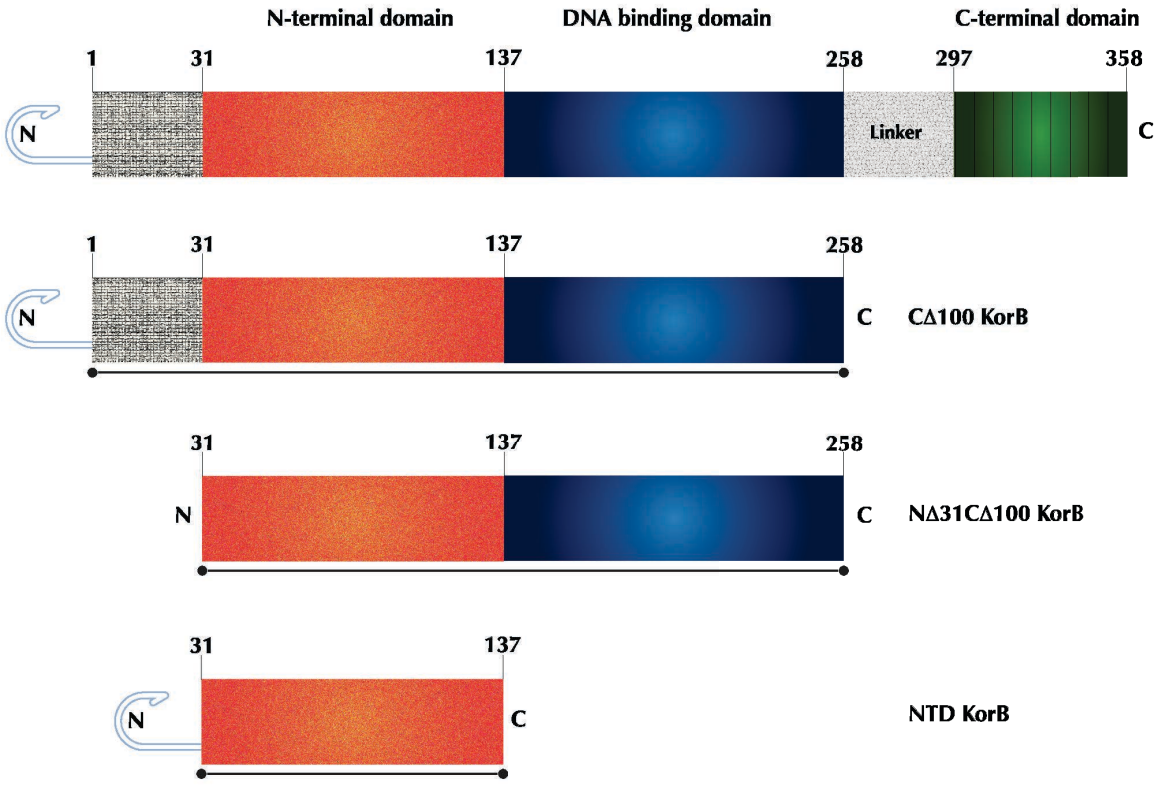

Figure 9: Illustration and nomenclature of all the mutants of KorB used in this study. (A) The wild-type (WT) KorB protein with N-terminal domain, DNA-binding domain and C-terminal domain is shown in orange, blue and green respectively. The N-terminal contains a His-tag (23 residues) shown with a hook and the flexible regions are shown in grey. **(B)** 100 residues from the C-terminal of the KorB WT are truncated and the resulting protein is named CΔ100 KorB. The CΔ100 KorB is without the linker and the CTD but contains N-terminal His-tag for purification with the same cleavage site as of KorB WT. **(C)** The His-tag along with first 31 residues from the N-terminal of the CΔ100 KorB are truncated and the resulting protein is named NΔ31CΔ100 KorB. **(D)** Removing the DNA-binding domain from the NΔ31CΔ100 KorB resulted in a protein named NTD KorB (with N-terminal His-tag).

Protein construct	Vector type	Residue number	Molecular weight (Da)	Affinity tag
KorB WT	pET28a	1-358	41537.8	N-terminal 6xHis-tag
CΔ100 KorB	pET28a	1-258	30591	N-terminal 6xHis-tag
NΔ31CΔ100 KorB	pET28a	31-258	25284.2	None
NTD KorB	pET28a	31-137	14083.4	N-terminal 6xHis-tag
KorA WT	pET28a	1-101	11305.9	None

Table 8: Protein construct table (I)

Protein construct	Antibiotic resistance	Cleavage tag (residue)	Molar extinction coefficient (M^{-1} cm^{-1})
KorB WT	Kanamycin	Thrombin (17/18)	23950
CΔ100 KorB	Kanamycin	Thrombin (17/18)	15470
NΔ31CΔ100 KorB	Kanamycin	None	15470
NTD KorB	Kanamycin	Thrombin (17/18)	9970
KorA WT	Kanamycin	None	13980

Table 9: Protein construct table (II)

2.2.1.3 Large scale expression of unlabelled KorB constructs

Following transformation as described in section 2.2.1.1, a single colony was picked with an inoculation loop to inoculate LB medium (10 mL) supplemented with Kanamycin (30 μg/mL). Cultures were then left to incubate overnight at 37 °C with shaking at 220 rpm. The overnight starter cultures were used to inoculate 1 L (2*500 mL) LB media (10 mL overnight culture per 1 L LB media) supplemented with Kanamycin (30 μg/mL). Once the culture reached an A_{600} reading of 0.6 [measured using the BioMate 3 (Thermo Scientific^{TM})], the temperature of the incubator was dropped to 18 °C. The cells were grown in 1 L (2*500 mL) LB media with Kanamycin to an A_{600} of about 0.6 at 37 °C. At this stage, the temperature was reduced to 18 °C and expression of T7 RNA polymerase was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was then left to incubate overnight for 16 hours.

2.2.1.4 Large scale expression of labelled KorB constructs

The over-expression and purification of isotopically labelled protein was performed as described in the previous section (2.2.1.3). The isotopic labelling of the protein sample was done by growing the bacteria in minimal medium, M9.

The recipe to make 1L of the M9 medium: mix the salts as described in table 10, top up to 970 mL, adjust the pH to 7.3 and autoclave the medium. Next, make up 30 mL of the nutrient mix (table 11), filter sterilise the solution with 0.2 μm and store at 4°C for future use. The metal mix (used in the nutrient mix) consists of 4 mM $ZnSO_4$, 1 mM $MnSO_4$, 4.7 mM, H_3BO_3 and 0.7 mM $CuSO_4$.

Chemical	Amount
Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g

Table 10: Minimal medium salt mix recipe

Chemical	Amount
Glucose (^{13}C, 99%)	2 g
$^{15}NH_4Cl$ (99%)	1 g
1 M $MgSO_4$	2 mL
1 M $CaCl_2$	10 μL
1 M $FeCl_3$ (dissolved in HCl)	400 μL
Thiamine (20 mg/mL)	1 mL
Metal mix	0.5 mL
Kanamycin (30 mg/mL)	1 mL

Table 11: Minimal medium nutrient mix recipe

2.2.1.5 Harvesting bacterial cells by centrifugation

The culture was centrifuged in Beckman Coulter Avanti® J20/I, rotor JLA 8.1 at 6,000 rpm for 20 min. The supernatant was discarded and pellet was suspended in 50 mM Tris HCl, pH 7.5, 0.1 mM EDTA (∼50 mL) and centrifuged again at 6,000 rpm for 20 min to ensure that the cells were media-free. The sample aliquots of before and after induction were analysed on a 4-12% Criterion ™ XT Bis-Tris SDS-PAGE gel with loading dye (Sigma-Aldrich) to check for expression.

2.2.1.6 Mechanical homogenisation of bacterial cells

The cell pellet from 1 L of culture was resuspended in ∼40 mL of 10 mM Imidazole, 50 mM Na phosphate, pH 8, and 300 mM NaCl containing Phenyl Methane Sulfonyl Fluoride, (PMSF) and a protease inhibitor tablet (EDTA free, from Roche). The cells were homogenised at 15,000 psi (pound per square inch) for five continuous cycles using a C3 EmulsiFlex cell homogeniser (AVESTIN Europe GmbH, Mannheim, Germany) at 4 °C. To pellet the cell debris, the homogenised solution was centrifuged at 25,000 rpm in Beckman Coulter Avanti® J series centrifuge using JA-25.50 rotor for 45 min. The pellet was discarded and the supernatant was filtered using a 0.45 μm filter, followed by a 0.2 μm Minisart® filter.

2.2.2 Protein purification

2.2.2.1 Initial protein purification by affinity chromatography

As the KorB proteins were His-tagged, the initial protein purification was done using a 5 mL HIS-Select Nickel Affinity Gel column (Sigma-Aldrich) at RT. The column was equilibrated with buffer containing 10 mM Imidazole, 50 mM Na phosphate, pH 8.0, 300 mM NaCl, and 10 μg/mL PMSF. Homogenised supernatant ($\sim$ 20 mL) was loaded onto a Ni-NTA column. The column was pre-equilibrated with 50 mL buffer containing 10 mM Imidazole. A linear gradient of Imidazole (20 mM to 250 mM) was applied to elute the proteins with N-terminal His-tag. The presence of protein in the fractions was confirmed using Bradford reagent (Bradford, 1976) at A_{595}. In order to confirm the presence and purity of the KorB samples, SDS-PAGE was performed. The protein samples were prepared for electrophoresis by mixing 1:1 with gel loading dye A precast criterion gel was run with a BioRad apparatus at 180 V for 45 minutes and the gel was stained at RT with Expedeon InstantBlue protein stain – a ready to use Coomassie protein stain for polyacrylamide gels; followed by gentle shaking of the gel in 30 mL of InstantBlue stain for 30 minutes. Fractions containing the protein were

dialysed into 10 mM Tris HCl, pH 7.5, 100 mM NaCl, and 0.1 mM EDTA and then concentrated on an Amicon Ultrafiltration cell using Millipore (MWCO: 3/10/30 kDa) ultrafiltration membrane. Further purification of the protein was performed using size exclusion chromatography.

2.2.2.2 Determination of protein purity by SDS-PAGE

Pre-and-post-induction cell samples and protein samples were prepared by mixing samples ($\sim$1-20 μL) with equal volume of 2x Laemmlli Concentrate buffer (Merck, USA) and samples were then loaded onto an 18/26-well pre-cast 4-12% Criterion^{TM} XT Bis-Tris protein gel (Bio-Rad Laboratories, USA). The SDS-PAGE gels were typically run at 180 V for 45 min in 1x XT-MES running buffer (Bio-Rad Laboratories, USA) or until the leading edge of the Bromophenol blue sample buffer had reached the other end of the gel plates. Prestained protein markers (Bio-Rad) were run alongside lysate and protein samples to compare the molecular weight against known standards. Following the electrophoresis run, the gels were stained with InstantBlue^{TM} Coomassie stain (Expedeon, UK) until protein bands were visible. The background stain on the gels was removed by rocking the gels in H_2O for an hour. Once ready, the images of the gels were captured with a FinePix AX650 digital camera (Fujifilm, Tokyo, Japan) and optimised with PhotoScape X 2.4.2.

2.2.2.3 Protein purification by size exclusion chromatography

Protein fractions purified from affinity chromatography were concentrated in Amicon concentrators (3 or 10 kDa MWCO) to $<$5 mL and centrifuged at 4,000 g for 10 min. Ni-NTA purified protein (with contaminants) was further purified using a Superdex 75, 26 cm, by 60 cm, size exclusion column with an $\ddot{A}$KTA Purifier P-900 (GE Healthcare, Chicago, USA) and the chromatogram was monitored at 280 nm. The column was equilibrated with size exclusion chromatography (SEC) buffer: 10 mM Tris HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA for 1.2 CVs. The flow rate was set to 2 mL/min.

Protein from affinity chromatography was concentrated to 4 mL and was loaded on to the column at RT. A 5 mL Hamilton syringe, pre-washed with SEC buffer was used to inject the concentrated protein sample into a 5 mL sample loop attached to the *Ä*KTA Purifier. Following sample injection, the column was washed with 1 column volume of buffer and the sample was collected in 12 mm tubes with 4 mL fraction volume. Eluted fractions containing the protein were pooled according to the highest absorbance peaks in the chromatogram. To gauge the purity of protein, samples were analysed by SDS-PAGE. Fractions with purified protein were pooled again concentrated using the Amicon (Merck) centrifugal filter units with molecular weight cut-off (MWCO) of 3/10/30 kDa in an Eppendorf 5810R bench-top centrifuge at 4,000 g to ∼2 mL. Quantification of protein concentration was done with BioMate 3 (Thermo Scientific™) spectrophotometer. All SEC experiments were performed at room temperature. Note: for labelled protein samples (^{15}N/^{13}C), the buffer used was phosphate buffer with 10 mM Na phosphate, pH 7, 100 mM NaCl and 0.1 mM EDTA. The proteins were further concentrated using the Thermo Scientific Pierce® concentrator having MWCO 9 kDa on a Thermo Scientific Heraeus Megafuge 11R at 2500 g to about 1 mL (10-22 mg/mL) in volume for protein crystallisation trials.

2.2.2.4 Determination of protein concentration by A_{280} measurement

The concentration of the purified proteins was determined by taking absorbance readings at 280 nm on a BioMate 3 UV spectrophotometer and computing the concentrations using the molar extinction coefficient (table 9) for the protein sequence [obtained from ProtParam (web.expasy.org/protparam)] into the Beer-Lambert equation:

$$\mathrm{A} = \varepsilon \mathrm{c\,l/d} \tag{2.1}$$

where A = observed absorbance at 280 nm, ε = molar extinction coefficient ($M^{-1}cm^{-1}$), c = concentration (M), l = path length (cm), d = dilution factor.

2.3 Biophysical characterisation of KorB constructs

2.3.1 Polydispersity analysis with analytical ultracentrifugation experiments

Analytical ultracentrifugation (for theoretical background of AUC, see section 3.3.2) experiments were performed on Beckman Coulter ProteomeLab™ XL-1 Protein Characterisation System (Pasadena, California, USA). The experiments were performed in a two-chamber centerpiece and the data were collected using absorbance and interference optics. Purified proteins were dialysed into 10 mM Tris HCl, pH 7.5, 100 mM NaCl, and 0.1 mM EDTA and 0.5 mL of each protein sample was used for the experiment. Three concentrations of proteins (A_{280} – 0.1, 0.5, 0.9) were used for the experiment. The samples were centrifuged overnight at 20 °C at 40,000 rpm in the T1i/50 rotor. The absorbance was measured continuously for 20 hours at 280 nm and 120 scans were collected. Experimental data were processed in SEDFIT (Vistica et al., 2004) program using the c(S) distribution based on Lamm equation (Schuck et al., 2000).

2.3.2 Binding constants and secondary structure estimation with circular dichroism experiments

For CD experiments, both proteins and DNA were diluted in 100 mM $NaClO_4$, 10 mM Tris HCl, pH 7 and 0.1 mM EDTA. NaCl was substituted with ($NaClO_4$) as perchlorate ions absorb much less in the 180-190 nm range than chloride (NaCl) ions (Kelly et al., 2005). All CD measurements were recorded at RT. Most of the CD experiments were performed using a Jasco J-1500 spectrophotometer (Essex, UK). Measurements were made between 340 nm and 190 nm with step size of 1 nm and bandwidth of 1-2 nm. Fixed quartz cuvettes (Hellma, UK) of path-length 10 mm were used with an averaging time of 1 s and 4-8 scans were recorded. The data from CD experiments were submitted to DICHROWEB, a web server that perform analyses of Circular Dichroism data, (Lobley et al., 2002; Whitmore and Wallace, 2004, 2008) for secondary structure estimation; the CDSSTR analysis program (Sreerama et al., 2000; Sreerama and Woody,

2000) was used with reference sets 3 and 6. Protein-DNA titration measurements were performed with a Chirascan Plus spectrometer (Applied Photophysics, Surrey, UK) on beamline B23 at the Diamond Light Source (Oxfordshire, UK). The protein and DNA concentrations used for CD titration experiments are given in table 12.

KorB construct	**Concentration (μM)**	**ds O_B DNA**	**Concentration (μM)**
NΔ31CΔ100	80	Full-length	10
NΔ31CΔ100	80	Half-length	20
CΔ100	60	Full-length	15
CΔ100	60	Half-length	20

Table 12: Concentration of KorB constructs and double stranded (ds) O_B DNA used for CD experiments.

DNA was measured between 240-340 nm and protein-DNA samples measured between 180-360 nm. For each sample, average of four scans were subtracted from average of four buffer scans. The spectra of buffer were taken under identical conditions and subtracted from that of the respective protein and / or DNA spectra. The data fitting for protein-DNA titrations was done with non-linear regression analysis to a modified quadratic equation

$$f = y_o + a * \frac{x + z + k - \sqrt{(x + z + k)^2 - 4xz}}{2} \tag{2.2}$$

using SigmaPlot 13 (Systat Software, San Jose, US). Here, y_o is the initial (non-zero) value of CD signal, a is the difference between the maximum and minimum CD signal *i.e.* max(y)-min(y), x is the protein concentration, z is the DNA concentration (constant), and K_d is the dissociation constant. Theoretical background of CD is given in section 6.2.2 and details on the above modified quadratic equation is given in section 6.3.

2.4 Protein-DNA interactions experiments

2.4.1 O_B DNA containing oligonucleotides

The modified 3′-fluorescein tagged oligonucleotides containing O_B DNA were ordered from Eurofins Genomics (Eurofins Scientific, Luxembourg). The double stranded full-length DNA (dsFL) is a 17 bp sequence containing the palindromic consensus operator (O_B) sequence for KorB binding. The double stranded half-length DNA (dsHL) is a 12 mer sequence containing one half of the palindromic O_B sequence. For dsFL DNA, the tagged strand and its untagged complementary strand were dissolved in filtered TE buffer (20 mM Tris, 1 mM EDTA) and annealed by heating to 70 °C for 5 minutes in a water bath before cooling to room temperature (24 μM). Likewise for dsHL DNA, the tagged and complementary strand were dissolved in the same TE buffer and mixed to a final concentration of 50 μM. The sequences of O_B DNA containing oligonucleotides used for KorB-DNA interactions are given in table 13. All DNA was stored at -20 °C and the DNA was wrapped inside aluminium foil.

O_B DNA	**Sequence**	**Concentration**
dsFL	5′-GGTTTAGCGGCTAAAGG-3′	24 μM
dsHL	5′-GCCTTTAGCGGC-3′	50 μM

Table 13: O_B DNA containing oligonucleotides. dsFL and dsHL stands for double stranded full-length O_B DNA and double stranded half-length O_B DNA respectively.

2.4.2 Microscale thermophoresis

All microscale thermophoresis (for theoretical background of MST, see section 6.2.3) experiments were performed at RT on a Monolith NT.115 (NanoTemper Technologies, München, Germany) containing a red/blue filter set. Blue-Red excites fluorescence with wavelengths 460-480nm and 600-650nm; also the filter detects wavelengths between 510-530 nm and 675-690 nm. The blue filter was used for labelled DNA samples

and the red one for NHS (fluorescent dye) tagged KorA. The NanoTemper 647-NHS dye contains 'NHS-ester chemistry', that reacts with 1°amine groups of solvent accessible lysine residues (usually) in the protein of interest to form stable dye-protein adducts. The NHS-ester chemical reaction is given in appendix on page 218.

The proteins conjugated with NHS label fluoresce with excitation and emission maxima of ∼650 and ∼670 nm respectively. The protein samples were centrifuged at 13,000 rpm in a bench-top centrifuge for 5 min to discard any precipitates. The samples were prepared in microfuge tubes (supplied in the MST kit) and 1 μL (10 mg/mL) of BSA was added so as to minimise adsorption of the protein sample to the tube walls. With capillary action, a sample volume of ∼5 μL was loaded into the NanoTemper standard glass capillary and microthermophoresis was carried out. MST experiments should contain a labelled molecule for the detection of the signal and one or more unlabelled binding partners. A capillary scan (capscan) was used to determine the minimum amount of fluorescently labelled partner. The capscans were carried out at 10% LED power and the binding experiments were done with 20% MST power. The LED power was adjusted (typically between 10-20%) so as to have fluorescence counts between 500-1000, minimum being 300 counts. The buffer used for protein-DNA measurements was 20 mM Na phosphate, pH 7.6, 100 mM NaCl. 1 μL of fluorescein labelled DNA was added to 200 μL of buffer and blue LED was used to excite the DNA at 10% power resulting in 500-600 counts. Dilution series with 8-16 points of unlabelled binding partner were prepared by diluting the partner in equal part buffer (1:1). Consistent pipetting of the samples should yield fluorescence count (for the labelled molecule) to be within ±10% for all concentration points in the titration. Most of the experiments were measured at varying MST power (10, 20 and 50%) and best results were obtained at 10 or 20%. BSA was used as a negative control for the MST experiments.

2.4.3 Fluorescence anisotropy

All the Fluorescence Anisotropy (for theoretical background of FA, see section 6.2.1.2) experiments were performed on a PerkinElmer luminescence spectrometer LS 50B (Waltham, Massachusetts, USA) and data were collected using the program FL WinLab 2.01 at RT. Fluorescently labelled DNA and unlabelled KorB constructs KorB WT, NTD KorB, CΔ100 KorB and NΔ31CΔ100 KorB were used for the anisotropy experiments.

Excitation maximum and emission maximum for fluorescein is 495 nm and between 516-520 nm respectively. The excitation wavelength was optimised by scanning in the range 400-500 nm (keeping the emission maximum at 518 nm) and emission wavelength in the range of 500-600 nm (keeping the excitation maximum at 495 nm) and staying below the limit of 1000 units of intensity (INT). Slit width of 5 nm was used for both the polarisers and an integration time of 5 s was used to measure the anisotropy values. Before measuring the anisotropy value at each protein concentration, fluorescence intensity was monitored for quenching and keeping the intensity within the spectrometer's detectable range of 100-1000 intensity units. To minimise the effect of random error both on intensity and anisotropy, the fluorescence intensity was maximised while keeping its intensity below 1000 intensity units.

The vials containing fluorescein were wrapped with aluminium foil to minimise the decay of fluorescence by photo-bleaching. The experiments were done in 20 mM Na phosphate, 100 mM NaCl throughout. 1 μL of fluorescein labelled DNA was added to 100 μL of buffer. Bovine serum albumin (BSA) was used as a negative control for the anisotropy experiments. Prior to being titrated into the fluorescein-labelled DNA, serial dilutions of the proteins were prepared. The averaged anisotropy values were determined from at least three sets of protein-DNA samples. The anisotropy data were fitted using the Hyperbola, Single Rectangular I, 3 Parameter equation (given below) in Sigmaplot 13.

$$f = y_o + \frac{a * x}{b + x} \tag{2.3}$$

In the equation, y_o is the initial anisotropy value, a is the difference between maximum and minimum values of anisotropy, x is the protein concentration, and b is the dissociation constant (K_d).

2.5 X-ray crystallisation trials

Initial protein crystallisation trials were done using a Mosquito® nanolitre crystallisation robot (TTP Labtech Limited, Melbourn, UK). Prior to setting up crystallisation trays, unlabelled proteins were purified to $>95\%$ homogeneity as analysed by SDS-PAGE (described in section 2.2.2.3). Proteins were dialysed into 10 mM Tris HCl, pH 7, 100 mM NaCl, 0.1 mM EDTA and concentrated to 10-22 mg/mL using concentrator having 3/10 kDa MWCO on a Eppendorf 5810R centrifuge with rotor S-4-104 at 4,000 rpm at 4 °C. First, 400 nL of protein was mixed with 400 nL of precipitant in a 96-well plate and hanging drop crystallisation method was used. Two protein crystal screening trays were setup using Molecular Dimensions i) JCSG- plusTM Screen Box 1 and Box 2 and ii) Morpheus® HT-96 Screen Box 1 and Box 2. Both screens are a 96 reagent, sparse-matrix screen optimised for protein crystal formation. Second, the sitting drop method was used for the trial experiments on a combinatorial crystallisation plate. Immediately before setting up the crystal trays, the protein was centrifuged at 13,000 for 5 minutes at 4°C; pellet was discarded (if any) and the protein was transferred to a sterile tube. 100 μL of sparse matrix solution was added to reservoir well for 96 conditions, 1-48 (Box 1), 49-98 (Box 2). 1 μL of protein was mixed with 1 μL of well solution to form a sitting drop on the tray at RT. The trays were sealed with crystal clear tape and were left undisturbed in a room maintained at 18 °C for vapour diffusion to take place and crystal growth is initiated. The preliminary precipitant condition was 0.1 M ammo-

nium sulfate, 0.1 M BIS-Tris, pH 5.5, 30% PEG 3350 with protein:DNA as 2:1.1, where 100 μm sized protein crystals were observed. The hit was used to further optimise the condition by varying the ratio of DNA and keep the protein constant and adding cryoprotectant in the well (glycerol and DMSO). The volume in the reservoir and volume of the sitting drop was also varied with respect to each other.

2.6 NMR experiments

The NMR spectra were collected at the Henry Wellcome Building NMR Centre at the University of Birmingham. The experiments were performed on a 500 MHz Bruker, a 600 MHz Varian, an 800 MHz Varian and a 900 MHz Bruker spectrometer. For NMR experiments, all $^{13}C/^{15}N$ protein samples were made in same batch and in 20 mM Na phosphate, pH 6, 100 mM NaCl and 0.1 mM EDTA. Processing of the spectra were done with NMRPipe software (Delaglio et al., 1995). The spectra were analysed using the software package NMRFAM-SPARKY (Lee et al., 2015). The sequence specific backbone assignment (^{1}HN, ^{15}N, ^{13}C', $^{13}C\alpha$) and side chain assignment ($^{13}C\beta$) for NTD KorB were obtained and assigned using the following experiments: HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH, CBCA(CO)NH (CB only). The protons for the side chains were obtained and assigned using the HBHACONH experiment. For theoretical background of these above mentioned NMR experiments, see section 4.3.

The secondary structure of the protein was predicted computationally using the online servers Jpred4 (Drozdetskiy et al., 2015) and Psipred 3.3 (Buchan et al., 2013). The secondary structure for the NTD KorB was calculated with sequence specific chemical shift values using TALOS+ (Shen et al., 2009) with a prediction accuracy of about 89%. The structural alignment for NTD KorB and ParB homologues was performed with ENDscript (Robert and Gouet, 2014). The resulting ucsf files were imported in SPARKY to be visualised and analysed. The ucsf files were visualised in SPARKY and the files were exported in postscript (.PS) file format. PS file format is a versatile file format for

producing high resolution vector files. Inkscape 0.92, a vector graphics editor was used to render the PS files for image production.

2.6.1 ^{15}N KorA and CΔ100 KorB HSQC titrations

A range of concentrations of CΔ100 KorB protein were titrated against a fixed concentration of ^{15}N KorA protein. The buffer used for both the proteins (^{15}N KorA and CΔ100 KorB) was 10 mM Na phosphate, pH 7, 100 mM NaCl, 0.1 mM EDTA. All titration experiments were done on a 900 MHz Bruker magnet at 35 °C. The volume used for each titration experiment was 600 μL with 10% of D_2O (540 μL of sample and 60 μL of D_2O). The range of ratios of KorA:CΔ100 KorB used for the experiments were 1:1, 1:2, 1:3, 1:4 and 1:5. The concentrations of both KorA and CΔ100 KorB used in the titrations are listed in table 30 (section 6.5.2).

2.6.2 NMR structure calculation and analysis

In order to calculate the NTD KorB structure, the protein sequence, a list of backbone chemical shifts assignments, a list of TALOS+ predicted backbone torsion angles and 3D-NOESY (^{13}C and ^{15}N) spectra were submitted as input to Ponderosa-C/S (Lee et al., 2014) server (hosted by National Magnetic Resonance Facility at Madison, NMRFAM), with its refinement option calculating the initial structure of NTD KorB.

Ponderosa Analyzer is a program within Ponderosa client which was used to evaluate the calculated protein structure. The Analyzer program was used to iteratively reduce the violations in the structure *i.e.* refine the distance and angle constraints and export the refined data to the Ponderosa server for iterative structure calculations. Further, the water refinement of the structure was done by the Ponderosa server to calculate the final 20 models of the NTD KorB (XPLOR-NIH, Schwieters et al., 2003).

The final ensemble structure of the NTD KorB with 20 models was analysed using the PSVS (protein structure validation software) suite (Bhattacharya et al., 2006) and to evaluate the structural integrity and stereo-chemical quality of the ensemble, multiple

analysis tools were employed including MolProbity (Davis et al., 2007), RAMPAGE (Lovell et al., 2003), ProSA-web (Sippl, 1993; Wiederstein and Sippl, 2007), Verify3-D (Bowie et al., 1991; Lüthy et al., 1992) and PROSESS (Berjanskii et al., 2010).

2.6.3 Molecular dynamics simulations

All molecular dynamics (MD) simulations were carried out using the LAMMPS molecular dynamics package (Plimpton, 1995) with AWSEM-MD program implemented inside the LAMMPS (Davtyan et al., 2012). The temperature in all the simulations was kept constant at 27 °C using the Nosé-Hoover thermostat algorithm. Simulations of the NTD KorB were performed using the model number 8 of the NTD KorB's ensemble structure as starting coordinates. The model number 8 of the NTD KorB contains residues 1–130 where first 23 residues are from the His-tag and the N-terminal domain corresponds to residue 31-137 of KorB (for NTD KorB sequence, see figure 85).

The structural knowledge of the 20 NTD KorB models was used to bias the simulations. Out of 20 models in the NTD KorB ensemble, 19 were used (excluding the model number 8) to generate overlapping fragment information called memories. The additional energy term associated with memories is called Fragment Memory. To include fragment memories from the 19 NTD KorB models, a fragment library was generated using the scripts in the AWSEM-MD package (awsem-md.org/index.html).

All the simulations were done on the super-computing facility at the University of Maryland's high-performance computing cluster, Deep Thought 2. As the MD simulations are extensive and resource hungry in nature, a limited amount of computing time was allocated on the Deep Thought 2 cluster (glue.umd.edu/hpcc/dt2.html). AWSEM package computes in serial mode and considering the time constraints on using the cluster, the MD simulations of the NTD KorB were performed for 110 ns. The simulation trajectories for the NTD KorB were visualised with VMD (Humphrey et al., 1996) and UCSF CHIMERA (Pettersen et al., 2004).

3

Biophysical characterisation of N-terminal of KorB

3 Biophysical characterisation of N-terminal of KorB

3.1 Results

3.1.1 Over-expression of N-terminal of KorB

The NTD KorB with amino acids 31-137 (MW: 14083 Da) was over-expressed in *E. coli* BL21 (DE3). The NTD KorB's expression was induced with IPTG and the cells were homogenised to release the protein into the solution.

3.1.2 Purification of N-terminal of KorB

3.1.2.1 Affinity chromatography

The initial NTD KorB purification was done using a HIS-select Nickel affinity gel column as the N-terminal His-tag of the protein has affinity towards the Ni-column. A linear gradient of Imidazole was used to elute NTD KorB from the column. In the eluted fractions, presence of the NTD KorB was confirmed using Bradford reagent and an SDS-PAGE gel was run to check the fractions for protein purity. The results of Ni-column purification and SDS-PAGE gel are shown in figure 10A and 10B respectively.

3.1.2.2 Size exclusion chromatography

The fractions from affinity chromatography with NTD KorB were concentrated and the protein was purified on a size exclusion column. Figure 11 represents the chromatogram and SDS-PAGE gel for the NTD KorB. SDS-PAGE gel was not able to resolve the species present in the fractions as single band was observed but we can observe two peaks in the chromatogram. Fractions 7 to 15 were pooled and concentrated and total yield of the NTD KorB obtained from 1 L of culture was 22 mg.

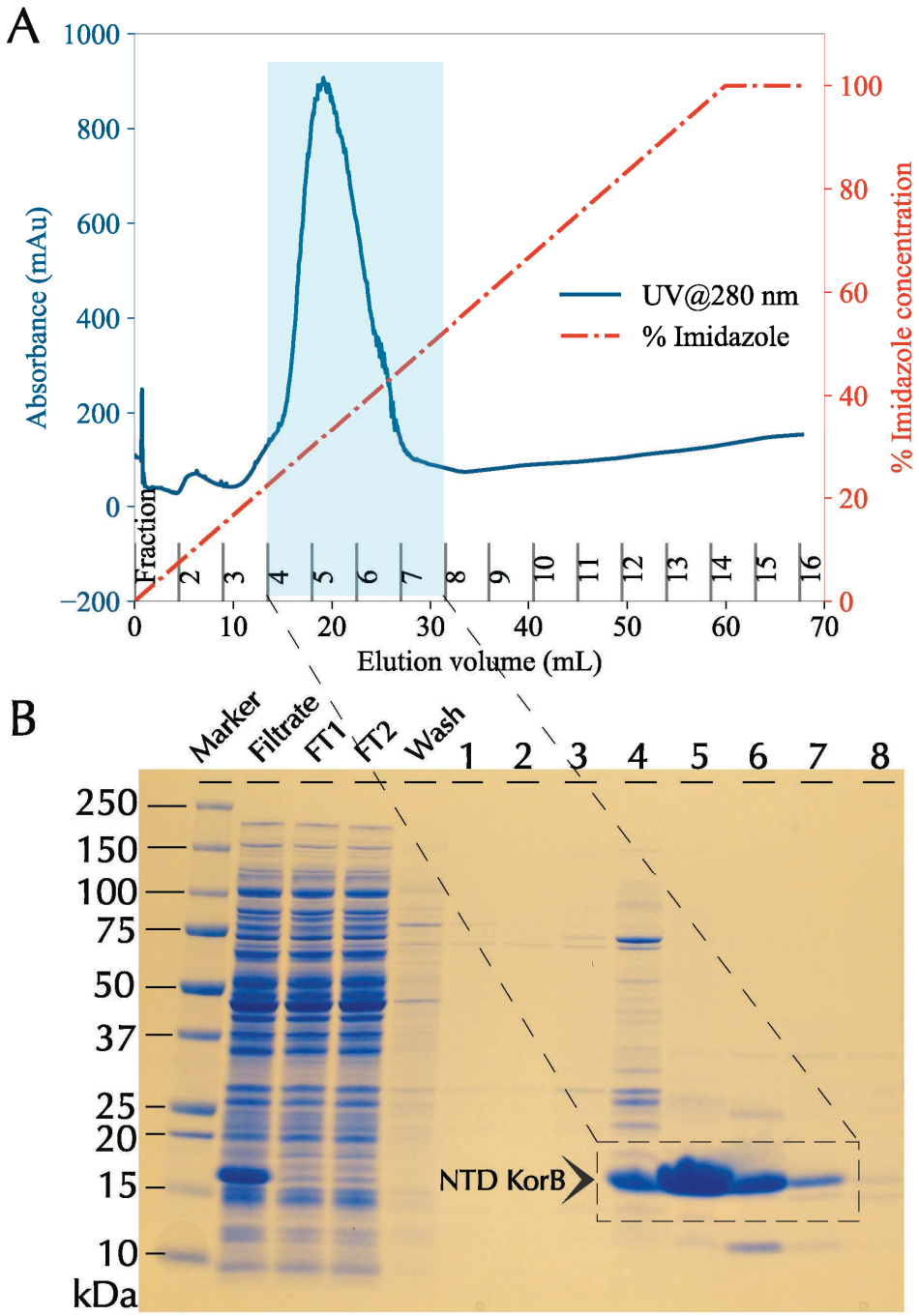

Figure 10: Purification of the NTD KorB by affinity chromatography and analysis by SDS-PAGE. (A) The chromatogram of the NTD KorB showing the peak between 10 and 30 mL of elution volume. Fractions 4 to 7 contain the purified NTD KorB (turquoise box). **(B)** SDS-PAGE gel with NTD KorB in the lanes labelled 4, 5, 6 and 7. Dashed box and arrowhead indicate the over-expression of the NTD KorB. From left to right: molecular weight marker (Marker), supernatant after filtration (Filtrate), flow-through from soluble lysate fraction (FT1 and FT2), wash with buffer (Wash) and eluted fractions (1-8). The fraction size is 4 mL, flow rate is 2 mL/min and length of the run is 60 mL. 20-400 mM Imidazole gradient is used for elution. NTD KorB is eluted from the column between 100 mM and 180 mM of Imidazole. NTD KorB in the pooled fractions is further purified to homogeneity by size exclusion chromatography.

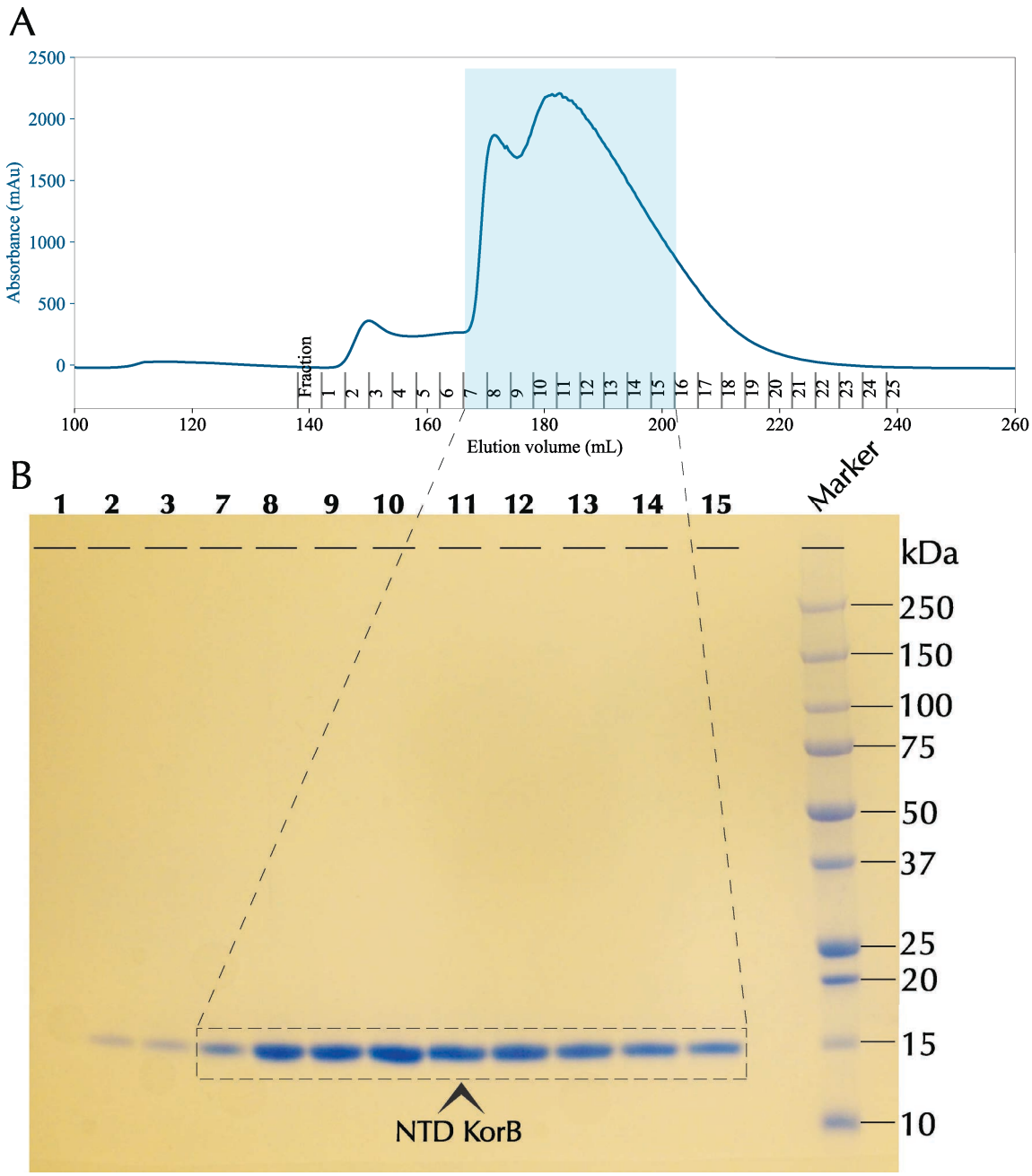

Figure 11: Purification of the NTD KorB by size exclusion chromatography and analysis by SDS-PAGE. (A) Chromatogram of the NTD KorB showing the absorbance trace (A_{280}) of eluent and peaks between 160 mL and 200 mL of elution volume. Fractions on the X-axis indicate each 4 mL of collected fraction. Void volume of the column is 100 mL. The flow rate is 2 mL/min and length of the run is 380 mL. Fractions 7 to 14 contain the purified NTD KorB (turquoise box). **(B)** Image of the SDS-PAGE gel with NTD KorB in the lanes labelled 7, 8, 9, 10, 11, 12, 13 and 14, showing presence of a single protein species in solution. From left to right - eluted fractions from size exclusion chromatography run (1-3 and 7-15) and molecular weight marker (Marker). Dashed box and arrowhead indicate the purified NTD KorB after size exclusion run.

3.1.3 Optimisation of thrombin digestion of N-terminal of KorB

NTD KorB contains a 23 residue N-terminal His-tag with thrombin cleavage site between residue 17 and 18. In order to remove the N-terminal tag from the NTD KorB, thrombin digestion was done. To remove the tag and to optimise the temperature, the protein was incubated with thrombin at 4 °C and room temperature (RT). Thrombin concentration was optimised for the digestion experiment. Thrombin was diluted 100 fold in the same buffer and the reaction was stopped at different times using the inhibitor, PMSF. Figure 12A shows the SDS-PAGE gel with the optimisation of thrombin digestion of the NTD KorB. After optimising the thrombin digestion, the N-terminal tag was removed by incubating the NTD KorB in presence of thrombin at 4 °C for 20 hr. The tag was successfully cleaved with thrombin. The SDS-PAGE gel for the thrombin digestion experiment for the NTD KorB is shown in figure 12B. After the affinity chromatography, the NTD KorB was immediately incubated with thrombin and then the mixture of the NTD KorB and thrombin was loaded on to the size exclusion column to further purify the NTD KorB.

3.2 Crystallisation trials of N-terminal of KorB

The crystallisation trays of the NTD KorB were set using two sparse screens with varied precipitants. In both the screens, precipitation was observed for almost one-third of the conditions and about 70 μm sized protein micro-crystals of the NTD KorB were observed. The sparse-matrix solution conditions that facilitated micro-crystal growth for the NTD KorB are noted in the appendix (page 231). In order to obtain bigger, diffraction quality protein crystals of the NTD KorB, the initial conditions were varied with trials varying NaCl concentrations (0.2-3 M) at different Tris-HCl, pH values (5.5-8.5) but the NTD KorB micro-crystals failed to grow in size with time.

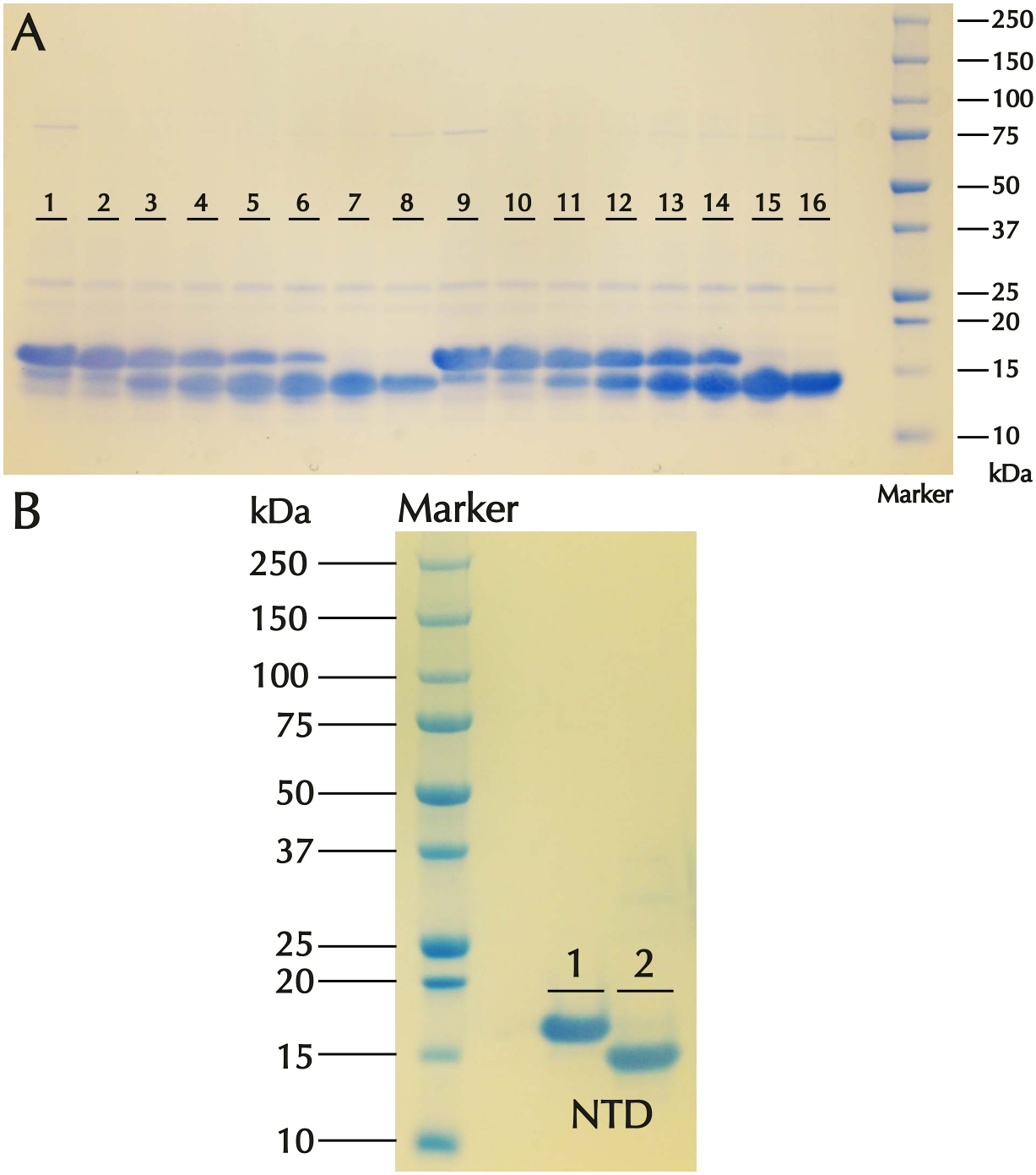

Figure 12: Thrombin digestion of the NTD KorB. (A) The N-terminal tag was removed by incubating NTD KorB in presence of thrombin at RT and 4 °C. SDS-PAGE gel showing the optimisation of thrombin digestion of the NTD KorB. NTD KorB in lanes 2-8 are incubated with thrombin at RT whereas the protein in lanes 9-16 are incubated with thrombin at 4 °C. From left to right - lane 1: stock NTD KorB kept at RT; lane 2: NTD KorB with thrombin at 0 hr (at RT); lane 3: NTD KorB with thrombin at 1 hr; at 2 hr (lane 4); at 4 hr (lane 5); at 5.5 hr (lane 6); at 20 hr (lane 7); at 25 hr (lane 8); lane 9: stock NTD KorB kept at 4 °C; lane 10: NTD KorB with thrombin at 0 hr (4 °C); lane 11: NTD KorB with thrombin at 1 hr (4 °C); at 2 hr (lane 12); at 4 hr (lane 13); at 5.5 hr (lane 14); at 20 hr (lane 15); at 25 hr (lane 16); and molecular weight marker (Marker). **(B)** SDS-PAGE gel of thrombin cleavage of the NTD KorB. The protein was incubated with thrombin at 4 °C for 20 hr. From left to right - molecular weight marker (Marker); lanes 1 and 2: NTD KorB before and after digesting the protein with thrombin respectively.

3.3 Biophysical characterisation of N-terminal of KorB

3.3.1 Circular Dichroism

Circular dichroism (CD) is a biophysical technique with a wide gamut of applications in different fields and is a relatively quick measurement method with easy sample preparation. (Siligardi et al., 2014). In biochemistry and structural biology, CD is used to determine the secondary structure of biological molecules especially proteins. The details about the phenomenon of CD will be explained in section 6.2.2.

CD of the NTD KorB was done to determine the percentage of secondary structure in the protein. Also the CD of the NTD KorB was measured in the presence of a solvent, 2,2,2-trifluoroethanol (TFE). TFE is often used in protein solutions to aid the formation of secondary structures in proteins and TFE has been proposed to induce α-helical character in proteins (Povey et al., 2007; Shiraki et al., 1995). Figure 13 shows the CD spectra of the NTD KorB in the presence and absence of TFE.

The CD data from NTD KorB measurements were deconvoluted from DICHROWEB webserver. The secondary structure percentages for the CD measurements are listed in table 14. The CD data suggest that the NTD KorB (with/without the tag) is partially unfolded. About two-third of the protein has secondary structure characteristic whilst rest is disordered. An increase in α-helical character was observed for the NTD KorB (with/without the tag) with the addition of 20% TFE.

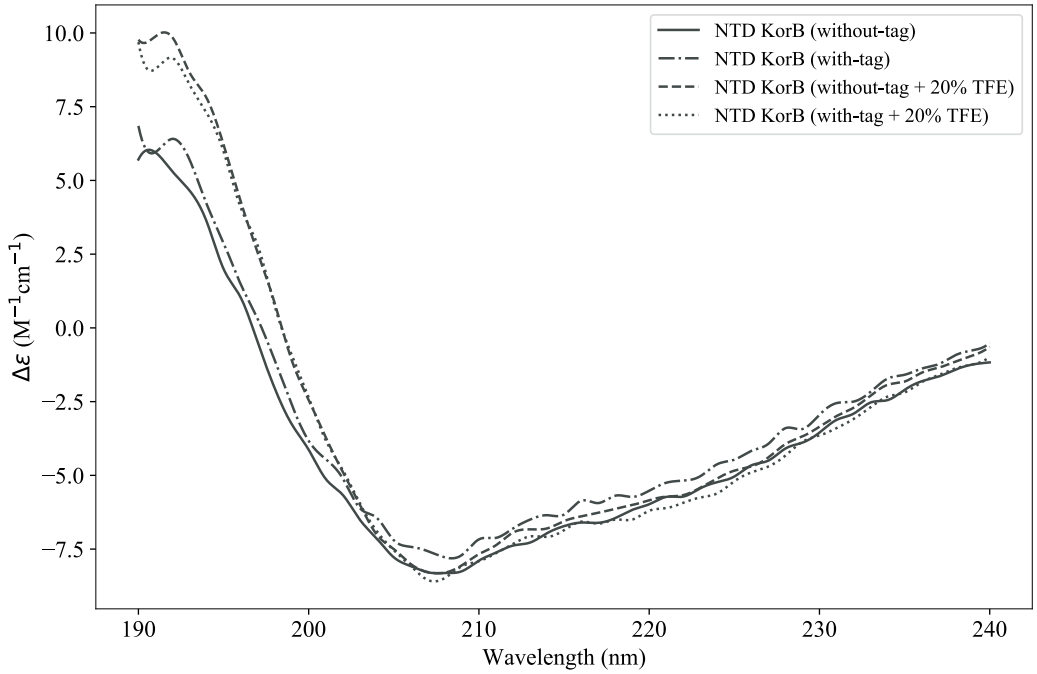

Figure 13: Circular dichroism of the NTD KorB. CD spectra of the NTD KorB (with/without the tag) and in the presence and absence of trifluoroethanol (TFE). Data are shown in the range of 190-240 nm. Delta epsilon ($\Delta\epsilon$) is also known as molar circular dichroism and is measured in per residue molar absorption units of CD.

Protein (condition)	**α-helix (%)**	**β-strand (%)**	**Turn (%)**	**Disorder (%)**
NTD KorB (without-tag)	54	14	10	21
NTD KorB (with-tag)	51	15	12	22
NTD KorB (without-tag + 20% TFE)	61	10	10	19
NTD KorB (with-tag + 20% TFE)	60	15	9	16

Table 14: Secondary structure deconvolution of the NTD KorB from the CD data. CD data deconvoluted for the NTD KorB from DICHROWEB analysis webserver. The data suggest that NTD KorB (with/without the tag) is partially unfolded. About two-third of the protein has secondary structure characteristic whilst rest is disordered. An increase in α-helical character is observed for the NTD KorB (with/without the tag) with the addition of 20% trifluoroethanol (TFE).

3.3.2 Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC, invented by Svedverg in 1923 and awarded Nobel prize in 1926) is a complementary technique used for structural analysis of proteins (The Nobel Prize, 1926). AUC is one of the best-known methods to identify polymers and especially proteins. The method is very good to identify mass, polydispersity, size and even conformational change or complex formation. The sedimentation velocity is one of the two types of approaches that can be used for structural analysis. A relatively high rotor speed helps in measuring the rate of sedimentation and yields sedimentation coefficient(s) which is used to calculate size of the molecule being studied with AUC. The other way is the sedimentation equilibrium that is used to investigate oligomeric states and equilibrium assembly constants (Lebowitz et al., 2009; Schuck et al., 2002).

The velocity AUC measurements were run for the NTD KorB at three different concentrations in order to observe the concentration dependent polydispersity of the NTD KorB. Figure 14 shows the analysed AUC data for the NTD KorB at three different concentrations – A_{280} of 0.1, 0.5 and 0.9. AUC measurements for three protein concentrations agree with each other and processed velocity run data suggest that the NTD KorB behaves as a single species in solution. The frictional ratio for the analysis was 1.39246 and buffer viscosity was 0.01002 Poise. The NTD KorB molecular mass calculated from the AUC data for different protein concentrations was 19.5 kDa at 1.2 mg/mL, 18.4 at 0.6 mg/mL and 15.7 at 0.2 mg/mL. The NTD KorB behaves as a single species in solution with apparent molecular mass on the higher side when compared to the expected molecular mass of 14.1 kDa as predicted from the protein sequence with ProtParam server (Gasteiger et al., 2005). For the NTD KorB, any oligomeric species or higher order aggregates were not observed.

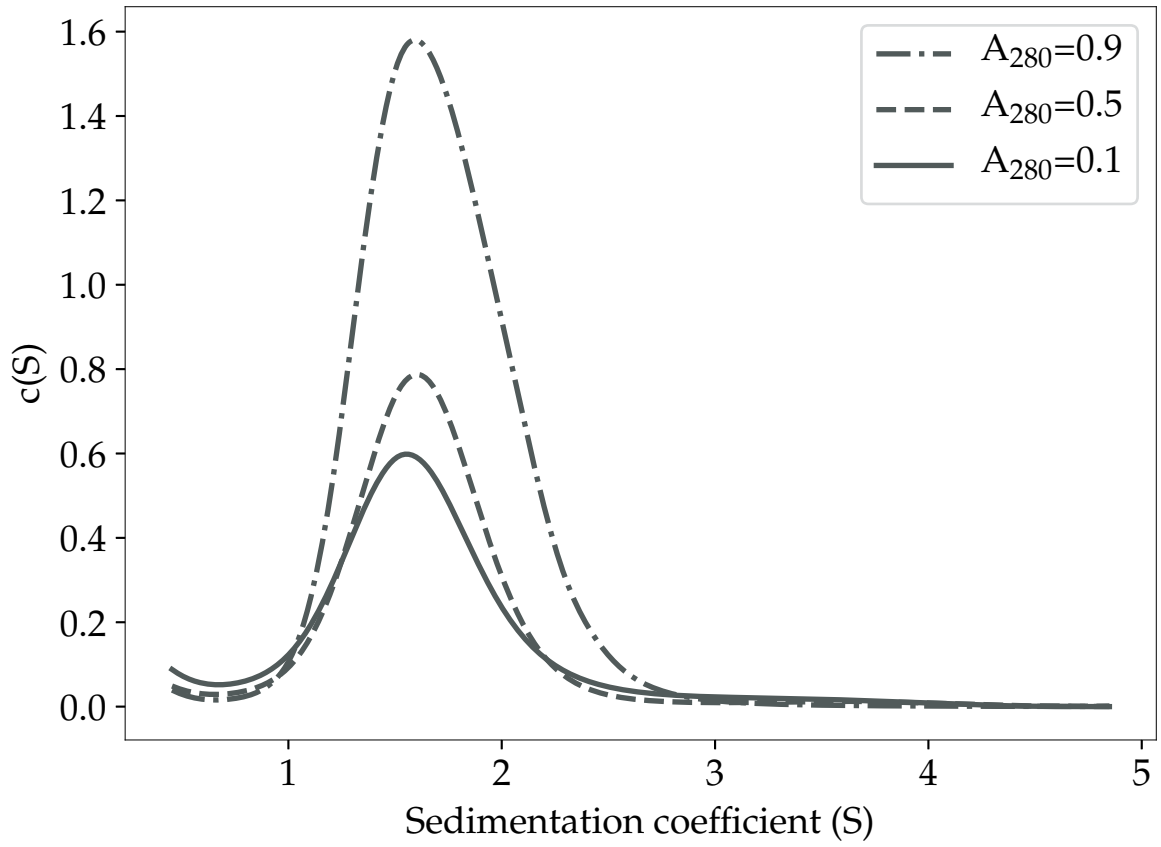

Figure 14: Analytical ultracentrifugation data analysis of the NTD KorB. The AUC velocity run data represent the distribution of sedimentation coefficient of the NTD KorB. Symmetric single peak for the NTD KorB at 1.2 mg/mL, 0.6 mg/mL and 0.2 mg/mL indicates that the protein behaves as a single species in solution with apparent molecular mass on the higher side when compared to the expected molecular mass of 14.1 kDa as predicted from the protein sequence with ProtParam server (Gasteiger et al., 2005). For NTD KorB, any oligomeric species or higher order aggregates are not observed.

3.3.3 Molecular mass estimation by Mass Spectrometry

In order to check the structural integrity of the NTD KorB protein, mass spectrometry (MS) was done on the NTD KorB. The purified NTD KorB was subjected to Electrospray Ionization-Mass Spectrometry (ESI-MS) for estimation of molecular mass and the protein samples were given to the mass spectrometry facility in the School of Chemistry at the University of Birmingham. Figures 15 and 16 show the mass spectrometry data for the NTD KorB before and after cleaving the N-terminal tag with thrombin respectively.

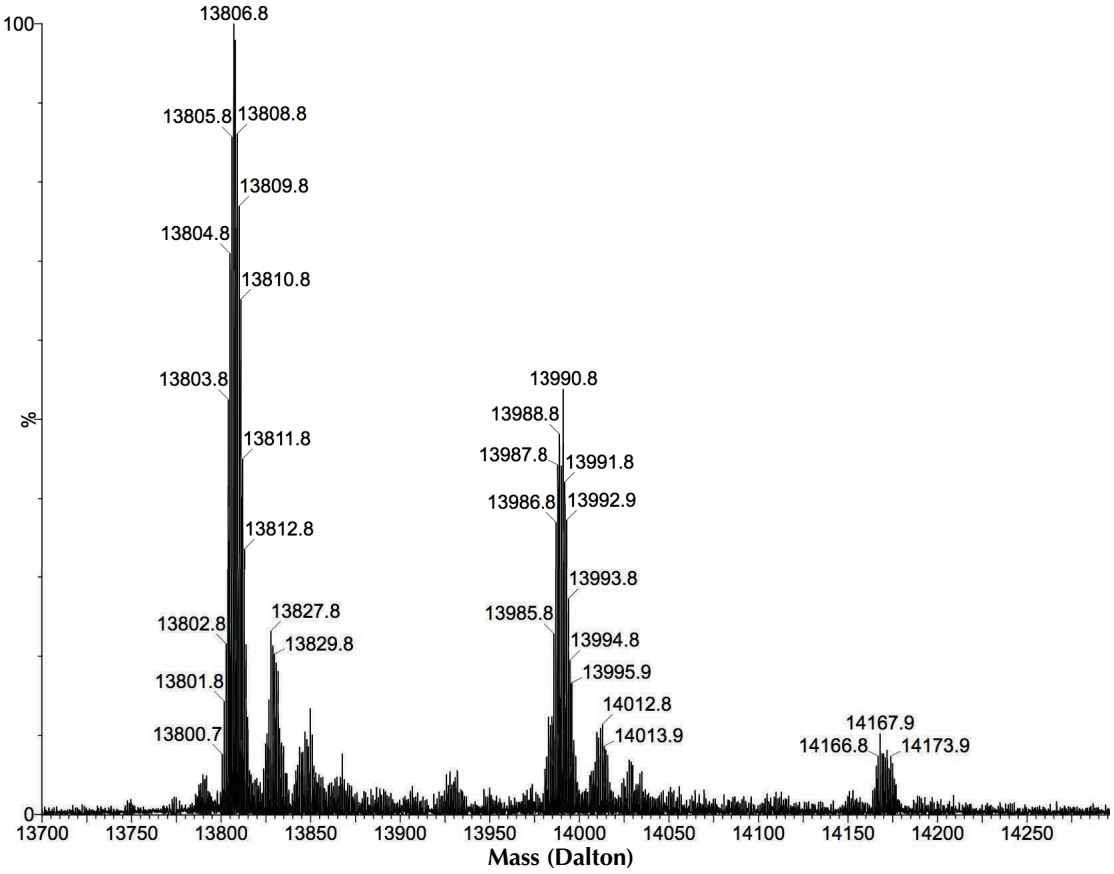

Figure 15: Molecular mass estimation of the NTD KorB (with tag) by mass spectrometry. The ESI-MS spectrum is shown for the NTD KorB before the N-terminal tag cleavage. The molecular mass calculated for the protein is 13,806.8 Da and the theoretical mass of the protein calculated from the amino acid sequence with ProtParam server (Gasteiger et al., 2005) is 14,083 Da.

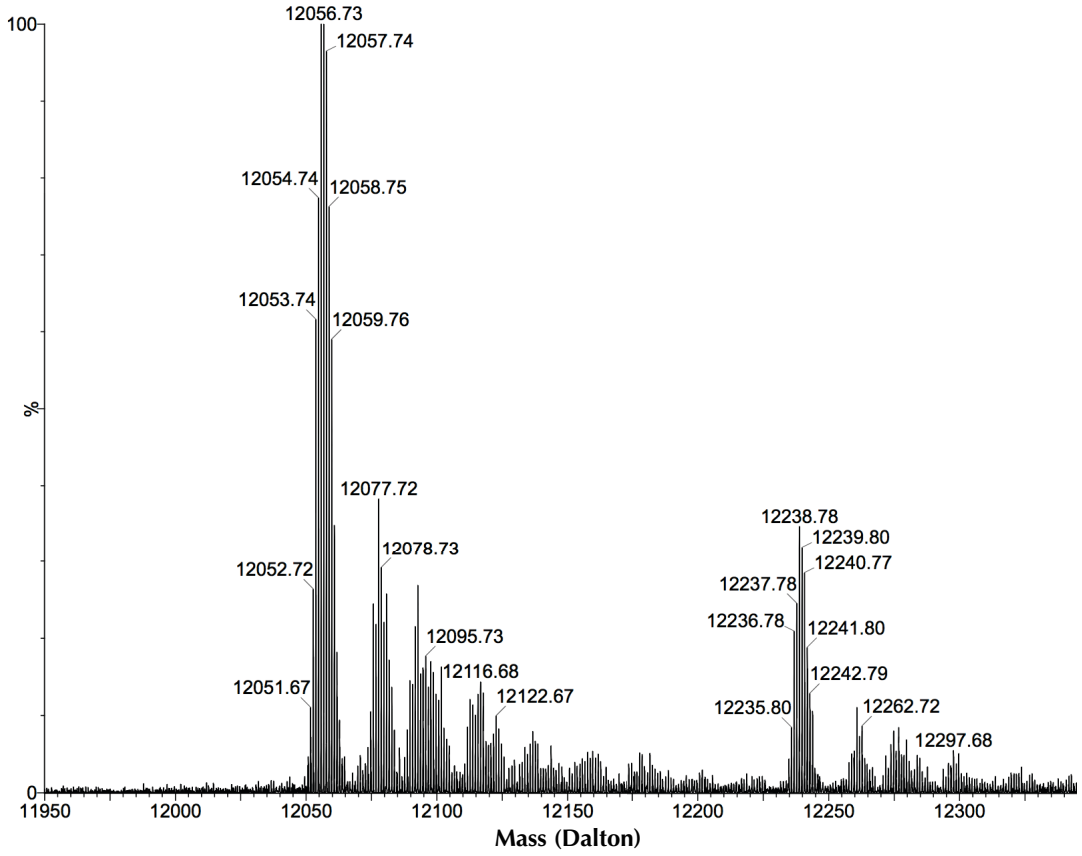

Figure 16: Molecular mass estimation of the NTD KorB (without tag) by mass spectrometry. The ESI-MS spectrum is shown for the NTD KorB after cleaving the N-terminal tag. There is reduction in the molecular mass of the NTD KorB as the first 17 residues were cleaved after digesting the protein with thrombin. The molecular mass calculated for the protein is 12,056.7 Da and the theoretical mass of the protein calculated from the amino acid sequence with ProtParam server (Gasteiger et al., 2005) is 12,234 Da.

3.4 Discussion

NTD KorB was expressed and purified to homogeneity. From 1 L of the culture, 22 mg of the NTD KorB was obtained, sufficient amount to perform the biophysical experiments. In the size exclusion chromatogram for the NTD KorB, two peaks were observed that might suggest the presence of more than one species. It is plausible that there is protein monomer-dimer equilibrium or the protein is modified.

Crystallisation trays of the NTD KorB were set using two sparse screens with varied precipitants and precipitation was observed for almost one-third of the conditions. The micro-crystals for the NTD KorB were observed but the micro-crystals failed to grow in size with time. The difficulty to crystallise the NTD KorB in a stable conformation could be attributed to its partial unfolded character, as indicated by the CD data.

AUC velocity run suggests that the NTD KorB behaves as a single species in solution with average molecular mass of three concentrations as 17.9 kDa and this observation is consistent with the size exclusion chromatography as NTD KorB primarily elutes as a single, monodisperse species with molecular mass of 17-18 kDa. Thus, from AUC velocity data and size exclusion chromatogram, it is conclusive that the NTD KorB behaves as a single species in solution but with higher molecular mass in comparison to the predicted molecular mass. The MS data indicated that the molecular mass of the NTD KorB agreed with the theoretical mass but not exactly. The implication of the difference in the protein mass will be discussed in the next chapter.

CD data predict that the NTD KorB with His-tag has 66% (α-helix: 51%; β-strand: 15%) secondary structure character. In absence of the tag, secondary structure prediction increases slightly to 68%. On the other hand, in the presence of TFE, the secondary structure of the NTD KorB increases to 75% and 71%. The CD data suggest that NTD KorB (with/without the tag) is partially unfolded. About two-third of the protein has secondary structure characteristic whilst rest is disordered. In the presence of TFE, a

slight increase in α-helical character was observed for the NTD KorB (with/without the tag). The TFE did increase the secondary structure of the protein but the effect did not aid in the crystal trials of the NTD KorB.

Since a structure for the NTD KorB was not elucidated from the crystal trials, the structure of the protein was approached with the help of solution state NMR spectroscopy. A structure of the NTD KorB will provide atomic level details. The structure of the NTD KorB can be compared with the N-terminal domains of the structures of the CBPs that have been characterised. The comparison between the structures can inform about the similarities/differences between the different proteins. This information can help to understand how CBPs interact with each other around the *parS* or O_B sites. Furthermore, the information from this study can be extended to other bacterial partitioning systems with keen emphasis on the bacterial chromosomal distribution during division.

In the next chapter, an ensemble structure of the N-terminal domain of KorB is calculated using the solution state NMR spectroscopy data. Further, Molecular Dynamics (MD) simulations on the 3D NTD KorB structure are performed using the coarse-grain simulation package, AWSEM. The details from the NMR and MD experiments about the flexibility of the NTD KorB are discussed suggesting important structural insights.

4

Structure of the N-terminal domain of KorB

4 Structure of the N-terminal domain of KorB

4.1 Introduction

In 1938 the phenomenon of Nuclear Magnetic Resonance (NMR) was first described by Isidor Rabi (Rabi et al., 1992) and in 1944 Isidor was awarded the Nobel Prize in Physics (The Nobel Prize, 1944). In 1952, Felix Bloch and Edward Mills Purcell shared the Nobel Prize in Physics for expanding the technique of NMR to liquids and solids (The Nobel Prize, 1952). Since then the technique has been improved drastically.

NMR spectroscopy is an extremely powerful analytical technique in the field of structural biology providing extensive information about the physical and the chemical properties of a molecule. NMR is used for the determination of high resolution 3-dimensional (3D) solution structures and chemical environment of molecules. NMR is unparalleled when it comes to determining the dynamics occurring at individual atoms and NMR is also used in quality control to determine the purity of a sample.

4.2 Protein sample preparation for NMR analysis

The preparation of protein sample for NMR data collection is taxing when compared to that of a typical protein purification strategy. Under natural conditions, the NMR compatible isotopes are present but at very low concentrations. However, to observe reasonable signal-to-noise ratio (SNR) in an NMR experiment, the sample has to be enriched with NMR active isotopes. A standard over-expression of a desired protein is achieved by growing the bacteria in a rich medium such as Lysogeny Broth (LB) which serves as a source for majority of the growth nutrients. On the other hand, minimal medium with single nitrogen and carbon source is provided for enrichment of NMR-compatible isotopes. The molecules for the protein production machinery must be syn-

thesised from the single sources provided and this results in impeding the growth of cells. Therefore, it becomes difficult for a bacterial cell to over-express protein in high quantity. For most of the KorB constructs including the NTD KorB, the yield of the protein in minimal medium was half that of the standard protein yield in LB.

4.3 NMR experiments

One-dimensional (1D) proton spectrum is the basic spectrum obtainable from an NMR spectrometer. The key problem with the analysis of large molecular systems by 1D NMR is the complexity of signal and the overlap of peaks. By introducing additional spectral dimensions, the 1D spectrum can be simplified and additional information is obtained. This led to the development of multidimensional NMR methods to study and analyse larger molecular species.

Researchers have devised different types of multidimensional (2, 3 or 4D) experiments to assign macromolecules like proteins. In 3D NMR experiments ('triple resonance') three different nuclei (^{1}H, ^{13}C, ^{15}N) are correlated and these experiments are the method of choice for sequentially assigning proteins. The protein has to be labelled with ^{15}N and ^{13}C for most of the 3D experiments including HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and HBHA(CO)NH. Table 15 lists the NMR experiments along with their salient features, which were used for structure calculation of the NTD KorB. The advantage of the 3D spectra is their simplicity and because of that they are used to assign the protein backbone using the sequential walk method.

4.3.1 Basic NMR experiment: 1D ^{1}H-NMR

A 1D NMR spectrum is the basic spectrum obtainable and the chemical environment of the molecule dictates the position of the peaks on the 1D scale. Simply put, for a particular molecule, this spectrum gives information about the chemical shift of protons in different chemical environments and these shifts can be used to elucidate the structure of small molecules. The signal from the water (55 M) dominate the 1D ^{1}H spectrum

and the contribution of water is suppressed with solvent presaturation or by gradient pulses. By observing the proton signals across the 0-10 ppm range, a 1D ^{1}H spectrum indicates whether a particular protein occupies a folded conformation and the signals from high-field methyl groups at < 0.5 ppm are indicative of the protons being close to the aromatic residues. The 1D spectrum can help in deciding, if one should label the protein for 2D experiments.

4.3.2 HSQC: acquiring fingerprint of a protein

HSQC stands for heteronuclear single quantum coherence spectroscopy and is a particularly significant experiment in the field of protein NMR. This experiment is used to obtain a 'fingerprint' of a protein. The data from the ^{15}N-^{1}H HSQC experiment provides information about the overall fold and dynamics of a protein in solution.

^{15}N-^{1}H HSQC is a classical NMR experiment correlating N and H nuclei. Ideally, this experiment detects all NH groups present in a protein. The spectrum is 2D, with one axis as ^{1}H and other axis corresponds to a heteroatom (a nucleus other than ^{1}H). The spectrum gives a peak for each ^{1}H-heteroatom pair being considered. Essentially the name HSQC, heteronuclear single quantum coherence describes that the chemical shifts of two nuclei (^{1}H and ^{13}C/^{15}N) are related. In this method, we observe signals from ^{1}H and the heteroatoms (^{13}C/^{15}N) directly attached via a single bond to the proton. The ^{1}H signals are observed directly, while the heteroatoms are observed via the scalar couplings, via one-bond scalar couplings for heteronuclear correlations as proton is relatively more sensitive than ^{13}C/^{15}N. It is not possible to detect Proline residues in an ^{15}N-^{1}H HSQC spectrum because of the absence of an amide proton. The HSQC method is the building block for most of the 3D or higher dimensional NMR spectra (Teng, 2007). The 3D experiments used in this study include HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HBHA(CO)NH, HNN, HCCH-TOCSY, ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC and are described in the succeeding sections.

Table 15: NMR experiments used for structure calculation of NTD KorB.

NMR experiment	**Minimal labelling**	**Dimensions acquired**	**Information gathered**
HSQC	^{15}N	2 (H, N)	Ideally a single amide peak per residue, additional peaks from sidechain of amine residues
HN(CA)CO	^{15}N, ^{13}C	3 (H, N, C′)	Two peaks are observed at NH frequency: CO*(i)*, CO*(i–1)*
HNCO	^{15}N, ^{13}C	3 (H, N, C′)	A single peak is observed at NH frequency: CO*(i–1)*
HN(CO)CA	^{15}N, ^{13}C	3 (H, N, C)	A single peak is observed at NH frequency: Cα*(i–1)*
HNCA	^{15}N, ^{13}C	3 (H, N, C)	Two peaks are observed at NH frequency: Cα*(i)*, Cα*(i–1)*
HN(CO)CACB	^{15}N, ^{13}C	3 (H, N, C)	Two peaks are observed at NH frequency: Cα*(i–1)*, Cβ*(i–1)*
HNCACB	^{15}N, ^{13}C	3 (H, N, C)	Four peaks are observed at NH frequency: Cα*(i)*, Cα*(i–1)*, Cβ*(i)*, Cβ*(i–1)*
HBHA(CO)NH	^{15}N, ^{13}C	3 (H, H, N)	Two peaks are observed at NH frequency: Hα*(i–1)*, Hβ*(i–1)*
HCCH–TOCSY	^{13}C	3 (H, H, C)	All the Hydrogen atoms attached to all the carbon atoms within a residue (apart from the aromatics)

NMR experiment	Minimal labelling	Dimensions acquired	Information gathered
^{15}N–NOESY–HSQC	^{15}N	3 (H, H, N)	All the Hydrogen atoms within a radius of 5–6 Å of the NH group observed.
^{13}C–NOESY–HSQC	^{13}C	3 (H, H, C)	All the Hydrogen atoms within a radius of 5–6 Å of the CH group observed. Aromatics are usually recorded separately.
HNN	^{15}N	3 (H, N, N)	Three peaks are observed at NH*(i)* frequency: HN*(i-1)*, HN*(i)*, HN*(i–1)*

Table 15: NMR experiments used for structure calculation of NTD KorB (continued).

4.3.3 HNCO and HN(CA)CO experiments

HNCO is the most sensitive 3D NMR experiment (Ikura et al., 1990). The magnetisation is transferred from the NH proton *i.e.* $^{1}H^{N}$ to the $^{15}N^{H}$ and then selectively to the ^{13}CO of previous residue (*i-1*) via the $^{15}N^{H}$-^{13}CO J-coupling, followed by evolution of the C′ chemical shift (figure 17A). Following the evolution, magnetisation is transferred in reverse order from ^{13}CO (*i-1* residue) to $^{15}N^{H}$ and then to $^{1}H^{N}$. The chemical shift is evolved for ^{13}CO, $^{15}N^{H}$ and $^{1}H^{N}$ resulting in a 3D spectrum. This experiment results in a single peak per residue in the spectrum correlating N*(i)*H*(i)* and C′*(i-1)*. Because of this experiment, the distinction between the *i* and the *i-1* ^{13}CO residue is achievable in the HN(CA)CO experiment.

The HN(CA)CO experiment is similar to the HNCO experiment and contains additional information (Grzesiek and Bax, 1992). The magnetisation is transferred from the NH proton *i.e.* $^{1}H^{N}$ to the $^{15}N^{H}$ and then via the $^{15}N^{H}$-$^{13}C\alpha$ J-coupling to both the ^{13}CO

of the residue *i* and to the CO of the previous residue (*i-1*) followed by evolution of the chemical shift (figure 17B). Following the evolution, the magnetisation is transferred in the reverse order from ^{13}CO (both *i* residue and *i-1* residue) to ^{13}Cα*(i)* and then to ^{15}N^{H}*(i)* and then to ^{1}H^{N}*(i)* . The chemical shifts are evolved for ^{13}CO, ^{15}N^{H} and ^{1}H^{N} resulting in a 3D spectrum. Two peaks are observed in this experiment at each NH chemical shift, one for the *i* ^{13}CO residue and other for the *i-1* ^{13}CO residue.

4.3.4 HN(CO)CA and HNCA experiments

The HN(CO)CA (Kay et al., 1990) experiment is used in conjunction with the HNCA experiment (Ikura et al., 1990). In HN(CO)CA, information about the *i-1* residue is collected. The magnetisation is transferred from the NH proton *i.e.* ^{1}H^{N} to the ^{15}N^{H} followed by transfer to ^{13}CO*(i-1)* (backbone carbonyl carbon; CO). It is then transferred to the ^{13}Cα of the previous residue (*i-1*) followed by evolution of the chemical shift (figure 17C). Following the evolution, the magnetisation is transferred in the reverse order to CO*(i-1)* to ^{15}N^{H}*(i)* and then to ^{1}H^{N}*(i)*. Here, the chemical shift is evolved for ^{13}Cα, ^{15}N^{H} and ^{1}H^{N}.The chemical shift is not evolved for ^{13}CO. At the end we get one peak for one *i-1* Cα residue correlated with each NH*(i)* resonance. It is majorly because of this experiment that the distinction between the *i* and the *i-1* Cα residue is achievable in the HNCA experiment.

The HNCA experiment is similar to the HN(CO)CA experiment but in HNCA, information about the Cα resonance of both the *i* residue and the *i-1* residue is collected. In this experiment, magnetisation is transferred from the NH proton *i.e.* ^{1}H^{N} to the ^{15}N^{H} followed directly to the ^{13}Cα of itself *i.e. i* and previous residue (*i-1*) followed by evolution of the chemical shift. Following the evolution, the magnetisation is transferred in the reverse order from ^{13}Cα (both *i* and *i-1*) to ^{15}N^{H} and then to ^{1}H^{N} (figure 17D). Here, the chemical shift is evolved for ^{13}Cα, ^{15}N^{H} and ^{1}H^{N} and in the process, information about two resonances are collected. At the end of it, we get a spectrum that is similar to the HN(CO)CA spectrum but we get two peaks in this spectrum, one from the Cα of *i*

residue and the previous residue. Note: Two peaks are observed because the coupling constant of NH to Cα for both residues *i* and *i-1* is similar; also, because the coupling constant of NH to C′ *(i-1)* is much bigger than to C′*(i)*, only one peak is observed via C′ *(i-1)*.

4.3.5 HNCACB and HN(CO)CACB experiments

The HNCACB experiment (Yamazaki et al., 1994) is similar to the HN(CO)CACB experiment (explained later) but has additional information. In this experiment information about the *i* residue and the *i-1* residue is collected. The magnetisation is transferred from $^{1}H^{N}$ to $^{15}N^{H}$ *i* and is further transferred to $^{13}C\alpha$ and $^{13}C\beta$ (figure 17F). At the end of it, we get two peaks for $^{13}C\alpha$ (*i* and *i-1*) and two peaks for $^{13}C\beta$ (*i* and *i-1*) for a single NH group. Simultaneous evolution of chemical shifts on $^{13}C\alpha$ and $^{13}C\beta$ makes them appear in a single dimension. We get four peaks in total in this spectrum. Two peaks (usually the stronger ones) from the $^{13}C\alpha$ and $^{13}C\beta$ of the residue *i* and two peaks (usually relatively weaker) from the $^{13}C\alpha$ and $^{13}C\beta$ of the residue *i-1*.

The HN(CO)CACB experiment is similar to the HNCACB experiment (Grzesiek and Bax, 1992). In this experiment information about the *i-1* residue is collected. The magnetisation is transferred from $^{1}H^{N}$ to $^{15}N^{H}$ *i* and is further transferred to $^{13}C\alpha$ and $^{13}C\beta$ via ^{13}CO (figure 17E). Since this is an out-and-back experiment, the magnetisation comes all the way back from $^{13}C\beta$ to $^{1}H^{N}$ for detection. Simultaneous evolution of chemical shifts on $^{13}C\alpha$ and $^{13}C\beta$ makes them appear in a single spectrum. We get two peaks in total in this spectrum. Both peaks are from the $^{13}C\alpha$ and $^{13}C\beta$ of the residue *i-1*. The chemical shift is evolved for $^{15}N^{H}$ and $^{1}H^{N}$ that makes the other two dimensions of the 3D spectrum. The chemical shift is not evolved for ^{13}CO.

The HN(CO)CACB_only experiment is a modified version of HN(CO)CACB experiment. The data are collected such that we get a single peak from $^{13}C\beta$ of *i-1* in this spectrum. The chemical shift is evolved for $^{15}N^{H}$ and $^{1}H^{N}$ that makes the other two dimensions of the 3D spectrum. The chemical shift is not evolved for ^{13}CO. This

experiment is helpful in accurately determining the $^{13}C\beta$ *i-1* in conjunction with the HN(CO)CACB spectrum and HNCACB spectrum.

4.3.6 HBHA(CO)NH experiment

In this experiment, information about the *i-1* residue is collected. The magnetisation is transferred from $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$ respectively. From $^{13}C\beta$, magnetisation is transferred $^{13}C\alpha$ and then to $^{15}N^{H}$ and $^{1}H^{N}$ via ^{13}CO. The observed peaks in the spectrum are from the $^{1}H\alpha$ and $^{1}H\beta$ of the residue *i-1* (Grzesiek and Bax, 1993).

4.3.7 HNN experiment

HNN is a 3D experiment to acquire sequential correlations between H^{N} and ^{15}N along the backbone of a protein molecule (Panchal et al., 2001). Self peaks and sequential peaks in the spectrum help to assign sequential residues in both directions along the protein backbone. In HNN spectrum, a combinations of positive and negative peaks facilitate sequential walk. In the HNN spectrum, three peaks are observed at NH*(i)* frequency: HN*(i-1)*, HN*(i)* and HN*(i–1)*. The schematic representation of HNN data performed on the NTD KorB are shown in figure 22A-C and the magnetisation pathway is shown in figure 22D. The expected peak pattern for triplet peaks are shown in appendix on page 218 (figure 81).

4.3.8 HCCH-TOCSY experiment

The spectrum resulting from HCCH-TOCSY experiment is used for side-chain assignment and is particularly useful for collating the ^{1}H information for the backbone assignment. The magnetisation is transferred from all the side-chain hydrogen atoms to their connected ^{13}C atoms, followed by isotropic ^{13}C mixing and magnetisation is transferred back for detection via the side-chain hydrogen nuclei (Bax et al., 1990; Olejniczak et al., 1992). The resulting spectrum consists of strips at each ^{13}C frequency in the side chain wherein all the side-chain hydrogen resonances are observable. The representation of

this experiment is shown in figure 27A and the data for the NTD KorB are shown in figure 27B-C.

4.3.9 NOESY experiments

For ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC (Marion et al., 1989a,b) experiments, the protein has to be minimally labelled either with ^{15}N or ^{13}C. In ^{15}N-NOESY-HSQC, magnetisation is exchanged using NOEs between all hydrogen atoms. Thereafter, magnetisation is transferred to the neighbouring ^{15}N atoms and back to ^{1}H atoms for detection. In ^{13}C-NOESY-HSQC, magnetisation is exchanged using the NOEs between all hydrogen atoms. Thereafter, magnetisation is transferred to the neighbouring ^{13}C atoms and back to ^{1}H atoms for detection. Depending on the ^{13}C frequency used for excitation during the ^{13}C-NOESY-HSQC experiment, transfer either occurs to/from the aliphatic ^{13}C nuclei or to/from the aromatic ^{13}C nuclei but not both instances. Before acquisition, the spectrum can be centred either on the ^{13}C aliphatic or on the ^{13}C aromatic residues, resulting in acquiring useful information about the aliphatic and aromatic carbons for the assignment.

4.4 Results

NTD KorB (1-130 aa) with a His-tag was expressed in minimal M9 medium and purified as described in chapter 3. The residues 1-23 in the NTD KorB construct come from the expression vector (N-terminal) and residues 24-130 correspond to the residues 31-137 of the wild-type KorB protein. During size exclusion chromatography, sodium phosphate buffer was used to purify the NTD KorB. All the NMR experiments were acquired in the same buffer (20 mM Na phosphate, pH 6, 100 mM NaCl, 0.1 EDTA). Initially, multiple 2D experiments (HSQCs) for NTD KorB were acquired at different pH values and temperatures to optimise the spectrum and the number of peaks. However, peak overlap was observed in the HSQC spectrum of the NTD KorB.

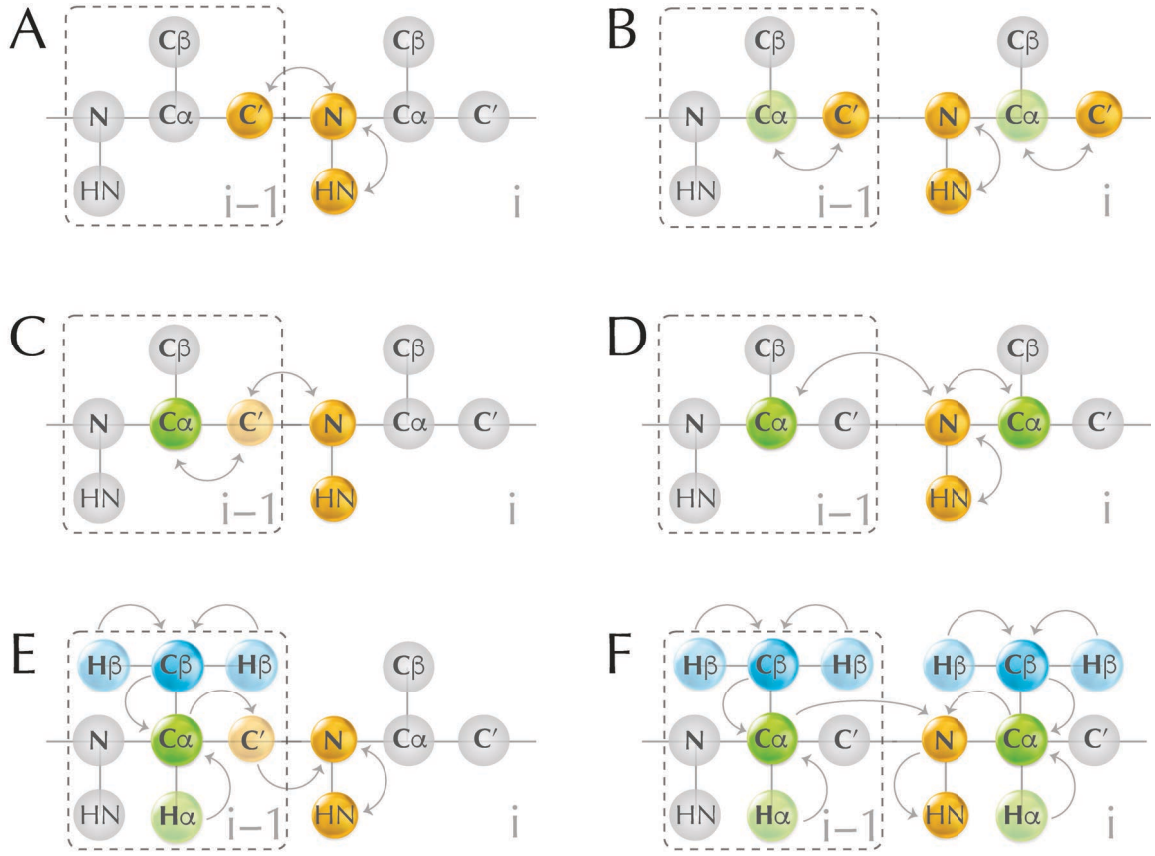

Figure 17: Schematic representing the magnetisation transfer pathway for 3D protein backbone assignment experiments. Following the classic NMR nomenclature, each experiment is depicted with a residue "*i*" and preceding residue "*i-1*". The individual atoms for *i-1* residue are shown inside the dashed box and rest of the atoms represent the residue "*i*". All the atoms are shown in grey but the atoms involved in magnetic transfer have been coloured. HN, N and C′ are represented with orange circles. Cα and Hα are represented with green circles. Cβ and Hβ are represented with blue circles. Solid coloured circles represent those atoms where magnetisation is transferred and data are collected indirectly, meanwhile atoms where magnetisation is transferred but absence of data collection are shown with diaphanous circles. Experiments for backbone assignments include: (A) HNCO, (B) HN(CA)CO, (C) HN(CO)CA, (D) HNCA, (E) HN(CO)CACB and (F) HNCACB and arrows depict the magnetisation transfer between the atoms. The experiments A to D are out and back type experiments and the magnetisation begins on atom HN. On the other hand, the magnetisation begins at the Hα and Hβ for experiments E and F. The illustration is adapted from Higman (2017).

Figure 18 shows the superimposition of HSQC of the NTD KorB before and after cleaving the tag (only a few peaks disappear in the cleaved spectrum). Each dimension for each HSQC experiment was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID (free induction decay). The HSQC spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions. All the following HSQC spectra mentioned later in the thesis were processed as described above.

Thereafter, complementary 3D NMR backbone spectra were used to assign the backbone of the NTD KorB. Figure 19 shows an extract from HNCO and HN(CA)CO experiments of the NTD KorB and the peak in the HNCO helps to distinguish the two C′ peaks from the *i* and the *i-1* residue in the HN(CA)CO. Figure 20 shows an extract from HN(CO)CA and HNCA experiments of the NTD KorB and the peak in the HN(CO)CA helps to distinguish the two Cα peaks in the HNCA. Figure 21 shows an extract from HN(CO)CACB and HNCACB experiments of the NTD KorB and the two peaks in HN(CO)CACB helps to distinguish between the four peaks in the HNCACB.

In general, flexible proteins suffer from narrow chemical shift dispersions (particularly ^{1}H). It is to be noted that ^{15}N chemical shift dispersions for flexible proteins are good and there are experiments such as HNN that exploit the broad chemical shift dispersion along ^{15}N to rapidly assign most of the backbone resonances. The HNN experiment is particularly useful to perform the sequential walk for backbone NH assignments. To aid the NTD KorB backbone assignment, the HNN experiment was acquired. Figure 22 shows data from the HNN experiment of the NTD KorB. In HNN, three peaks were observed, each indicating an NH peak for the residue *i-1*, *i* and *i+1* and these peaks were used to walk across prolines and confirm the backbone assignment.

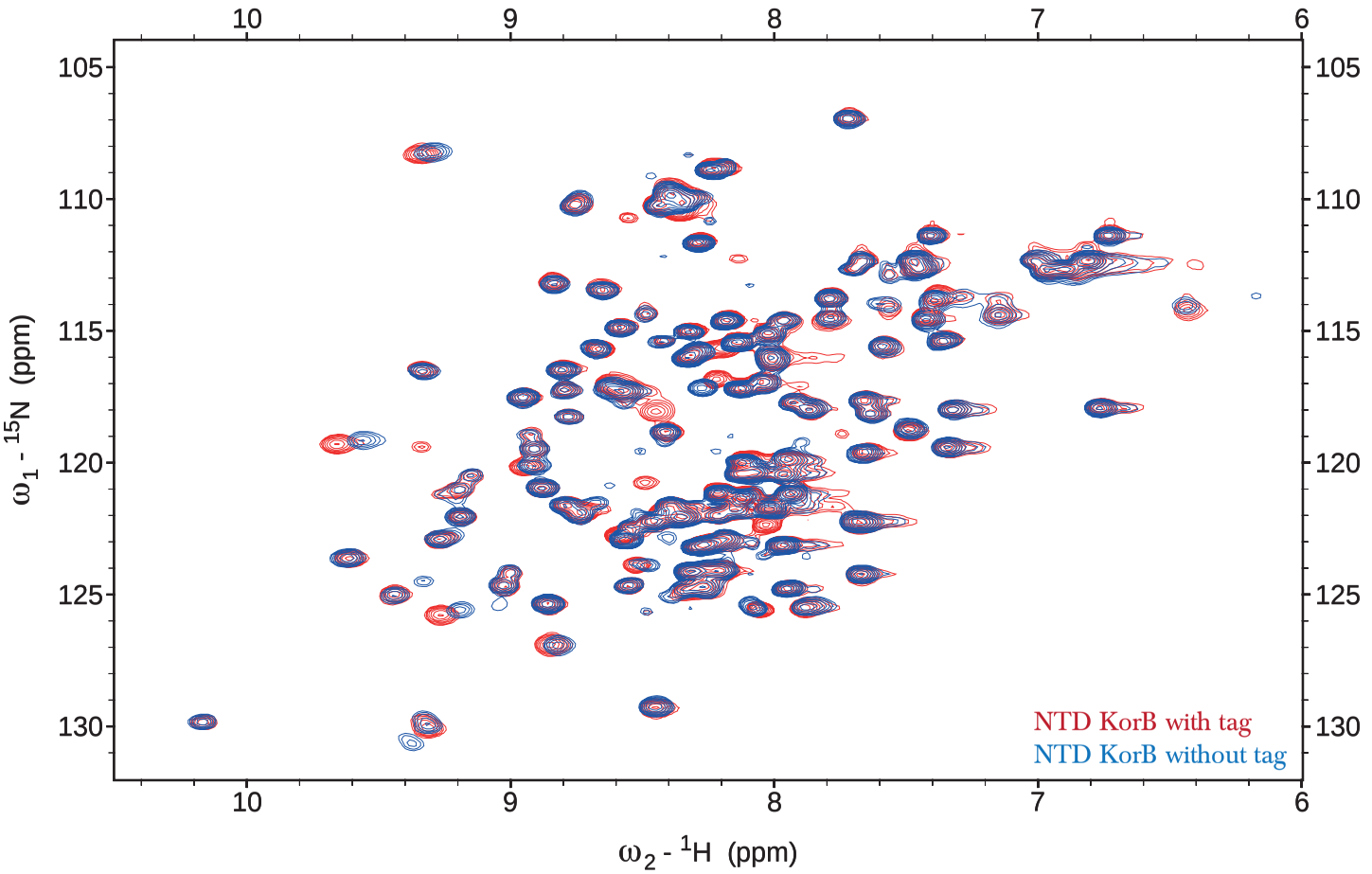

Figure 18: Overlay of ^{1}H-^{15}N HSQC spectrum of the NTD KorB before and after cleaving the N-terminal tag. The NTD KorB contains a 23 residue N-terminal His-tag with thrombin cleavage site between residue 17 and 18 and thrombin digestion was done in order to remove the N-terminal tag from the protein. Both HSQC spectra were taken in 20 mM Na phosphate, pH 6, 100 mM NaCl, 0.1 EDTA with 0.5 mM NTD KorB. The spectra were acquired on a Bruker 900 MHz at 35 °C. Red contoured spectrum is of the NTD KorB before cleaving the N-terminal tag and the blue contoured spectrum is the one after tag removal. All the following HSQC spectra mentioned later in the thesis were processed as described in section 4.4.

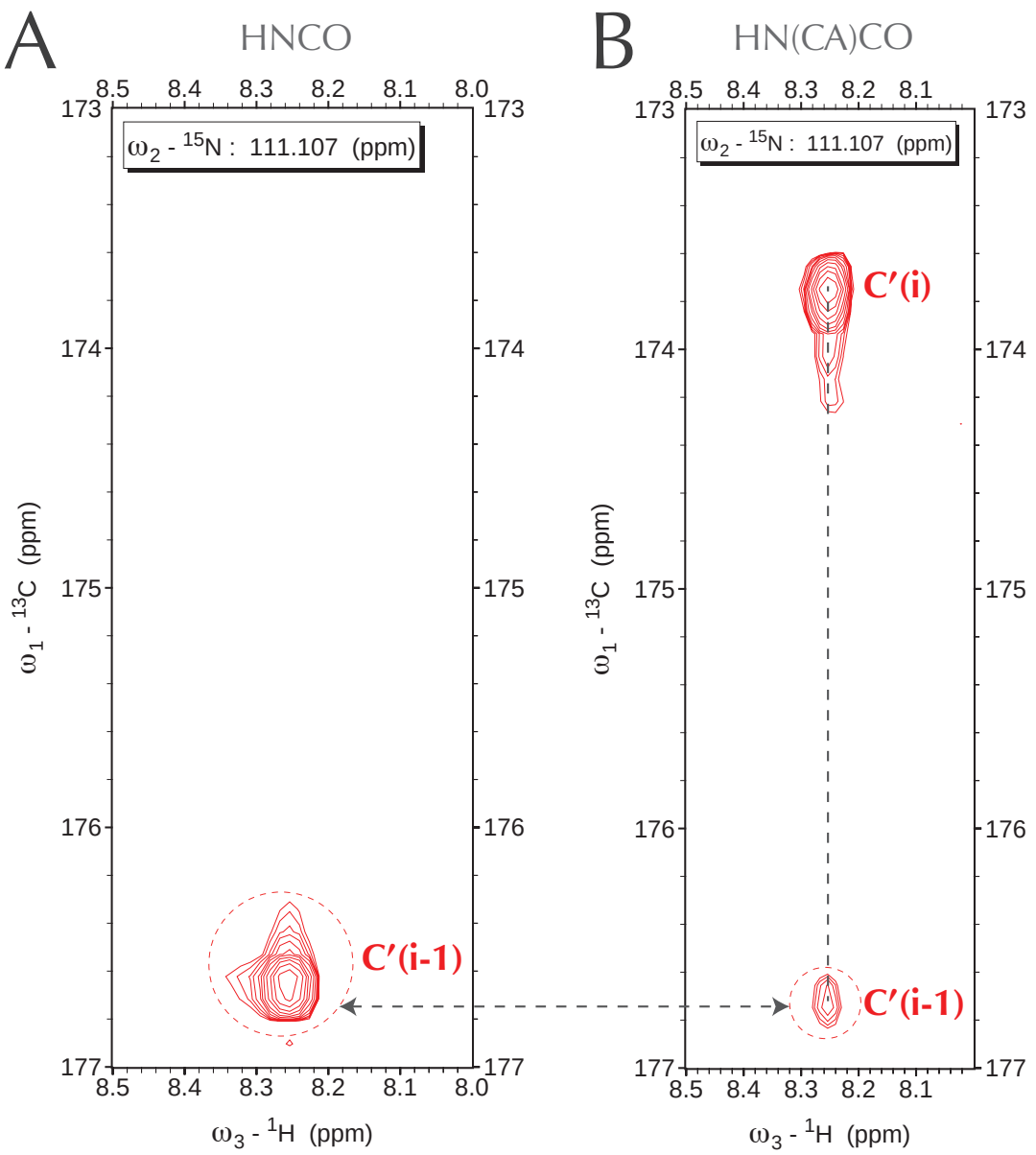

Figure 19: Extract from HNCO and HN(CA)CO experiments performed on the NTD KorB. Shown are the slices from two-dimensional planes for HNCO and HN(CA)CO experiments recorded on the NTD KorB, at ^{15}N chemical shift of 111.107 ppm. The amide proton ($^{1}H^{N}$) chemical shifts are on the X-axis and CO chemical shifts are on the Y-axis. The dashed lines represent the resonances associated with residue *i-1* / *i* and the connection between the two experiments. **(A)** HNCO experiment informs only about the CO chemical shift for the residue *i-1* whereas **(B)** HN(CA)CO informs about two associated CO resonances, one of them is CO for the residue *i-1* at the same chemical shift as observed in HNCO and the other one corresponding to the CO of residue *i*. The peak in the HNCO helps in making the distinction between the two resonances in the HN(CA)CO as shown by the horizontal dashed line. Each dimension for both the experiments was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID. To obtain the spectra shown above, the spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions.

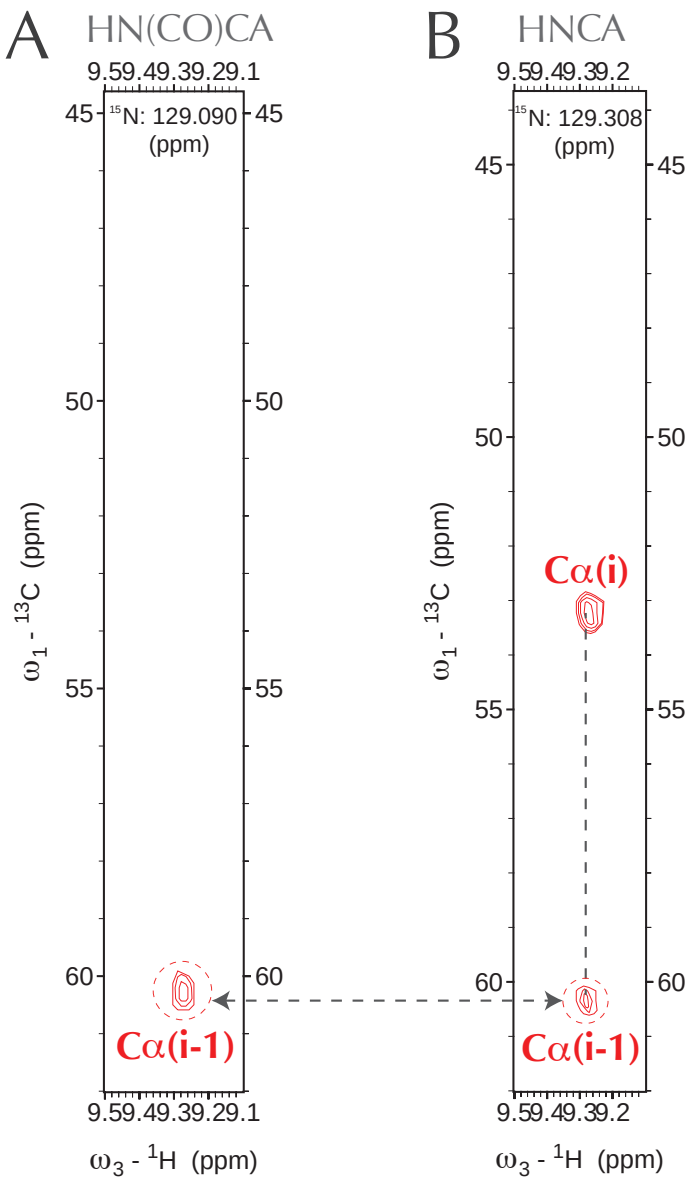

Figure 20: Extract from HN(CO)CA and HNCA experiments performed on the NTD KorB. Shown are the slices from two-dimensional planes for HN(CO)CA and HNCA experiments recorded on the NTD KorB, at ^{15}N chemical shift of 129.090 ppm. The amide proton ($^{1}H^{N}$) chemical shifts are on the X-axis and $^{13}C\alpha$ chemical shifts are on the Y-axis. The dashed lines represent the resonances associated with residue *i-1*/*i* and the connection between the two experiments. **(A)** HN(CO)CA experiment informs only about the $^{13}C\alpha$ chemical shift for the residue *i-1* whereas **(B)** HNCA informs about two associated Cα resonances, one of them is $^{13}C\alpha$ for residue *i-1* at the same chemical shift as observed in HN(CO)CA and the other one corresponding to the $^{13}C\alpha$ of residue residue *i*. The peak in the HN(CO)CA helps in making the distinction between the two resonances in the HNCA as shown by the horizontal dashed line. Each dimension for both the experiments was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID. To obtain the spectra shown above, the spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions.

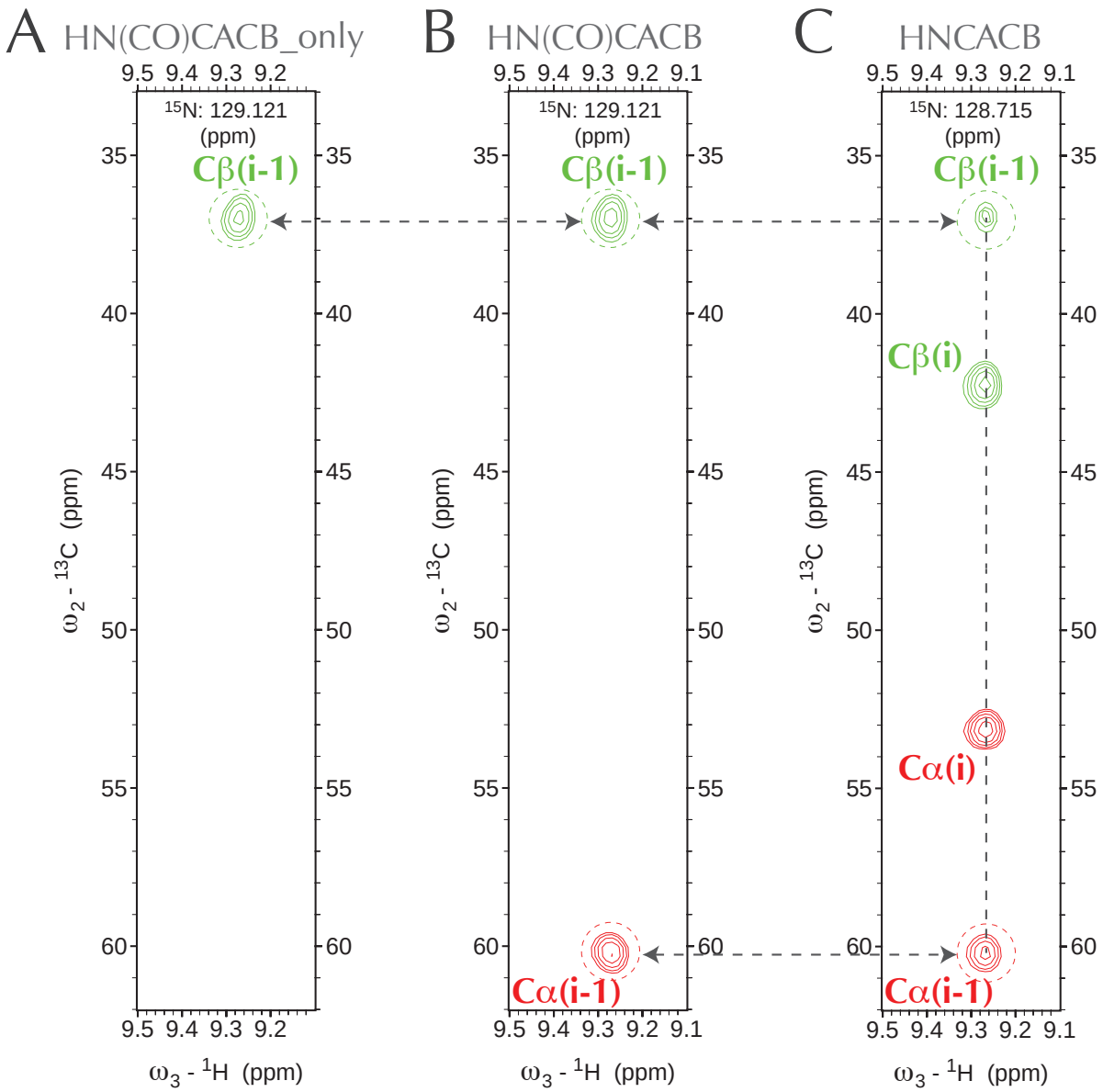

Figure 21: Extract from HN(CO)CACB and HNCACB experiments performed on the NTD KorB. Shown are the slices from two-dimensional planes for HN(CO)CACB and HNCACB experiments recorded on the NTD KorB, at ^{15}N chemical shift of 129.121 ppm. The amide proton ($^{1}H^{N}$) chemical shifts are on the X-axis and $^{13}C\alpha/\beta$ chemical shifts are on the Y-axis. The dashed lines represent the resonances associated with residue *i-1*/*i* and the connection between the two experiments. **(A)** HN(CO)CACB_only experiment informs only about the $^{13}C\beta$ chemical shift for the residue *i-1*. **(B)** HN(CO)CACB experiment informs about the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts for the residue *i-1* whereas **(C)** HNCACB informs about four associated resonances, a) $^{13}C\alpha$ and b) $^{13}C\beta$ for the residue *i-1*, at the same chemical shifts as observed in HN(CO)CACB, the other two shifts *viz.* corresponding to the c) $^{13}C\alpha$ and d) $^{13}C\beta$ of residue residue *i*. The peak in the HN(CO)CACB helps in making the distinction between the two resonances in the HNCACB as shown by the horizontal dashed line. Each dimension for both the experiments was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID. To obtain the spectra shown above, the spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions.

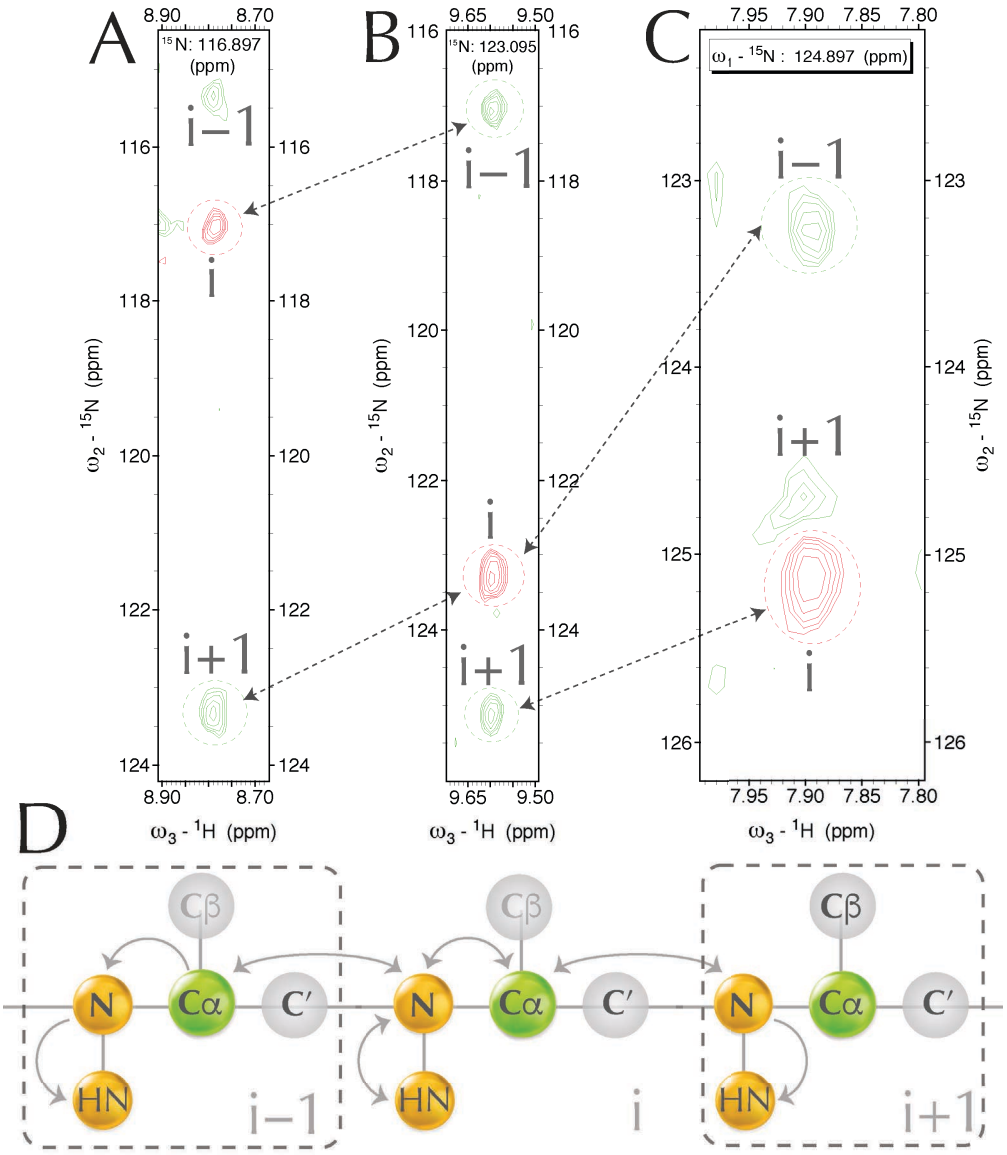

Figure 22: Extract from HNN experiment performed on the NTD KorB. This experiment is particularly useful to perform a sequential walk for backbone assignment. Shown are the slices of two-dimensional planes of the HNN experiment for the residue *i-1*, *i*, *i+1*. The proton (^{1}H) chemical shifts are on the X-axis and ^{15}N chemical shifts are on the Y-axis. **(A)** Slice of two-dimensional planes at ^{15}N (ω_1 - 116.897 ppm) plane in the HNN experiment. **(B)** Slice of two-dimensional planes at ^{15}N (ω_1 - 123.095 ppm) plane in the HNN experiment. **(C)** Slice of two-dimensional planes at ^{15}N (ω_1 - 124.897 ppm) plane in the HNN experiment. **(D)** Magnetisation transfer pathway is shown for an HNN experiment. Within **A, B, and C**, three peaks are observed, each indicating an NH peak for the residue *i-1*, *i* and *i+1*. The connectivity between the slices are shown with dashed arrows. Each dimension for both the experiments was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID. To obtain the spectra shown above, the spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions.

4.4.1 NMR peak assignment for NTD KorB

4.4.1.1 Manual backbone assignment

The backbone assignment of the NTD KorB was carried out manually, with the sequential walk methodology that uses the chemical shifts and the sequential connectivities of the residues to yield information about the type of the residue and the residue's position in the sequence. To assign the chemical shifts of the NTD KorB backbone residues, a ^{1}H-^{15}N HSQC spectrum and several 3D spectra including HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and HBHA(CO)NH were used in the sequential walk methodology. Out of these spectra, the primary information is contained in HNCACB and HN(CO)CACB. The peaks for the residue *i* and residue *i-1* in HNCACB are discerned with the help of the complementary experiment HN(CO)CACB. Figure 23 shows the representative sequential walk methodology for backbone assignment. Further, the backbone sequential walk was corroborated with the help of HNN experiment.

4.4.1.2 Fragment assigning

During the backbone sequential walk, NMRFAM SPARKY's inbuilt 'reposition sequence' extension was initially used to a) match, b) connect and c) assign connected residues for NTD KorB. Since this proved challenging in assigning all the residues of the NTD KorB, the connected residues in the protein sequence were corroborated with the help of another NMRFAM SPARKY extension specifically developed with intrinsically disordered proteins (IDPs) in mind. ncIDP-assign (Tamiola and Mulder, 2011) is an effective plugin for NMRFAM SPARKY that facilitates in sequentially assigning protein residue resonances for IDPs.

The ncIDP-assign (neighbour-corrected IDP chemical shift assignment, version 1.2b) consists of an ncIDP repositioner and an ncIDP spin graph and the combination of the

two tools drastically reduces the degree of protein assignment ambiguities. ncIDP-assign uses a new random-coil chemical shift library of IDPs and on the basis of matching tripeptides the plugin predicts the chemical shifts of a queried protein (Tamiola et al., 2010). Thus the plugin effectively matches a set of connected resonances to the appropriate position in the protein sequence.

Using the ncIDP-assign plugin in the NMRFAM SPARKY program, the chemical shifts of the NTD KorB were assigned and out of 130 residues, 105 (80%) residues were assigned. The sequential walk for NTD KorB often breaks because there are many proline residues, 10 proline out of 130 residues. However, in most cases, the neighbouring residues to prolines have been assigned. Further, data from the HNN experiment were used to walk across prolines and confirm the backbone assignment. The residues 1-23 in the NTD KorB construct come from the expression vector (N-terminal) and residues 24-130 correspond to the residues 31-137 of the wild-type KorB protein. The assigned ^{1}H-^{15}N HSQC spectrum of the NTD KorB showing the residue specific peaks is shown in figure 24. Figure 25 illustrates the residues of the NTD KorB that have been connected using the 3D NMR experiments in the SPARKY program.

The assignment for NTD KorB starts at residue H20 and the residues assigned prior to H20 can be discarded because of the lack of connectivity during the sequential walk (possibly because the peaks for the N-terminal residues 1-20 are missing). Table 16 lists the chemical shifts matched for each protein residue in the NTD KorB. The chemical shifts for connected residues in the NTD KorB were processed through TALOS+ to empirically predict the backbone torsion angles and secondary structural elements in NTD KorB. Further, side-chain assignment of the NTD KorB were considered using the H(CCO)NH-TOCSY, C(CO)NH-TOCSY and HCCH-TOCSY spectra (see section 4.4.1.3).

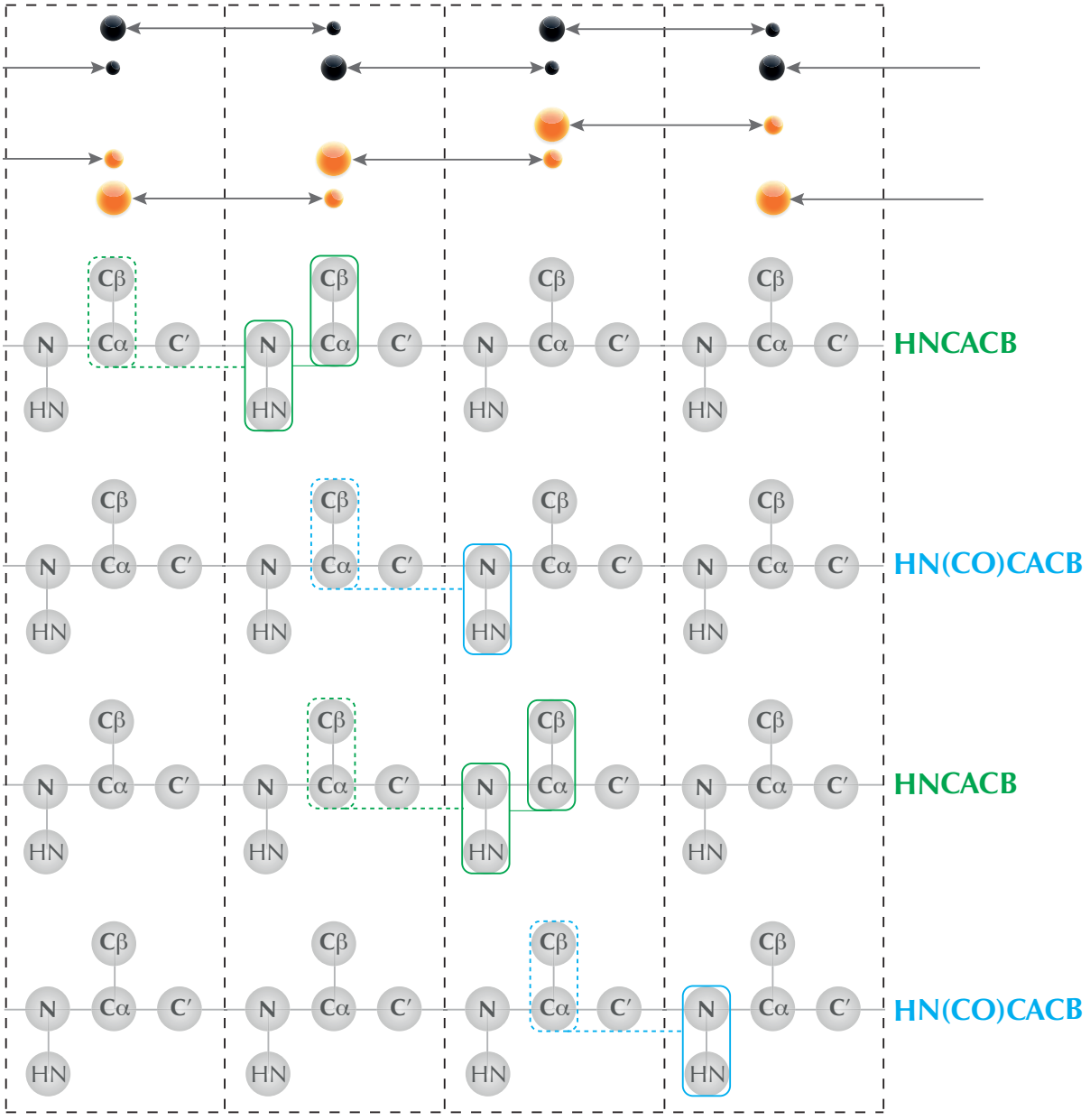

Figure 23: Schematic representing the sequential walk methodology for backbone assignment. The HNCACB experiment gives peaks for the Cα and Cβ nuclei for the residue *i* (shown in solid blue lines) and similarly Cα and Cβ peaks are observed for the preceding residue *i-1* (segmented blue lines). The peaks for the residue *i* and residue *i-1* in HNCACB are distinguished with the help of the complementary experiment HN(CO)CACB. As the HN(CO)CACB experiment gives peaks for the Cα and Cβ chemical shift as a positive phase peak for the residue *i-1*. The figure is adapted from Higman, 2017.

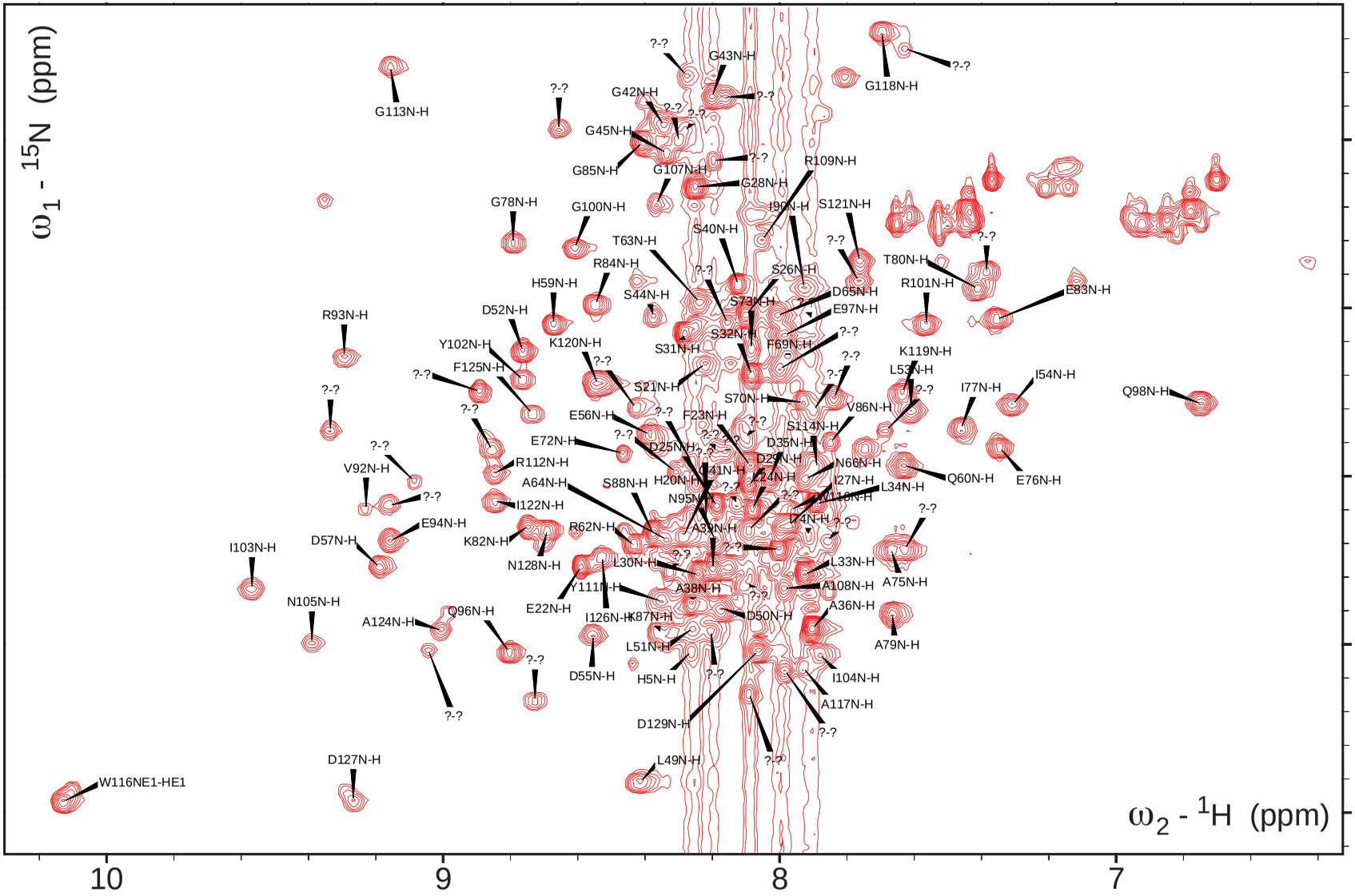

Figure 24: The assigned ^{1}H-^{15}N HSQC spectrum of the NTD KorB. The assigned ^{1}H-^{15}N HSQC spectrum of the NTD KorB with backbone assignments. The spectrum was acquired at 600 MHz for NTD KorB construct (31-137) with an N-terminal tag (23 additional residues). The HSQC spectrum is taken in 20 mM Na phosphate, pH 6, 100 mM NaCl, 0.1 EDTA with 0.5mM NTD KorB at 35 °C.

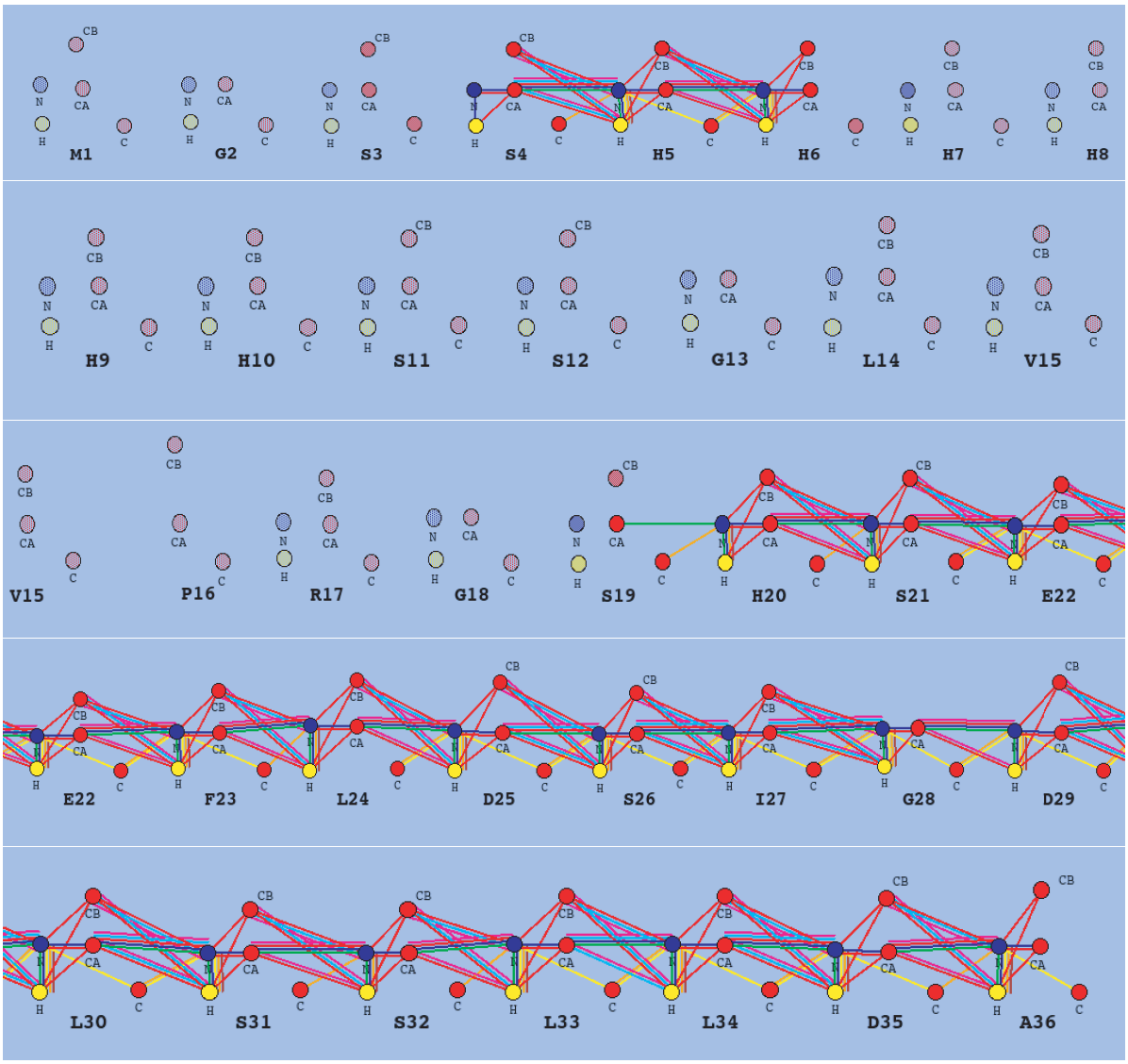

Figure 25: NTD KorB assignment panel.

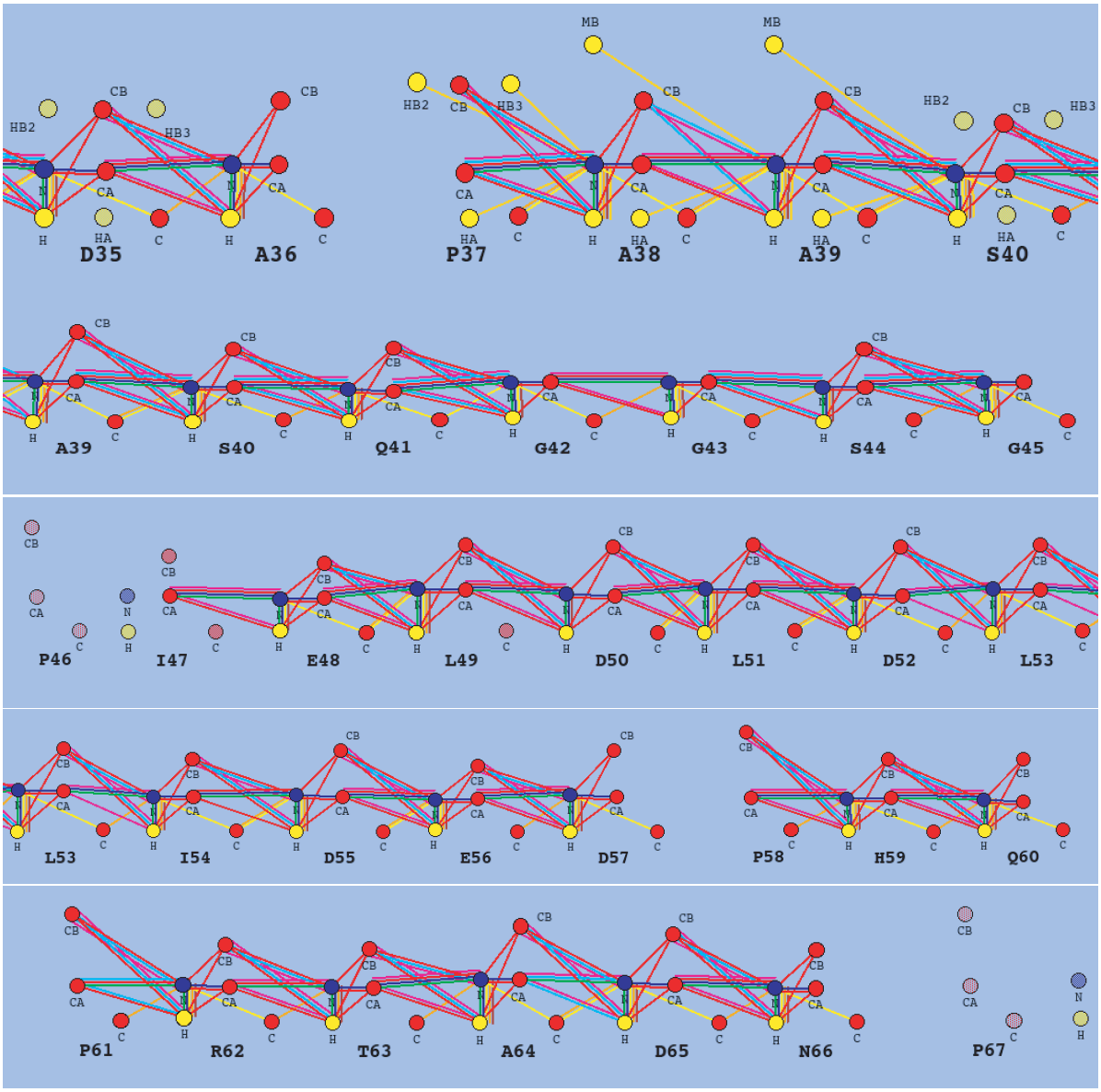

Figure 25 (continued): NTD KorB assignment panel (2).

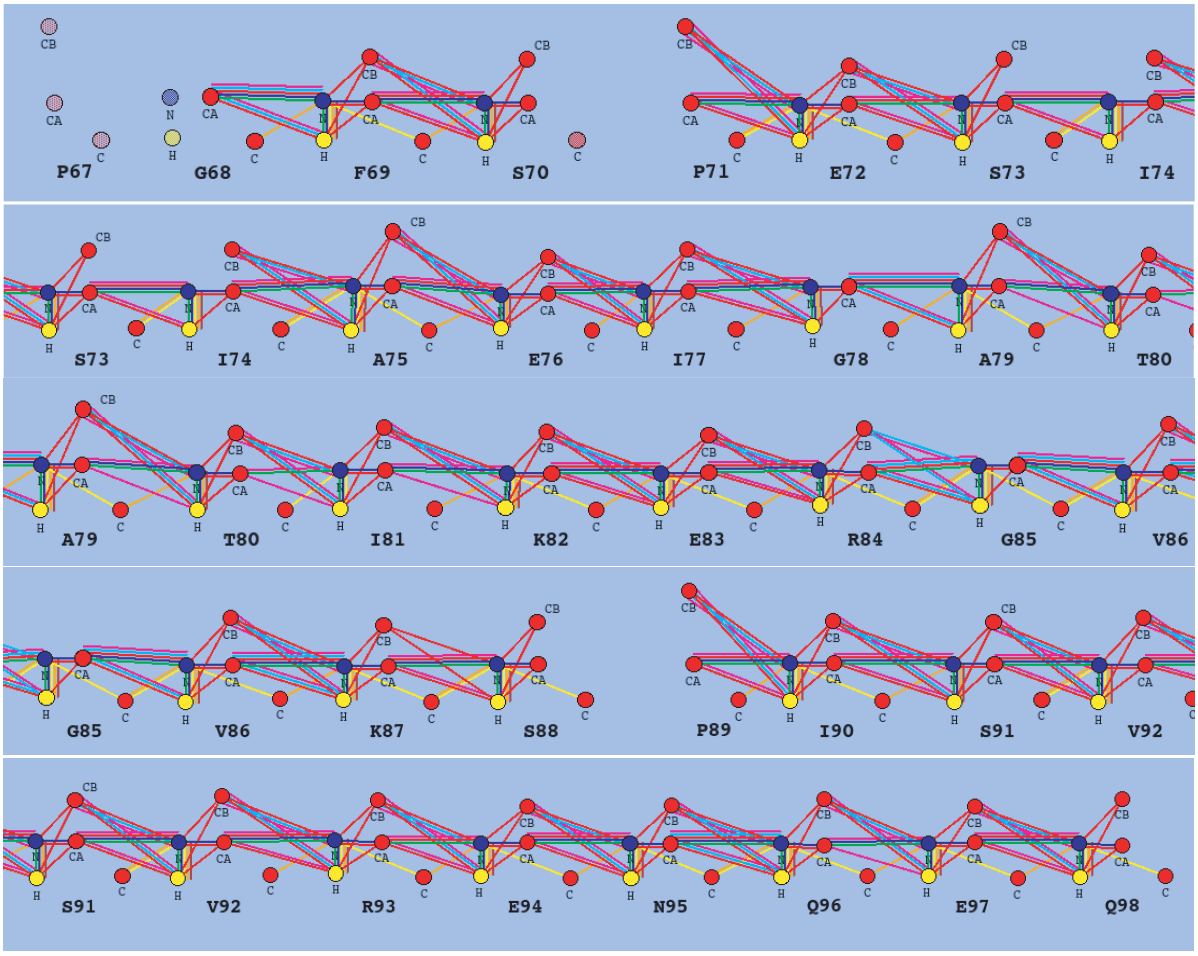

Figure 25 (continued): NTD KorB assignment panel (3).

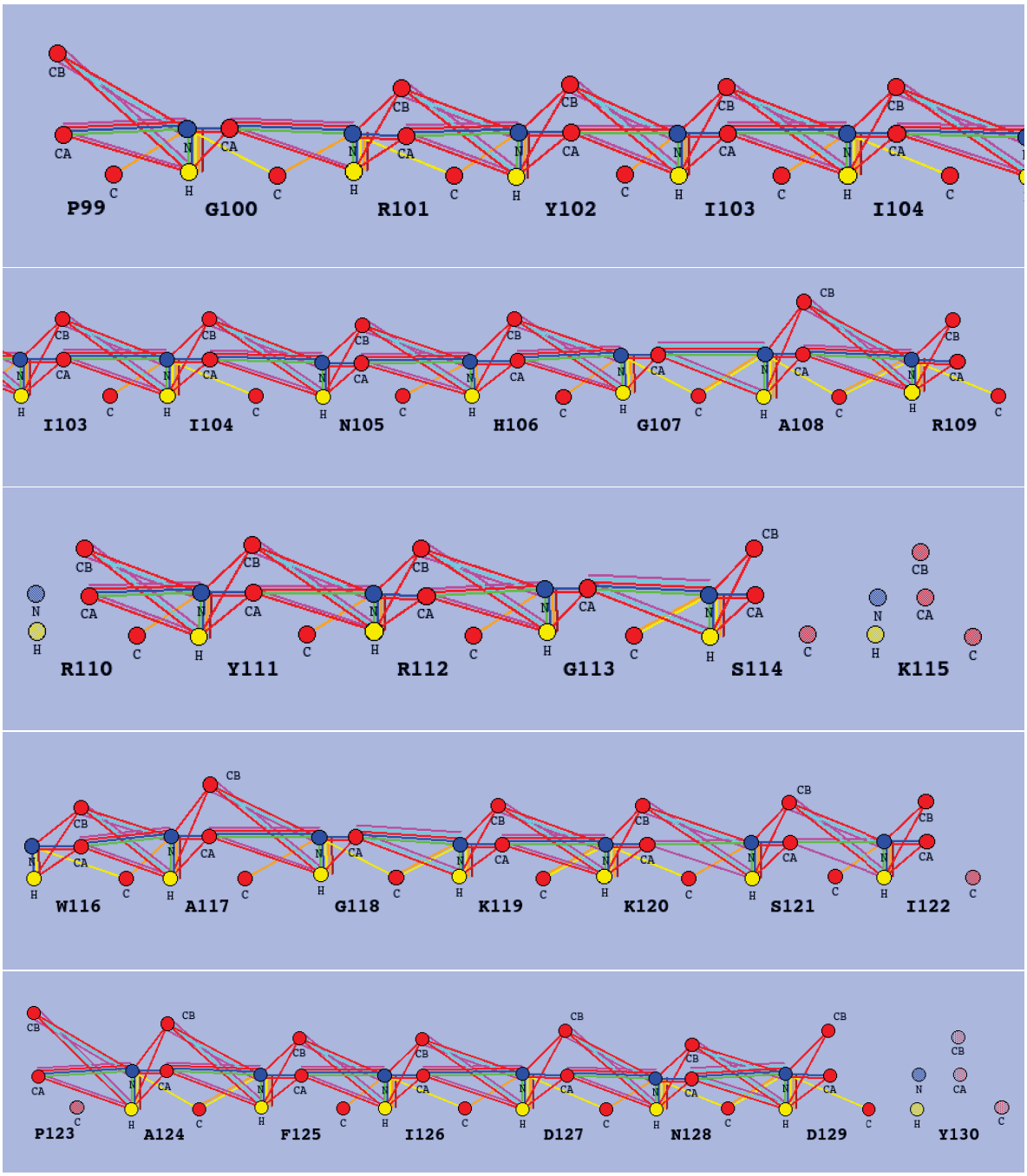

Figure 25 (continued): NTD KorB assignment panel (4).

		C	CA	CB	H	N
2	GLY	172.459	44.319			
4	SER	174.371	61.671	69.855	8.102	115.064
5	HIS	177.239	52.565	19.106	8.266	125.166
10	HIS	174.571	55.733	29.228		
11	SER	174.52	58.568	63.833	8.216	116.64
12	SER	173.748	56.964	29.505	8.584	122.56
19	SER	175.128	53.38	69.855	8.102	115.063
20	HIS	174.571	55.821	29.216	8.25	120.135
21	SER	174.52	58.568	63.833	8.216	116.646
22	GLU	176.251	56.917	29.559	8.585	122.568
23	PHE	175.786	55.821	29.559	8.09	119.645
24	LEU	175.771	54.294	40.95	8.168	122.914
25	ASP	176.204	54.505	41.168	8.175	121.596
26	SER	174.606	57.534	41.168	8.104	115.131
27	ILE	176.699	56.282	38.638	7.978	121.363
28	GLY	173.741	45.376		8.25	111.349
29	ASP	176.644	54.14	41.292	8.094	120.202
30	LEU	178.005	55.752	42.078	8.241	122.832
31	SER	175.146	59.689	63.54	8.284	115.698
32	SER	174.676	58.821	31.897	8.087	116.929
33	LEU	177.402	55.48	41.889	7.93	122.849
34	LEU	176.995	55.302	42.047	7.896	120.789
35	ASP	175.276	53.98	41.213	8.032	120.024
36	ALA	175.348	50.663	18.419	7.91	124.441
37	PRO	176.915	63.185	32.021		
38	ALA	177.784	52.48	19.107	8.261	123.709
39	ALA	177.915	52.621	19.156	8.199	122.721
40	SER	174.692	58.401	63.859	8.129	114.275
41	GLN	176.471	56.07	29.374	8.295	121.705
42	GLY	174.632	45.455		8.341	109.549
43	GLY	174.21	45.257		8.2	108.678
44	SER	174.825	58.425	64.14	8.373	115.195
45	GLY	173.748	44.439		8.308	110.297
46	PRO	175.721	60.808	38.539		
47	ILE	175.877	59.121	38.781	8.065	124.429
48	GLU	175.309	55.871	28.579	8.225	123.827
49	LEU	175.897	52.889	45.058	8.425	128.93
50	ASP	177.404	54.557	40.75	8.171	123.867
51	LEU	179.36	57.906	41.625	8.26	124.443

Table 16: Chemical shift table for NTD KorB.

52	**ASP**	176.898	55.761	40.233	8.775	116.248
53	**LEU**	174.804	54.754	43.221	7.615	117.987
54	**ILE**	173.748	60.625	40.456	7.308	117.794
55	**ASP**	175.707	52.65	44.499	8.571	124.663
56	**GLU**	175.76	56.623	30.706	8.386	118.658
57	**ASP**	175.752	51.846	41.946	9.191	122.596
58	**PRO**	177.453	63.826	31.937		
59	**HIS**	174.261	55.473	29.084	8.649	115.625
60	**GLN**	172.437	53.566	29.182	7.617	119.569
61	**PRO**	177.361	63.241	31.846		
62	**ARG**	176.426	55.049	31.224	8.424	121.755
63	**THR**	174.891	60.862	65.056	8.236	114.71
64	**ALA**	177.327	53.634	18.822	8.336	121.828
65	**ASP**	175.952	54.301	40.983	8.015	115.162
66	**ASN**	173.489	51.568	39.121	7.929	120.019
67	**PRO**	177.731	44.611	31.845		
68	**GLY**	173.97	45.027		8.651	109.445
69	**PHE**	175.404	57.48	39.727	7.974	116.347
71	**PRO**	177.276	63.816	31.855	3.73	23.893
72	**GLU**	176.832	57.188	29.609	8.467	119.213
73	**SER**	175.442	58.552	66.083	8.088	116.113
74	**ILE**	177.195	61.478	35.313	7.922	121.757
75	**ALA**	179.593	53.807	30.907	7.669	122.097
76	**GLU**	178.779	59.369	29.468	7.349	119.108
77	**ILE**	178.106	63.357	37.546	7.466	118.538
78	**GLY**	175.378	47.973		8.799	112.921
79	**ALA**	180.859	55.467	17.903	7.673	124.047
80	**THR**	178.71	48.785	68.659	7.419	114.363
81	**ILE**	178.674	64.757	37.941	8.299	124.526
82	**LYS**	177.536	60.254	33.114	8.741	121.393
83	**GLU**	177.948	58.518	31.089	7.353	115.167
84	**ARG**	176.9	56.105	32.374	8.551	114.849
85	**GLY**	174.034	45.303		8.4	109.905
86	**VAL**	176.221	62.132	32.731	7.862	119.004
87	**LYS**	176.257	56.161	20.613	8.339	124.767
88	**SER**	176.956	63.111	62.168	8.372	121.567
89	**PRO**	177.61	62.827	31.446		
90	**ILE**	173.078	61.208	38.108	7.943	114.371
91	**SER**	175.486	56.357	64.325	8.058	112.247
92	**VAL**	172.464	58.43	35.614	9.237	120.969
93	**ARG**	176.21	52.458	33.316	9.302	116.389

Table 16 (continued): Chemical shift table for NTD KorB (2).

94	**GLU**	175.886	57.363	30.254	9.159	121.825
95	**ASN**	176.043	52.259	38.411	8.188	121.473
96	**GLN**	176.87	57.919	28.55	8.804	125.124
97	**GLU**	176.44	57.073	30.745	7.992	115.763
98	**GLN**	170.027	52.057	29.704	6.752	117.755
99	**PRO**	177.818	63.738	31.846		
100	**GLY**	172.954	44.705		8.486	111.249
101	**ARG**	176.217	54.97	31.976	7.567	115.408
102	**TYR**	175.067	57.32	42.865	8.763	117.035
103	**ILE**	175.926	59.206	40.751	9.463	122.601
104	**ILE**	175.343	64.338	38.784	7.912	124.683
105	**ASN**	176.141	55.361	35.697	9.404	124.872
106	**HIS**	177.007	60.282	37.12	7.903	117.754
107	**GLY**	174.06	45.413		8.368	111.882
108	**ALA**	178.088	52.691	19.294	7.992	123.247
109	**ARG**	174.682	62.248	69.708	8.058	112.984
110	**ARG**	179.239	59.406	30.311	7.802	119.081
111	**TYR**	176.056	48.415	38.958	8.365	123.612
112	**ARG**	180.847	60.123	29.809	8.865	119.796
113	**GLY**	174.169	47.2		9.171	107.741
114	**SER**	177.818	63.771	62.631	7.902	119.58
115	**LYS**	176.239	28.582		8.614	113.218
116	**TRP**	175.876	56.721	30.227	7.967	120.93
117	**ALA**	176.63	52.749	19.424	7.929	125.537
118	**GLY**	175.624	45.726	33.602	7.698	106.778
119	**LYS**	175.783	54.708	32.569	7.638	117.448
120	**LYS**	177.055	56.302	33.602	8.525	117.136
121	**SER**	171.747	56.772	66.751	7.759	113.447
124	**ALA**	175.581	50.615	23.709	9.003	124.381
125	**PHE**	175.002	56.86	41.424	8.734	118.035
126	**ILE**	175.42	60.274	37.242	8.518	122.24
127	**ASP**	175.018	53.253	42.554	9.269	129.5
128	**ASN**	174.702	53.357	39.812	8.692	121.587
129	**ASP**	180.592	55.939	42.558	8.066	125.189

Table 16 (continued): Chemical shift table for NTD KorB (3).

After assigning the residues for the NTD KorB, TALOS+ was used to empirically predict the backbone torsion angles and secondary structure from chemical shift assignments. The secondary structure of the NTD KorB was also predicted with Jpred and Psipred (using the protein sequence). The alignment of NTD KorB secondary structures obtained from TALOS+, Jpred and Psipred is given in figure 26. TALOS+ predicts two helices, α1, in the residue range of 72-83 and α2, towards the C-terminus of the NTD KorB (105-113). Similarly, the two α-helices predicted by both Jpred and Psipred were in agreement with the TALOS+ predicted helices. But the TALOS+ predicted helices are a residue shorter in length in comparison to helices predicted by Jpred and Psipred. Both Jpred and Psipred predict four β-strands around residue 46-49, 87-94, 101-106 and 123-127. From TALOS+, the region of 122-126 aa is in the β-strand conformation but the length of the strand is shorter than the β-strand predicted by Jpred and Psipred.

Figure 26: Comparison of the NTD KorB secondary structure from TALOS+, Jpred and Psipred. TALOS+ (Shen et al., 2009), a web-based tool was used to empirically predict secondary structure for NTD KorB from chemical shift assignments of the protein. Both Jpred (Drozdetskiy et al., 2015) and Psipred (Buchan et al., 2013) are sequence based protein secondary structure prediction servers. NTD KorB is partially folded with two major helices (α1 and α2) predicted by TALOS+, Jpred and Psipred. α-helix (H) and β-strand (E) are represented with red and blue respectively.

4.4.1.3 Side-chain assignment

Residue connectivities satisfied from ncIDP-assign plugin were considered for further side-chain assignment using the H(CCO)NH-TOCSY, C(CO)NH-TOCSY and HCCH-TOCSY. Both H(CCO)NH-TOCSY and C(CO)NH-TOCSY in conjunction with HCCH-TOCSY were utilised during the assignment process and were useful for confirming Hydrogen and Carbon side-chain assignments. HCCH-TOCSY spectrum is particularly utilised for side-chain assignment and is useful for incorporating all the backbone assignment information. Figure 27 shows the extract from the HCCH-TOCSY experiment performed on the NTD KorB. At each carbon frequency of a given residue (Cα, Cβ, Cγ, Cδ...), all the hydrogen atoms attached to all the carbon atoms in the same side chain are observed. Eventually, the side-chain peaks were assigned using the HCCH-TOCSY as the spectrum contains chemical shift information about both Hydrogen and Carbon nuclei.

4.4.1.4 NOESY assignment

NOESY experiments are crucial as the information from the NOESY spectra are used to gather restraints for NMR structure calculations. Ideally in a NOESY spectra, all the Hydrogen atoms within a range of 5-6 Å should give rise to peaks in the spectrum. Notably, this results in ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC spectra that are notoriously crowded in nature. In a single strip for both the experiments (^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC), multiple peaks should be observed. Figure 28 shows a slice of the ^{15}N-NOESY-HSQC experiment performed on the NTD KorB. For assignment of NOESY spectra, the AUDANA (Automated Database-Assisted NOE Assignment) algorithm inside the Ponderosa program was employed (Lee et al., 2016b). The algorithm automates the assignment of 3D NOESY spectra and helps to determine the 3D structure of a protein from NMR data.

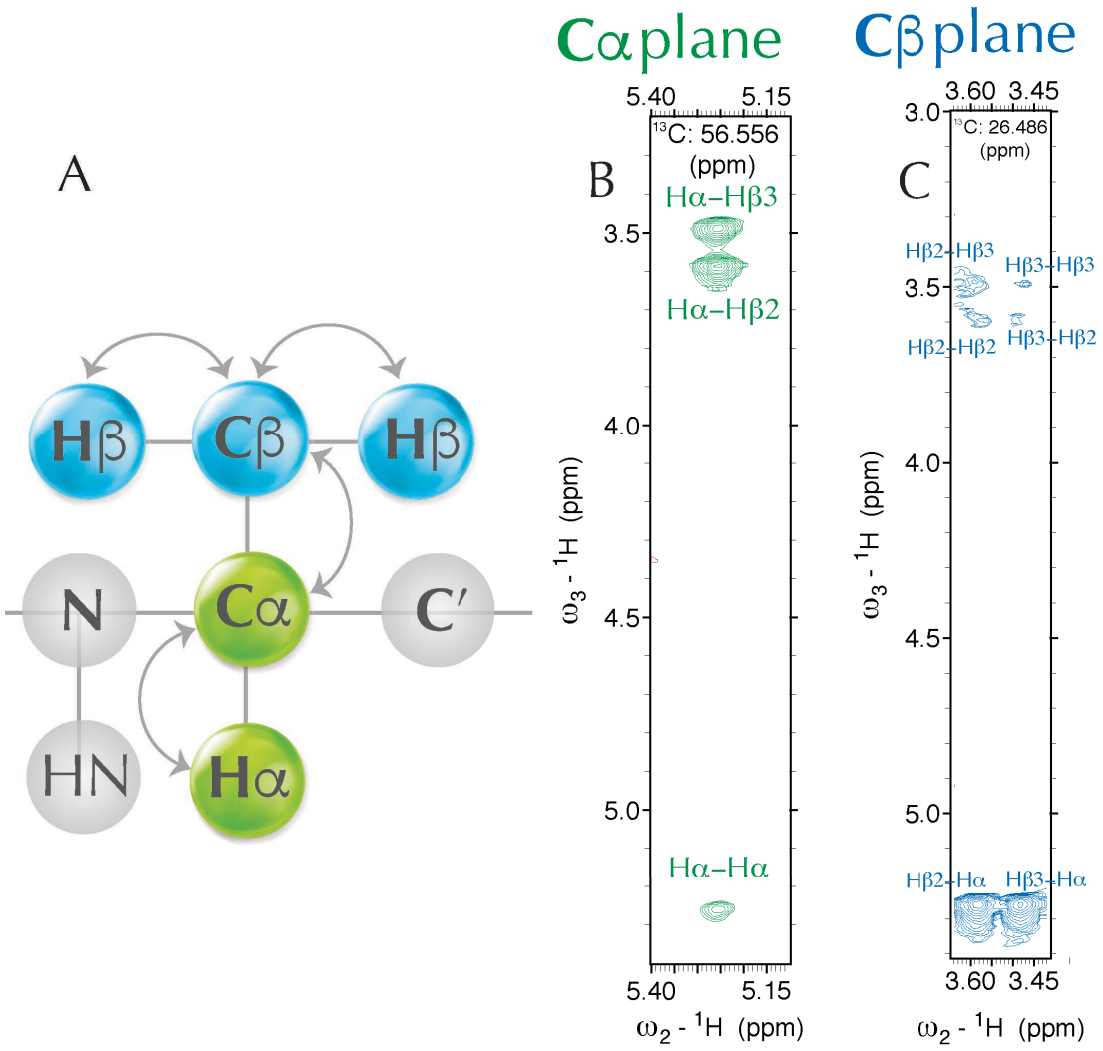

Figure 27: Extract from HCCH-TOCSY performed on the NTD KorB. This experiment is particularly used for side-chain assignment and is useful for incorporating all the backbone assignment information. At each carbon frequency (Cα, Cβ, Cγ, Cδ...), all the hydrogen atoms attached to all the carbon atoms in the same side chain are observed. Shown are the extract of two-dimensional planes for ^{13}Cα plane and ^{13}Cβ plane in the HCCH-TOCSY experiment. The proton (^{1}H) chemical shifts are both on the X-axis and Y-axis. **(A)** Magnetisation transfer pathway is shown for a residue with side-chain containing only Cα and Cβ for HCCH-TOCSY experiment. **(B)** Slice of two-dimensional planes for ^{13}Cα (chemical shift at 56.556 ppm) plane in the HCCH-TOCSY experiment. At the bottom of the Cα plane, a single peak is observed that represents Hα (attached to the Cα) and two other peaks Hβ2 and Hβ3 (attached to the Cβ). **(C)** Slice of two-dimensional planes for ^{13}Cβ (chemical shift at 24.486 ppm) plane in the HCCH-TOCSY experiment. At the bottom of the Cβ plane, two peaks are observed a) Hβ2-Hα and Hβ3-Hα. On the other side, there are cross peaks for Hβ2 and Hβ3 and peaks for Hβ2-Hβ3 and Hβ3-Hβ2. Each dimension for both the experiments was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID. To obtain the spectra shown above, the spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions.

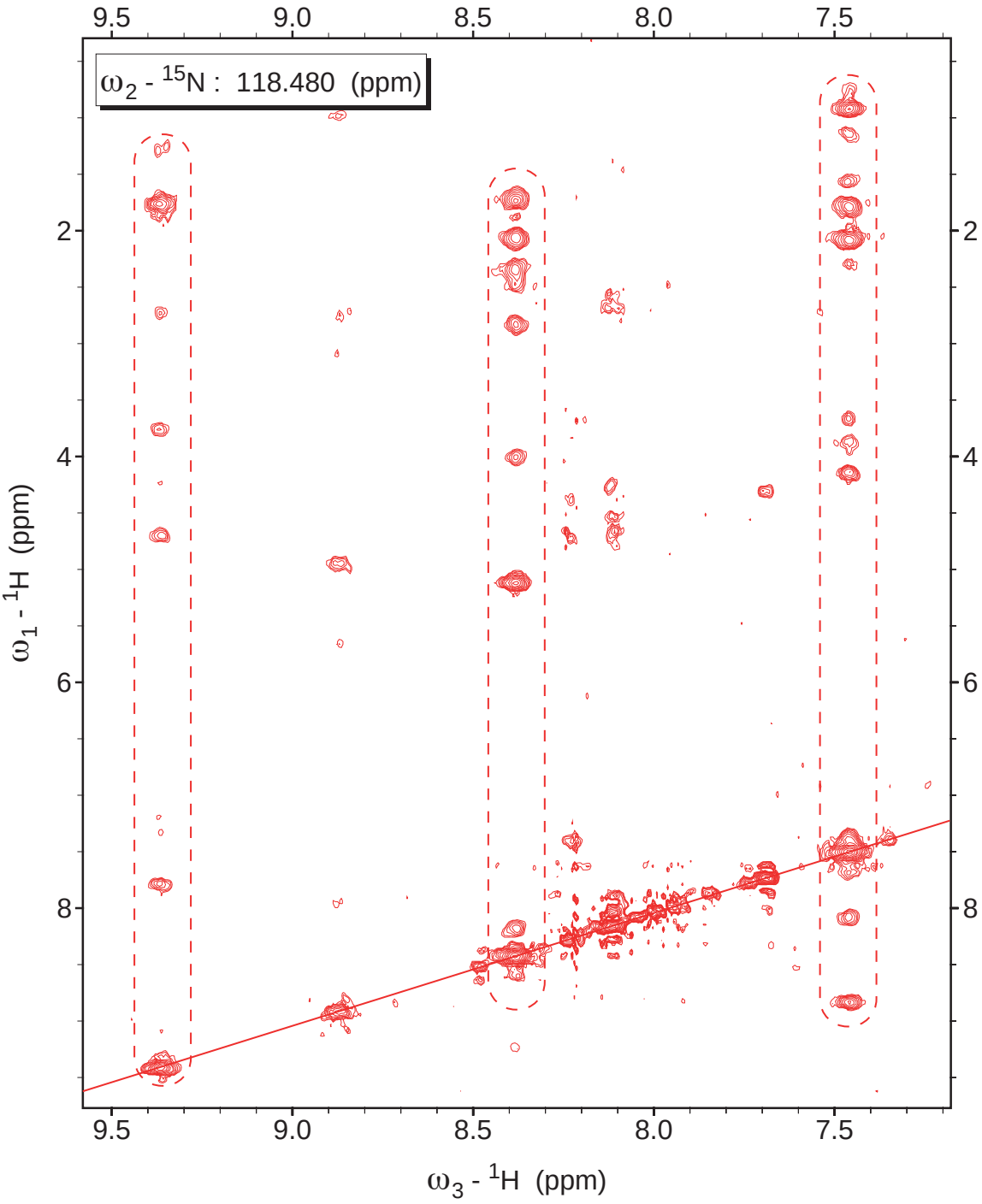

Figure 28: Extract of a ^{15}N-NOESY-HSQC performed on the NTD KorB. NOESY spectra are particularly used to obtain distance restraints for NMR structure calculations. Shown is a slice of a two-dimensional ^{1}H plane with ^{15}N chemical shift of 118.480 ppm in the ^{15}N-NOESY-HSQC experiment. Each strip (shown with segmented line box) contains NOEs from one particular NH group to all other Hydrogen atoms in the radius range of 5-6 Å and the solid diagonal line represents amide cross speaks. Each dimension for both the experiments was solvent filtered, apodised with a standard sine-bell window function and zero filled prior to Fourier transformation of the FID. To obtain the spectra shown above, the spectra were phased visually with NMRDraw and polynomial baseline correction was applied in all dimensions.

4.4.2 Structure calculation of the NTD KorB

Determination of protein structure with solution state NMR spectroscopy is labour intensive and computationally demanding. The difficulty in accurately assigning the resonances increases with the introduction of through-space (NOESY) experiments. With NOESY experiments, the total number of peaks depend on multiple factors including the conformation of the protein in solution at a particular pH as well as the length of the protein sequence.

Ponderosa-C/S is a structure determination software package which takes the refined peak lists for backbone and NOE experiments. Figure 29 illustrates the work flow that Ponderosa uses, shown as a flowchart for structure calculation (Lee et al., 2016a, 2011, 2016b, 2014). Ponderosa uses assignment information to pick NOE peaks from 3D NOESY spectra. This picked data is used as input for the AUDANA algorithm to assign the NOEs and for calculating the NMR structure on the NMRFAM-hosted Ponderosa Server.

In order to calculate the NTD KorB structure, the protein sequence, the assigned backbone chemical shifts, corresponding side-chain information and the NOESY spectra were submitted to the Ponderosa program. On the servers, the program automates the assignment of supplied NOE spectra and generates distance constraints, followed by high-temperature molecular dynamics (iterative in nature). During the final step, the standard explicit water refinement method ('constraints only for the final step') was used followed by simulated annealing to generate the water refined NTD KorB ensemble structure. Figure 30 shows the 20 water refined NTD KorB models calculated using Ponderosa. Majority of the NTD KorB is disordered especially the N-terminal of the protein and two α-helices (α-helix 1 and 2) are observed towards the C-terminal of the protein (residues 73-114).

To understand the variability in orientation of the two α-helices (core region) to-

wards the C-terminal of the NTD KorB, the 20 conformers of the ensemble were superimposed and an average 'sausage' view of the ensemble was computed with ENDScript (Robert and Gouet, 2014). Figure 31A represents the α-helices (73-114 aa) towards the C-terminal of the NTD KorB and figure 31B shows the flexibility of the NTD KorB ensemble as represented by the 'sausage' view. The sausage view helps to visualise the variability in position of individual residue relative to each other and the width of the sausage is proportional to the RMSD between different models. The majority of the NTD KorB is disordered with reduced variability observed towards the C-terminal of the NTD KorB protein (table 17).

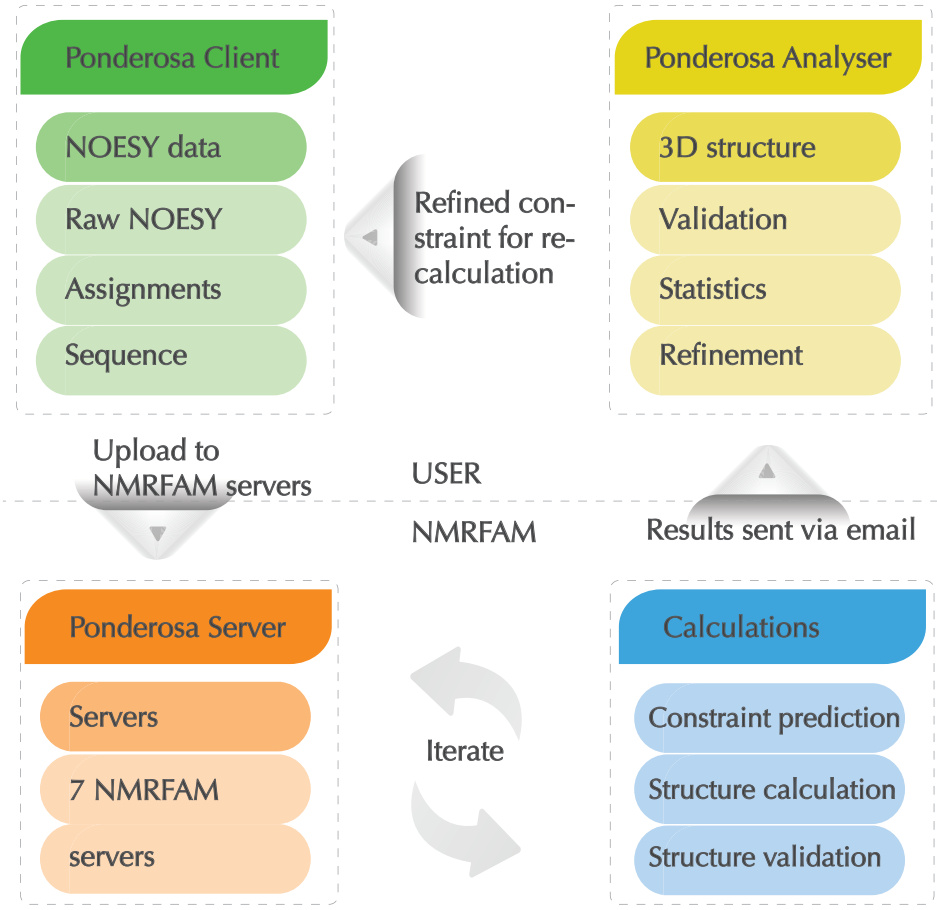

Figure 29: Schematic representing the Ponderosa flowchart for NTD KorB structure calculation. Ponderosa uses backbone and side-chain assignment information to pick NOE peaks from 3D NOESY (^{13}C and ^{15}N) spectra. This picked data is used as input for a series of Automated Database-Assisted NOE Assignments (AUDANA) and calculating the NMR structure. The figure is adapted from Lee et al., 2014.

Figure 30: NTD KorB water refined ensemble structure. Ribbon representation of the NTD KorB ensemble structure (calculated using Ponderosa). The α-helices are shown in blue and the disordered region is shown in grey. Majority of the NTD KorB is disordered especially the N-terminal part and two α-helices (α-helix 1 and α-helix 2) are observed towards the C-terminal of the protein. NTD KorB ensemble structure is superimposed on the α-helix 1 of the protein. NTD KorB ensemble structure is rendered with POV-Ray in VMD.

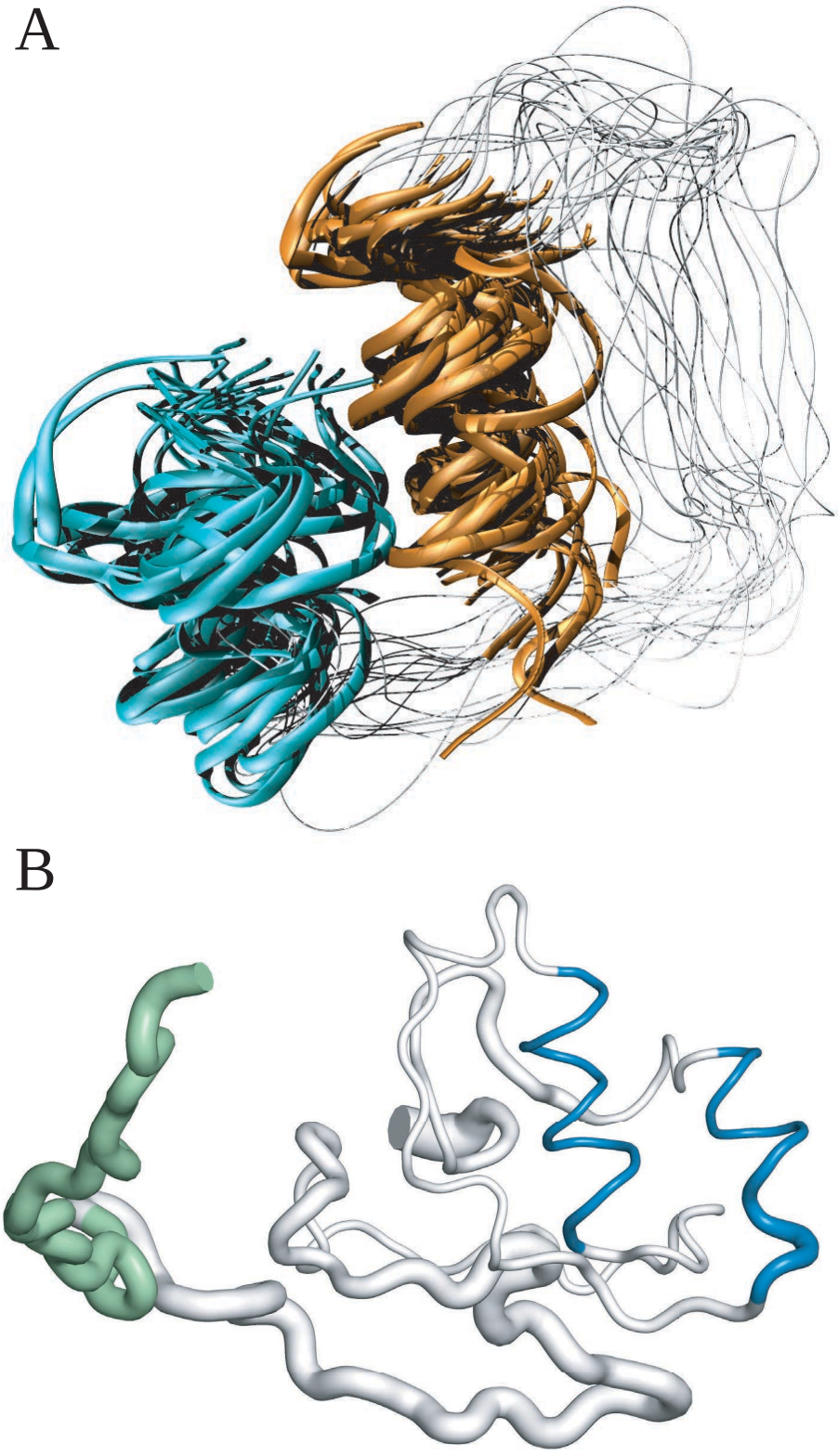

Figure 31: Representation of C-terminal α-helices and flexibility of the NTD KorB ensemble structure. **(A)** Ribbon representation of the two α-helices towards the C-terminal in the NTD KorB ensemble structure. The residue 73-114 of the NTD KorB are superimposed and the structures are rendered with POV-Ray in VMD. **(B)** All 20 conformers of the NTD KorB ensemble are superimposed and an average 'sausage' view of the ensemble is computed with ENDScript (Robert and Gouet, 2014). The sausage view helps to visualise the relative flexibility of individual residue relative to each other. For NTD KorB, the N-terminal tag is shown in green and the two α-helices towards the C-terminal are shown in blue.

4.4.3 NTD KorB ensemble validation

In order to evaluate the structural integrity and stereo-chemical quality of the NTD KorB ensemble structure, multiple analysis tools were employed including MolProbity, RAMPAGE, ProSA-web and Verify3-D.

Verify3-D is a web-based diagnostic tool that examines the reliability of an input protein model by evaluating the suitability of an atomic structure (3D) with the amino acid sequence of the protein (1D). Verify3-D analyses the compatibility of the protein sequence in its current 3D environment by spatially assigning a structural class (α-helix/β-sheet/loop) to its environment. The 3D environment can be from a theoretical model or an experimentally derived structure. The result of the program is a residue-wise Verify3-D profile with an average score for the model structure and an assessment of the model with a comment "pass/warning/fail". Residues are considered reliable with a score ≥ 0.2.

Protein Structure Analysis (ProSA) is a web-based diagnostic tool that examines for possible errors in a protein structure model. ProSA-web is based on statistical analysis of all the protein structures (solved by XRC and NMR) available in the PDB database. The overall model quality is measured by a Z-Score and generally, a model with positive value of Z-Score will indicate errors/problems in the model. The overall quality score for a protein structure model is shown inside a plot that displays the computed score for all the experimentally solved protein chains in the present PDB database versus number of residues in the protein. Figure 32 shows the ProSA plot for NTD KorB and the Z-Score for model8 of NTD KorB ensemble is -2.73.

The Ramachandran plot (see figure 33) of the NTD KorB ensemble shows that 91.2% (2335/2560) of all residues are in favoured regions, 97.6% (2499/2560) of all residues are in allowed regions and 2.4% (61/2560) residues are in outlier region. According to PROSESS, (PROtein Structure Evaluation Suite and Server), NTD KorB (model8)

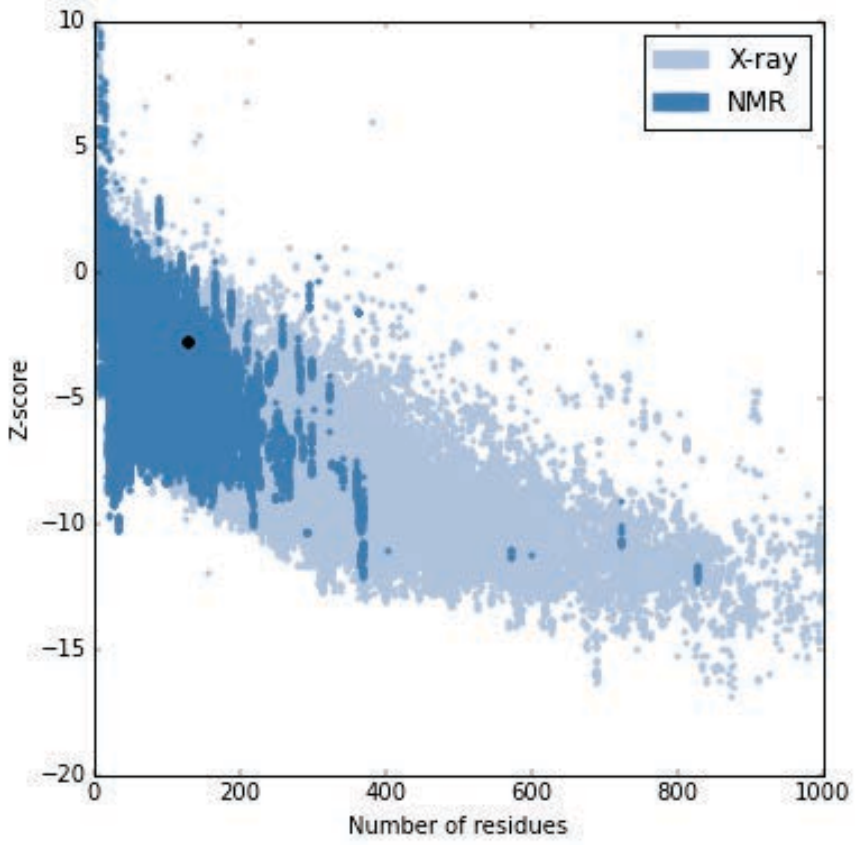

Figure 32: Protein Structure Analysis (ProSA) of NTD KorB. The black dot represents the NTD KorB inside a plot that displays the computed score for all the solved protein chains either with X-ray crystallography or NMR spectroscopy in the present PDB database. The overall model quality is measured by a Z-Score and generally, a model with positive value of Z-Score indicates errors in the model. The Z-Score for NTD KorB conformer was -2.73 (model8 in the NTD KorB ensemble).

consists of 13% α-helix, 1% β-strand, 15% turn and 87% random coil. Further, the NTD KorB ensemble was validated with the Protein Structure Validation Server (PSVS) and the corresponding validation report is given in table 17. Out of the 20 calculated NMR structures for the NTD KorB, the model with best structural characteristics (model8) was selected for MD calculations. The structural integrity for the NTD KorB's model8 is verified with RAMPAGE and ProSA and is described in section 4.5 along with MD calculation results.

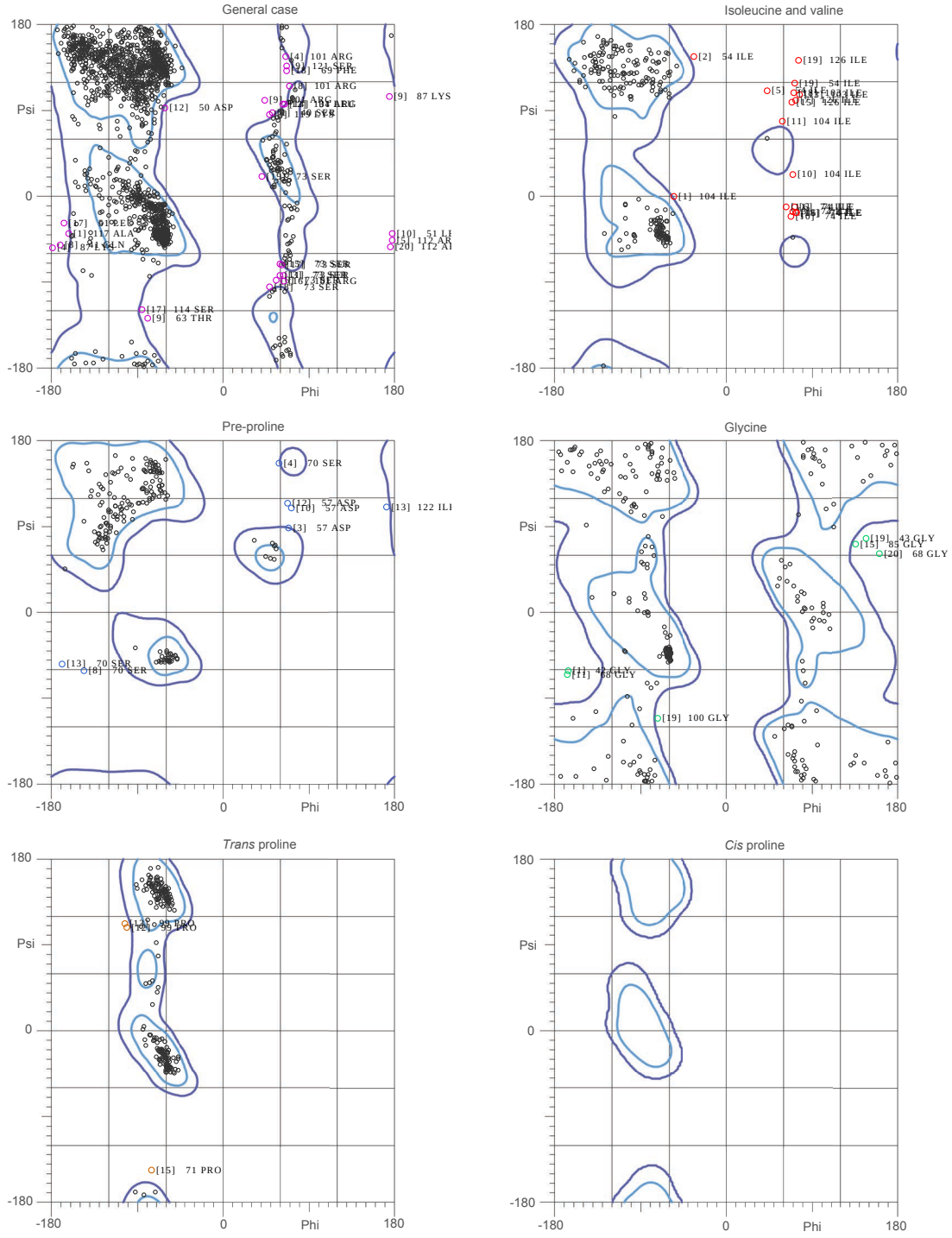

Figure 33: The Ramachandran plots of the NTD KorB ensemble. Overlay of the Ramachandran plots for 20 NTD KorB conformers. Highly favoured regions are shown with light blue contours whilst the allowed regions are represented with dark blue. Disallowed residues are shown with points in magenta and the corresponding model is noted in square brackets. 91.2% (2335/2560) of all residues are in favoured regions and 97.6% (2499/2560) of all residues are in allowed regions and 2.4% (61/2560) residues are in outlier region. The plot is generated with MolProbity. The outlier residues with phi and psi values and the residues that may be wrongly assigned (based on the plot) are listed in appendix on page 225 and page 226 respectively.

Backbone deviation (RMSD, Cα atoms) for all models	
All residues (backbone)	13.9 Å
Ordered residues (74-83 and 108-113)	2.7 Å
Core residues (73-114)	2.9 Å
Bond statistics for all models	
Number of close contacts	14
(within 1.6 Å for H atoms, 2.2 Å for heavy atoms)	
RMS deviation for bond angles:	0.7°
RMS deviation for bond lengths:	0.008 Å
Z-score	
Verify-3D	-5.46
ProsII	-2.69
Procheck (phi-psi)	-1.42
Procheck (all)	-1.36
MolProbity clash score	-0.63
Ramachandran Plot Summary (Molprobity)	
Favoured regions	91.6%
Allowed regions	6.2%
Disallowed regions	2.2%

Table 17: Validation report of NTD KorB from Protein Structure Validation Server (PSVS). NTD KorB structure quality assessment summary by PSVS. It is required that one should resist depositing a structure to the PDB with Z scores $<$ -5 for Procheck scores and Molprobity clash scores. Z scores $>$ -3 for both Procheck and Molprobity clash score are recommended.

4.5 Molecular Dynamics simulation of NTD KorB

4.5.1 Introduction

Molecular dynamics (MD) simulations are computational experiments performed to predict the properties of a system in question, be they physical or chemical. MD simulation is now a widely used approach that exploits the advances in the empirical force fields to study properties of molecular systems. In the last three decades, the development of simulations as a 'computational microscope' has provided a platform to study biological phenomena at atomic (or near-atomic) level (Lee et al., 2009; Perilla et al., 2015). MD simulation experiments are computationally demanding and before running a simulation, the size of biomolecules to be simulated and the time-scale to be covered should be considered (Lee et al., 2009). Figure 34 represents the range of time scales and length scales used for different MD simulation techniques.

MD simulation has emerged as a standard method to explore biological systems at atomic and molecular level. For an MD simulation, there are many approaches to treat interatomic interactions, ranging from detailed quantum mechanical calculations to coarse-grained approaches where some of the atomic details are sacrificed in order to consider the larger size of the system. Within single molecules, quantum mechanical calculations help to study chemical reactions at electronic level. The all-atom or atomistic simulations permit the study of dynamics of small proteins, whilst coarse-grained simulations allow the study of large macromolecular (subcellular organelles) systems (Kmiecik et al., 2016; Perilla et al., 2015).

4.5.2 Results for NTD KorB simulations

In order to investigate the flexibility of NTD KorB, MD simulations were employed. The NTD KorB is majorly disordered, therefore AWSEM simulation package was employed (Davtyan et al., 2012). The AWSEM force-field is useful to simulate protein structures where structural information about that protein is known. On the other hand

simulations can also be run using information solely on the sequence of the protein. In the AWSEM force field, a protein residue is represented with a three-bead model to simulate larger biomolecules for a given amount of time (as compared to atomistic simulations).

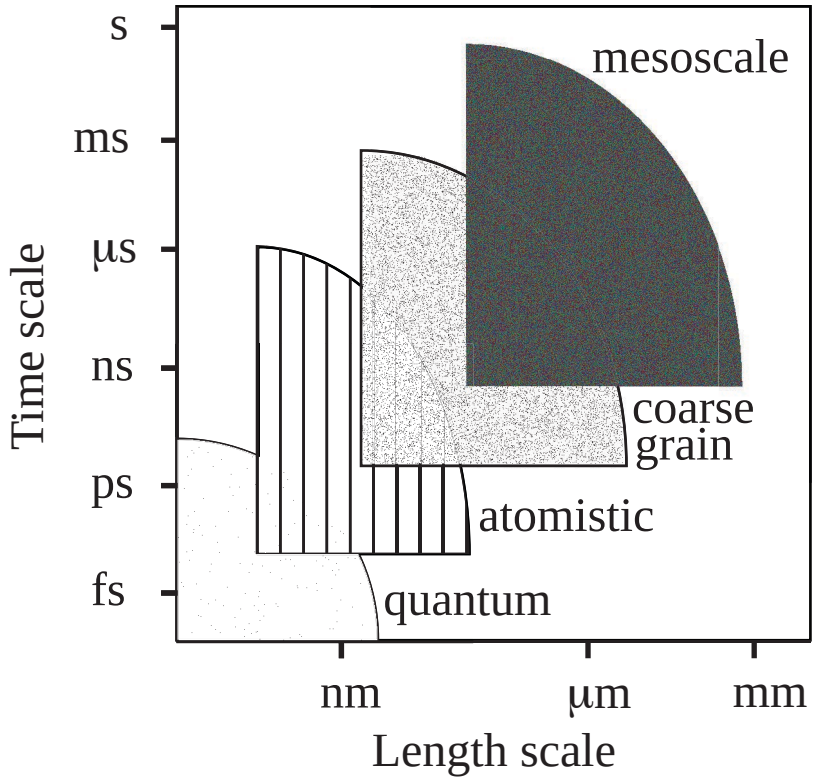

Figure 34: Schematic representation of the time scale and length scale for different MD simulation techniques. The plot shows the approximate ranges of time scales and system sizes (length scale) used for molecular modelling at different resolutions (quantum, all-atom or atomistic, coarse-grained and mesoscale). Within single molecules, quantum mechanical calculations help to study chemical reactions at electronic level. The all-atom or atomistic simulations permit the study of dynamics of small proteins whilst coarse-grained simulations allow for the study of large macromolecular systems (lipid vesicles). The figure is adapted from Kmiecik et al., 2016.

One of the representative conformers of NTD KorB (model8) was used to simulate the protein's motions. Figure 35 shows the Ramachandran plot of model8 of the NTD KorB ensemble. Collectively, 99.2% of NTD KorB's residues are in the allowed region of Ramachandran plot and a single residue has been identified as an outlier. As NMR-derived structural data was available for NTD KorB, so the protein was simu-

lated in the presence and absence of fragment memory. The role of fragment memory is to use experimentally resolved structural information to bias local interactions within a molecule. The fragment memory uses short (9 residues or shorter), overlapping residues in the protein as fragments to bias the simulation, which are called memories and the corresponding energy term in the AWSEM package is called Fragment Memory. The structural knowledge of the NTD KorB ensemble structure was used to bias the simulations and a fragment library was generated containing memories from the 19 NTD KorB models.

During the MD run, the structural information from the NMR-derived ensemble (19 conformers) was used as memories to bias the simulation. A 110 ns MD simulation was run for NTD KorB. The change in the RMSD and R_g of NTD KorB from the starting structure is shown in the figure 36. In the presence of fragment memory, the RMSD appears to increase rapidly in the first 30 nanoseconds and does not increase as much for the rest of the simulation. On the other hand, RMSD is smaller for NTD KorB simulated in the absence of fragment memory. The change in R_g from the starting structure, over the course of the simulation is limited and varies between 25 and 28 Å. The first and final structure of the simulation is overlaid in figure 37. The RMSD (Cα atoms) between the two structures is 16.2 Å. The N-terminal of NTD KorB is highly mobile compared to the C-terminal of the protein. The region spanning the residues 73-113 (core region) for the NTD KorB construct (with tag) includes two α-helices and is the structured region of the protein. The flexibility of the core region was reviewed by superimposing the core region (73-113 aa) from the 110 ns trajectory run with the structure extracted at every 10 ns. Figure 38 shows the superimposition of structured region (core) of NTD KorB. The RMSD (Cα atoms) of the core region is 2.5 Å and the region is observed to exhibit limited motion compared to the N-terminal of NTD KorB.

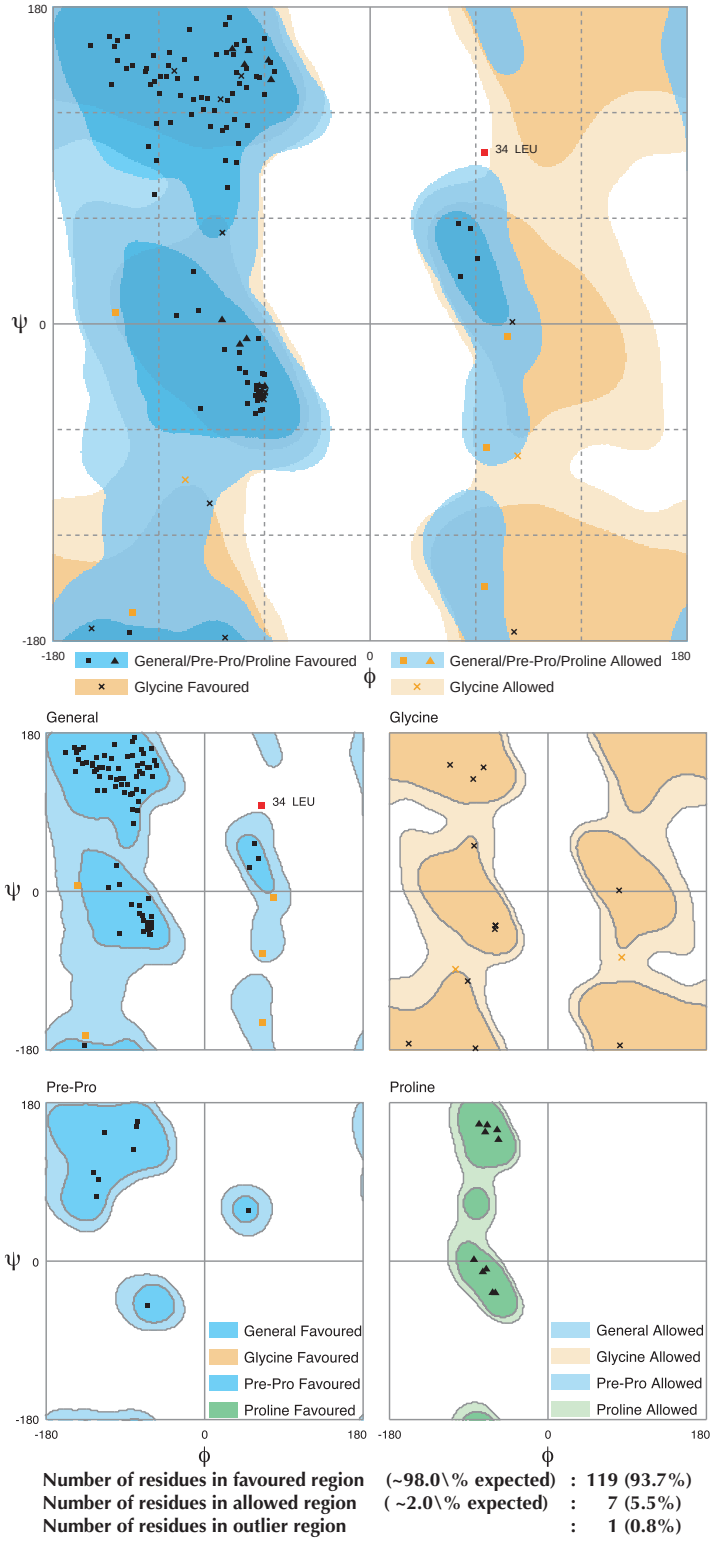

Figure 35: The Ramachandran plot of NTD KorB. (A) The Ramachandran plot of NTD KorB (model8) showing the protein residues. The number of residues in favoured region are 119/130 (93.7%), number of residues in allowed region are 7/130 (5.5%) and a single residue (0.8%) is in outlier region. **(B)** Individual Ramachandran plots showing General, Pre-Proline, Glycine and Proline plots with the possible ψ and ϕ dihedral angles and allowed and disallowed regions of complete amino acids of NTD KorB.

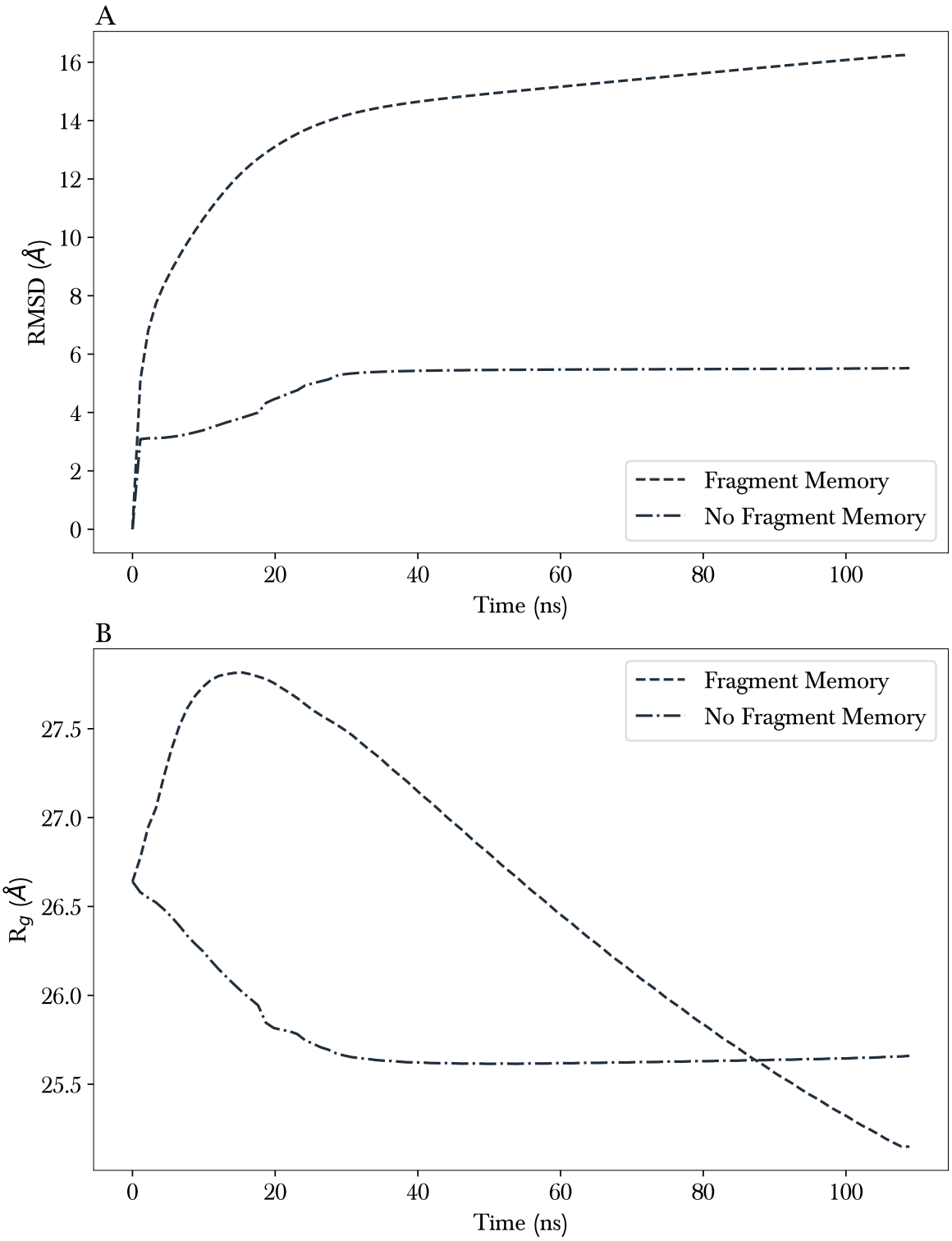

Figure 36: The change in RMSD and R_g of NTD KorB during the 110 ns run. (A) The RMSD (Cα atoms) of NTD KorB from the starting structure in the presence and absence of fragment memory are plotted against time. With fragment memory, the RMSD appears to increase rapidly in the first 30 ns and does not increase as much for the rest of the simulation, whilst the RMSD is smaller in the absence of fragment memory. **(B)** The radius of gyration (R_g) of NTD KorB in the presence and absence of fragment memory are plotted against time. The change in R_g is limited, both in the presence and absence of fragment memory and varies between 25-28 Å for the 110 ns simulation.

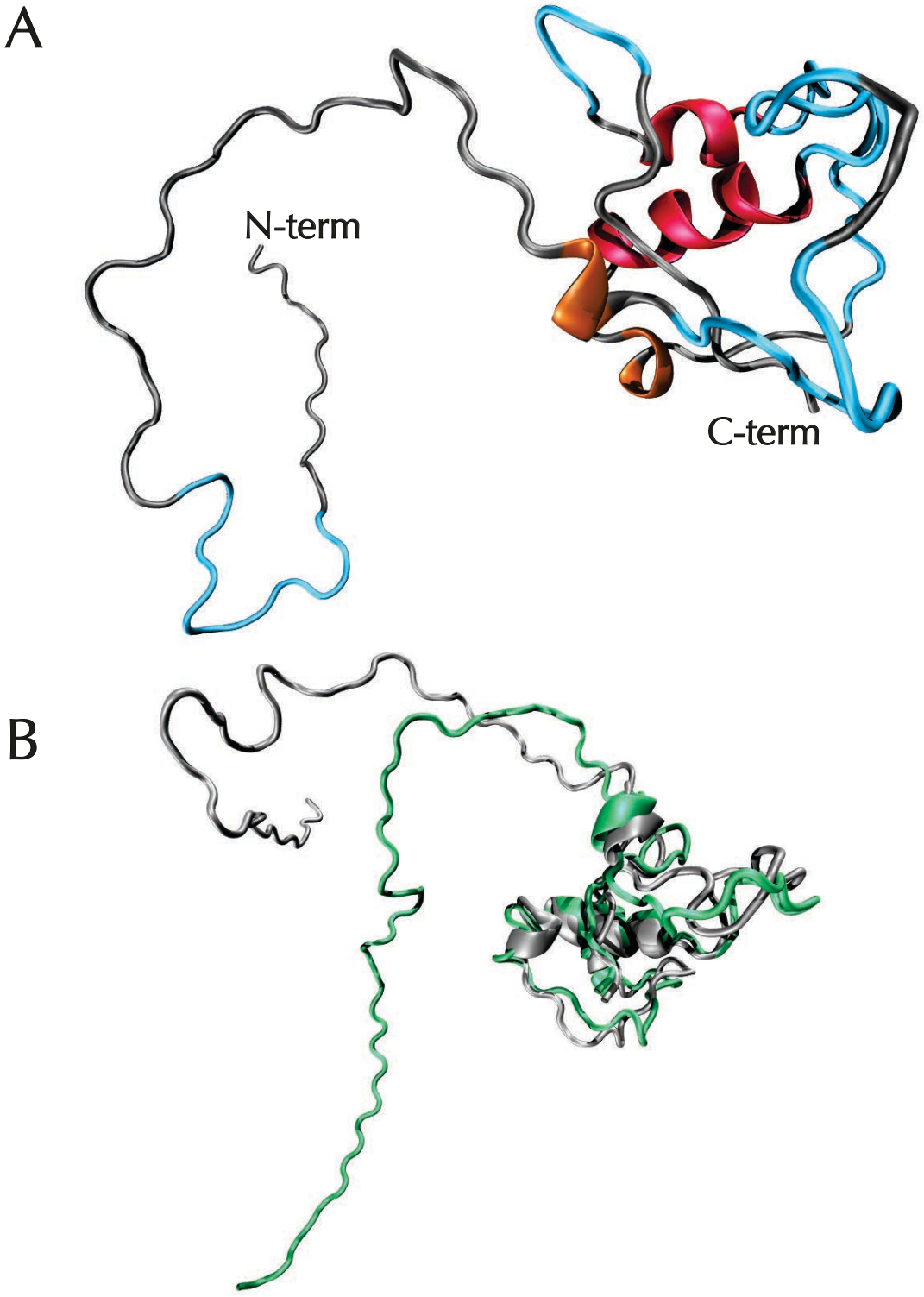

Figure 37: Structure of NTD KorB used for MD simulations. (A) Ribbon representation of the NTD KorB structure (model8, first conformation of NTD KorB with tag) with secondary structure represented by red for α-helix, orange for 3_{10} helix, cyan for turn and grey for coil. **(B)** Overlay of NTD KorB first conformation (t=0 ns) and final conformation (t=110 ns). The first conformation of NTD KorB (with tag) is coloured grey and the final conformation is coloured green. The RMSD (Cα atoms) between the two structures is 16.2 Å. The N-terminal of NTD KorB is highly flexible compared to the C-terminal of the protein. The structures are rendered with POV-Ray in VMD.

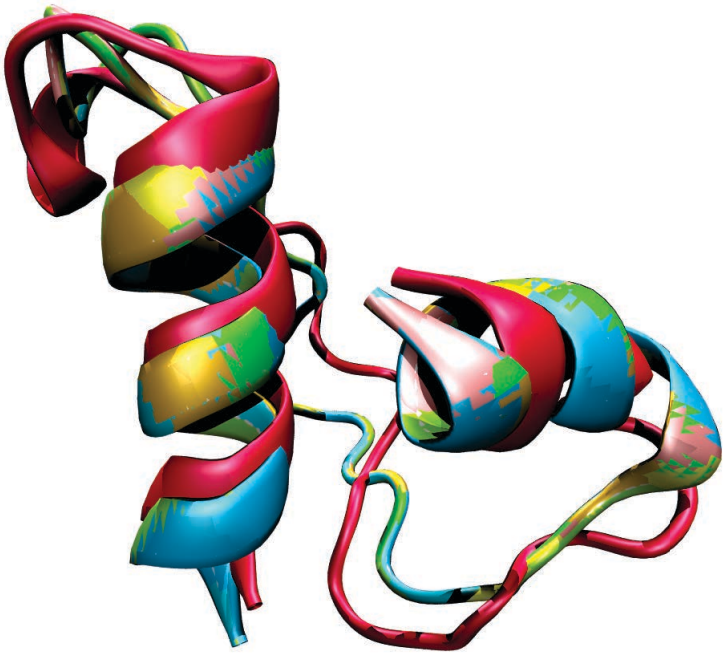

Figure 38: Superimposition of structured region of NTD KorB following the MD trajectory of 110 ns. Ribbon representation of superimposition of structured region of N-terminal domain of KorB is shown. The overlay is of 110 ns trajectory run with NTD KorB's structure extracted at every 10 ns and superimposed. The region spans residues 73-113 (core region) for the NTD KorB construct (with tag) and includes two α-helices. The reference NTD KorB structure shown in red. The RMSD (Cα atoms) for the core region is 2.5 Å. The structures are rendered with POV-Ray program in VMD.

4.6 Discussion

The solution structure has been calculated for the N-terminal domain of KorB using solution state NMR spectroscopy. Only after combining a suite of complementary structural and biophysical techniques (NMR, CD, AUC and MD) was it possible to study and gather structural information about the inherently dynamic nature of NTD KorB.

80% of the NTD KorB's backbone was assigned with confidence and the solution state NMR data yielded 20 conformers compatible with the restraints used for structure refinement. The less than ideal completeness of NMR peak assignments was due to relatively high number of proline residues (10 Pro residues) in a 130 residue protein

and the intrinsic flexibility of the protein. The assignment walk was proving difficult because of the proline residues.

The number of amide peaks in HSQC spectrum of the NTD KorB is more than the number of NH groups in the protein. These peaks seem reproducible between different NTD KorB preparations and so are unlikely to be due to protein degradation products. Figure 18 shows the overlay of HSQC spectra of NTD KorB before and after cleaving the N-terminal tag. Visually, most of peaks in both spectra appear to have the same NH shifts whilst a few of the peaks have moved, possibly because of the absence of first 17 N-terminal residues comprising of the tag (1-17 aa). However, there are no additional peaks in the spectrum of the protein with the tag. This demonstrates that the N-terminal tag is highly flexible and does not contribute to any peaks in the HSQC spectrum. Also, it is plausible to suggest that the tag does not contribute to the protein structure and does not reduce the significance of the NTD KorB ensemble structure.

The NTD KorB including the His-tag consists of 130 residues and ideally one should observe the same number of peaks in the HSQC spectrum apart from the ten Prolines. We have established that most of the peaks from the N-terminal (1-23 aa) tag are not visible in the spectrum. However we still observe about 130 peaks in the spectrum, thereby indicating possibility of interconverting protein conformational states or a covalent modification on the protein. There is limited evidence for exchange between different protein conformational states so whether these additional peaks in HSQC are because of very slowly exchanging conformers (e.g. cis-trans proline) or partial covalent modification is not clear. Using ponderosa, NTD KorB's ensemble structure was calculated compatible with the restraints used for structure refinement. NTD KorB is mainly disordered and consists of two α-helices (α1 and α2) towards the C-terminal region (core region). While trying to assign the additional peaks, the assignments overlapped with already assigned α-helix 1 of NTD KorB.

Comparison of secondary structure of NTD KorB from the chemical shift data using TALOS+ and computationally predicted Jpred and Psipred data show agreement

with all of the methods predicting two main α-helices in the protein. On the other hand, Jpred and Psipred predict β-strand but there is no evidence for β-strand in the NMR spectra. TALOS+ predicts NTD KorB to be primarily disordered for the first N-terminal 70 residues. This TALOS+ prediction is in agreement with the bioinformatics analysis by Rajasekar et al., (2010) using the database of intrinsically unstructured proteins, DisProt (Sickmeier et al., 2007) which suggests that the first 64 amino acids of KorB are disordered.

Structural comparisons of the crystal structure of four N-terminal domains of Spo0J monomer structures from *H. pylori* reveals that the N-terminal exhibit multiple conformations (Chen et al., 2015). As evident from the figure 39, a poor fit was observed for the over-laying NTD Spo0J structures, with an RMSD (Cα atoms) range of 11.1 Å. Various loop regions are observed in the N-terminal suggesting high flexibility of the N-terminal domains of Spo0J.

The structure of NTD KorB is mainly disordered and consists of two α-helices (α1 and α2) towards the C-terminal region (core region). Comparison of NTD KorB secondary structure with two structural homologues is shown in figure 40A. These are 1VZ0 and 4UMK and are the crystal structures of Spo0J from *T. thermophilus* and *H. pylori* respectively.

The sequence of NTD KorB along with secondary structure elements was aligned with homologous protein sequences with 3D structures in the PDB. The aligned sequences suggest that the two α-helices of NTD KorB, although shorter in length, match the helices in the homologue proteins. Superimposing structures of N-terminals of KorB, 1VZ0 and 4UMK (figure 40B) do not match each other (Cα RMSD for the fit is 20.9 Å) but the helices from the homologues reasonably fit with the two helices of NTD KorB.

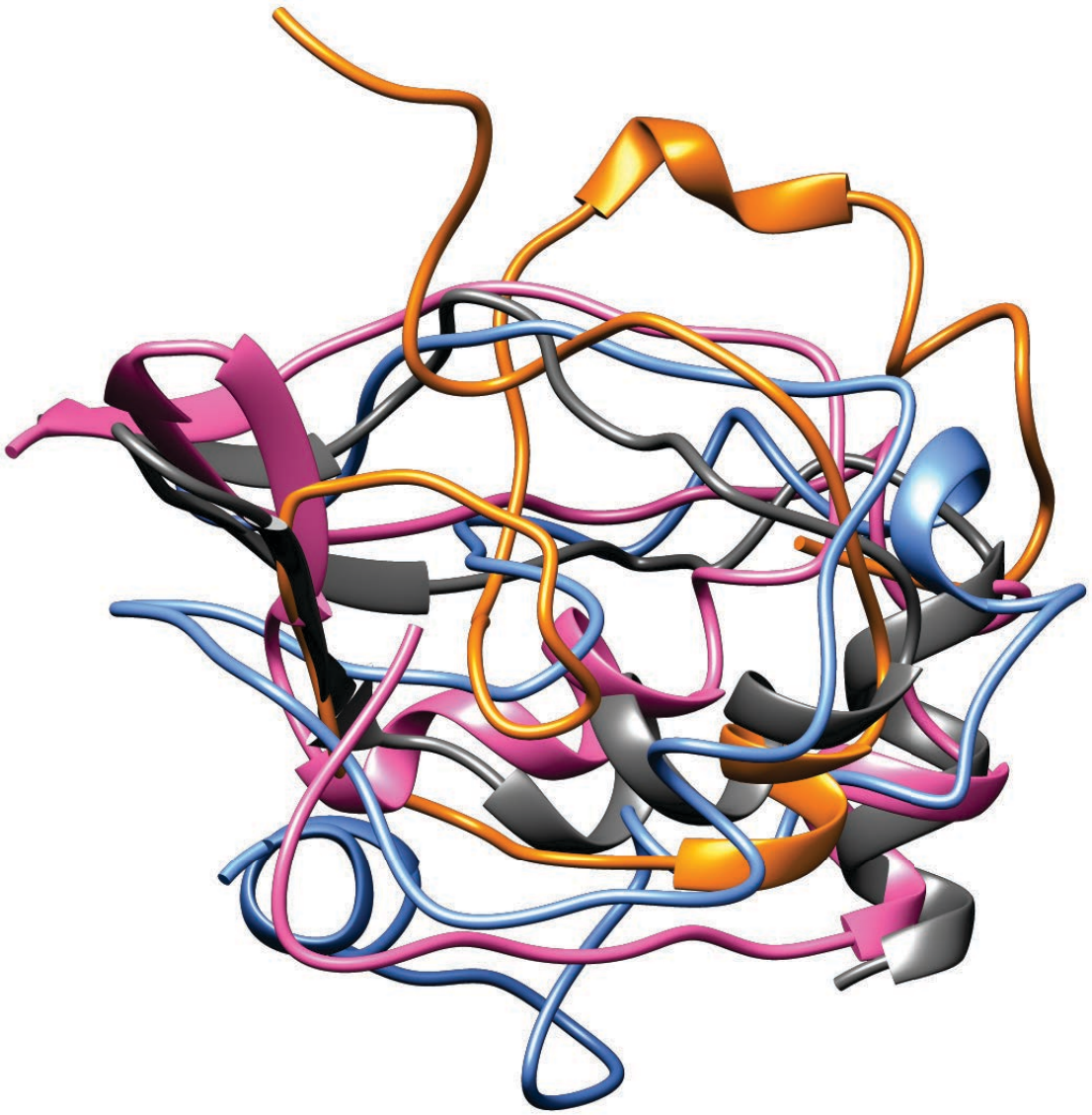

Figure 39: Structural comparisons of NTD Spo0J monomers from *H. pylori*. Ribbon representation of N-terminal domain of Spo0J monomers (excluding the DNA-binding domain) from *H. pylori* is shown. Four Spo0J monomers (from the same crystal structure, PDB ID: 4UMK) with chains A (grey, 55-111), B (pink, 45-111), C (blue, 35-111), and D (orange, 51-111) are superimposed for structural comparison. The structures were overlaid with chain A as the reference structure. The RMSD (Cα atoms) range for the fit is 11.1 Å and the four monomers do not superimpose. Various loop regions are observed in the N-terminal suggesting high flexibility of the N-terminal domains of Spo0J. The structures are rendered with POV-Ray program in Chimera.

In terms of biological significance, the protein-protein interactions in ParBs are thought to occur via an N-terminal 'arginine patch' RRXR which is highly conserved in the ParB family (Chen et al., 2015). TALOS+ predicts the arginine patch (109-RRYR-112) to be part of a helix (α-helix 2) in the NTD KorB. It has previously been hypothesised that adjacent and traverse protein-protein interactions lead to the oligomerisation of (Spo0J from *H. pylori*) on DNA (Chen et al., 2015). Using simulation studies, a similar model was presented by Broedersz et al., (2014) suggesting that the NTD of CBPs is involved in 1D spreading and 3D bridging over plasmid DNA. The aforementioned simulation study proposes that these interactions can be important for the assembly of a higher-order protein-DNA complex resulting in ParB spreading on to DNA (Broedersz et al., 2014).

The NTD KorB ensemble structure described here might reasonably well represent the internal dynamics, but the expectation that those conformers reflect the protein's internal dynamics *in vivo* is not necessarily met. Most of the NMR parameters measured during the experiments can be considered as an average of dynamics over different time scales. It can be considered as a biased expectation for a single protein ensemble structure to reflect all the range of motions exhibited by the protein. This is reflected from the fact that during the process of structure refinement, the emphasis is to calculate similar structures as represented by a low backbone RMSD value (Ángyán et al., 2010). This presents as an attempt to represent NMR derived protein structures as single conformers, similar to X-ray crystallography structures that represent a static protein state. MD simulation data for NTD KorB suggests that the two α-helices towards the C-terminus exhibit limited mobility and internal motion in comparison to rest of the protein. This can be seen in the overlay of the helices every 10 ns of the simulation (see figure 38). NTD KorB solution structure along with MD simulation data should be regarded in future investigations of KorB.

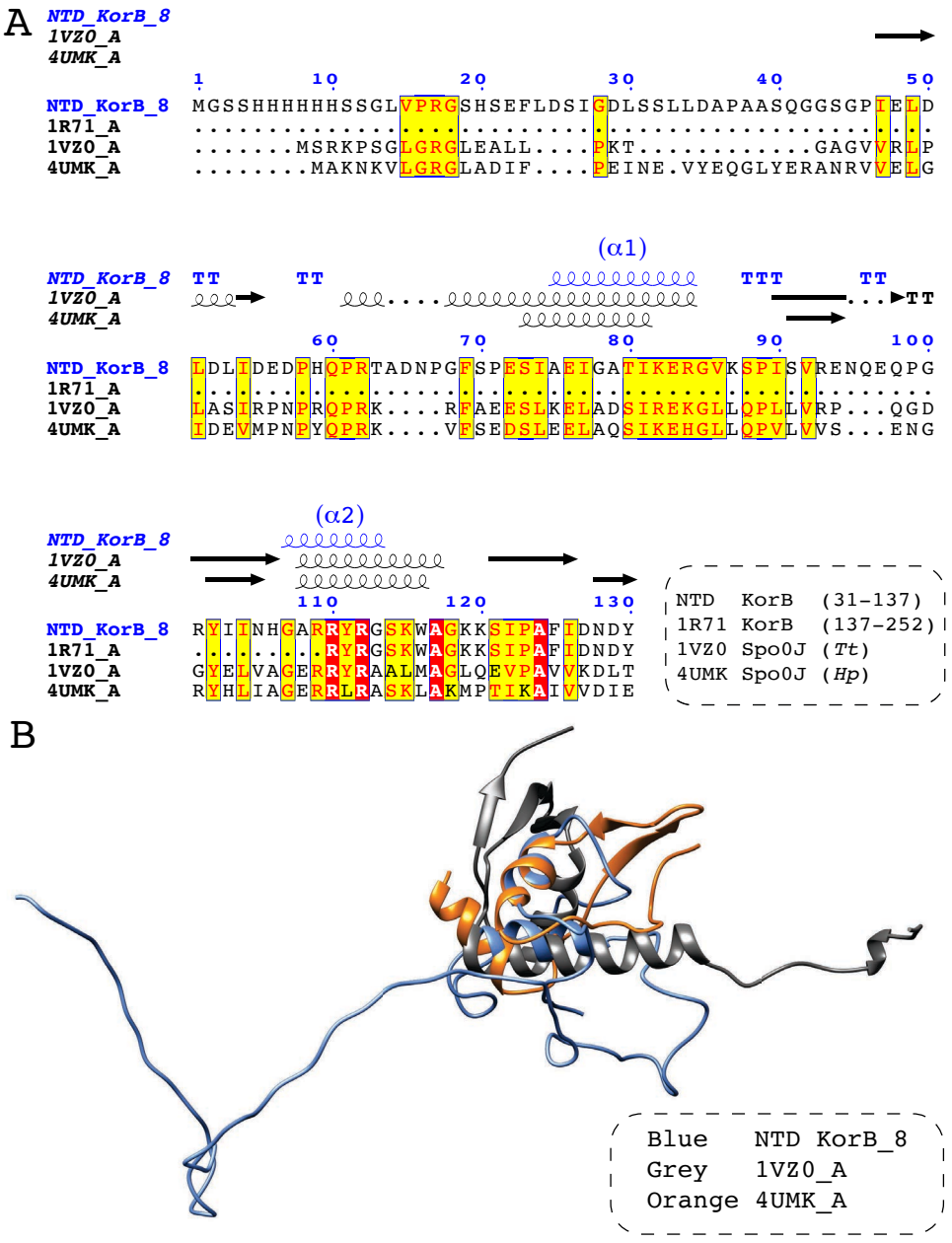

Figure 40: Comparison of NTD KorB structure with KorB homologues. (A) NTD KorB secondary structure compared with KorB structural homologues. The structure of NTD KorB containing residue 31-137 of KorB was searched with BLAST+ against PDB to detect protein homologues. Homologues include 1VZ0 and 4UMK and these are the crystal structures of Spo0J from *T. thermophilus (Tt)* (Leonard et al., 2004) and *H. pylori (Hp)* (Chen et al., 2015) respectively. The sequence of NTD KorB along with secondary structure elements was aligned with homologous protein sequences with 3D structures in the PDB. The aligned sequences are coloured according to the degree of similarity and the secondary structures i.e. α-helices are represented by helix, β-strands are represented by black dashed arrow and turns are represented by blue coloured letter 'T'. The structural alignment is performed with ENDscript. **(B)** Superimposed structures of NTD KorB, 1VZ0 and 4UMK. The superimposition is done keeping NTD KorB as the reference structure. The 1VZ0 and 4UMK structures are incomplete but the helices reasonably fit with the two helices of NTD KorB. The structures are rendered with POV-Ray in Chimera.

5

Structural characterisation of DNA binding domain of KorB

5 Structural characterisation of DNA-binding domain of KorB

In order to characterise the DNA-binding domain of KorB, 100 residues from the C-terminal of the KorB WT were truncated and the resulting protein was named CΔ100 KorB. The CΔ100 KorB contains an N-terminal His-tag for purification and a thrombin cleavage site between residue numbers 17 and 18. The N-terminal His-tag along with the first 31 residues from the N-terminal of the CΔ100 KorB were truncated and the resulting protein was named NΔ31CΔ100 KorB. The CΔ100 KorB consists of 1-258 aa and NΔ31CΔ100 KorB consists of 31-258 aa and both constructs are illustrated in figure 41.

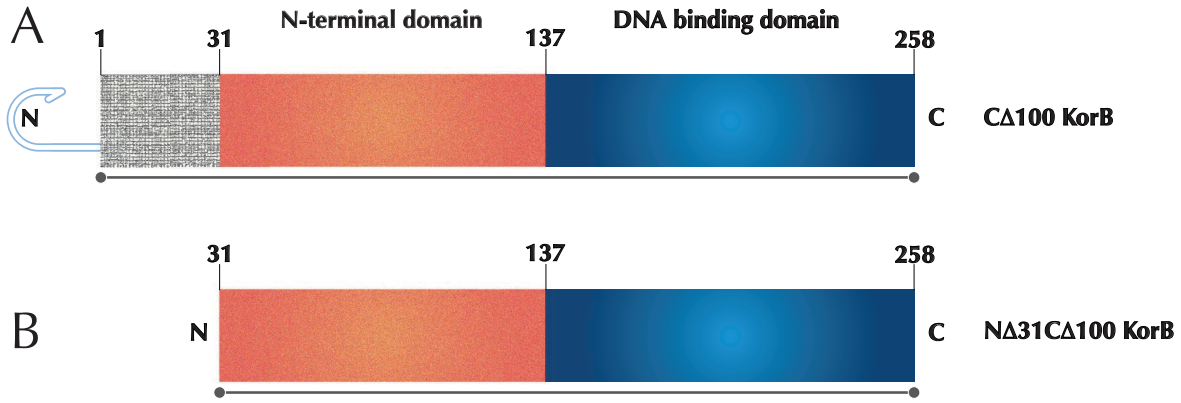

Figure 41: Illustration of KorB mutants studied in this chapter. The N-terminal domain and DNA-binding domain are shown in orange and blue respectively. **(A)** 100 residues from the C-terminal of the KorB WT are truncated and the resulting protein is named CΔ100 KorB. The N-terminal contains a His-tag (shown with a hook) for purification and the flexible region of the protein is shown in grey. **(B)** The N-terminal His-tag along with first 31 residues from the N-terminal of the CΔ100 KorB are truncated and the resulting protein is named NΔ31CΔ100 KorB.

The results for over-expression and purification of CΔ100 KorB and NΔ31CΔ100 KorB are given in section 5.1 and section 5.2 respectively and the biophysical characterisation of both the proteins is given in section 5.3.

5.1 Results for CΔ100 KorB

5.1.1 Over-expression of CΔ100 KorB

CΔ100 KorB with residue 1-258 and molecular weight of 30.6 kDa contains a 23 aa His-tag on the N-terminal. CΔ100 KorB was expressed in *E. coli* BL21 cells, induced using IPTG, and the cells were homogenised using the emulsiflex to release the CΔ100 KorB into solution.

Modified pET28a vector with the CΔ100 KorB gene cloned under the control of T7 promoter was used to over-express the protein. The cloned gene also contains an N-terminal His-tag followed by a thrombin cleavage site at residues 17 and 18. One μL of plasmid DNA (40 ng/μL) in nuclease free water was used to transform competent *E. coli* cells using the heat-shock method. The Kanamycin resistance gene in the vector is used to screen for transformed bacterial colonies. Bacterial colonies ($\sim$200) were observed on the agar plate with Kanamycin, upon overnight incubation at 37 °C. Two L (2*1 L) of cells were grown in LB media with Kanamycin to an A_{600} of $\sim$0.6 at 37 °C, measured using the BioMate 3 spectrophotometer (ThermoFisher). At this stage, expression of T7 RNA polymerase was induced with 1 mM IPTG. After the induction, the culture was grown overnight at 18 °C to yield maximum protein.

5.1.2 Purification of CΔ100 KorB

5.1.2.1 Affinity chromatography

The initial purification of the CΔ100 KorB was done using a 5 mL HIS-select Nickel affinity gel column at RT as CΔ100 KorB contains a His-tag on the N-terminal. The column was equilibrated with buffer containing 10 mM Imidazole, 50 mM Na phosphate, pH 8.0, 300 mM NaCl, and 10 μg/mL PMSF. The supernatant after centrifugation ($\sim$50

mL) was loaded onto the Ni-NTA column. The column was washed with 50 mL buffer containing 20 mM Imidazole. A step gradient of 50 mM and 300 mM Imidazole (50 mL each) was applied to elute the protein. Presence of protein in the fractions was confirmed using Bradford reagent followed by SDS-PAGE gel analysis. CΔ100 KorB was purified by affinity chromatography and the SDS-PAGE gel is shown in figure 42. Fractions containing the protein were pooled and concentrated using the Amicon (Merck) centrifugal filter units with MWCO of 10 kDa in an Eppendorf 5810R bench-top centrifuge at 4,000 g. Further purification of the CΔ100 KorB was performed using size exclusion chromatography.

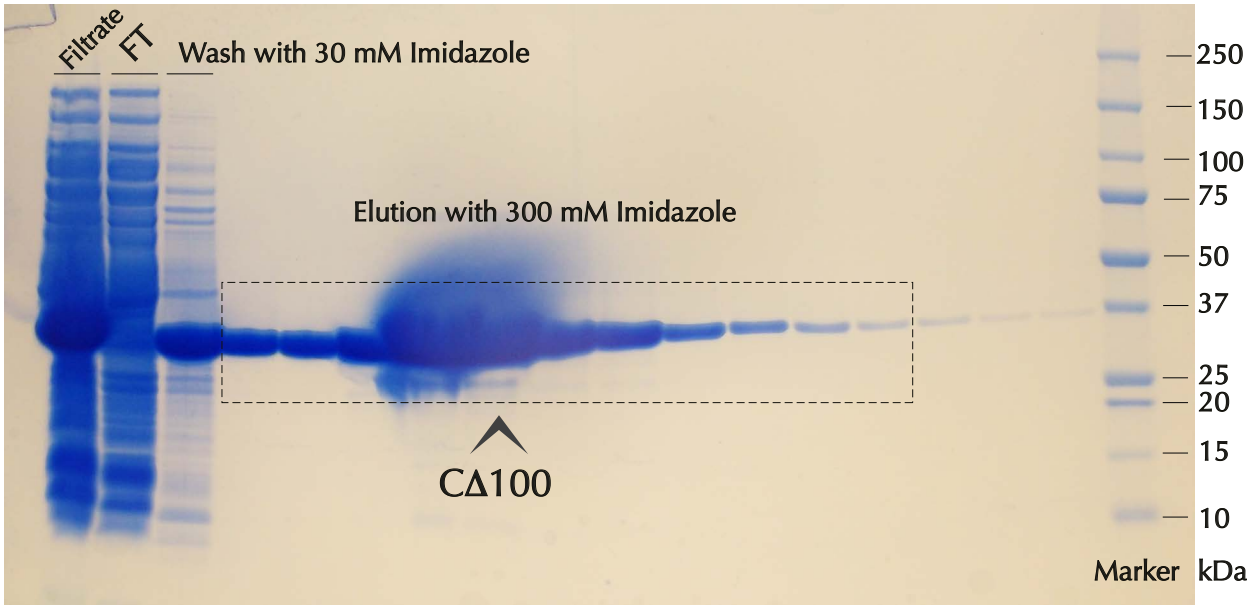

Figure 42: Purification of CΔ100 KorB by affinity chromatography and analysis by SDS-PAGE. SDS-PAGE gel with dashed box and arrowhead indicate the CΔ100 KorB protein. From left to right: supernatant after filtration (Filtrate), flow-through from soluble lysate fraction (FT), wash with buffer containing 30 mM Imidazole (Wash), rest are the elution fractions containing 300 mM Imidazole and molecular weight marker (Marker). The fraction size is 2 mL, flow rate is 1 mL/min and length of the run is 50 mL. Step gradient of Imidazole is used for elution. Most of the CΔ100 KorB eluted off from the column with 300 mM of Imidazole. The fractions in the dashed box are pooled and are further purified to homogeneity by size exclusion chromatography.

5.1.2.2 Size exclusion chromatography

After affinity chromatography, the CΔ100 KorB was further purified using the size exclusion chromatography. The column was equilibrated with 10 mM Tris HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA and 5 mL of CΔ100 KorB was loaded on to the column at 4 °C. For ^{15}N labelled protein expression of CΔ100 KorB, Tris was substituted with phosphate in the buffer. Void volume of the column was 100 mL and 4 mL of fractions were collected during the run. The size exclusion chromatogram for CΔ100 KorB is given in figure 43A, and figure 43B shows the SDS-PAGE gel confirming the presence of a single protein species in solution. CΔ100 KorB elutes as a single peak -with shoulders- indicating the presence of a monomeric species. After the size exclusion run, the CΔ100 KorB was again concentrated to ∼2 mL. The total yield of CΔ100 KorB obtained from 1 L of culture was 25 mg.

5.1.3 Optimisation of thrombin digestion of CΔ100 KorB

To remove the N-terminal His-tag from CΔ100 KorB and to optimise the temperature, the protein was incubated with thrombin at 4 °C and RT. Thrombin was diluted 10 fold in the same buffer and reaction was stopped at different times using the inhibitor, PMSF. Figure 44A shows the SDS-PAGE gel with optimisation of thrombin digestion of CΔ100 KorB. It was observed that CΔ100 KorB degrades with passage of time even at 4 °C, so the purification has to be done quickly and thrombin digestion was preferred at 4 °C immediately after running the Ni-NTA column. The SDS-PAGE gel for thrombin digestion of CΔ100 KorB is given in figure 44B. The tag was removed by incubating CΔ100 KorB in presence of thrombin at 4 °C for 1 hr. CΔ100 KorB was immediately incubated with thrombin after affinity chromatography and mixture of CΔ100 KorB and thrombin was loaded on to the size exclusion column to further purify the CΔ100 KorB.

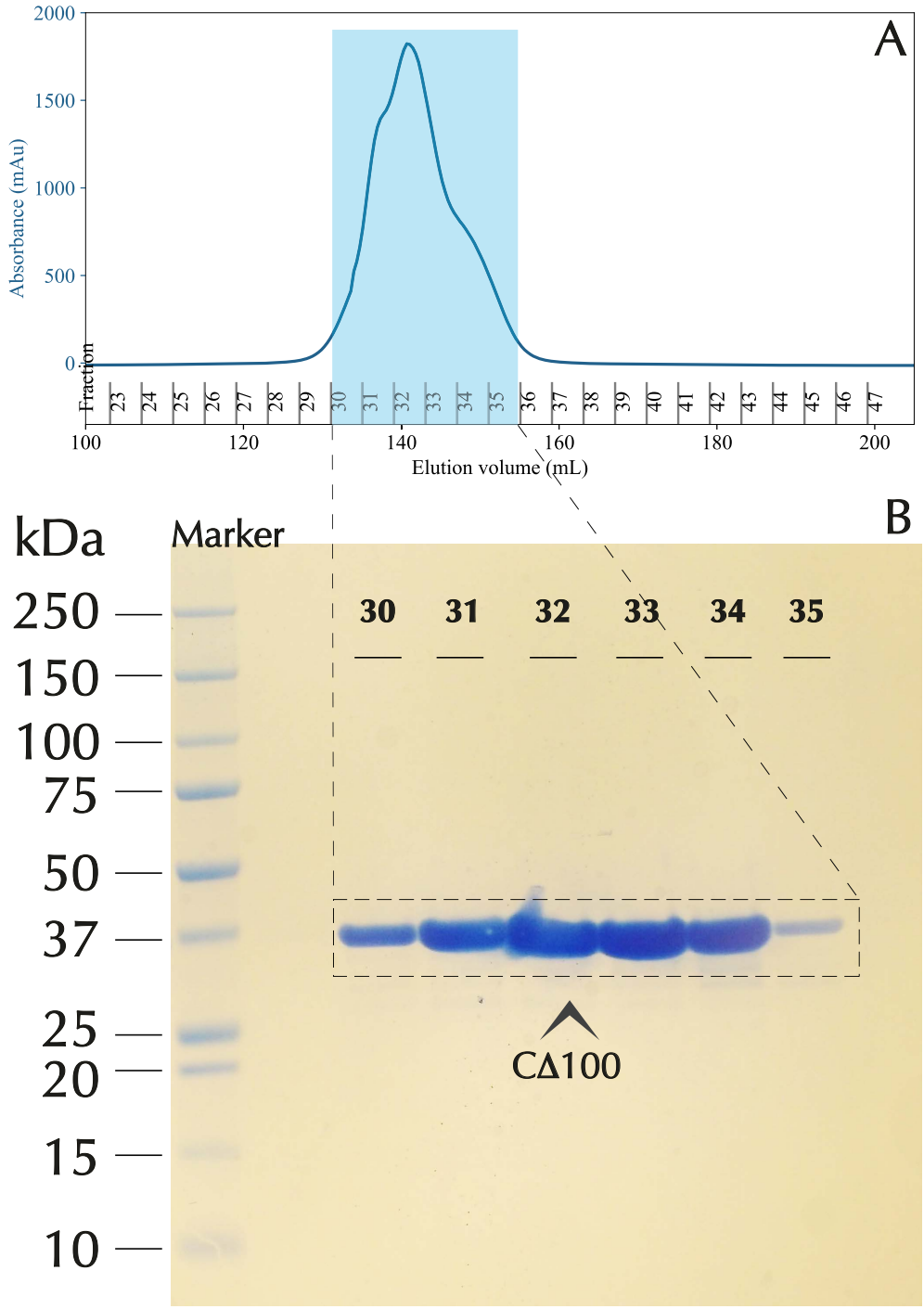

Figure 43: Purification of CΔ100 KorB by size exclusion chromatography and analysis by SDS-PAGE. (A) Chromatogram of CΔ100 KorB showing the absorbance trace (A_{280}) of eluent and peaks between 130 mL and 160 mL of elution volume. Fractions on the X-axis indicate 4 mL (each) of collected fraction. Void volume of the column is 100 mL. The flow rate of 2 mL/min is used and length of the run is 380 mL. Fractions labelled 30 to 35 contain the protein (turquoise box). CΔ100 KorB elutes as a single peak -with shoulders- indicating the presence of a monomeric species as seen in the gel image. **(B)** Image of the SDS-PAGE gel with protein in the lanes labelled 30, 31, 32, 33, 34 and 35 showing presence of a single protein species in solution. From left to right: molecular weight marker (Marker), elution fractions from size exclusion chromatography. Dashed box and arrowhead indicate the purified CΔ100 KorB after size exclusion run.

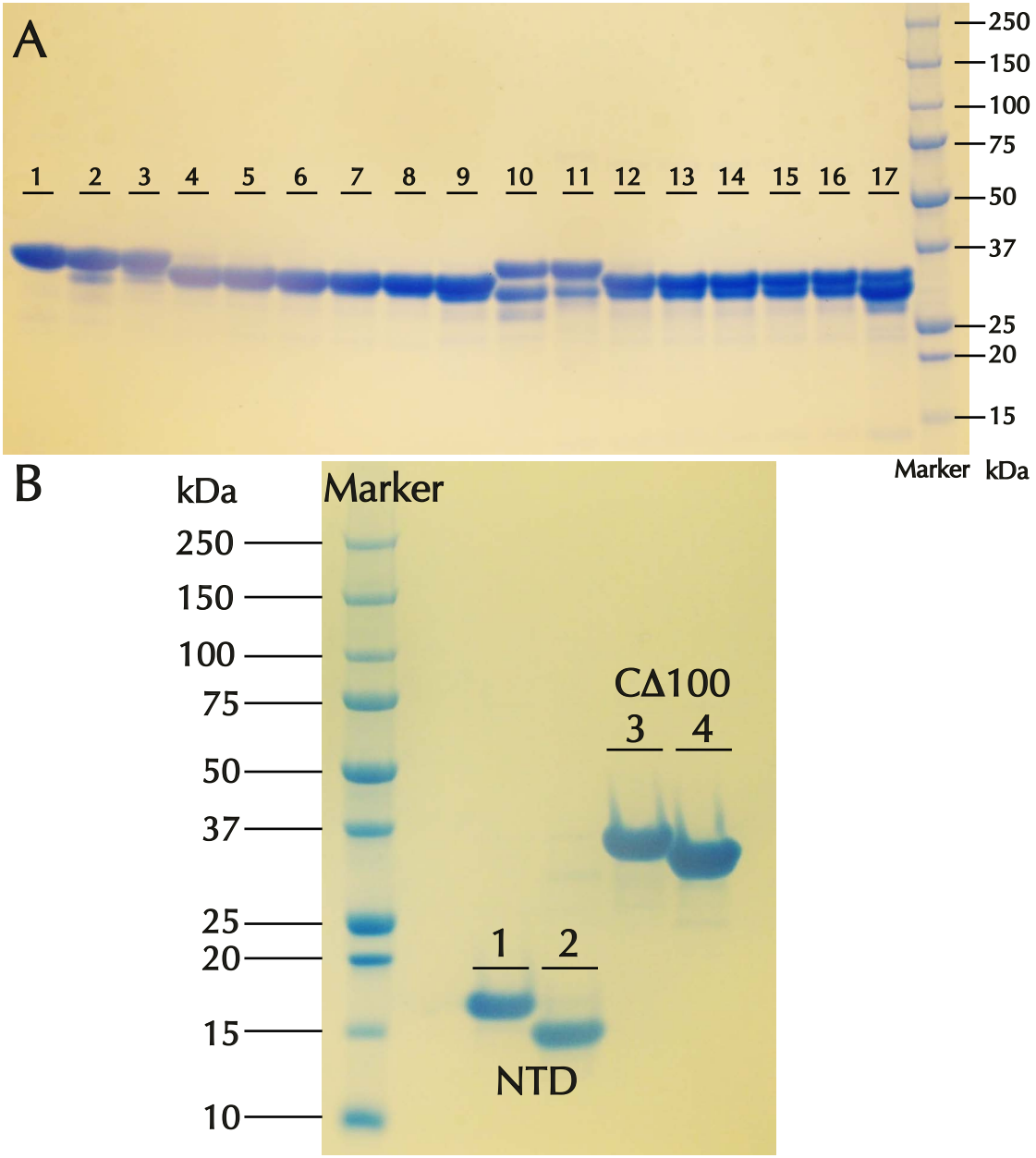

Figure 44: Thrombin digestion of CΔ100 KorB. (A) The N-terminal tag is removed by incubating CΔ100 KorB in presence of thrombin at RT and 4 °C. SDS-PAGE gel showing the optimisation of thrombin digestion of CΔ100 KorB. CΔ100 KorB in lanes 3-9 is incubated with thrombin at 4 °C whereas the protein in lanes 11-17 is incubated with thrombin at RT. From left to right – 1 and 2: CΔ100 KorB kept at -80 °C and 4 °C respectively; 3: CΔ100 KorB with thrombin at 0 hr (at 4 °C); 4: protein with thrombin after 1 hr; 5: protein after 2 hr; 6: protein after 4 hr; 7: protein after 6 hr; 8: protein after 8 hr; 9: protein after 50 hr; 10: CΔ100 KorB kept at RT; 11: CΔ100 KorB with thrombin at 0 hr (at RT); 12: CΔ100 KorB with thrombin after 1 hr; 13: protein after 2 hr; 14: protein after 4 hr; 15: protein after 6 hr; 16: protein after 8 hr; 17: protein after 50 hr and molecular weight marker (Marker). **(B)** SDS-PAGE gel of thrombin cleavage of CΔ100 KorB. The protein is incubated with thrombin at 4 °C for 1 hr. From left to right – molecular weight marker (Marker); 1 and 2: NTD KorB before and after thrombin digestion respectively (see figure 12); 3 and 4: CΔ100 KorB before and after thrombin digestion respectively.

5.1.4 Stability of CΔ100 KorB at room temperature

During the purification, the CΔ100 KorB was observed to be unstable and appears to degrade rapidly at RT. In order to check the stability of CΔ100 KorB at RT, the protein was left undisturbed at RT for multiple days and the protein lysis was stopped by storing the protein at -80 °C. The stability of CΔ100 KorB at RT is shown in figure 45. Considering the instability and rapid degradation of CΔ100 KorB, the protein was purified within a day. In order to maintain integrity of protein for future experiments, the protein was flash frozen immediately after concentrating the size exclusion fractions.

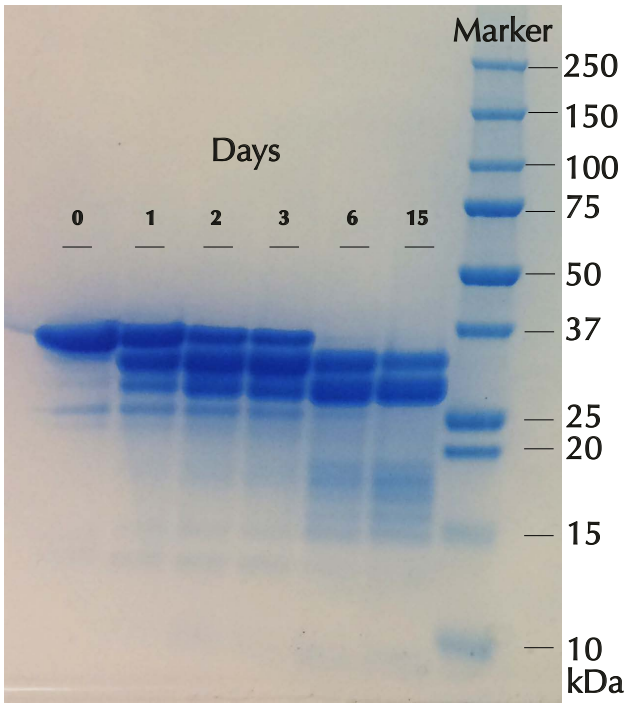

Figure 45: Stability of CΔ100 KorB at room temperature. SDS-PAGE gel for CΔ100 KorB left at room temperature for 15 days. From left to right – day 0: single band is observed with limited contamination as the protein is kept at -80 °C; day 1: protein left for a day at RT, protein starts to degrade and multiple bands are observed; day 2 and day 3: CΔ100 KorB continues to degrade; day 6: most of the CΔ100 KorB is degraded and two species are observed; day 15: most of the CΔ100 KorB is degraded and the two stable species are observed; molecular weight marker (Marker). CΔ100 KorB appears to degrade rapidly at RT as multiple bands are observed within the first three days. Leaving the protein for longer duration (more than 6 days), resulted in two stable species, as seen on the gel.

5.1.5 Molecular mass estimation of CΔ100 KorB by mass spectrometry

In order to check the structural integrity of CΔ100 KorB and estimation of molecular mass, the purified CΔ100 KorB was subjected to Electrospray Ionization-Mass Spectrometry (ESI-MS). CΔ100 KorB protein samples were given to the mass spectrometry facility in the School of Chemistry at the University of Birmingham. Figure 46 shows the mass spectrometry data for CΔ100 KorB.

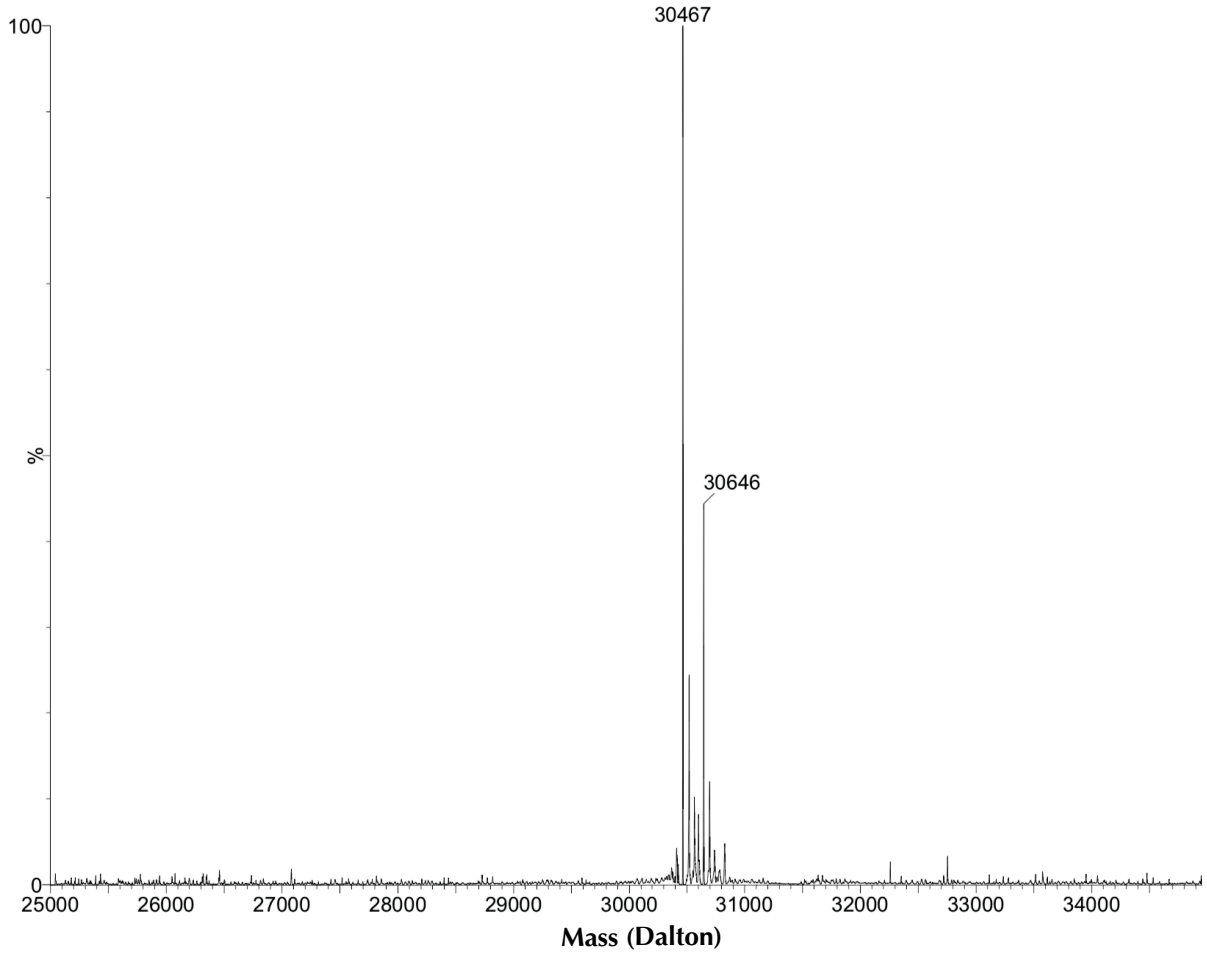

Figure 46: Molecular mass estimation of CΔ100 KorB by mass spectrometry. The purified CΔ100 KorB is subjected to MS for estimation of molecular mass. ESI-MS spectrum is shown for CΔ100 KorB and the molecular mass observed for the protein is 30,467 Da and the theoretical mass of the protein calculated from the amino acid sequence with ProtParam server (Gasteiger2005) is 30,591 Da.

5.2 Results for NΔ31CΔ100 KorB

5.2.1 Gradient PCR

The digestion problem of N-terminal of CΔ100 KorB was partially circumvented by removing the first 30 residues from the CΔ100 KorB. The CΔ100 KorB was further sub-cloned in a modified pET28a vector. The primers were designed so that the first 30 residues of CΔ100 KorB in addition to the N-terminal His-tag were removed. The primer attachment to the modified pET28a vector with cloned CΔ100 KorB is shown in figure 47. NΔ31CΔ100 KorB protein was cloned using the quick change mutagenesis protocol.

Figure 47: Schematic representation of primer attachment to the pET28a vector, showing the positions of T7 promoter, lac operator, ribosome binding site (RBS), start site (ATG[1]), hexa his-tag, thrombin cleavage site, start site for KorB WT[§], and both forward and reverse primers.

For quick change mutagenesis, gradient polymerase chain reaction (PCR) was employed. The contents of the PCR reaction mix along with the concentrations and volumes used for the experiment are listed in table 18. Primer annealing temperature conditions used were 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, and 65 °C. Bands were observed at temperatures 45 °C, 50 °C, 55 °C. Reaction mix for those three temperature conditions were pooled and digested with enzyme Dpn1 for 20 min at 37 °C. Dpn1 enzyme digests

only methylated DNA and digests the template DNA in the reaction mix. Two μL of reaction mix was used to transform 45 μL of XL 10 gold ultra-competent cells (Agilent Technologies) and bacterial colonies were observed on the plate.

Sample	**Concentration**	**Volume (μL)**	**Control (μL)**
Vector DNA	40 ng/μL	6	0
Forward primer	100 μM	1.5	1.5
Reverse primer	100 μM	1.5	1.5
Fusion enzyme buffer	5x	60	60
dNTPs	50 mM	6	6
$MgCl_2$	50 mM	6	6
Fusion polymerase enzyme	2 units/μL	5	5
H_2O	55.5 mol/L	214	220

Table 18: PCR reaction mix used for quick change mutagenesis protocol to clone the NΔ31CΔ100 KorB protein.

5.2.2 Colony PCR

For colony PCR, five colonies were randomly selected. Using an autoclaved toothpick, a single colony was picked, streaked on an agar plate (with Kanamycin) and the toothpick was washed with 30 μL of nuclease free water. The water with the cells was heated and kept at 100 °C for 10 min. High temperature released the plasmid DNA into the solution and the sample was centrifuged at 10,000 rpm on benchtop centrifuge for 5 min. 2.5 μL of the supernatant was used as DNA template for colony PCR. The PCR products were run on 1% agarose gel and the truncated product was 159 bp smaller in size in comparison to the reference sequence. The DNA from the colonies number 1 and 2 were sequenced and the sequence was confirmed.

5.2.3 Over-expression of NΔ31CΔ100 KorB

NΔ31CΔ100 KorB with residue 31-258 and molecular weight of 25.28 kDa was expressed in *E. coli* BL21(DE3). NΔ31CΔ100 KorB gene was cloned under the control of T7 promoter in a modified pET28a vector. One μL of plasmid DNA in nuclease free water was used to transform competent *E. coli* BL21 cells using the heat-shock method.

The Kanamycin resistance gene in the vector was used to screen for transformed bacterial colonies. Bacterial colonies ($\sim$330) were observed on an agar plate with Kanamycin, upon overnight incubation at 37 °C. Two L (2*1 L) of cells were grown in LB media with Kanamycin to an A_{600} of $\sim$ 0.6 at 37 °C, measured using the BioMate 3 spectrophotometer (ThermoFisher). At this stage, expression of T7 RNA polymerase was induced with 1 mM IPTG. After the induction, the culture was grown overnight at 18 °C to yield maximum protein.

The culture was centrifuged in Beckman Coulter Avanti® J20/I, rotor JLA 8.1 at 6,000 rpm for 20 min. The supernatant was discarded and pellet was suspended in 50 mM Tris HCl, pH 7.5, 1 mM EDTA ($\sim$50 mL) and centrifuged again at 6,000 rpm for 20 min to ensure that the cells were media-free. Samples were saved before and after induction and were analysed on a 4-12% Criterion^{TM} XT Bis-Tris SDS-PAGE gel with loading dye (Sigma-Aldrich) to check for induction.

5.2.4 Purification of NΔ31CΔ100 KorB

5.2.4.1 Ion chromatography

The cell pellet was resuspended in $\sim$40 mL of 50 mM Tris HCl, pH 7.5, 1 mM EDTA, and a cocktail of inhibitors (DNase, RNase, Mg, Mn) and a protease inhibitor tablet (EDTA free, from Roche) were added. The cells were homogenised using the C3 EmulsiFlex cell homogeniser (AVESTIN Europe GmbH, Mannheim, Germany). To pellet the cell debris, the homogenised solution was centrifuged at 25,000 rpm in Beckman Coulter Avanti® J series centrifuge using JA-25.50 rotor for 45 min. The pellet was discarded

and the supernatant was filtered using the 0.45 μm and 0.2 μm Minisart® filter.

Ammonium sulphate was added to the supernatant according to this webpage: encorbio.com/protocols/AM-SO4.htm. 40% ammonium sulphate was added to the supernatant and the solution was centrifuged at 25,000 rpm in Beckman Coulter Avanti® J series centrifuge using JA-25.50 rotor for 5 min. The pellet was discarded and ammonium sulphate was added upto 75%. The sample was centrifuged again and the supernatant was discarded. Pellet obtained after precipitating the protein with 75% ammonium sulphate was resuspended in buffer. The protein sample was dialysed in 2 L of 20 mM Tris, pH 8.5, EDTA 0.1 mM, NaCl 20 mM + PMSF 1μl/mL of buffer overnight. The sample was loaded on to the HiTrap Q HP, 5 mL prepacked, Q Sepharose anion exchange column. Before that, the Q column was equilibrated with 5 CV of the dialysis buffer. After loading the column with protein, the column was washed with 4 CV of buffer. The protein was eluted using a salt gradient, low salt buffer – 50 mL of 20 mM Tris, pH 8.5, EDTA 0.1 mM, NaCl 20 mM + PMSF 1 μL/mL of buffer and high salt buffer – 50 mL of Tris 20 mM, pH 8.5, EDTA 0.1, NaCl 500 mM + PMSF 1 μL/mL of buffer. The salt gradient was performed on an $\ddot{A}$KTA Prime Plus with 5 mL fraction.

Presence of protein in the fractions was confirmed using Bradford reagent followed by analysis with SDS-PAGE gel. Fractions containing the protein were pooled and concentrated using the Amicon (Merck) centrifugal filter units with MWCO of 10 kDa in an Eppendorf 5810R bench-top centrifuge at 4,000 g. Fractions containing the protein were pooled according to the chromatogram given in figure 48 (fractions 25-33). The chromatogram for NΔ31CΔ100 KorB is given in figure 48A, and figure 48B shows the SDS-PAGE gel confirming the presence of NΔ31CΔ100 KorB in solution. But as seen on the SDS-PAGE gel, the NΔ31CΔ100 KorB sample contains other species as well. Further purification of the NΔ31CΔ100 KorB was performed using the size exclusion chromatography.

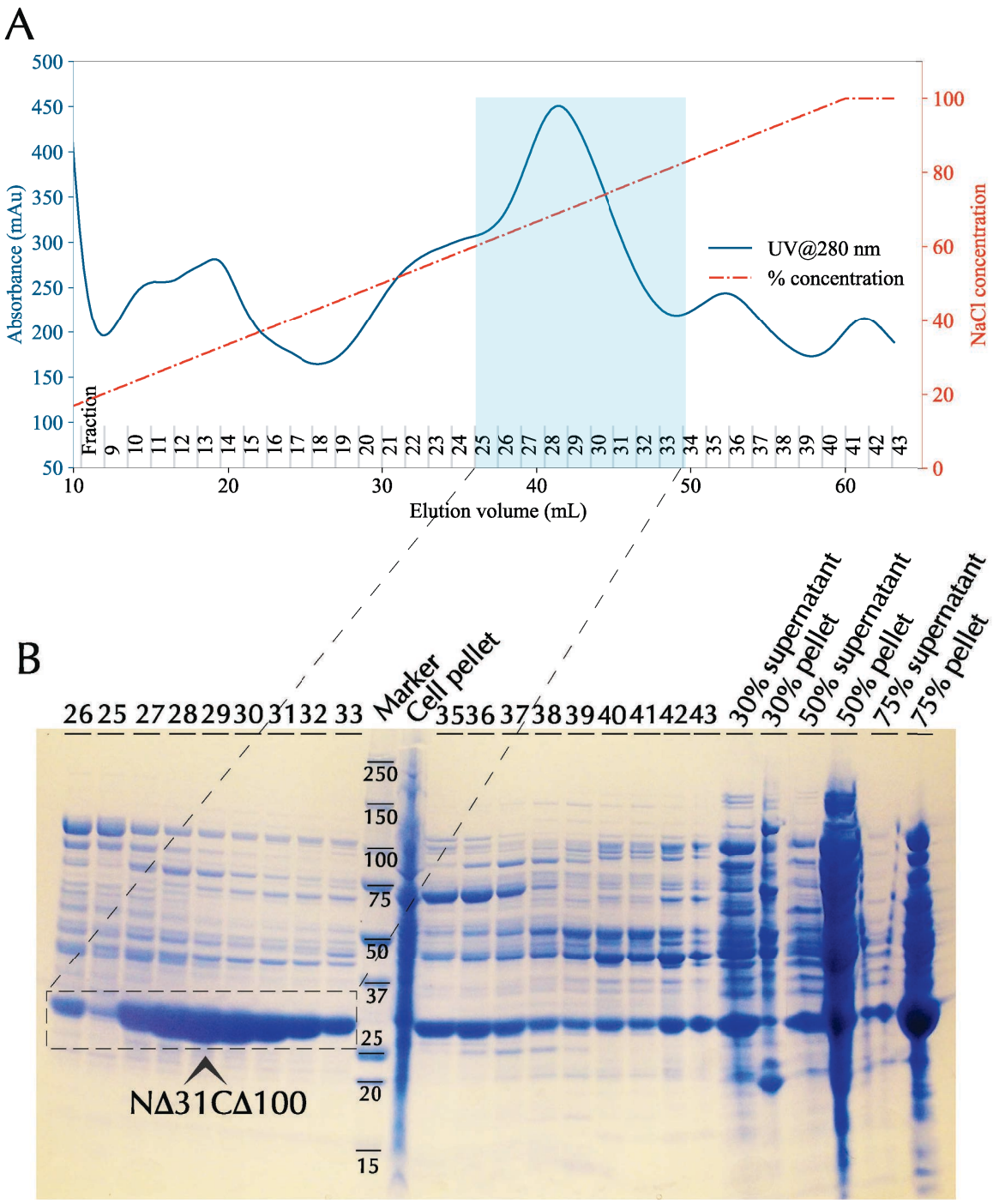

Figure 48: Purification of NΔ31CΔ100 KorB by Ion chromatography and analysis by SDS-PAGE. (A) Chromatogram of NΔ31CΔ100 KorB showing the peak between 35 and 50 mL of elution volume (turquoise box). Fractions 25 to 33 contains the NΔ31CΔ100 KorB as shown by dashed rectangular box. **(B)** SDS-PAGE gel with protein in the lanes labelled 25 to 33. Dashed box and arrowhead indicate the overexpression of NΔ31CΔ100 KorB. From left to right: lanes labelled 25 to 33 contain the protein of interest but with impurities, molecular weight marker (Marker), 35 to 43 lanes labelled depict the rest of the elution profile. The fraction size is 4 mL, flow rate is 2 mL/min and length of the run is 60 mL. 20 - 400 mM NaCl gradient is used for elution. NΔ31CΔ100 KorB in the pooled fractions is further purified to homogeneity by size exclusion chromatography.

5.2.4.2 Size exclusion chromatography

Size exclusion chromatography was done using a Superdex 75, 16 cm by 60 cm column using an *Ä*KTA Purifier P-900 and the chromatogram was monitored at 280 nm. The column was equilibrated with 10 mM Tris HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA and 5 mL of NΔ31CΔ100 KorB was loaded on to the column at 4 °C. For ^{15}N labelled protein prep of NΔ31CΔ100 KorB, Tris was substituted with phosphate in the buffer. A flow rate of 2 mL/min was used and the length of the run was 380 mL. Void volume of the column was 100 mL and 4 mL of fractions were collected during the run.

Fractions containing the protein were pooled according to the chromatogram given in figure 49. NΔ31CΔ100 KorB elutes as a single peak –with shoulders– indicating the presence of a monomeric species. The size exclusion chromatogram for NΔ31CΔ100 KorB is given in figure 49A, and figure 49B shows the SDS-PAGE gel confirming the presence of a single dominant protein species (but with contamination) in solution. NΔ31CΔ100 KorB was again concentrated using the Amicon (Merck) centrifugal filter units with MWCO of 10 kDa in an Eppendorf 5810R bench-top centrifuge at 4,000 g to ∼2 mL. The total yield of the NΔ31CΔ100 KorB obtained from 1 L of culture was 20 mg.

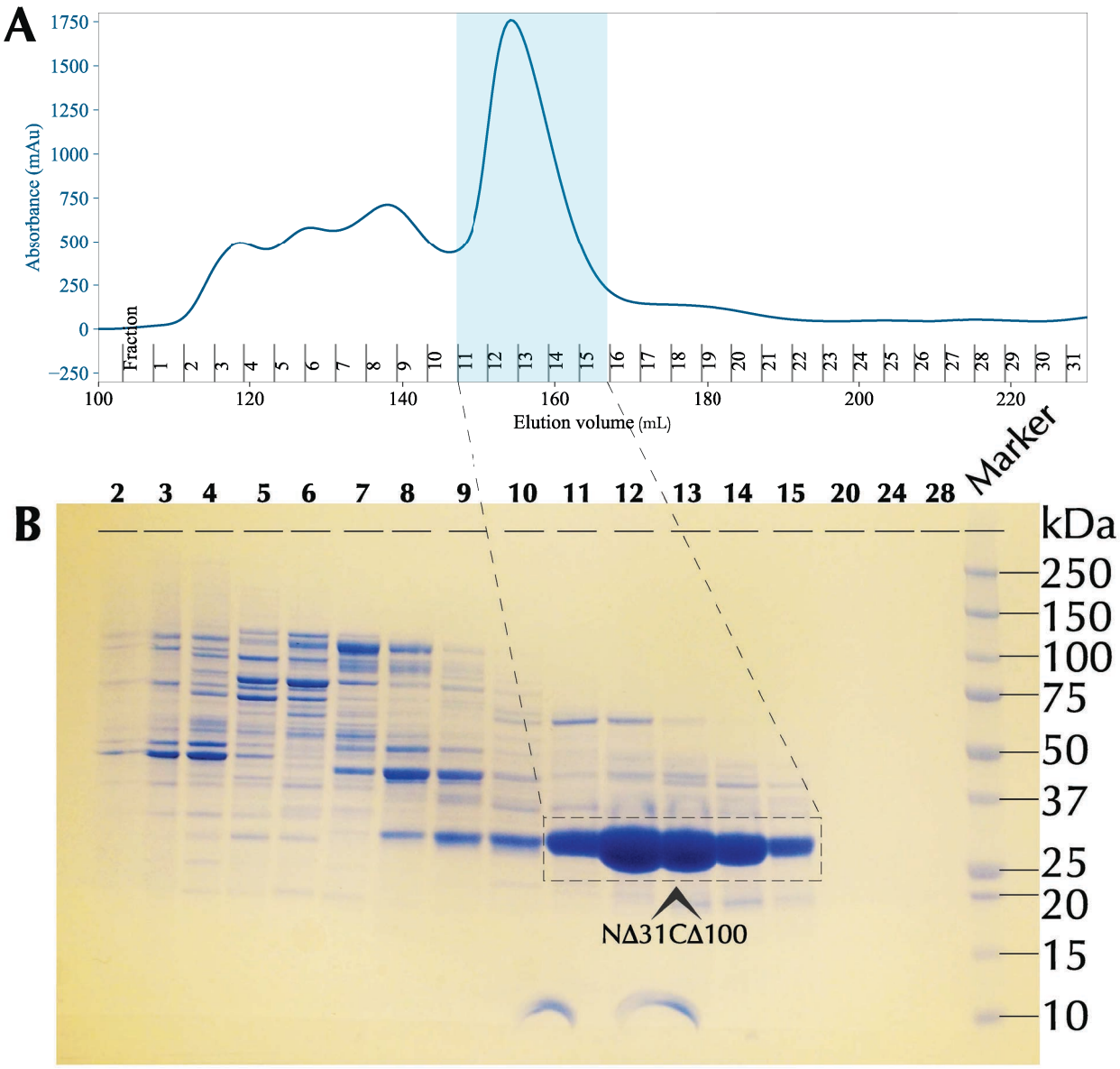

Figure 49: Purification of the NΔ31CΔ100 KorB by size exclusion chromatography and analysis by SDS-PAGE. (A) Chromatogram of the NΔ31CΔ100 KorB showing the absorbance trace (A_{280}) of eluent and peaks between 150 mL and 170 mL of elution volume. Fractions on the X-axis each indicate 4 mL of collected fraction. Void volume of the column is 100 mL The flow rate is set to 2 mL/min and length of the run is 380 mL. Fractions 11 to 15 contain the protein. **(B)** Image of the SDS-PAGE gel with protein in the lanes labelled 11, 12, 13, 14 and 15, showing presence of a single protein species as the dominant protein (but with contamination) in solution. From left to right – 2-9 elution fractions from size exclusion chromatography run, 10 to 15 lanes represent the protein and molecular weight marker (Marker) on the right. Dashed box and arrowhead indicate the purified NΔ31CΔ100 KorB after size exclusion run.

5.2.5 Stability of the NΔ31CΔ100 KorB at room temperature

The stability of NΔ31CΔ100 KorB was checked by keeping the protein undisturbed at RT for 15 days. The stability of NΔ31CΔ100 KorB at RT is shown with an SDS-PAGE gel in figure 50. No sign of protein degradation was observed within the first three days. A stable species was observed from day 6 onwards as seen in figure (dashed box). The protein remains stable if frozen instantly. The protein was purified within two days of homogenising the cells and in order to maintain integrity of protein for future experiments, the protein was flash frozen immediately after concentrating the size exclusion fractions.

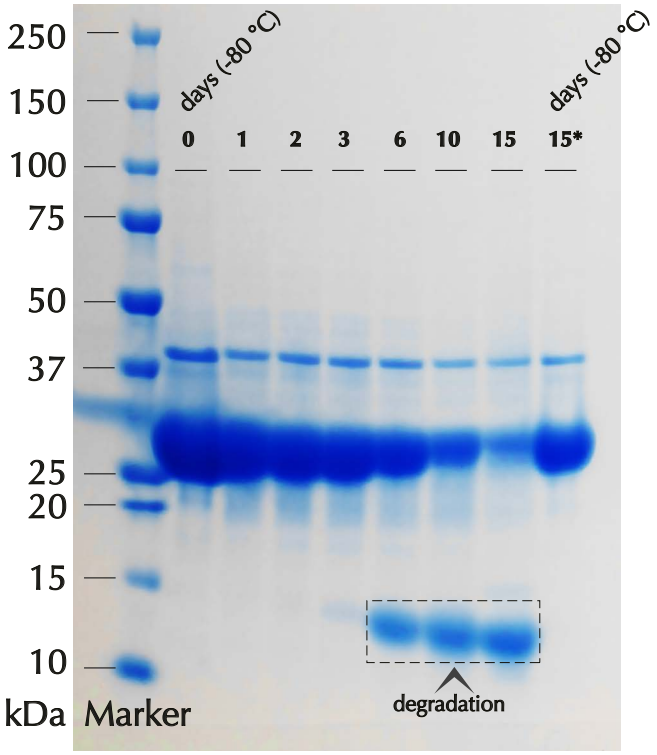

Figure 50: Stability of the NΔ31CΔ100 KorB at room temperature. From left to right: molecular weight marker (Marker); day 0: NΔ31CΔ100 KorB frozen at -80 °C; protein kept undisturbed at RT for n days (n= 1, 2, 3, 6, 10, 15); day 15*: same stock of NΔ31CΔ100 KorB kept at -80 °C. The degradation of NΔ31CΔ100 KorB is stopped by storing the protein at -80 °C. No sign of protein degradation is observed for the first three days. By day 15, most of NΔ31CΔ100 KorB is degraded in the solution. And a stable species is observed from day 6 onwards as seen on the gel (dashed box).

5.2.6 Molecular mass estimation of NΔ31CΔ100 KorB by mass spectrometry

In order to check the structural integrity of NΔ31CΔ100 KorB and estimation of molecular mass, the purified NΔ31CΔ100 KorB was subjected to Electrospray Ionization-Mass Spectrometry (ESI-MS). NΔ31CΔ100 KorB protein samples were given to the mass spectrometry facility in the School of Chemistry at the University of Birmingham and figure 51 shows the mass spectrometry data for NΔ31CΔ100 KorB.

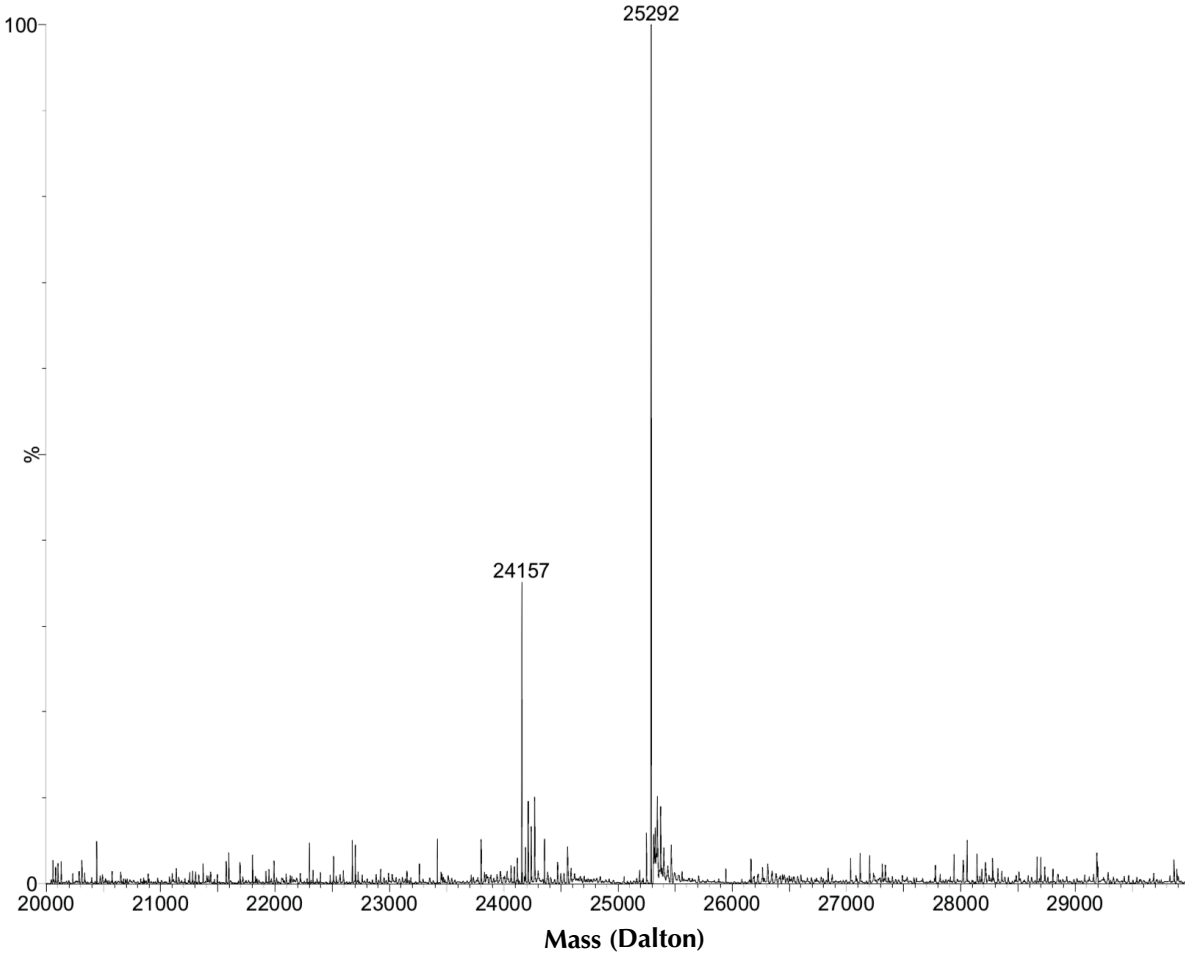

Figure 51: Molecular mass estimation of NΔ31CΔ100 KorB by mass spectrometry. The purified NΔ31CΔ100 KorB is subjected to MS for estimation of molecular mass. ESI-MS spectrum is shown for NΔ31CΔ100 KorB and the molecular mass observed for the protein is 25,292 Da and the theoretical mass of the protein calculated from the amino acid sequence with ProtParam server (Gasteiger2005) is 25,284 Da.

5.3 Biophysical characterisation

Biophysical characterisation of both CΔ100 KorB and NΔ31CΔ100 KorB was performed using circular dichroism and the CD data for both proteins were deconvoluted from DICHROWEB analysis webserver to get the secondary structure estimation.

5.3.1 Circular dichroism of DNA-binding domain of KorB

In order to compare the secondary structure elements of CΔ100 KorB and NΔ31CΔ100 KorB, circular dichroism experiments were performed for both proteins. Figure 52 shows the CD data for CΔ100 KorB and NΔ31CΔ100 KorB. The CD spectra for both proteins are shown in the range of 190-240 nm. The spectra for both proteins were individually deconvoluted from DICHROWEB webserver to get the secondary structure estimation. The CD data suggest an increase in α-helical character for NΔ31CΔ100 KorB as compared to secondary structure estimate for CΔ100 KorB.

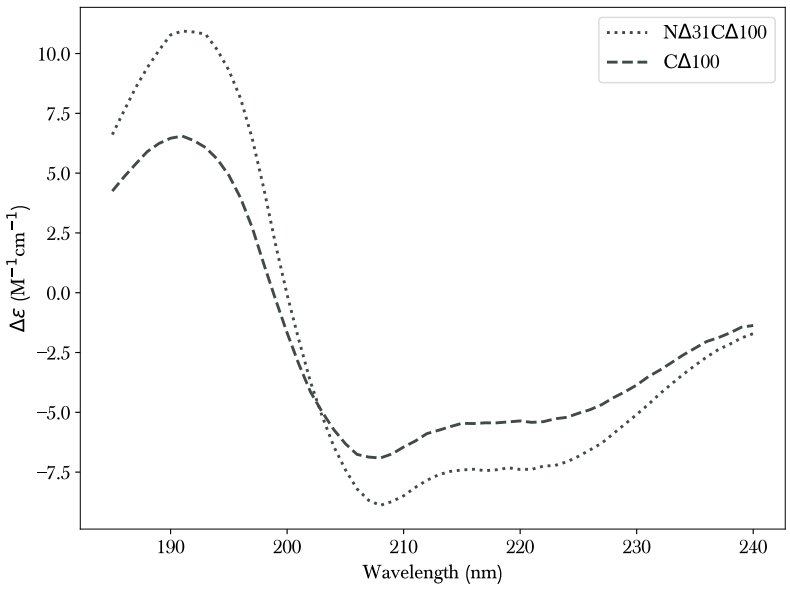

Figure 52: Circular dichroism data comparison of CΔ100 KorB and NΔ31CΔ100 KorB. CD spectra of CΔ100 KorB and NΔ31CΔ100 KorB, shown in the range of 190-240 nm. Delta epsilon ($\Delta\epsilon$) is also known as molar circular dichroism and is measured in per residue molar absorption units of circular dichroism.

The CD spectra for both proteins were individually deconvoluted from the web-server DICHROWEB to get the secondary structure estimation. Table 19 lists the percentage of secondary structure elements *viz.* α-helix, β-strand, turn and disorder character for CΔ100 KorB and NΔ31CΔ100 KorB. The CD data suggest an increase in α-helical character for NΔ31CΔ100 KorB and decrease in turn, disorder and β-strand character for NΔ31CΔ100 KorB when compared to CΔ100 KorB secondary structure estimates.

KorB construct	**α-helix (%)**	**β-strand (%)**	**Turn (%)**	**Disorder (%)**
NΔ31CΔ100 KorB	81	5	5	8
CΔ100 KorB	63	9	10	18

Table 19: Secondary structure deconvolution from CD data for CΔ100 KorB and NΔ31CΔ100 KorB. CD data deconvolution for CΔ100 KorB and NΔ31CΔ100 KorB from DICHROWEB analysis webserver. The data suggest an increase in α-helical character and decrease in turn, disorder and β-strand character for NΔ31CΔ100 KorB when compared to CΔ100 KorB.

5.4 Crystallisation trials of CΔ100 KorB and NΔ31CΔ100 KorB

To elucidate the structure of DNA-binding domain of KorB, crystallisation trials were setup both for CΔ100 KorB and NΔ31CΔ100 KorB. The crystallisation trays for both proteins were set using two sparse screens with varied precipitants. The best condition from the sparse matrix screen that facilitated micro-crystal growth for both CΔ100 KorB and NΔ31CΔ100 KorB was 0.1 M ammonium sulfate, 0.1 M Bis-tris, pH 5.5, 30% (w/v) PEG 3350 and about 100 μm sized protein crystals were observed for both proteins. In order to obtain bigger, diffraction quality protein crystals for both proteins, the initial conditions were varied with trials varying salt concentrations (0.1-0.2 M) and varying PEG 3350 (20-30%) in the presence and absence of 5% glycerol, but the micro-crystals failed to grow in size with time.

5.5 Discussion

Two deletion mutants of KorB (CΔ100 KorB and NΔ31CΔ100 KorB) capable of binding DNA were over-expressed and purified to homogeneity. From 1 L of culture, approximately 25 mg of CΔ100 KorB and 20 mg of NΔ31CΔ100 KorB were obtained, sufficient amount to perform the CD experiments and DNA-binding experiments discussed in the next chapter. In the size exclusion chromatogram, CΔ100 KorB was observed as a single species, devoid of any impurities, whilst impurities can be seen in the NΔ31CΔ100 KorB purification SDS-PAGE gel. Both proteins behave as a monomer on the size exclusion column.

Crystallisation trials of both deletion mutants were set using two sparse screens with varied precipitants. In all the four screens, precipitation was observed for almost one-third of the conditions and micro-crystals were observed for both mutants but the crystals failed to grow in size with time. The difficulty to crystallise these deletion mutants can be because of the disorder in the protein, possibly due to the flexible N-terminal domain. The NΔ31CΔ100 KorB was observed to be more stable at RT when compared to CΔ100 KorB. This can be explained with the removal of the first 53 residues from the N-terminal of CΔ100 KorB as those residues are possibly disordered.

CD data predict that NΔ31CΔ100 KorB has 86% (α-helix: 81%; β-strand: 5%) secondary structure character, whilst 72% (α-helix: 63%; β-strand: 9%) secondary structure character was observed for CΔ100 KorB. The data suggest an increase in α-helical character for NΔ31CΔ100 KorB and appears to be more ordered when compared to CΔ100 KorB possibly because of the loss of the first 53 disordered residues from the N-terminal of CΔ100 KorB. In the next chapter, using a combination of techniques (including fluorescence anisotropy, circular dichroism and microscale thermophoresis), the DNA-binding of both the deletion mutants capable of binding DNA along with the wild-type KorB are described and discussed.

6

Interaction of KorB with DNA and KorA

6 Interaction of KorB with DNA and KorA

6.1 Introduction

To obtain data on structural and functional aspects of biomolecules, there has been a great push in applying a wide range of biophysical techniques *e.g.* to probe the workings of proteins in physiologically relevant conditions. These include Fluorescence Polarisation (FP), Förster Resonance Energy Transfer (FRET), Fluorescence Anisotropy (FA), CD, NMR and many more (Ishima and Torchia, 2000; Kelly et al., 2005; Lakowicz, 2006). In addition a relatively new technique, Microscale Thermophoresis (MST, discussed in section 6.2.3) has been extensively used to study protein-protein and protein-DNA interactions (Seidel et al., 2013). FA, CD, NMR and MST were extensively used in this body of work to study protein-protein and protein-DNA interactions and are discussed below in detail.

Following the law of mass action with binding partners L (ligand) and P (protein) forming complex LP, the binding of a ligand-protein complex can be described as

$$\mathrm{L} + \mathrm{P} \rightleftharpoons \mathrm{LP} \tag{6.1}$$

and the dissociation constant K_d at equilibrium is defined as

$$K_d = \frac{[\mathrm{L}]_{free}\,[\mathrm{P}]_{free}}{[\mathrm{LP}]} \tag{6.2}$$

where $[\mathrm{L}]_{free}$ and $[\mathrm{P}]_{free}$ are the concentrations of free, unbound partners *i.e.* L and P and [LP] is the concentration of the complex.

If [P], [L] and [LP] represent molar concentrations of protein, ligand and complex respectively, then

$$[L] = [L]_{free} + [LP] \tag{6.3}$$

$$[P] = [P]_{free} + [LP] \tag{6.4}$$

and equation 6.2 adjusted accordingly results in

$$K_d = \frac{([L] - [LP])([P] - [LP])}{[LP]} \tag{6.5}$$

In this chapter, equation 6.5 serves as the main expression for the KorB-DNA binding studies using FA, CD and MST.

6.2 Theoretical background of techniques

6.2.1 Fluorescence Anisotropy

6.2.1.1 Introduction

Fluorescence anisotropy (FA) is used to determine biomolecular interactions and measure dissociation constants of interactions such as protein with DNA. FA has been described as a phenomenon where unequal intensity of light is emitted by a fluorescent chemical compound *i.e.* fluorophore along different axes of polarisation (Lakowicz, 2006, page: 353-378). FA measurements of dissociation constants is based on the change in rotational time and tumbling rates of the molecules on complex formation. In order to calculate the dissociation constant, the unlabelled partner is titrated against the labelled partner (fluorophore). A bound fluorophore will tumble at a slower rate in comparison to an unbound state. If binding is observed, the anisotropy increases from

an initial value (free to partially bound) and saturates at fully bound state. The change in anisotropy during the titration is used to calculate the degree of binding.

FA is a dimensionless quantity as it is a measurement of ratio of intensities. The difference in the intensity of fluorescence at two angles perpendicular to each other is normalised by the total intensity (equation 6.6). The magnitude of FA is thus independent of the concentration of the fluorophore and does not depend on the intensity of the sample.

6.2.1.2 Theoretical background

Light can be understood as a wave travelling in space and can have multiple orientations. Once the light passes through a polarisation filter, it is polarised. In an FA experiment, vertically polarised light is used to excite the sample in the cuvette. The sample is a fluorophore attached to a biomolecule (in this study fluorophore is Fluorescein attached at the 3′ end of DNA). Light excites the fluorophore and light of the same orientation is emitted. Molecules in solution are in motion and since the fluorophore is tumbling with the DNA, the orientation of emitted light changes too. The strength of the emitted signal changes according to the orientation of the emitted light, when passed through another polarisation filter at the same angle. If the orientation of emitted light is perpendicular to the filter then the signal is completely blocked, however the signal is maximum if it is parallel to the filter. Upon protein binding to DNA, the molecular weight of the complex increases and the tumbling rate decreases. Because of the slower movement, the polarisation of light is conserved for a longer time. An increase in molecular weight results in an increased anisotropy signal.

In this study, Fluorescein was used because of its relatively broad emission spectrum and good water solubility. The instrumentation and the phenomena of fluorescence anisotropy is highlighted in figure 53A. The sample in the cuvette is excited with vertically polarised light (Polariser1) oriented parallel to the Z-axis. The intensity of the emitted light is measured after it passes through another polariser (2). When the emis-

sion polariser (2) is oriented parallel (‖) and perpendicular (⊥) to the direction of the polarised excitation light, the intensity is called $I_{||}$ and $I_{\perp}$ respectively. The mathematical relationship between emitted light intensities and anisotropy is given in equation 6.6. In figure 53B, a fluorophore is attached to one of the binding partners (Partner1) and because of its small size, the molecule tumbles rapidly. When a large binding partner (Partner2) binds Partner1, the overall complex tumbles slowly and an increase in anisotropy is observed. Mathematically FA is defined as

$$\mathrm{FA} = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{6.6}$$

A similar and related phenomenon to FA is Fluorescence Polarisation (FP) and FP is defined as

$$\mathrm{FP} = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{6.7}$$

FA and FP are related as

$$\mathrm{FA} = \frac{2 * FP}{3 - FP} \tag{6.8}$$

FA and FP contain the same information and the use of FA is encouraged because FA is a normalised measurement *i.e.* the difference in intensities ($I_{||} - I_{\perp}$) is normalised by the total intensity ($I_{||} + 2I_{\perp}$), resulting in simplification of equations.

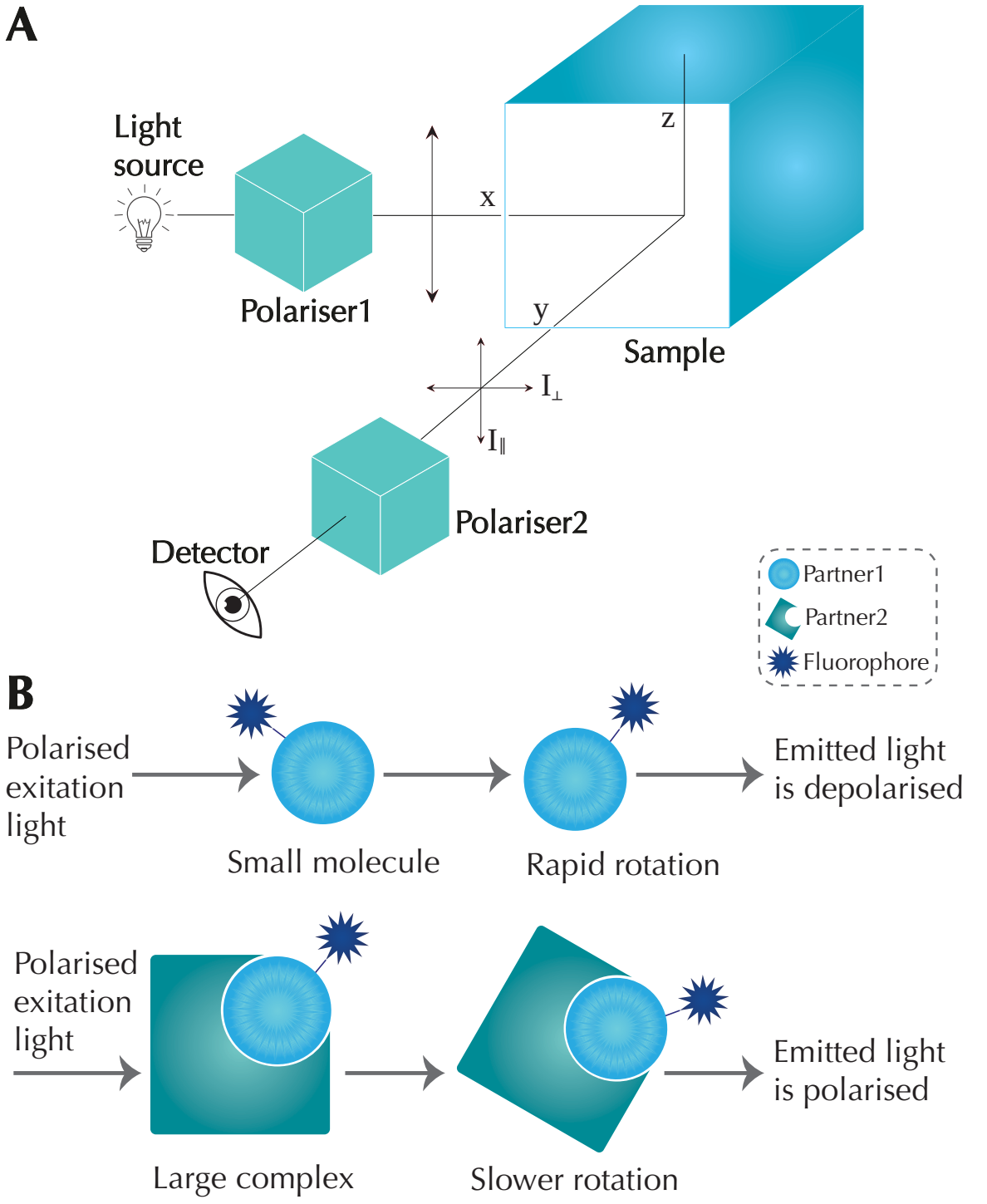

Figure 53: Schematic representing the technology of Fluorescence Anisotropy. (A) The sample in the cuvette is excited with vertically polarised light (Polariser1) orienting parallel to the Z-axis. The intensity of the emitted light is measured with the help of another polariser (Polariser2). When the emission polariser (Polariser2) is oriented parallel ($\|$) and perpendicular ($\perp$) to the direction of the polarised excitation light, the intensity is called $I_{\|}$ and $I_{\perp}$ respectively. The mathematical expression for anisotropy is given in equation 6.6. **(B)** Illustration of principle of FA where a bound fluorophore tumbles at a slower rate in comparison to an unbound state, leading to an increase in anisotropy.

6.2.2 Circular Dichroism

6.2.2.1 Introduction

Circular dichroism (CD) is a biophysical technique with a wide gamut of applications in different fields. In biochemistry and structural biology, CD is used to determine the secondary structure of biological molecules especially proteins. The technique of CD measures the differential absorption of left– and right–circularly polarised light and the results are given as angle of rotation in mdeg (Cary and Kneale, 2009).

6.2.2.2 Theoretical background

Plane polarised light forms the basis of CD spectroscopy. Electromagnetic radiation (EMR) can be imagined as a wave travelling in space composed of magnetic and electric fields propagating perpendicular to each other and to the direction of the propagation of the wave. When the oscillations of a light wave are confined to a single plane (*e.g.* by using a polariser), then that resultant light wave is termed as plane polarised light. If we take two, in-phase plane polarised light waves with equal amplitude, that are also perpendicular to each other, then the resultant light wave would be linearly polarised at 45°.

If two plane polarised light waves are out of phase, the resultant wave is not linearly polarised. If one of the plane polarised waves is out of phase by a quarter-wave ($\pi/_4$) with respect to the other wave, then the resultant wave would be a helix and the light wave is known as circularly polarised light (CPL). We can get helical light waves that are either left-handed circularly polarised (L-CPL) or right-handed circularly polarised (R-CPL). In circular polarisation of a wave, the strength of the electric field does not change but the direction of the wave changes in rotatory manner.

CD is based on differential absorption of the circularly polarised light from optically active molecules. Figure 54A shows the resulting circularly polarised wave (red)

because of the vectors that match the maximum of the vertical component (blue) and the horizontal component (green). CD measures the difference in absorption (δA) of left-handed circularly polarised light (A_L) and right-handed circularly polarised light (A_R). The circular dichroism of a sample of molar concentration [C] with absorbances for the two components, (A_L and A_R) is represented as (Atkins and De Paula, 2006, page: 488-489):

$$\delta\epsilon = \epsilon_L - \epsilon_R = \frac{A_L - A_R}{[C]L} \tag{6.9}$$

where L defines the path length of the cuvette used in the experiment.

6.2.2.3 Determination of secondary structure of protein

CD is a technique that analyses the conformation of proteins (Greenfield, 1996). Both X-ray crystallography (XRC) and NMR provide atomic level structural information on proteins whereas CD is a low resolution and semi-quantitative structural/biophysical technique that describes the overall secondary structural features of proteins. EMR can be characterised based on either wavelength or frequency. The EMR is exploited in CD spectroscopy to determine the secondary structure of macromolecules.

In proteins, the peptide bond is the chromophore (light-absorbing group) studied in the far-UV region (190-250 nm) and the nature of the CD signal is an indication of the conformation of the molecule. The characteristic CD signal observed for proteins with predominantly α-helix, anti-parallel β-sheet, type I β-turn, poly (Pro) II helix and unfolded structures is given in figure 54B (Kelly et al., 2005).

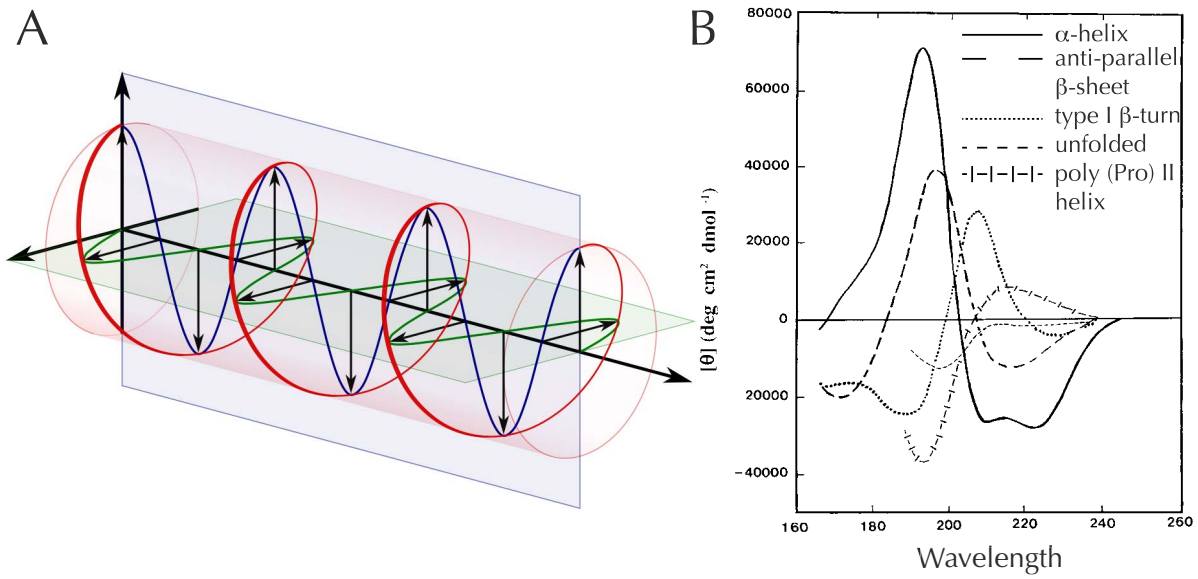

Figure 54: Schematic representing the phenomenon of circular dichroism and its application to proteins. (A) From the point of view of the source, circularly polarised wave (shown in red) is defined as left-handed or anti-clockwise. The vertical component of the wave with its propagation plane are shown in blue while the horizontal wave component and its plane are illustrated in green. Relative to the direction of propagation of the wave, the vertical component lags the leftward horizontal component by one quarter of the wavelength. The quadrature phase difference results in zero magnitude for the horizontal component and correspondingly maximum magnitude for the vertical component and vice versa. This results in the circularly polarised wave with vectors that match the maximum of the vertical and the horizontal component. **(B)** In proteins, different secondary structure elements give rise to characteristic CD spectra. A characteristic CD signal is observed for proteins with predominantly α-helix (solid line), anti-parallel β-sheet (long dashed line), type I β-turn (dotted line), poly (Pro) II helix (cross dashed line) and unfolded structures (short dashed line). The image is reproduced with permission of the rights holder, Elsevier (Kelly et al., 2005).

6.2.3 Microscale Thermophoresis

6.2.3.1 Introduction

Microscale thermophoresis (MST) is a recent technology to study interaction of biological molecules under close-to-native conditions (Jerabek-Willemsen et al., 2011; Wienken et al., 2010). The method is suitable for quantifying the binding affinities of ligands with very small amounts of protein (as low as nM) in solution (Jerabek-Willemsen et al., 2014; Seidel et al., 2013). In addition, MST can be used to test potential aggregation of biomolecules. MST is based on the phenomenon of directed movement of a particle subjected to microscopic temperature gradient. With MST, both labelled (fluorescently labelled) and label-free molecules can be used in free solution to measure thermophoresis. MST is a contact-free, optical method with minimal sample contamination and for experimentation low sample volume is required.

6.2.3.2 Theoretical background of thermophoresis

A wide range of biomolecular interactions can be measured with MST as the method monitors changes in the hydration shell, size or charge of molecules. Since ligand binding changes the hydration shell and other properties, thermophoresis changes can be used to determine binding affinities. Heating an aqueous solution locally leads to the movement of molecules along the temperature gradient and the flow of molecules is impeded by mass diffusion. At steady state, both effects balance each other resulting in a stationary spatial distribution of concentration. The relative concentration (c_{hot}/c_{cold}) is given by (Wienken et al., 2010)

$$\frac{c_{hot}}{c_{cold}} = exp[-S_T * (T - T_0)] \tag{6.10}$$

where $T - T_0$ represents the change in temperature (rise) and S_T is the Soret coefficient. S_T depends on nature of the molecule and is a measure of the strength of molecular thermophoretic flow when compared to ordinary diffusion. The theory of thermophoresis is still being debated but all theoretical explanations agree that S_T is influenced by the size of molecule and multiple surface parameters, such as charge and hydrophobicity.

Experimental evidence (Duhr and Braun, 2006) suggests that the Soret coefficient can be expressed as (Wienken et al., 2010)

$$S_T = \frac{A}{kT}\left\langle -\Delta s_{hyd}(T) + \lambda_{DH}\frac{\beta\sigma_{eff}^2}{4T\epsilon\epsilon_0}\right\rangle \tag{6.11}$$

where A is the surface area of molecule, s_{hyd} is the area-specific hydration entropy ($s_{hyd} = S_{hyd}/A$), ϵ is the dielectric constant, σ_{eff} is effective charge density, λ_{DH} is the Debye-Hueckel screening length and β is a temperature derivative arising from the ionic Gibbs-free enthalpy. Based on the equation, the Soret coefficient of a molecule depends on the its size (red term), entropy of i) ionic shielding (blue term) and ii) hydration (green term).

MST is capable of measuring biomolecular binding events with dissociation constants typically in the nM to mM range (K_d = 1 nM - 500 mM). With MST it is possible to measure binding of macromolecular complexes such as ribosomes (2.5 MDa) and smaller molecular weight molecules such as single ions (40 Da) to a target. MST experiments require limited amount of labelled compound (>1 nM) which is titrated against an unlabelled molecule with concentration in the range of $\pm$ 20-fold of expected dissociation constant for the binding. The technology of the MST is illustrated in figure 55 and a typical MST trace is described in figure 56.

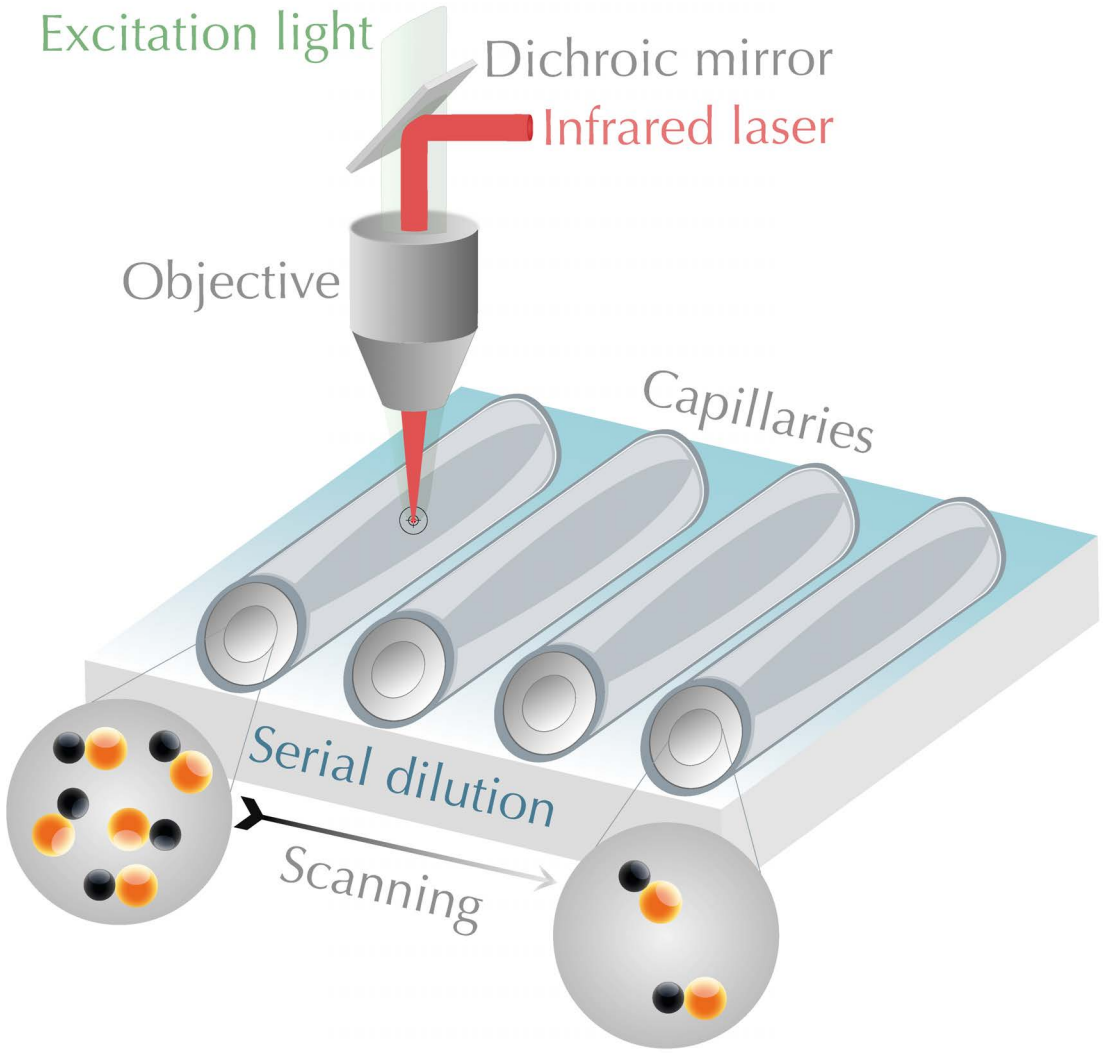

Figure 55: Schematic representing the instrumentation of MST. The Monolith NT.115 Blue-Red excites fluorescence with wavelengths 460-480 nm and 600-650 nm. Also, it detects wavelengths between 510-530 nm and 675-690 nm. Using capillary action, $\sim$5 μl of sample volume is sucked into a capillary. Then, the capillary is placed on a temperature-controlled tray and thermophoresis signal is recorded. An objective (single optical unit) is used to excite fluorophoes with a focused LED (green coloured beam) and detect the fluorescence. A dichroic mirror is placed above the optical unit that transmits the fluorescence excitation/emission and reflects the Infrared (IR)-laser. A part of the sample inside the capillary is heated the IR-laser to established the temperature gradient. The thermophoresis for the fluorescent molecules is detected with a charge-coupled device (CCD) camera housed in the optical unit. The figure is adapted from Jerabek-Willemsen et al., 2014.

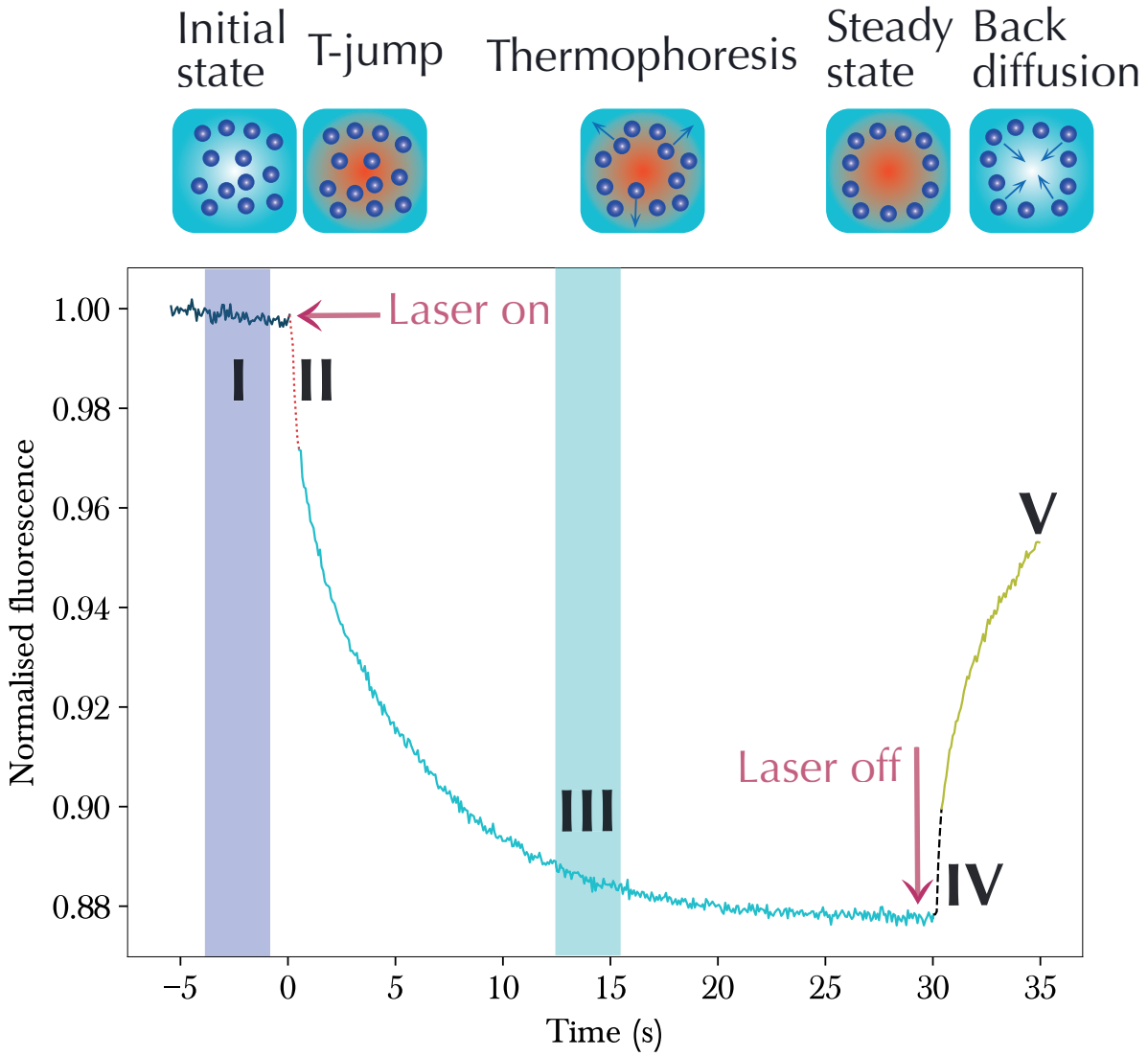

Figure 56: Schematic representing a typical MST trace. A typical MST signal trace recorded for a given capillary. The molecules in the sample are distributed homogeneously and a constant-initial fluorescence is detected for ∼5 s (I). As soon as the IR-laser is turned on, a temperature jump or T-jump (II) is observed. Depending on the fluorophore used, the T-jump can be either positive or negative. The T-jump results because of a sudden (∼100 ms time scale) change in fluorescence, followed by thermophoresis-induced motion (diffusion) of molecules (III). In this instance, the decrease in fluorescence is measured for ∼30 s, till steady-state is achieved (IV). Afterwards the IR-laser is turned off resulting in an inverse T-Jump (IV). Stage V is marked by the back-diffusion of molecules as a result of pure mass diffusion. The fluorescence signal at stage III is normalised to the signal at stage I to glean information about the molecular interactions. The nature of (bio)molecular interactions can influence the change in T-jump and thermophoresis.

6.3 Data analysis

All the KorB-DNA interaction experiments (FA, CD and MST) used addition of protein at every titration step while the amount of DNA remained constant. The titration data thus obtained were used to calculate the KorB-DNA dissociation constants. The KorB-DNA interactions (monophasic or single binding event) observed with FA, CD and MST were fitted using non-linear regression to a predefined hyperbola equation, 'Single Rectangular I, 3 Parameter' in Sigmaplot to calculate the dissociation constant (K_d).

$$f = y_o + \frac{a * x}{K_d + x} \tag{6.12}$$

where y is the signal at x and f is fitted to y, y_o is the initial (non-zero) value of the signal (from FA, CD or MST), a is the difference between the maximum and minimum signal *i.e.* max(y)–min(y), x is the protein concentration and K_d is the KorB-DNA dissociation constant. Equation 6.12 is correct if DNA concentration is considered to be negligible and protein concentration is in excess. Only monomeric protein concentration was used for dissociation constant calculations.

From raw anisotropy data, averaged anisotropy values (mean of five data measurements at a given protein concentration) were plotted against protein concentration. The raw data for KorB-DNA binding were fitted with the hyperbola equation 6.12 to calculate the dissociation constants. The same equation 6.12 was used to calculate the dissociation constants from CD data. However in case of CD experiments, the DNA concentration in the cuvette was comparable to the protein concentration and the DNA concentration was in the range of dissociation constant. In this scenario, the hyperbola equation 6.12 cannot be used as the presence of DNA cannot be neglected. In order to correct for the DNA concentration, the standard quadratic equation was modified

(6.13) and both protein and DNA concentrations were used to calculate the dissociation constant.

$$f = y_o + a * \frac{x + z + K_d - \sqrt{(x + z + K_d)^2 - 4xz}}{2z} \tag{6.13}$$

Here, y_o is the initial (non-zero) value of CD signal, a is the difference between the maximum and minimum CD signal *i.e.* max(y)–min(y), x is the protein concentration, z is the DNA concentration (constant), and K_d is the dissociation constant.

For single step binding events recorded with MST, equation 6.12 was used to calculate the dissociation constants but most of the KorB constructs bind O_B in a two-step manner. A predefined 'Double Rectangular, 5Parameter' equation in Sigmaplot was modified to fit the data and calculate the two dissociation constants for the two distinct binding events.

$$f = y_o + \frac{a * x}{K_{d1} + x} + \frac{c * x}{K_{d2} + x} \tag{6.14}$$

y_o is the initial MST signal value, a and c are the difference between maximum and minimum MST signal for first and second binding interaction respectively, x is the protein concentration, and K_{d1} and K_{d2} are the dissociation constants for first and second interaction respectively.

A couple of assumptions are employed while fitting the data with non-linear regression. The first is that the independent variable is known precisely and all the errors are in the dependent variable. Secondly, the dependent variable is measured independently of the independent variable. A number of statistical values were calculated from Sigmaplot, which are useful to validate the fitting and various terms are described below. t is the ratio of the coefficient (K_d) to the standard error in this parameter (a mea-

sure of precision) and higher values of t suggest greater precision. P is the probability of the event occurring by random chance and correlates to obtaining the t value by chance. R^2 is the coefficient of regression, a measure of how well the regression model fits the actual data. The value of R^2 varies between 0 and 1. A higher value of R^2 is predictive of a stronger relationship between independent and dependent variable. R^2 determines the degree of certainty in making predictions from a fitted model. Standard error* is the error in the estimation for the value of R^2. In this chapter, the dissociation constants (K_d) are expressed in μM unless stated otherwise.

6.4 Results of KorB-DNA interactions

6.4.1 Fluorescence Anisotropy

KorB constructs *viz.* KorB wild-type (WT), CΔ100, NΔ31CΔ100 and N-terminal domain of KorB were individually titrated against fluorescently labelled (fluorescein at the 3′ end) double stranded O_B DNA (dsHL and dsFL); dsFL and dsHL stands for double stranded full-length O_B DNA and double stranded half-length O_B DNA respectively. The sequences of O_B DNA containing oligonucleotides used for KorB-DNA interactions are given earlier in table 13. The concentration range of the KorB constructs along with the concentration of O_B DNA used in the fluorescence anisotropy study are listed in the table 20.

During the course of titration, two measurements were recorded, a) fluorescence intensity and b) anisotropy. The fluorescence intensity was observed to decrease due to the addition of protein solution. When the fluorescence intensity was measured with respect to the total volume of solution added, the intensity decreased linearly with dilution rather than protein concentration *i.e.* it was a dilution effect, not a quenching effect. In contrast, an increase in the anisotropy was recorded upon addition of KorB constructs as the protein-DNA complex tumbled relatively slowly in comparison to the free DNA in solution. The anisotropy is independent of the fluorescence intensity.

KorB construct	Concentration range (μM)	O_B DNA	Concentration (μM)
Wild-type KorB	0-11	dsFL	0.24
Wild-type KorB	0-9	dsHL	0.5
CΔ100 KorB	0-61	dsFL	0.24
CΔ100 KorB	0-31	dsHL	0.5
NΔ31CΔ100 KorB	0-27	dsFL	0.24
NΔ31CΔ100 KorB	0-29	dsHL	0.5
NTD KorB	0-55	dsFL	0.24
NTD KorB	0-55	dsHL	0.5

Table 20: KorB and O_B DNA concentrations used in the FA study.

The maximum difference in anisotropy values, a for each KorB construct is shown in table 21. The difference in the anisotropy values seems to vary between the constructs and appears to have no particular correlation with protein size, DNA type or dissociation constant.

Figure 57A and 57B show the data for wild-type KorB titration against dsFL O_B and dsHL O_B DNA respectively. Figure 58A and 58B show the data for CΔ100 KorB titration against dsFL O_B and dsHL O_B DNA respectively. Figure 59A and 59B show the data for NΔ31CΔ100 KorB titration against dsFL O_B and dsHL O_B DNA respectively. The dissociation constants with pertinent statistics for KorB WT, CΔ100, NΔ31CΔ100 with O_B DNA (dsHL and dsFL) are listed in table 21. As it was postulated, an increase in anisotropy was not observed when NTD KorB was titrated even in a large excess ($>$50 μM) with O_B DNA (both dsHL and dsFL). From the FA data, the CΔ100 KorB and NΔ31CΔ100 KorB bound to the O_B more weakly when compared to the KorB WT.

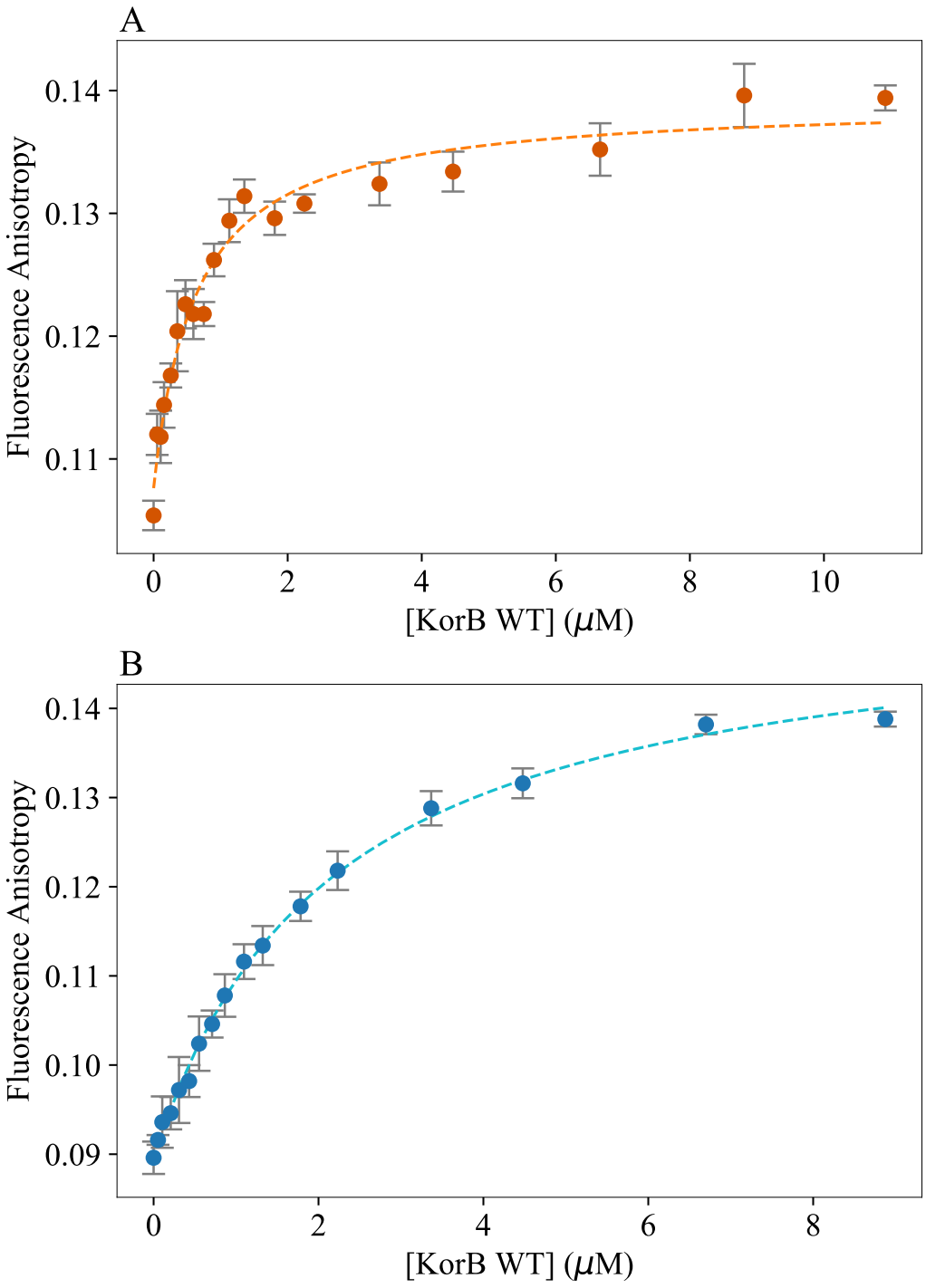

Figure 57: KorB WT-O_B DNA interaction studied with anisotropy. (A): KorB WT was titrated against dsFL O_B DNA and anisotropy was recorded. Anisotropy values as mean of five data points (dark orange circle) ± standard deviation (grey bar) are plotted against increasing KorB WT concentration. **(B)**: KorB WT was titrated against dsHL O_B DNA. Anisotropy values as mean of five data points (blue circle) ± standard deviation (grey bar) are plotted against increasing KorB WT concentration. For both the experiments, the anisotropy values were monitored every 5 s for every titration step of protein addition. The data for both the experiments were fitted using equation 6.12 and the fit is shown with dashed lines (orange for KorB WT-dsFL and cyan for KorB WT-dsHL interaction respectively).

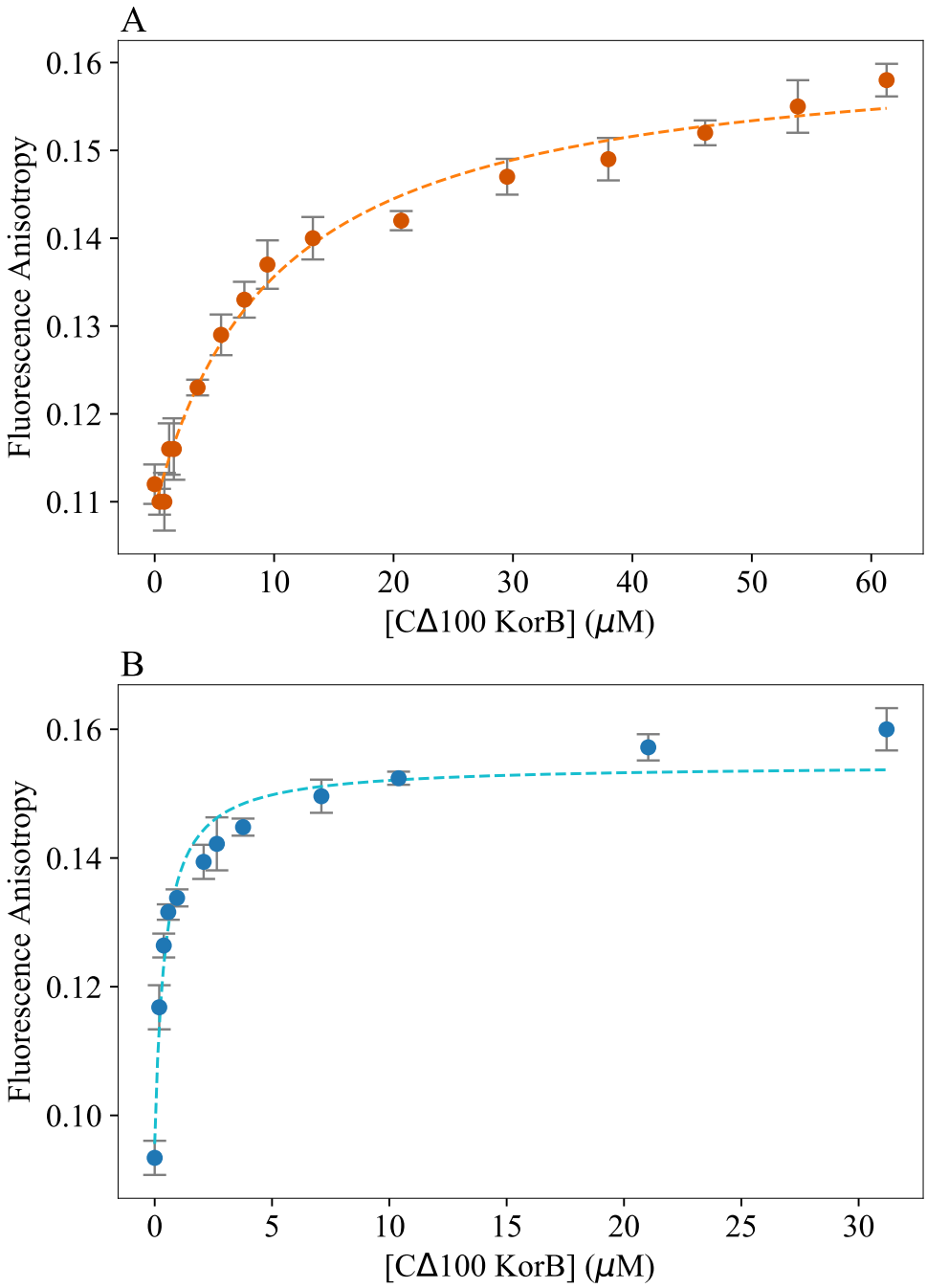

Figure 58: CΔ100 KorB-O_B DNA interaction studied with anisotropy. (A): CΔ100 KorB was titrated against dsFL O_B DNA and anisotropy was recorded. Anisotropy values as mean of five data points (dark orange circle) ± standard deviation (grey bar) are plotted against increasing CΔ100 KorB concentration. **(B)**: CΔ100 KorB was titrated against dsHL O_B DNA. Anisotropy values as mean of five data points (blue circle) ± standard deviation (grey bar) are plotted against increasing CΔ100 KorB concentration. For both the experiments, the anisotropy values were monitored every 5 s for every titration step of protein addition. The data for both the experiments were fitted using equation 6.12 and the fit is shown with dashed lines (orange for CΔ100 KorB-dsFL and cyan for CΔ100 KorB-dsHL interaction respectively).

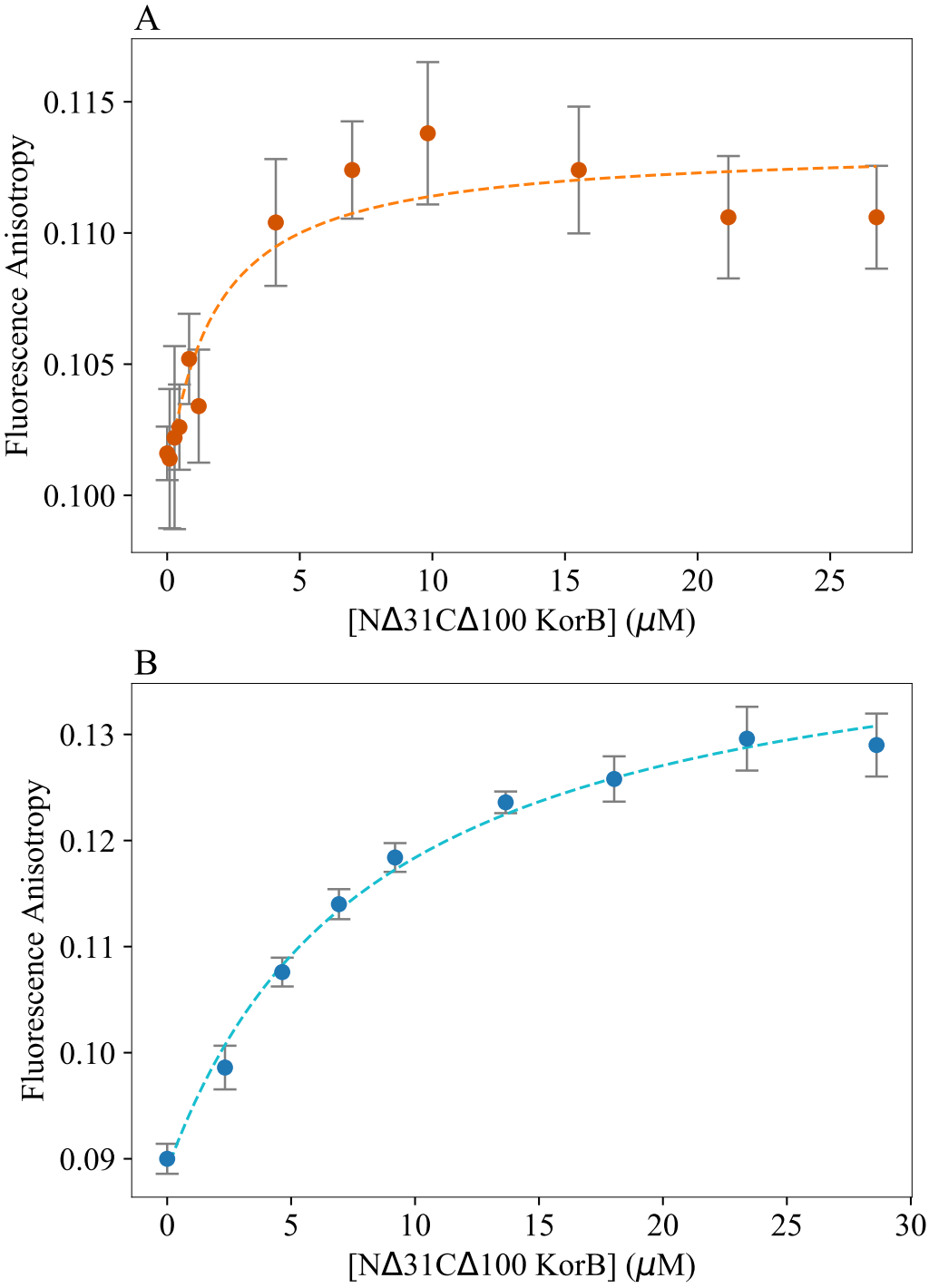

Figure 59: NΔ31CΔ100 KorB-O_B DNA interaction studied with anisotropy. (A): NΔ31CΔ100 KorB was titrated against dsFL O_B DNA and anisotropy was recorded. Anisotropy values as mean of five data points (dark orange circle) ± standard deviation (grey bar) are plotted against increasing NΔ31CΔ100 KorB concentration. **(B)**: NΔ31CΔ100 KorB was titrated against dsHL O_B DNA. Anisotropy values as mean of five data points (blue circle) ± standard deviation (grey bar) are plotted against increasing NΔ31CΔ100 KorB concentration. For both the experiments, the anisotropy values were monitored every 5 s for every titration step of protein addition. The data for both the experiments were fitted using equation 6.12 and the fit is shown with dashed lines (orange for NΔ31CΔ100 KorB-dsFL and cyan for NΔ31CΔ100 KorB-dsHL interaction respectively).

KorB construct	**O_B DNA**	**y_0**	**a**	**K_d (μM)**	**$\pm$Std. Error**	**t**	**P**	**R^2**	**$\pm$Std. Error***
Wild-type	dsFL	0.107	0.032	**0.63**	**0.1**	**6.3**	**<0.0001**	0.937	0.0024
	dsHL	0.089	0.063	**2.2**	**0.14**	**15.8**	**<0.0001**	0.984	0.0021
CΔ100	dsFL	0.109	0.053	**10.1**	**1.5**	**6.7**	**<0.0001**	0.987	0.0021
	dsHL	0.096	0.059	**0.44**	**0.05**	**8.9**	**<0.0001**	0.945	0.0044
NΔ31CΔ100	dsFL	0.101	0.013	**1.8**	**0.93**	**1.9**	**0.0842**	0.910	0.0016
	dsHL	0.089	0.054	**8.4**	**0.98**	**8.6**	**<0.0001**	0.970	0.0024
NTD KorB	dsFL	-	-	-	-	-	-	-	-
	dsHL	-	-	-	-	-	-	-	-

Table 21: Binding parameters of KorB constructs with O_B DNA calculated from the fluorescence anisotropy data. The dissociation constants (K_d) with standard error (Std. Error) for KorB wild-type (WT), CΔ100, NΔ31CΔ100 with O_B DNA (dsHL and dsFL) are calculated using the equation 6.12. t is the ratio of K_d to standard error, a measure of precision and P is the probability of obtaining the value of t by chance. The Std. Error, t and P values for respective K_d values are shown in bold. y_0 is the initial anisotropy value and a is the difference between maximum and minimum anisotropy. R^2 is the coefficient of regression and standard error (Std. Error*) is the error in estimation of R^2. Addition of NTD KorB (even in excess) to O_B DNA did not increase the anisotropy, suggesting the absence of binding to O_B as depicted by dashes. BSA was used as a negative control for the anisotropy experiments.

6.4.2 Circular Dichroism

To determine the binding affinities of KorB mutants with O_B DNA, CD spectra of O_B DNA were examined in the absence and presence of increasing concentrations of two KorB mutants *viz.* CΔ100 KorB and NΔ31CΔ100 KorB. The concentrations of KorB constructs and double stranded O_B DNA used for CD experiments are listed in table 22.

CD spectra for DNA (± protein) were measured between 340-240 nm whereas spectra for the protein alone were measured in the range of 340-185 nm. Around 280 nm (near-UV region of the spectrum), signals from the DNA dominate and the measured signals for protein are weaker (10-100 fold) (Hyde et al., 2017).

KorB construct	**KorB concentration (μM)**	**Double stranded O_B DNA**	**O_B DNA concentration (μM)**
NΔ31CΔ100	0-80	FL	10
NΔ31CΔ100	0-80	HL	20
CΔ100	0-60	FL	15
CΔ100	0-60	HL	20

Table 22: Concentration of KorB constructs and double stranded O_B DNA used for CD experiments.

NΔ31CΔ100 KorB-O_B interaction: The binding of NΔ31CΔ100 KorB was tested with both dsHL and dsFL O_B DNA. Figure 60 and figure 62 show the titration of the protein NΔ31CΔ100 KorB with dsHL and dsFL DNA respectively. In figure 60 the titration shows an overlay of NΔ31CΔ100 KorB-dsHL O_B DNA spectra with increasing protein concentration. The CD spectrum of 20 μM of dsHL O_B DNA was recorded between 340-240 nm. Then 5 μM of NΔ31CΔ100 KorB was added to a new sample of 20 μM DNA *i.e.* DNA:protein-1:0.25 and the spectrum was recorded. Subsequently

different amounts of protein was added to the DNA samples, making DNA-protein ratios – 1:0.5, 1:1, 1:1.5, 1:2, 1:3 and 1:4. The CD spectra of the NΔ31CΔ100 KorB [n] were subtracted from those of the equivalent [(dsHL O_B DNA + NΔ31CΔ100 KorB) (1:n)] for the corresponding n ratios (spectra not shown) to get the change in DNA with increasing protein concentration.

In order to calculate the stoichiometry and the dissociation constants for KorB-DNA interactions, the ellipticity at 280 nm (maximum signal but corrected for protein concentration) was plotted against the corresponding protein concentration. Figure 61A shows the change in slope which indicates a stoichiometry of dsHL DNA:NΔ31CΔ100 KorB to be 1:1. Figure 61B shows the fitted data using hyperbola equation (6.12) for dsHL DNA and NΔ31CΔ100 KorB.

In figure 62, the titration shows an overlay of NΔ31CΔ100 KorB-O_B dsFL DNA spectrum with increasing protein concentration. CD spectrum for 10 μM of O_B dsFL DNA was recorded between 340-240 nm. Then 5 μM of NΔ31CΔ100 KorB was added to another aliquot containing DNA at 20 μM *i.e.* DNA:protein-1:0.5 and the spectrum was recorded. Subsequently different amounts of protein was added to the DNA samples, making DNA-protein ratios – 1:1, 1:2, 1:4, 1:6 and 1:8. The respective protein spectra were subtracted from the DNA+protein spectra (not shown) to get the change in DNA signal with increasing protein concentration.

The dissociation constants calculated with hyperbola equation for NΔ31CΔ100 KorB (monomer) with dsHL and dsFL O_B DNA were 21$\pm$3.5 μM and 13.7$\pm$3.7 μM respectively, whilst dissociation constants for NΔ31CΔ100 KorB (using a modified quadratic equation, monomer) with dsHL and dsFL O_B DNA were 5.8$\pm$0.8 μM and 6.2$\pm$2.2 μM respectively. When the binding of dsFL was considered with NΔ31CΔ100 KorB concentration as a dimer, the calculated K_d with modified quadratic equation was reduced to 1.1$\pm$0.7 μM. Figure 63A shows the change in slope that indicates a stoichiometry of dsFL DNA:NΔ31CΔ100 KorB to be 1:2. Figure 63B shows the fitted data using hyperbola equation (6.12) for dsFL DNA and NΔ31CΔ100 KorB.

In figure 64, CD spectrum for titration of dsFL O_B with CΔ100 KorB in the range of 240-320 nm is shown. 15 μM of dsFL O_B DNA was recorded for 340-240 nm. Then 15 μM of CΔ100 KorB was added to another aliquot of DNA *i.e.* DNA:protein-1:1 and spectrum was recorded. Additional recorded spectra were overlaid with subsequent samples of the protein and DNA making DNA:protein 1:2, 1:4, 1:8, 1:10, 1:16, 1:20 and 1:30. The CD spectra of the CΔ100 KorB [n] were subtracted from those of the equivalent [(dsFL O_B DNA + CΔ100 KorB) (1:n)] for the corresponding n ratios (spectra not shown). The CD signal below 250 nm becomes more negative at higher concentration of the proteins because the protein shows maximum negative ellipticity around 220 nm. The inset (figure 64) shows the zoomed region between 270 and 290 nm where the DNA signal dominates.

In figure 65A, the stoichiometry of the dsFL O_B DNA interaction with CΔ100 KorB is shown to be 1:4 for DNA:protein (monomer). The dissociation constant for the interaction is 27.3$\pm$9.4 μM and calculated with non-linear regression analysis - modified quadratic equation 6.13 given on page 153 and the fit is shown in figure 65B.

In figure 66 the titration shows an overlay of CΔ100 KorB-dsHL DNA spectrum with increasing protein concentration. Figure 67 shows the interaction of the dsHL O_B DNA with CΔ100 KorB. Figure 67A shows the stoichiometry for DNA:protein to be 1:2 and figure 67B shows the fit and the constant was calculated to be 9.1 $\pm$6.1μM.

The dissociation constants were calculated with hyperbola equation for CΔ100 KorB (monomer) with dsHL and dsFL O_B DNA were 17.7$\pm$3.8 μM and 27.6$\pm$9.3 μM respectively, where as the dissociation constants for CΔ100 KorB (modified quadratic equation, monomer) with dsHL and dsFL O_B DNA were 6.4$\pm$3.6 μM and 22.6$\pm$10.2 μM respectively. When the binding of dsFL was considered with CΔ100 KorB as a dimer, the calculated K_d with modified quadratic equation was reduced to 5.9$\pm$1.7 μM. Dissociation constants with pertinent statistics for NΔ31CΔ100 and CΔ100 with O_B DNA (dsHL and dsFL) are listed in table 23.

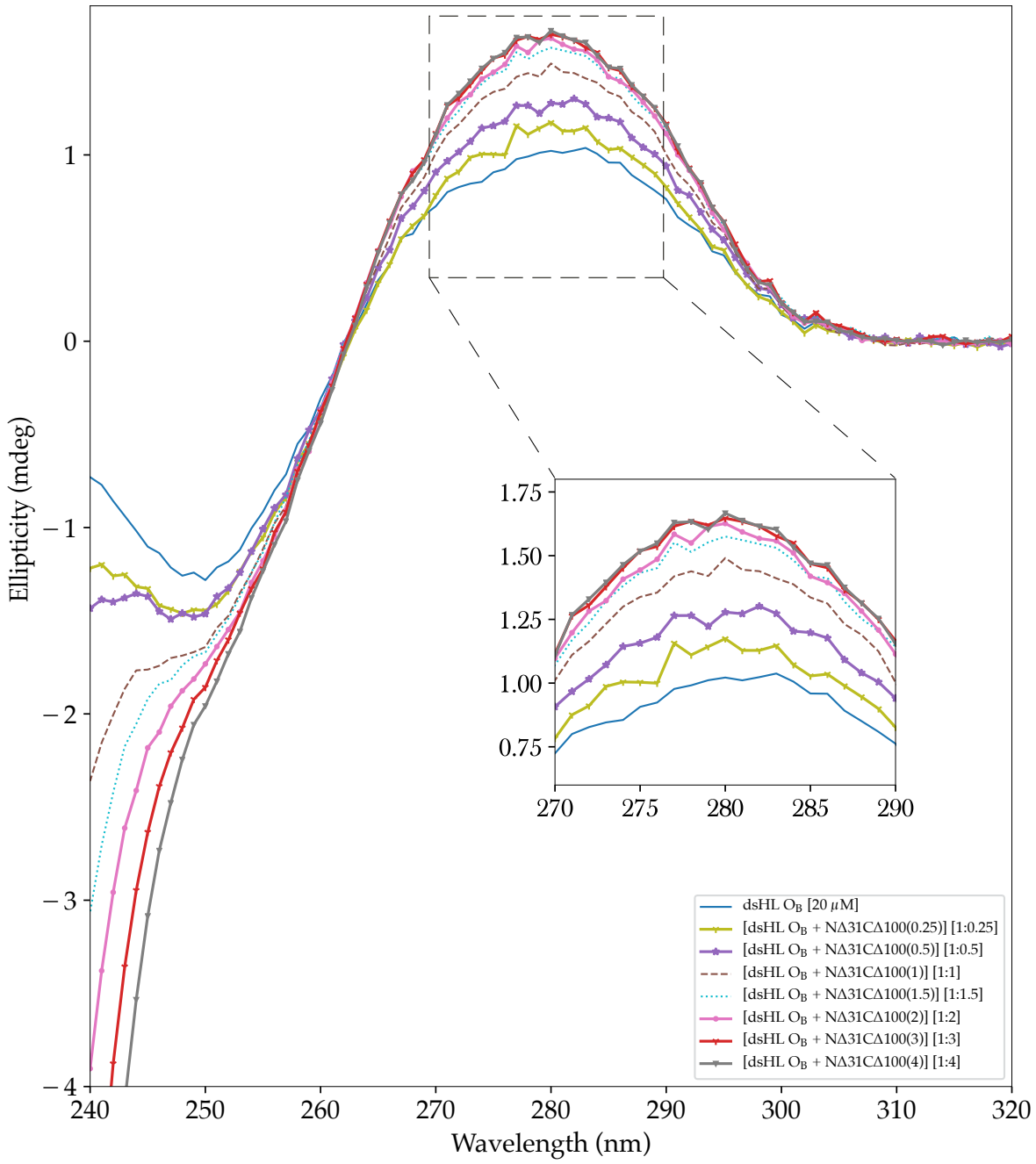

Figure 60: CD titration of the dsHL O_B DNA interaction with NΔ31CΔ100 KorB protein. The figure shows an overlay of protein-DNA CD spectra of [dsHL O_B+NΔ31CΔ100 KorB] [1:n] at various n molar ratios. Circular dichroism ellipticity is on the Y-axis and the wavelength (nm) is on the X-axis. The inset shows the DNA signal (zoomed in the range of 270-290 nm) upon protein binding (DNA signal is not corrected for protein addition).

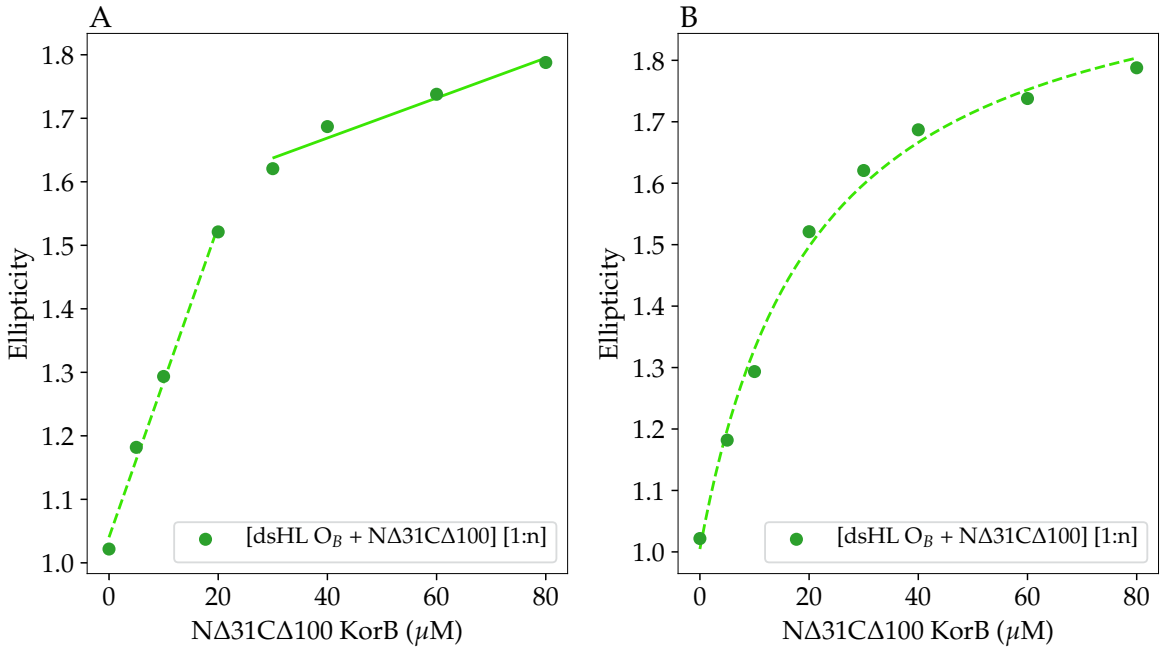

Figure 61: Interaction of the dsHL O_B DNA with NΔ31CΔ100 KorB protein. The ellipticity values at 280 nm (from figure 60 but corrected for protein concentration) are plotted against the corresponding protein concentration of NΔ31CΔ100 KorB. The DNA concentration used for this experiment is 20 μM and the concentration of NΔ31CΔ100 KorB is for monomeric form. **(A)** Stoichiometry of the dsHL O_B interaction with NΔ31CΔ100 KorB. The dashed line and solid line show the fit for increasing ellipticity signal and saturating ellipticity signal respectively. The change in slope indicates a stoichiometry of DNA:NΔ31CΔ100 KorB to be 1:1. **(B)** The data are fitted using hyperbola equation (6.12) with NΔ31CΔ100 KorB as monomer. The dissociation constant for the interaction is 21.6±3.5 μM when calculated with hyperbola equation (6.12) and the dissociation constant is 5.8±0.8 μM calculated with modified quadratic equation (6.13).

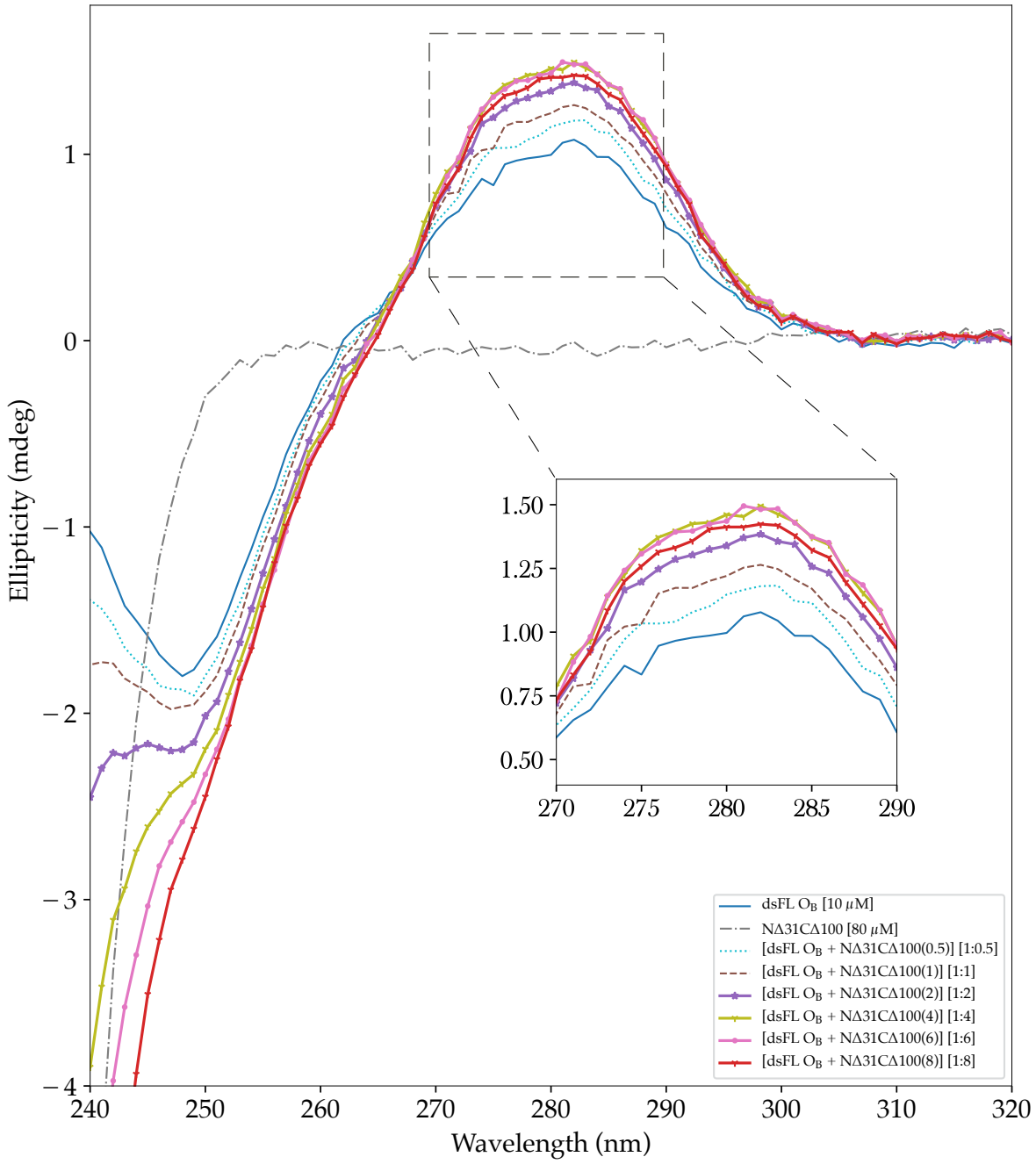

Figure 62: CD titration of the dsFL O_B DNA interaction with NΔ31CΔ100 KorB protein. The figure shows an overlay of protein-DNA CD spectra of [dsFL O_B DNA+NΔ31CΔ100 KorB] [1:n] at various n molar ratios. Circular dichroism ellipticity is on the Y-axis and the wavelength (nm) is on the X-axis. The inset shows the DNA signal (zoomed in the range of 270-290 nm) upon protein binding.

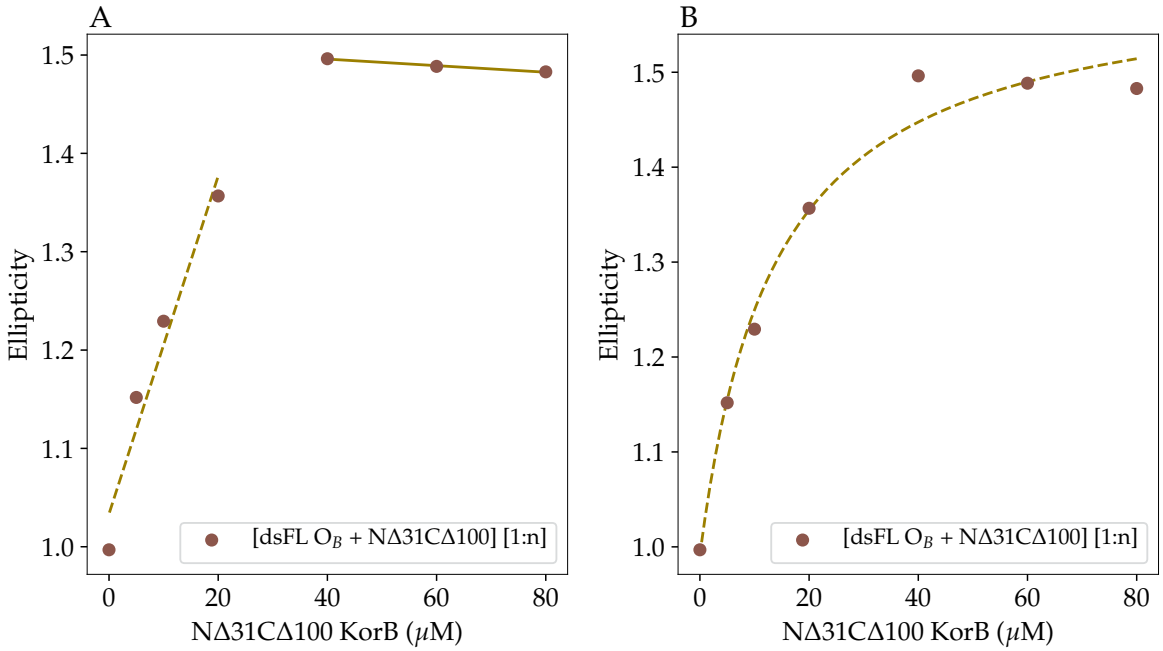

Figure 63: Interaction of the dsFL O_B DNA with NΔ31CΔ100 KorB protein. The ellipticity values at 280 nm (from figure 62 but corrected for protein concentration) are plotted against the corresponding protein concentration of NΔ31CΔ100 KorB. The DNA concentration used for this experiment is 10 μM and the concentration of NΔ31CΔ100 KorB is for monomeric form. **(A)** Stoichiometry of the dsFL O_B interaction with NΔ31CΔ100 KorB. The dashed line and solid line show the fit for increasing ellipticity signal and saturating ellipticity signal respectively. The change in slope possibly indicates a stoichiometry of DNA:NΔ31CΔ100 KorB to be 1:2. **(B)** The data are fitted using hyperbola equation (6.12) with NΔ31CΔ100 KorB as monomer. The dissociation constant for the interaction is 13.7±3.7 μM when NΔ31CΔ100 KorB is considered a monomer and calculated with hyperbola equation (6.12). The dissociation constant is 1.1±0.7 μM when NΔ31CΔ100 KorB is considered a dimer and calculated with modified quadratic equation (6.13).

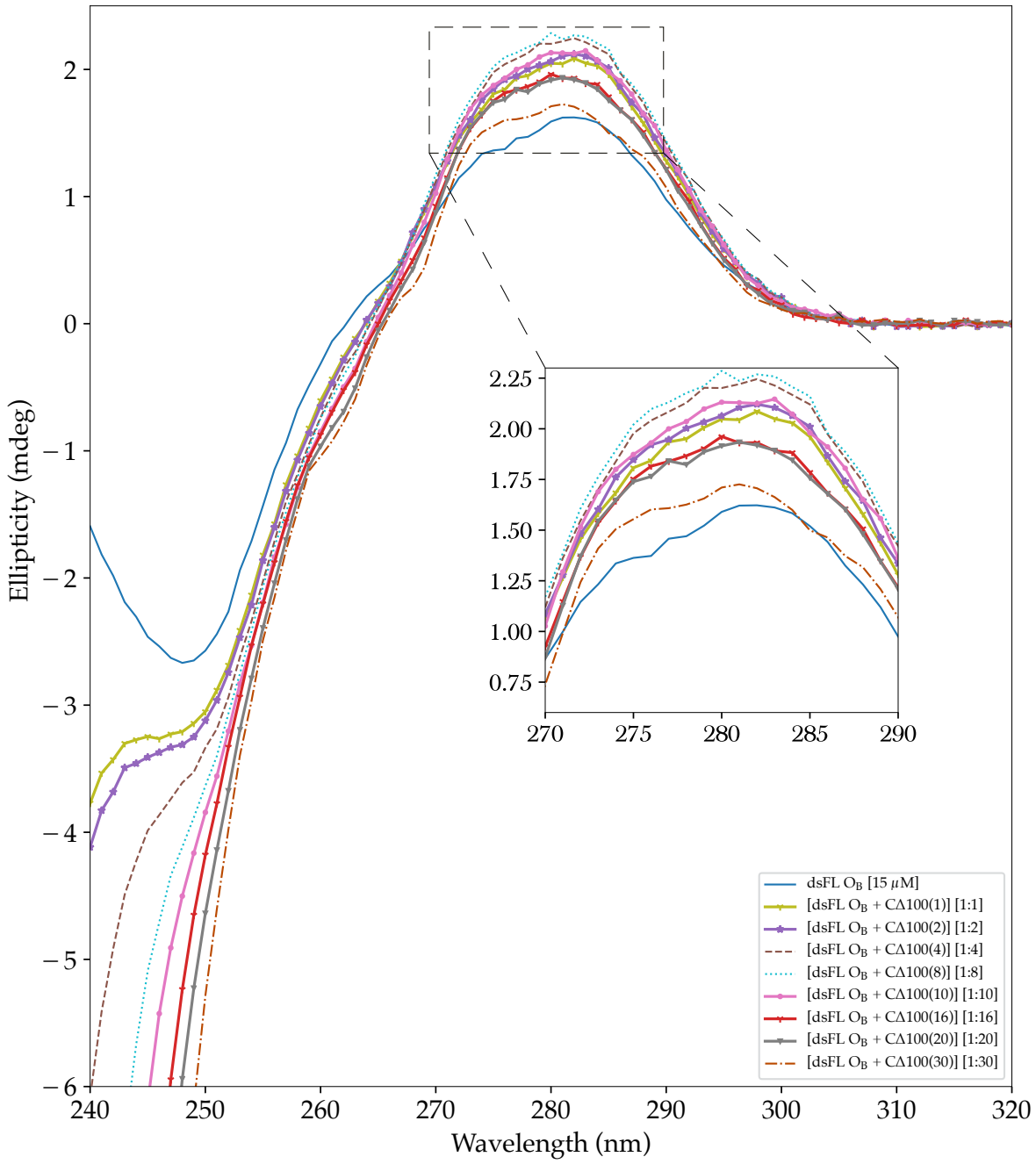

Figure 64: CD titration of the dsFL O_B DNA interaction with CΔ100 KorB protein. The figure shows an overlay of protein-DNA spectrum with increasing protein concentration. CD spectra of [dsFL O_B+CΔ100 KorB] [1:n] at various n molar ratios. Circular dichroism ellipticity is on the Y-axis and the wavelength (nm) is on the X-axis. The inset shows the DNA signal (zoomed in the range of 270-290 nm) at 280 nm upon protein binding.

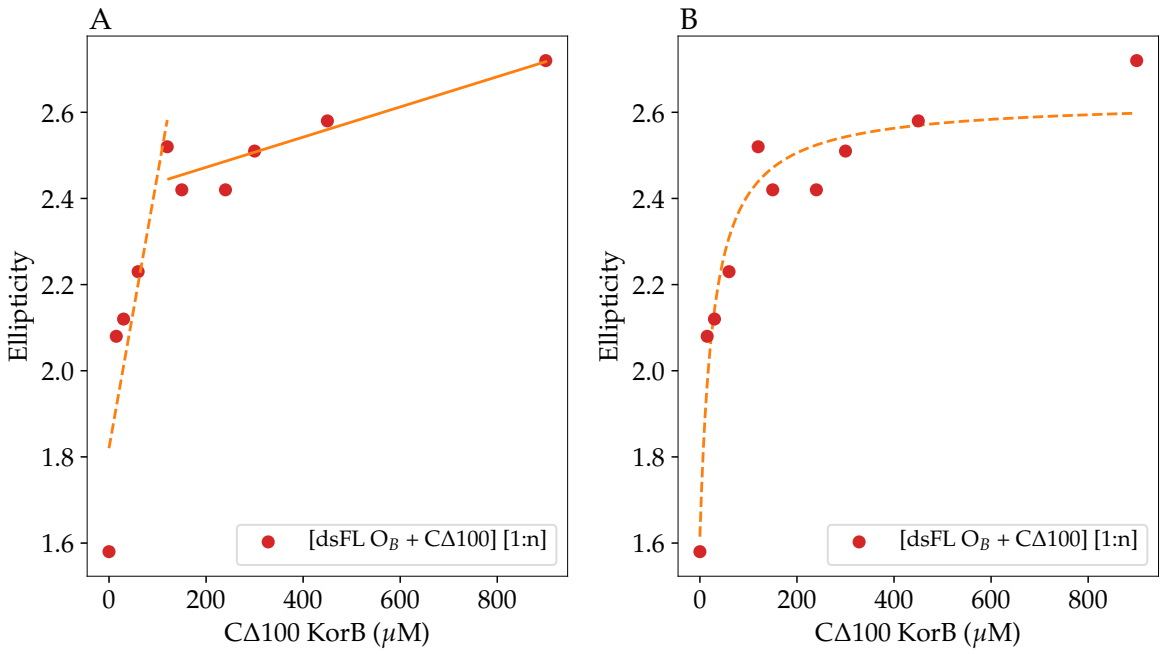

Figure 65: Interaction of the dsFL O_B DNA with CΔ100 KorB protein. The ellipticity values at 280 nm (from figure 64 but corrected for protein concentration) are plotted against the corresponding protein concentration of CΔ100 KorB. The DNA concentration used for this experiment is 15 μM and the concentration of CΔ100 KorB is for monomeric form. **(A)** Stoichiometry of the dsFL O_B interaction with CΔ100 KorB. The dashed line and solid line show the fit for increasing ellipticity signal and saturating ellipticity signal respectively. The change in slope indicates a stoichiometry of DNA:CΔ100 KorB (dimer) to be 1:4. **(B)** The data are fitted using hyperbola equation (6.12) with CΔ100 KorB as monomer. The dissociation constant for the interaction is 27.6±9.3 μM when CΔ100 KorB is considered a monomer and calculated with hyperbola equation (6.12). The dissociation constant is 9.7±5.9 μM when CΔ100 KorB is considered a dimer and calculated with modified quadratic equation (6.13).

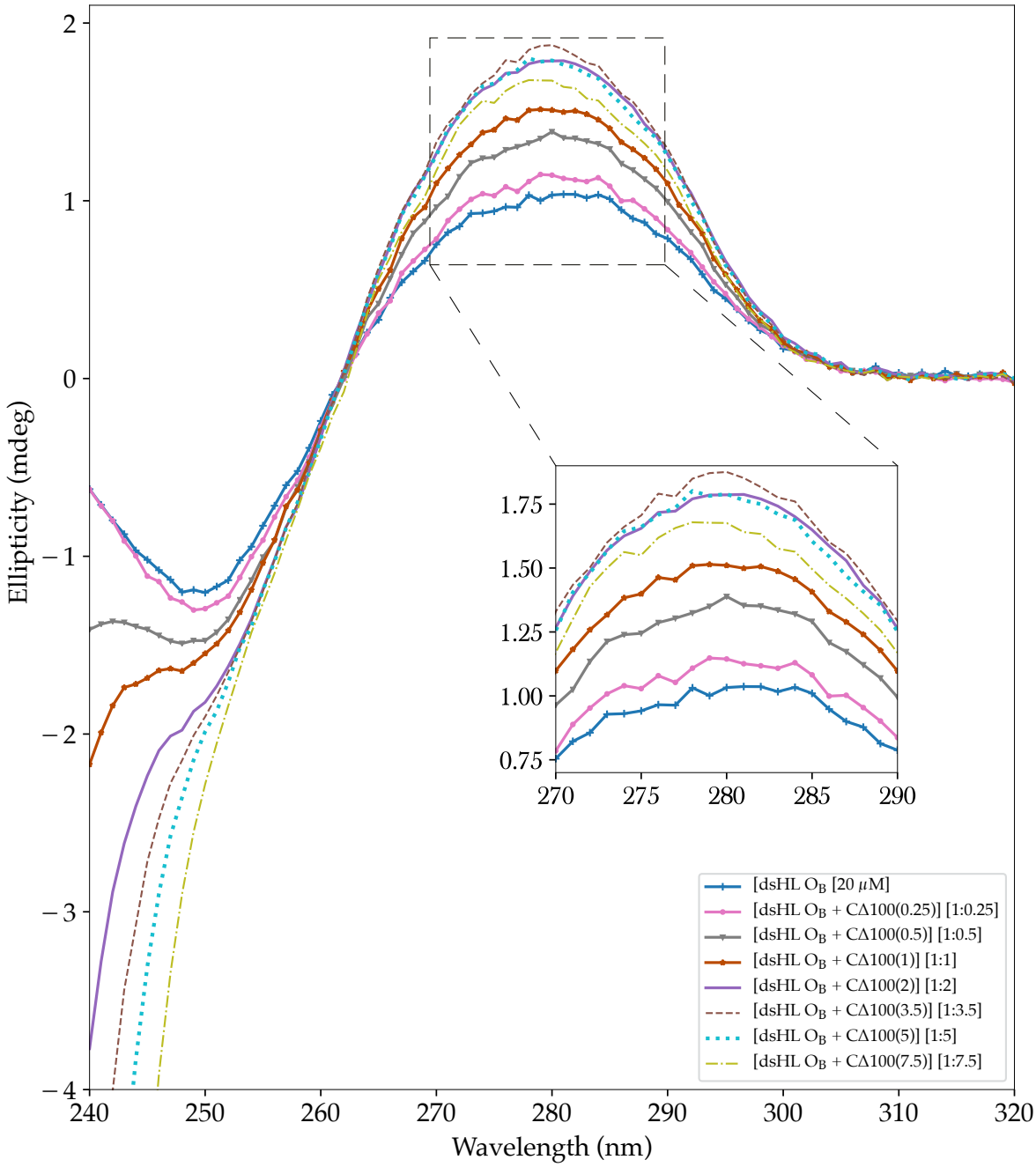

Figure 66: CD titration of the dsHL O_B DNA interaction with CΔ100 KorB protein. The figure shows an overlay of protein-DNA spectrum with increasing protein concentration. CD spectra of [dsHL O_B+CΔ100 KorB] [1:n] at various n molar ratios. The ellipticity (Y-axis) is plotted against the wavelength (X-axis). The inset shows the DNA signal (zoomed in the range of 270-290 nm) at 280 nm upon protein binding.

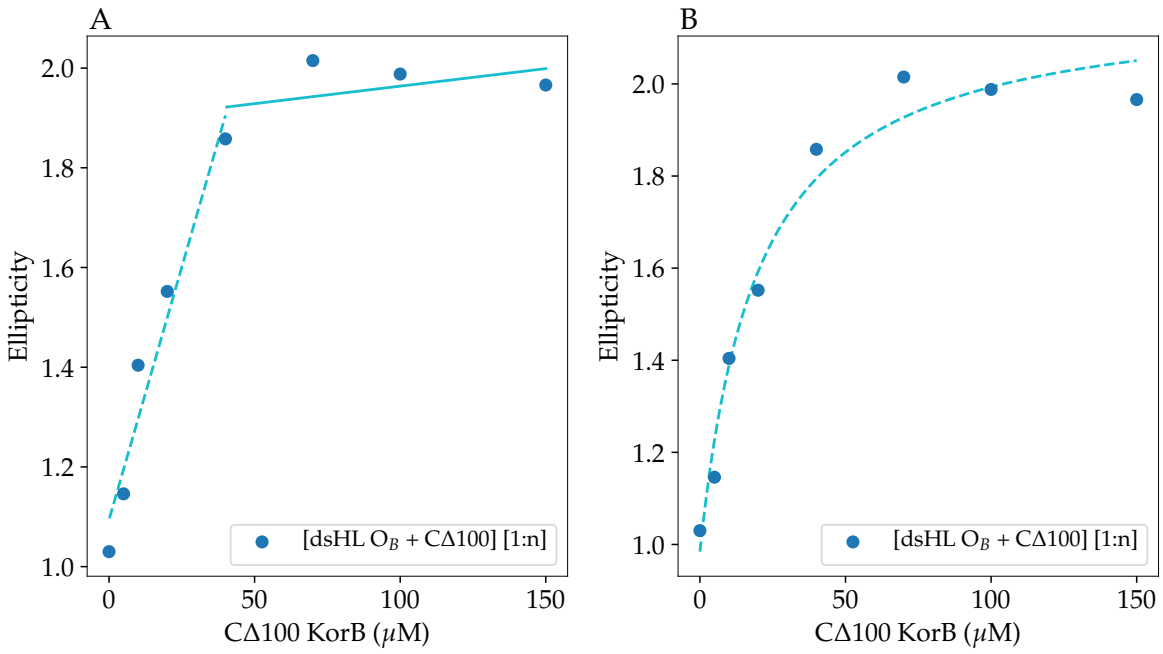

Figure 67: Interaction of the dsHL O_B DNA with CΔ100 KorB protein. The ellipticity values at 280 nm (from figure 66 but corrected for protein concentration) are plotted against the corresponding protein concentration of CΔ100 KorB. The DNA concentration used for this experiment is 20 μM and the concentration of CΔ100 KorB is for monomeric form. **(A)** Stoichiometry of the dsHL O_B interaction with CΔ100 KorB. The dashed line and solid line show the fit for increasing ellipticity signal and saturating ellipticity signal respectively. The change in slope indicates a stoichiometry of DNA:CΔ100 KorB to be 1:2. **(B)** The data are fitted using hyperbola equation (6.12) with CΔ100 KorB as monomer. The data are fitted with CΔ100 KorB as monomer. The dissociation constant for the interaction is 17.7±3.8 μM when calculated with hyperbola equation (6.12) and 6.4±3.6 μM when calculated with modified quadratic equation (6.13).

KorB construct	Equation	Protein species	O_B DNA	y_0	a	K_d (μM)	±Std. Error	t	P	R^2	±Std. Error*
NΔ31CΔ100	Hyperbola	Monomer	dsHL	1.009	1.005	**21**	**3.5**	**6**	**0.0019**	0.993	0.0279
			dsFL	0.099	0.613	**13.7**	**3.7**	**3.8**	**0.0200**	0.983	0.0309
NΔ31CΔ100	Mod. Quad.	Monomer	dsHL	1.022	0.002	**5.8**	**0.8**	**7.3**	**0.0007**	0.994	0.0118
		Monomer	dsFL	0.997	0.006	**6.8**	**2.2**	**3.1**	**0.0377**	0.989	0.0256
		Dimer	dsFL	1.016	0.005	**1.1**	**0.7**	**1.6**	**0.1767**	0.989	0.0242
CΔ100	Hyperbola	Monomer	dsHL	0.926	1.025	**17.7**	**3.8**	**4.7**	**0.0095**	0.991	0.040
			dsFL	1.613	1.015	**27.6**	**9.3**	**2.9**	**0.0212**	0.943	0.0892
CΔ100	Mod. Quad.	Monomer	dsHL	0.966	0.002	**6.4**	**3.6**	**1.8**	**0.1516**	0.983	0.0518
		Monomer	dsFL	1.639	0.004	**22.6**	**10.2**	**2.2**	**0.0621**	0.932	0.0973
		Dimer	dsFL	1.675	0.004	**9.7**	**5.9**	**1.7**	**0.1425**	0.919	0.1063

Table 23: Binding parameters of KorB constructs with O_B DNA calculated using the circular dichroism data. Dissociation constants (K_d) with standard error (Std. Error) for KorB constructs (NΔ31CΔ100 and CΔ100) with O_B DNA (dsHL and dsFL) are presented. The K_d values for NΔ31CΔ100 and CΔ100 with O_B DNA are calculated using hyperbola equation 6.12 for monomer proteins. The modified quadratic (Mod. Quad.) equation 6.13 is used for both the proteins with monomer concentrations used with dsHL, and monomer and dimer concentrations used with dsFL O_B DNA. t is the ratio of K_d to standard error, a measure of precision and P is the probability of obtaining the value of t by chance. The Std. Error, t and P values for respective K_d values are shown in bold. y_0 is the initial CD signal value and a is the difference between maximum and minimum CD signal. R^2 is the coefficient of regression and standard error (Std. Error*) is the error in estimation of R^2.

6.4.3 Microscale Thermophoresis

Scans of the samples in the capillary tubes used for MST experiments were made to check that the samples had the same amount of fluorescence and not affected by protein aggregation. Afterwards, the scans were optimised by observing the shape of the fluorescence signal. Figure 68A shows capillary scan data for each capillary used in a typical MST experiment. Positioned on the MST tray were 11 capillaries in the slots numbered 5 and 15. KorB was serially diluted (1:1) from capillary 5 (108 μM) to 15 (0.106 μM).

Each capillary contains KorB WT and labelled dsHL O_B DNA. The DNA used in the MST study was labelled with fluorescein at the 3′ end. Each capillary contains the same amount of DNA (250 nM). A series of aliquots of KorB at different concentrations was added. The concentration of KorB in each capillary is as follows: 5 (108 μM), 6 (54 μM), 7 (27 μM), 8 (13.5 μM), 9 (6.75 μM), 10 (3.375 μM), 11 (1.688 μM), 12 (0.844 μM), 13 (0.422 μM), 14 (0.211 μM), 15 (0.106 μM).

Raw fluorescence counts from the fluorophore are shown with respect to each capillary (figure 68A). Native fluorescence counts are acceptable between 300 and 1000 and within 10% of each other. The translucent blue strip in the image represents $\pm$ 10% of 400 and ideally all of the counts from each capillary should be in that range. Figure 68B overlays the shapes of the 11 capillary scans. Relative fluorescence is on the Y-axis and relative capillary position is on the X-axis and all the scans were overlaid relative to each other. Once the capillary scans were confirmed to be in agreement after overlaying them, MST signal was measured for protein-DNA binding. Figure 69A and figure 69B shows the original trace and relative MST trace respectively for the KorB WT and dsHL O_B DNA interaction.

At times ‘outliers’ are encountered in the measured data due to various reasons. The presence of water/sample/dirt on the outside of a capillary can locally perturb

the infrared laser. If the time trace appears to be 'bumpy', the change in the signal is because of the convective flow of aggregates inside the capillary. An outlier in fluorescence signal is because of the deviation in concentration. In this study, the outliers (if observed) were ignored from data analysis with the corresponding protein data point concentration.

The MST results presented here suggest that MST is suitable for the quantification of KorB-DNA dissociation constants and are of great importance in basic scientific research. Binding of KorB constructs *viz.* WT, CΔ100 KorB and NΔ31CΔ100 KorB were tested with full-site and half-site O_B DNA with MST. KorB WT bound to dsFL in a two step manner (biphasic). Here, the two binding steps are named as binding event 1 (BE1) and binding event 2 (BE2).

KorB WT-DNA interaction: KorB WT binds to half-site (dsHL) O_B in a single step whereas binds to full-site (dsFL) O_B in two steps. Binding of KorB WT to both half-site (dsHL) and full-site (dsFL) O_B are shown in figure 70 and figure 71 respectively. Each capillary contains the same amount of DNA (250 nM for dsHL and 120 nM for dsFL).

For KorB WT and dsFL interaction, KorB WT was serially diluted (1:1) from capillary 1 (108 μM) to 12 (0.106 μM), *i.e.* 1 (108 μM), 2 (54 μM), 3 (27 μM), 4 (13.5 μM), 5 (6.75 μM), 6 (3.375 μM), 7 (1.688 μM), 8 (0.844 μM), 9 (0.422 μM), 10 (0.211 μM), 11 (0.106 μM), 12 (0.026 μM).

The dissociation constants for KorB WT dsHL and dsFL O_B DNA are given in table 24 and table 25 respectively. The MST trace and binding profile for the KorB WT-dsHL interaction are shown in figure 70A and figure 70B respectively. The binding for KorB WT-dsHL was observed as a single step and the K_d for the interaction was 10.2$\pm$3.3 μM.

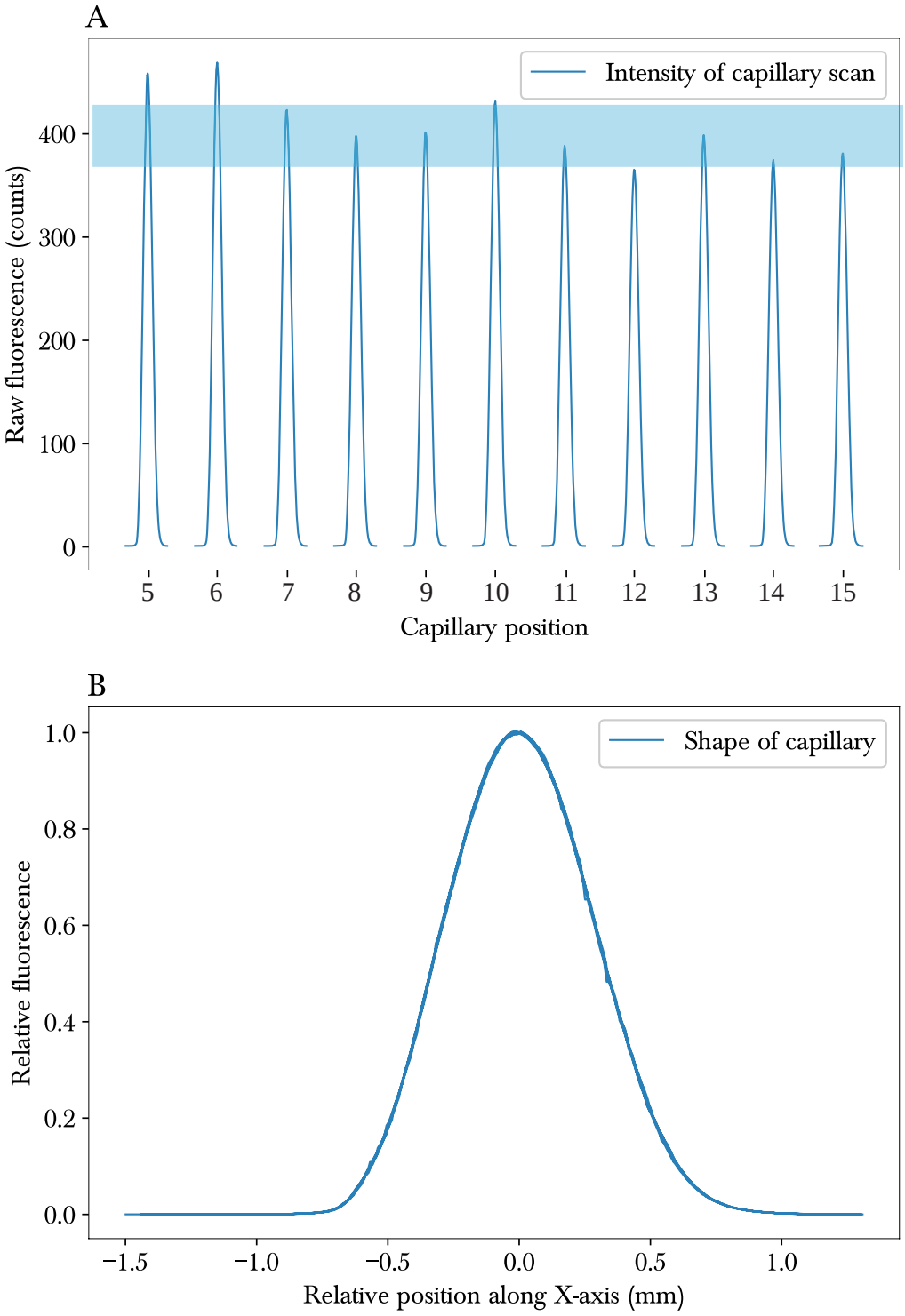

Figure 68: Capillary scan optimisation. (A) Graph representing the raw fluorescence counts for each capillary (position given as numbers on X-axis) used for KorB WT and dsHL O_B DNA binding experiment. Each capillary contains KorB WT and labelled dsHL O_B DNA. Each capillary contains the same amount of DNA (250 nM). KorB was serially diluted (1:1) from capillary 5 (108 μM) to 15 (0.106 μM). That is 5 (108 μM), 6 (54 μM), 7 (27 μM), 8 (13.5 μM), 9 (6.75 μM), 10 (3.375 μM), 11 (1.688 μM), 12 (0.844 μM), 13 (0.422 μM), 14 (0.211 μM), 15 (0.106 μM). Native fluorescence counts are acceptable between 300 and 1000 and within 10% of each other (shown with the teal band). **(B)** Overlaid shapes of the capillary scan. All the scans are overlaid relative to each other.

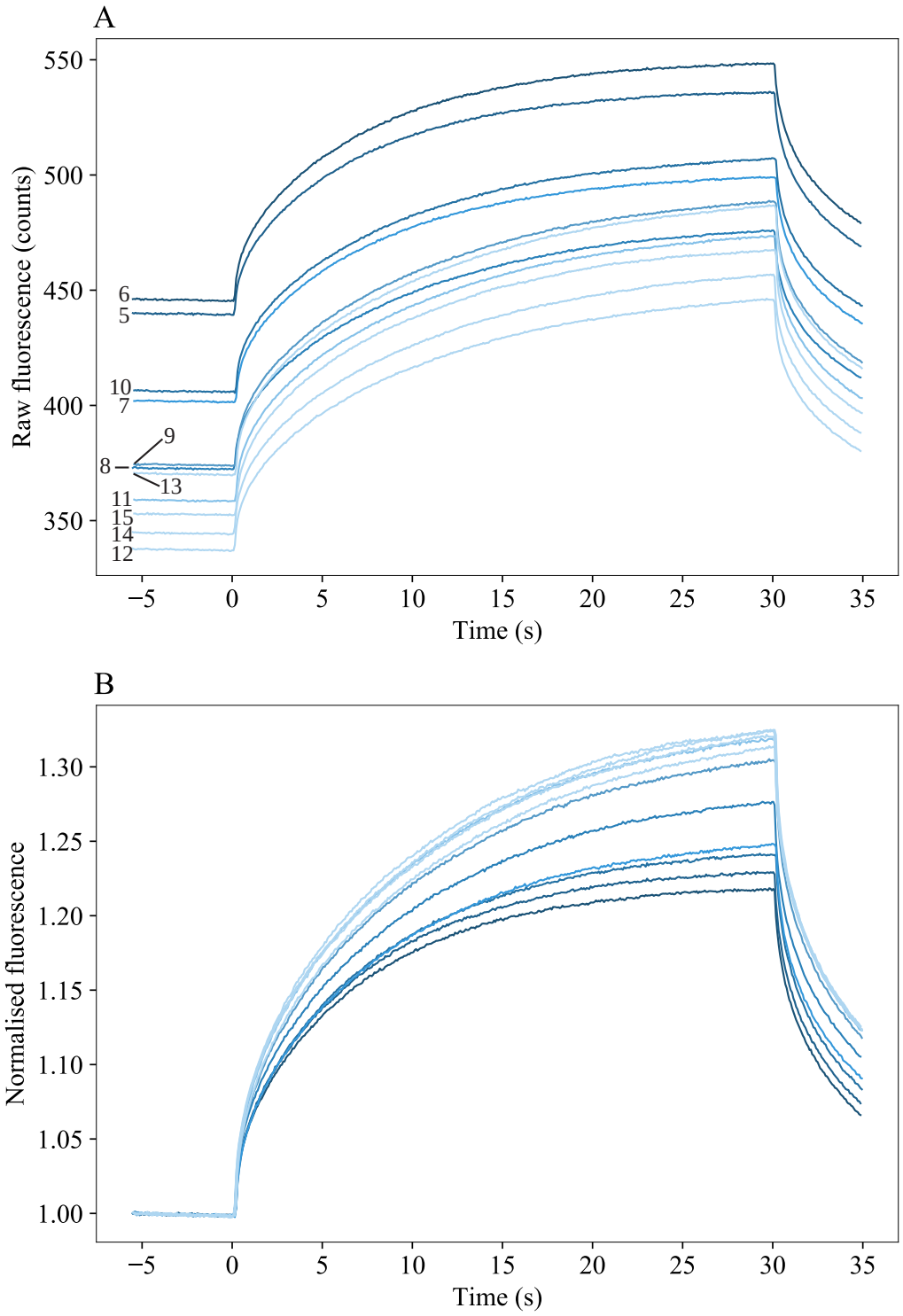

Figure 69: Native and normalised MST trace. (A) The original MST trace of KorB WT and dsHL O_B DNA interaction. Each trace represents the signal observed in the individual capillary. Each trace is numbered and the numbers represent the capillary position given in figure 68A. **(B)** The MST traces from all the capillaries have been normalised for signal intensity.

The KorB WT binds to dsFL O_B DNA in a two-step manner (biphasic fashion) and the two binding events are shown in figure 72. The overall binding profile for KorB-dsFL interaction is shown in figure 72A and the individual binding events are shown in figure 72B (BE1) and 72C (BE2) respectively. The overall fitting for the interaction was achieved by first individually calculating the dissociation constant for BE1 with the hyperbola equation 6.12 (K_d = 0.8±0.3 μM) and this value was incorporated in the modified hyperbola equation 6.15 to calculate constants for BE1 and BE2. The dissociation constant for BE1 for KorB WT-dsFL interaction was 0.8 (fixed) μM and for BE2 the dissociation constant was 9.4±2.3 μM.

CΔ100 KorB-DNA interaction: For CΔ100 KorB and dsHL interaction, CΔ100 KorB was serially diluted (1:1) from capillary 1 (1600 μM) to 12 (0.782 μM) and for CΔ100 KorB and dsFL interaction, CΔ100 KorB was serially diluted (1:1) from capillary 1 (210 μM) to 13 (0.052 μM).

The binding profile of CΔ100 KorB and half-site (dsHL) O_B DNA is shown in figure 73 and the binding of CΔ100 KorB and full-site (dsFL) O_B is shown in figure 74. The binding parameters for CΔ100 KorB dsHL and dsFL O_B DNA are given in table 26 and table 27 respectively. CΔ100 KorB appears to bind DNA in a two step manner. With both DNA species (dsFL and dsHL), there are two binding events – one with a higher dissociation constant than the other. For the first binding event (BE1), binding of CΔ100 KorB to dsFL (K_d = 1.1±0.3 μM) is tighter than to dsHL (K_d = 5.6±1.9 μM). Similarly for BE2, binding of CΔ100 KorB to dsFL (K_d = 110±11 μM) is tighter than to dsHL (K_d = 2250±250 μM).

Binding of NΔ31CΔ100 KorB was also tested with both dsHL and dsFL DNA but no changes in the signal were observed for the interaction.

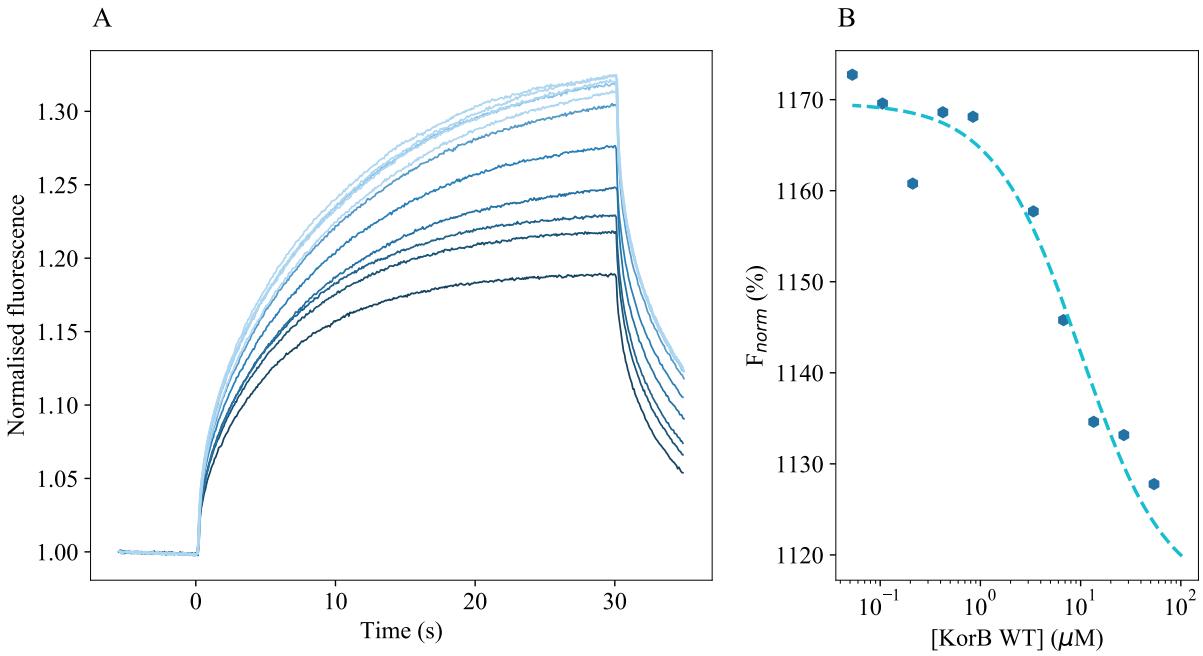

Figure 70: Binding of KorB WT and dsHL O_B DNA monitored with MST. (A) Normalised MST trace of KorB WT and dsHL O_B DNA. The laser is switched on for an interval of 30 s at time 0 s and the binding is observed. **(B)** Normalised fluorescence is plotted against KorB WT (dimer) and the fit calculated with equation 6.12 is shown with dashed line. The binding parameters thus obtained are presented in table 24. The KorB WT binds to dsHL O_B DNA in a single step and the dissociation constant for the interaction is 10.2±3.3 μM.

Parameter	**Coefficient**	**±Std. Error**	**t**	**P**
y_0	1169.6	2.2	531.6	<0.0001
a	-54.8	4.6	-11.8	<0.0001
K_d	**10.2 μM**	**3.3 μM**	**3.1**	**0.0152**
		Fit		
R	R^2	Adj. R^2	±Std. Error*	
0.9812	0.9628	0.9534	4.3062	

Table 24: Binding parameters of KorB WT with dsHL O_B DNA calculated from the MST data. The dissociation constant (K_d) with standard error (Std. Error) for KorB WT with dsHL O_B DNA has been calculated using the equation 6.12. The Std. Error, t and P values for respective K_d values are shown in bold.

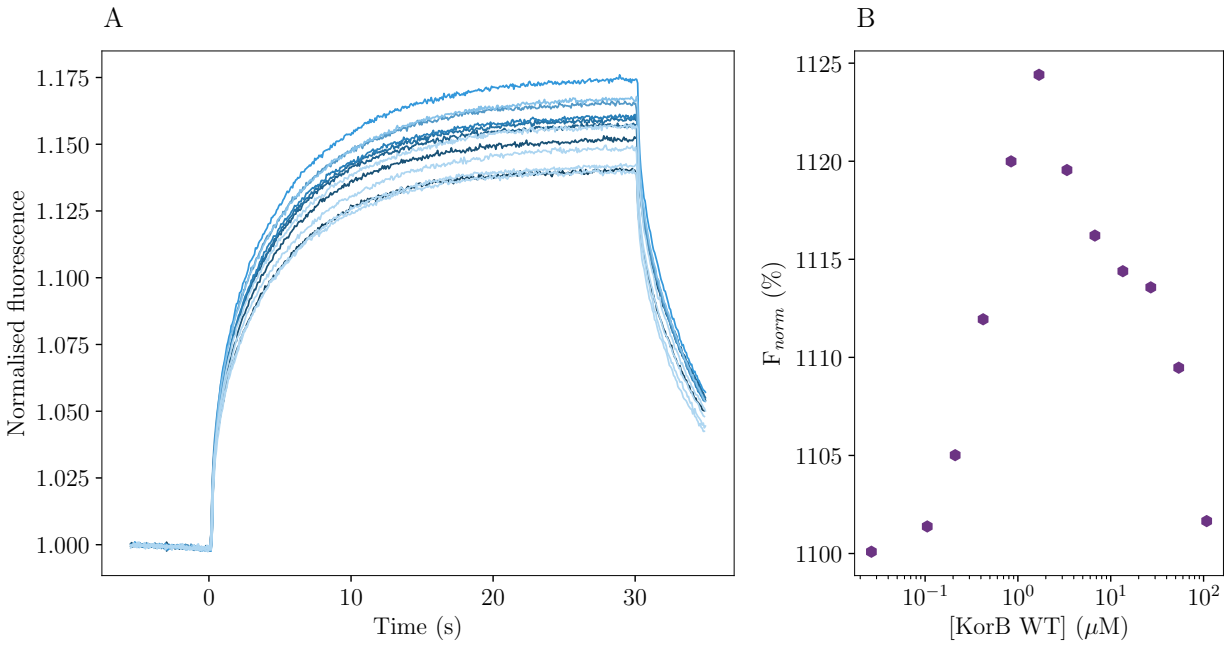

Figure 71: Binding of KorB WT and dsFL O_B DNA monitored with MST. (A) Normalised MST trace of KorB WT and dsFL O_B DNA. The normalised MST trace for each capillary is plotted against time. Each capillary contains the same amount of DNA (120 nM). The laser is switched on for an interval of 30 s at time 0 s and the protein-DNA binding is observed. **(B)** Normalised fluorescence is plotted against KorB WT. The KorB WT binds to dsFL O_B DNA in a biphasic fashion and the two binding events are shown in figure 72.

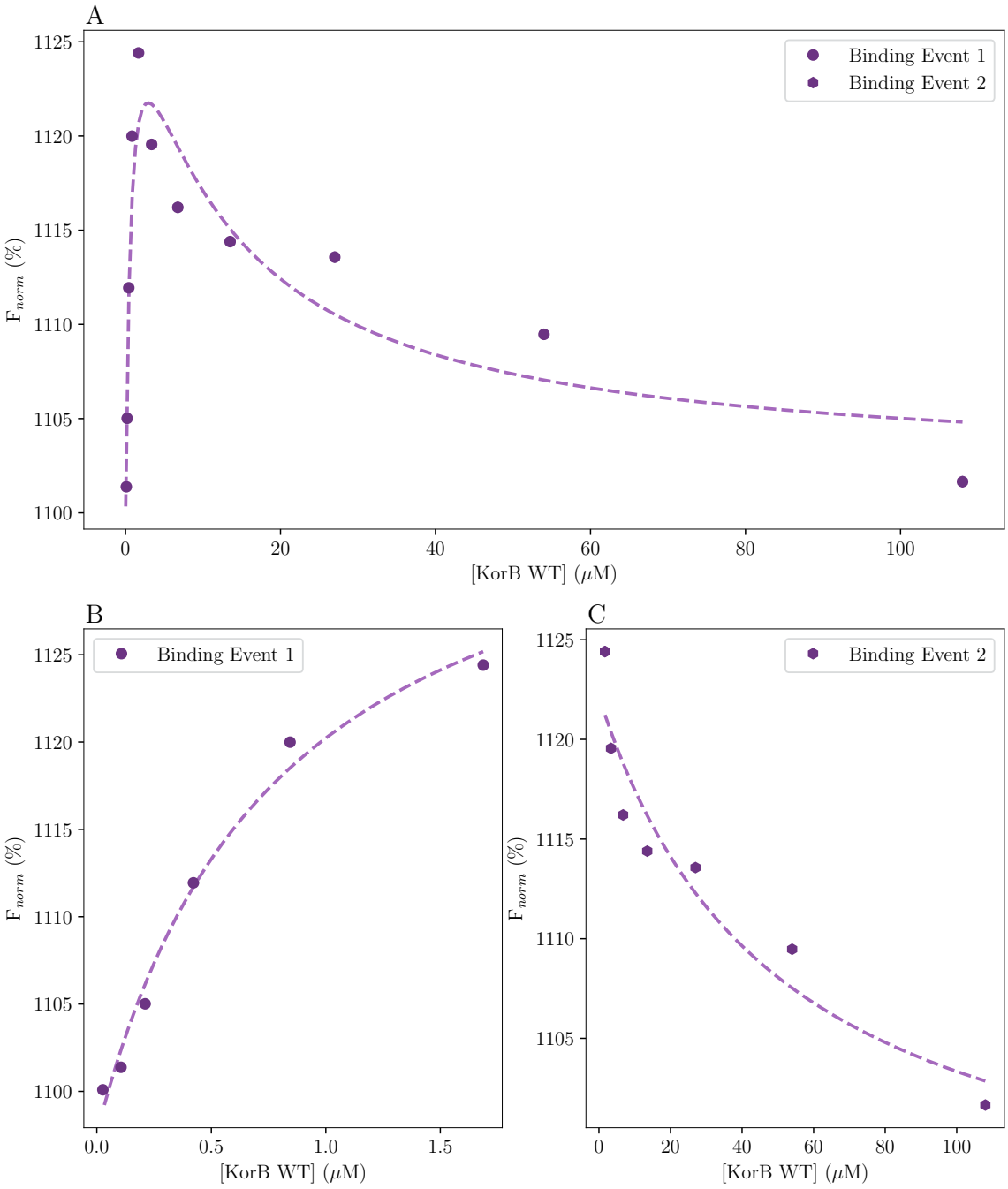

Figure 72: Binding of KorB WT and dsFL O_B DNA monitored with MST. (A) The binding profile for KorB WT and dsFL O_B DNA. The data points for Binding Event 1 (BE1) and Binding Event 2 (BE2) are represented with hexagons and circles respectively. **(B)** MST data and the fit for Binding Event 1. The dissociation constant for BE1 is 0.8±0.3 μM, calculated with equation 6.12. **(C)** MST data and the fit for Binding Event 2. The data points are fitted using equation 6.15 with K_{d1} fixed as 0.8 μM. The dissociation constant for BE2 is 9.4±2.3 μM. The dashed purple line represents the fit.

Parameter	**Coefficient**	**±Std. Error**	**t**	**P**
a	40	0.0	fixed	<0.0001
K_{d1}	**0.8 μM**	**fixed**	**fixed**	**fixed**
c	-37.0	0.0	fixed	<0.0001
K_{d2}	**9.4 μM**	**2.3 μM**	**4.1**	**0.0025**
y_0	1099.2	1.3	844.7	<0.0001
Fit				
R	R^2	Adj. R^2	±Std. Error*	
0.9474	0.8975	0.8873	2.7054	

Table 25: Binding parameters for KorB WT and dsFL O_B DNA from MST data.

The following equation was used to calculate the dissociation constants for KorB WT and dsFL O_B DNA for full range of data.

$$f = y_o + \frac{a * x}{K_{d1} + x} + \frac{c * x}{K_{d2} + x} \tag{6.15}$$

Here, y_o is the initial MST signal value, a and c are the difference between maximum and minimum values of MST signal for first curve and second curve respectively, x is the dimer protein concentration and K_{d1} and K_{d2} are the dissociation constants for first curve and second curve respectively.

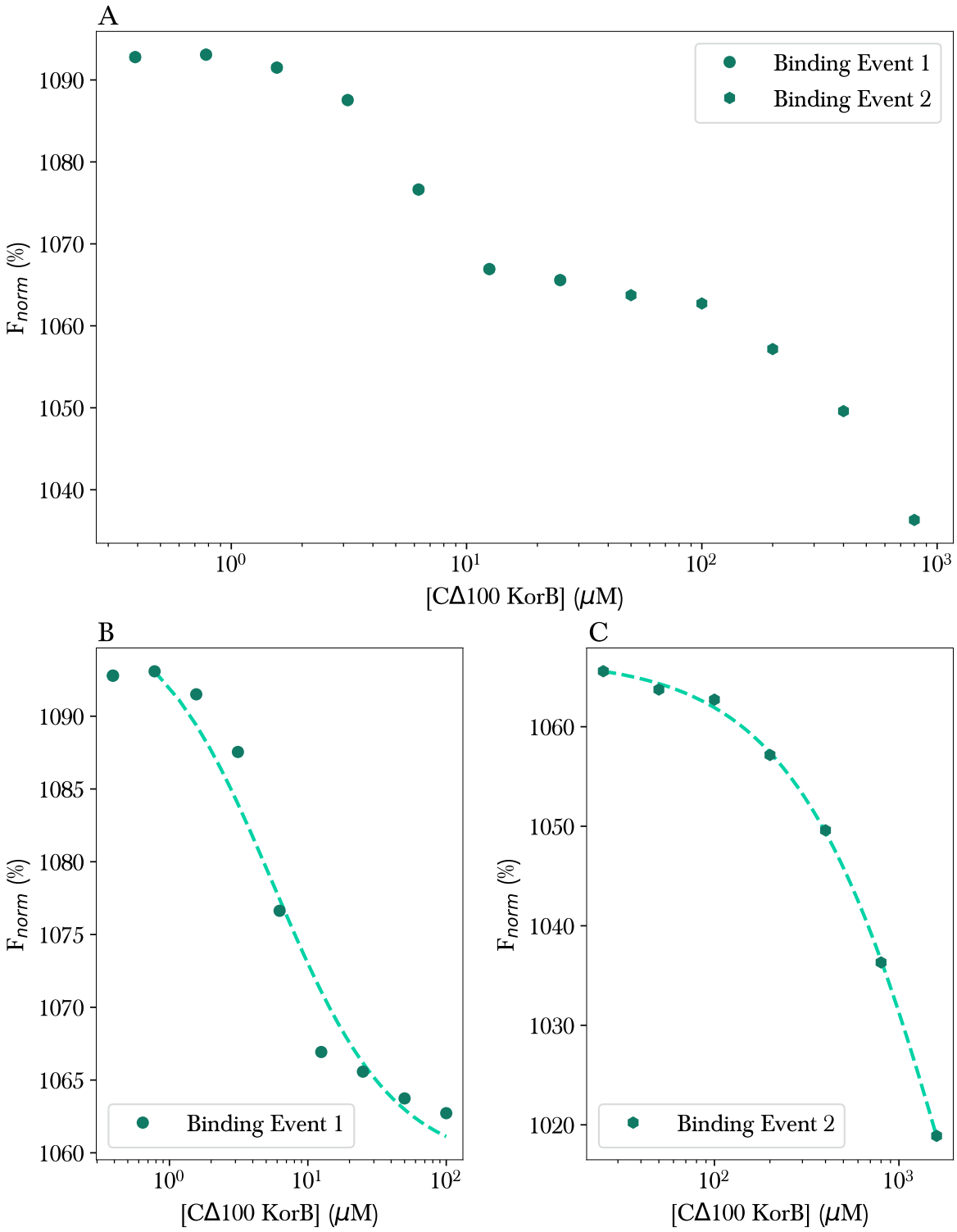

Figure 73: Binding of CΔ100 KorB and dsHL O_B DNA monitored with MST. (A) The binding profile for CΔ100 KorB and dsHL O_B DNA. The data points for Binding Event 1 (BE1) and Binding Event 2 (BE2) are represented with circles and hexagons respectively. The data points are fitted with equation 6.16. **(B)** MST data and the fit for Binding Event 1. The dissociation constant for BE1 is 5.6±1.9 μM. **(C)** MST data and the fit for Binding Event 2. The dissociation constant for BE2 is 2250±250 μM. The dashed green line represents the fit.

Parameter	**Coefficient**	**±Std. Error**	**t**	**P**
Binding Event 1 ($R^2 = 0.9687 \pm 2.7373$)				
y_0	1097.8	2.7	411.5	<0.0001
a	-38.8	2.9	-13.6	<0.0001
K_{d1}	**5.6 μM**	**1.9 μM**	**2.9**	**0.0254**
Binding Event 2 ($R^2 = 0.9994 \pm 0.533$)				
y_0	1066.9	0.4	2807.4	<0.0001
a	-115.4	7.9	-14.7	0.0001
K_{d2}	**2250 μM**	**250 μM**	**9**	**0.0009**

Table 26: Binding parameters for C△100 KorB and dsHL O_B DNA from MST data.

The following equation was used to calculate the dissociation constant for C△100 KorB and dsHL O_B DNA interaction.

$$f = y_o + \frac{a * x}{K_d + x} \tag{6.16}$$

Here, y_o is the initial MST signal value, a is the difference between maximum and minimum value of MST signal, x is the monomer protein concentration and K_d is the dissociation constant.

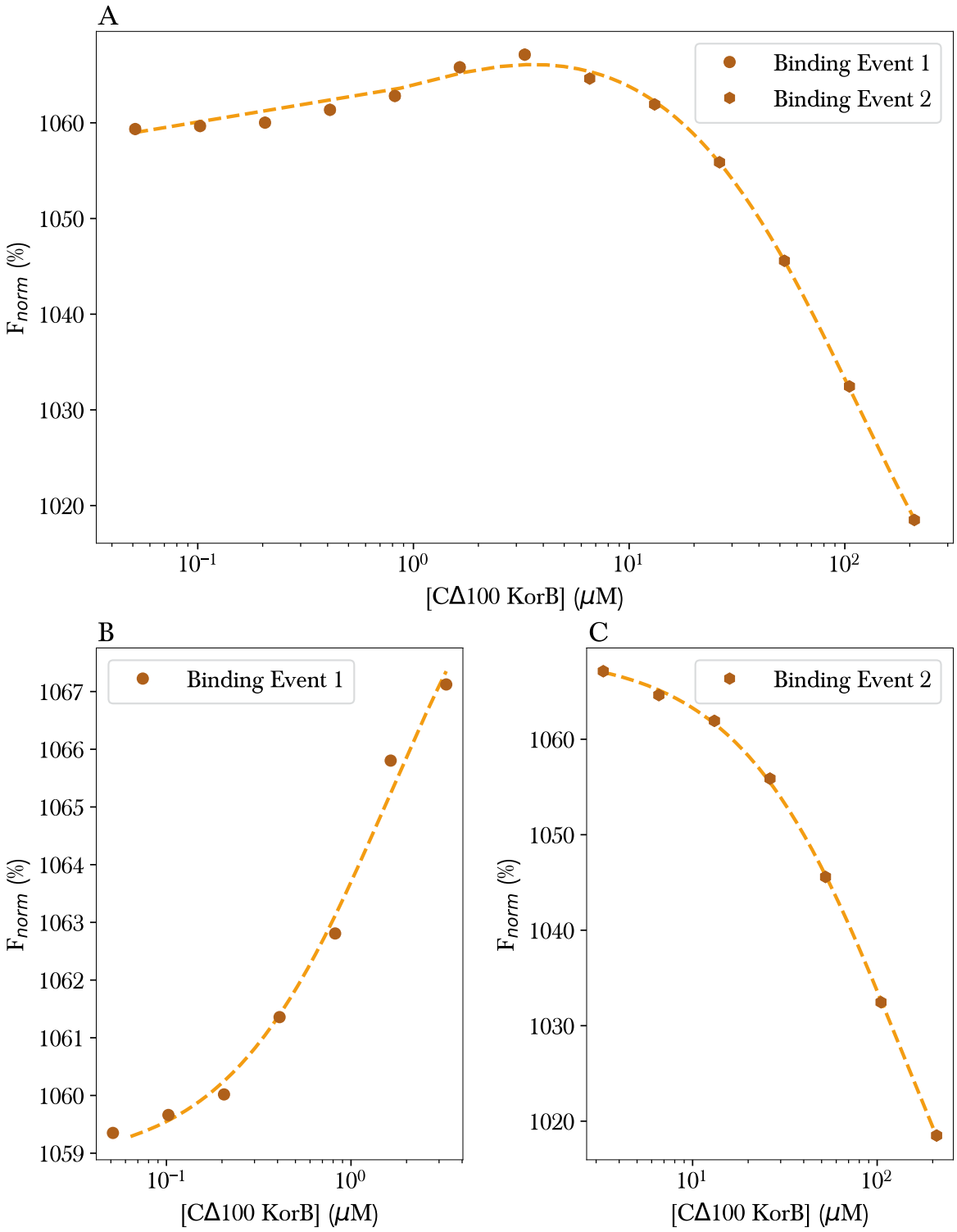

Figure 74: Binding of CΔ100 KorB and dsFL O_B DNA monitored with MST. (A) The binding profile for CΔ100 KorB and dsFL O_B DNA. The data points for Binding Event 1 (BE1) and Binding Event 2 (BE2) are represented with circles and hexagons respectively. The data points are fitted with equation 6.17. **(B)** MST data and the fit for Binding Event 1. The dissociation constant for BE1 is 1.1±0.3 μM. **(C)** MST data and the fit for Binding Event 2. The dissociation constant for BE2 is 110±11 μM. The dashed orange line represents the fit.

Parameter	**Coefficient**	**±Std. Error**	**t**	**P**
a	13.4	1.1	12.1	<0.0001
K_{d1}	**1.1 μM**	**0.3 μM**	**3.7**	**0.0059**
c	-81.6	2.5	-33.2	<0.0001
K_{d2}	**110 μM**	**11 μM**	**10**	**<0.0001**
y_0	1058.5	0.5	2024.9	<0.0001
Fit				
R	R^2	Adj. R^2	±Std. Error*	
0.9994	0.9988	0.9982	0.6108	

Table 27: Binding parameters for CΔ100 KorB and dsFL O_B DNA from MST data.

The following equation was used to calculate the dissociation constants for CΔ100 KorB and dsFL O_B DNA for full range of data.

$$f = y_o + \frac{a * x}{K_{d1} + x} + \frac{c * x}{K_{d2} + x} \tag{6.17}$$

Here, y_o is the initial MST signal value, a and c are the difference between maximum and minimum values of MST signal for first curve and second curve respectively, x is the monomer protein concentration and K_{d1} and K_{d2} are the dissociation constants for first curve and second curve respectively.

6.5 Results of KorB-KorA interactions

As mentioned in section 1.6, KorB and KorA interact with each other cooperatively. The interaction of KorB and KorA WT was studied using the microscale thermophoresis and interaction of CΔ100 KorB-KorA was studied using 2D NMR (HSQC) titrations.

6.5.1 Microscale Thermophoresis

KorB WT-KorA WT interaction: In order to probe the interaction of KorA and KorB, KorA was fluorescently labelled with the accompanying MST-NHS kit. KorB was serially diluted in two fold steps from 32 μM to 0.0156 μM and aliquots were added to a constant concentration of KorA. The concentration of KorB in each capillary is given in table 28. The normalised MST trace for KorA and KorB binding is shown in figure 75A and the resulting binding profile is shown in figure 75B. It can be deduced from the binding profile that KorB binds to KorA in a single step *i.e.* monophasic fashion. The data point were fit using the equation 6.12 and the dissociation constant for KorB-KorA interaction is 0.34$\pm$0.1 μM. The binding parameters for KorB and KorA binding are presented in table 29.

Lane	**KorB WT (μM)**
1	32
2	16
3	8
4	4
5	2
6	1
7	0.5
8	0.25
9	0.125
10	0.0625
11	0.0313
12	0.0156

Table 28: Concentration of KorB in each capillary.

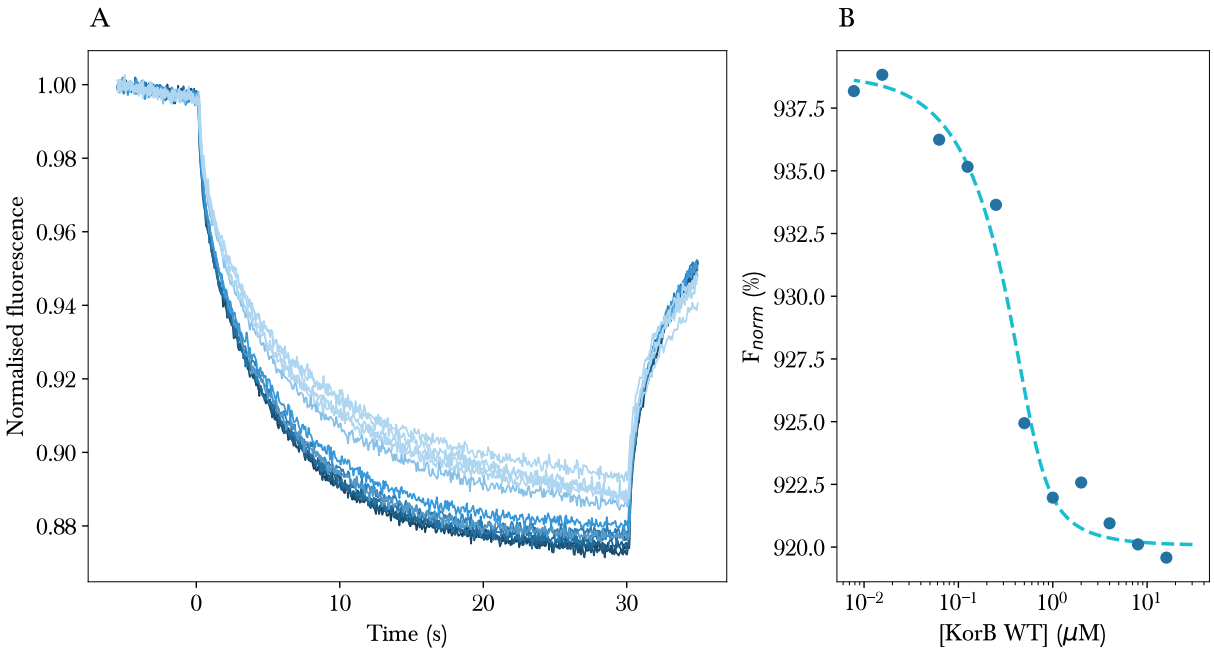

Figure 75: Binding of KorB WT and KorA monitored with MST. (A): Normalised MST trace of KorB WT and KorA. The laser was switched on for an interval of 30 s at time 0 s and the binding was observed. **(B)**: Normalised fluorescence was plotted against KorB WT and the fit calculated with equation 6.12 is shown with dashed line. The binding parameters thus obtained are presented in table 29. The KorB binds to KorA in a single step and the dissociation constant for KorB-KorA interaction is 0.34±0.1 μM.

Parameter	**Coefficient**	**±Std. Error**	**t**	**P**
y_0	939.6	1.1	853.9	<0.0001
a	-20.8	1.2	-16.8	<0.0001
K_d	**0.34 μM**	**0.1 μM**	**3.9**	**0.0035**
		Fit		
R	R^2	Adj. R^2	±Std. Error*	
0.9850	0.9702	0.9636	1.5373	

Table 29: Binding parameters for KorB WT and KorA from MST data. The equation 6.12 was used to calculate the dissociation constant for KorB and KorA interaction.

6.5.2 ^{15}N KorA and CΔ100 KorB HSQC titrations

A range of concentrations of CΔ100 KorB protein were titrated against a fixed concentration of ^{15}N KorA protein (dimer, 0.25 mM). The range of ratios of KorA:CΔ100 KorB used for the experiments were 1:1, 1:2, 1:3, 1:4 and 1:5 and the details about the protein concentrations are listed in table 30. ^{1}H-^{15}N HSQCs of KorA and CΔ100 KorB were superimposed to identify peaks that have either shifted or disappeared because of the addition of KorB binding to the KorA protein. Figure 76 shows the HSQC of ^{15}N KorA at 0.25 mM. Figure 77 and figure 78 show the overlay of ^{15}N KorA and CΔ100 KorB with minimum (KorA:KorB::1:1) and maximum (KorA:KorB::1:5) concentration of CΔ100 KorB respectively. On comparison of KorA and KorA-KorB overlaid HSQCs, it is evident that certain peaks in KorA HSQC disappear upon KorB binding but no peaks were shifted. And because of that, the data for the middle concentration points (1:2, 1:3, 1:4) are not shown because the spectra of those three middle points are identical to the spectrum with 1.25 mM KorB.

KorA (dimer, mM)	**KorA (mM) and CΔ100 KorB (monomer, mM)**	**KorA:KorB**	**Figure number**
0.25	0.25 and 0	1:0	76
0.25	0.25 and 0.25	1:1	77
0.25	0.25 and 0.5	1:2	Not shown
0.25	0.25 and 0.75	1:3	Not shown
0.25	0.25 and 1	1:4	Not shown
0.25	0.25 and 1.25	1:5	78

Table 30: KorA and CΔ100 KorB concentrations used in the HSQC titrations.

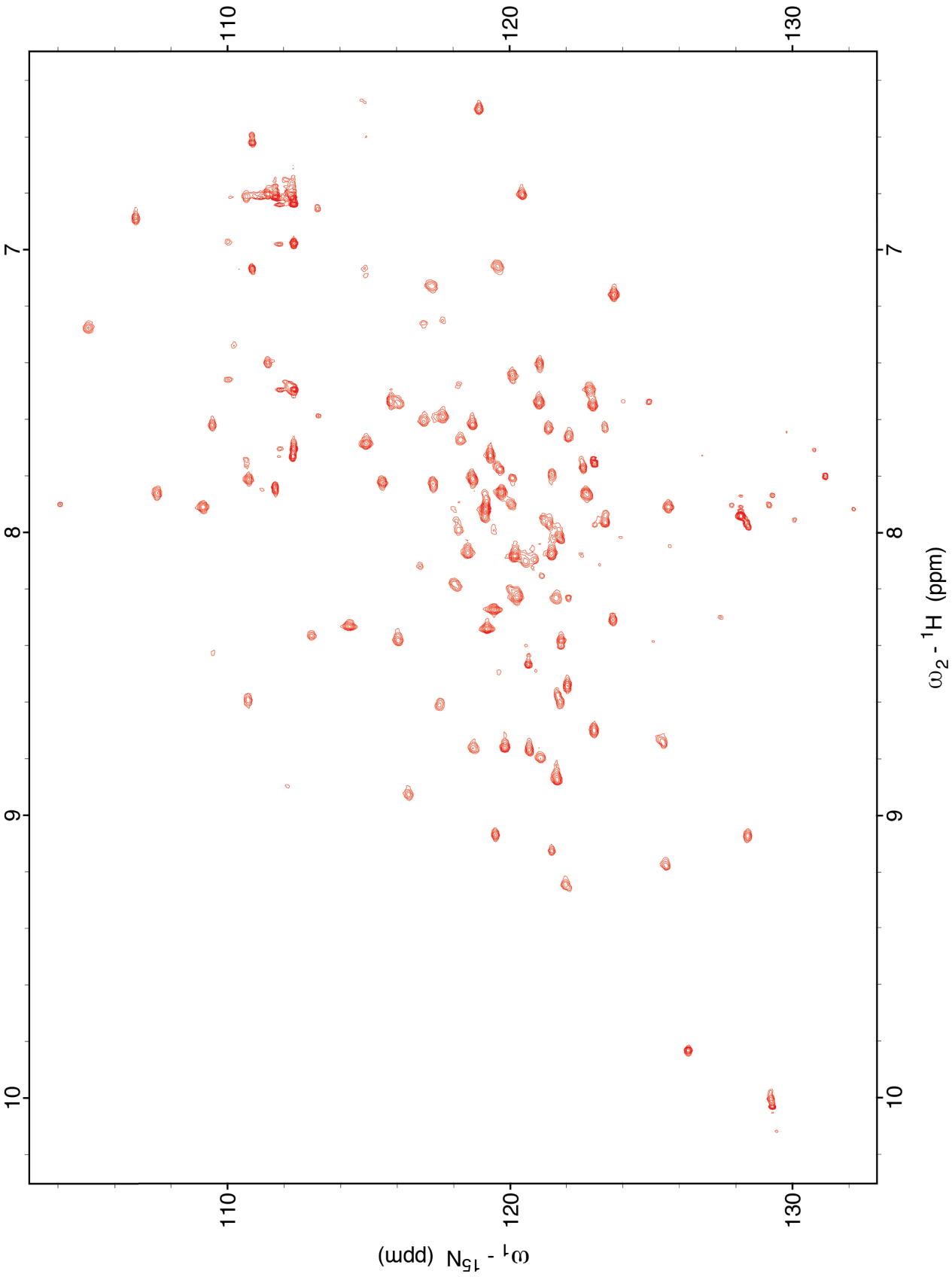

Figure 76: HSQC of 0.25 mM ^{15}N KorA.

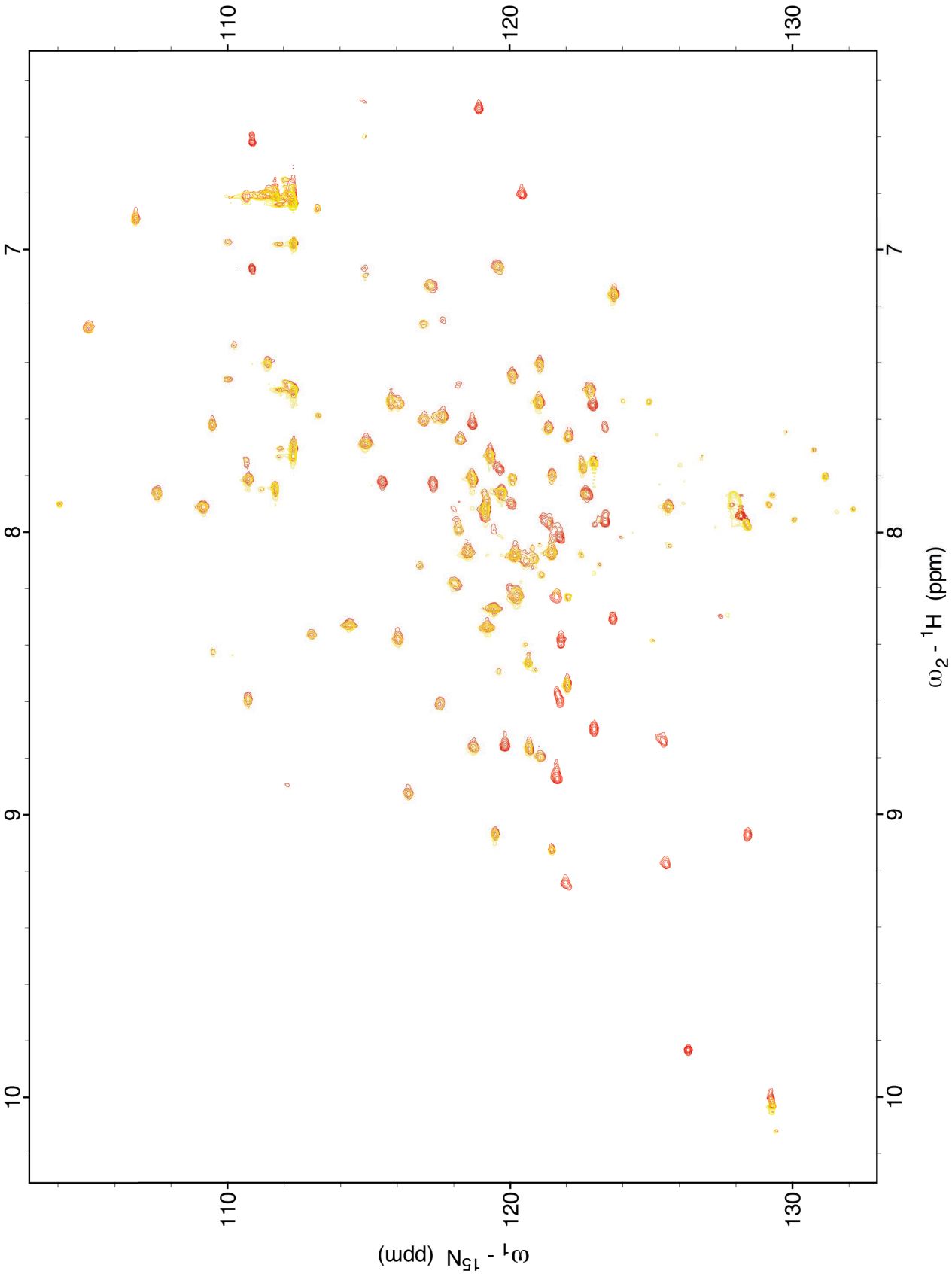

Figure 77: Overlay of HSQC of 0.25 mM ^{15}N KorA and 0.25 mM ^{15}N KorA with 0.25 mM CΔ100 KorB. The red coloured spectrum represents the HSQC of 0.25 mM ^{15}N KorA. The yellow coloured HSQC spectrum is obtained with the addition of 0.25 mM CΔ100 KorB to the 0.25 mM ^{15}N KorA.

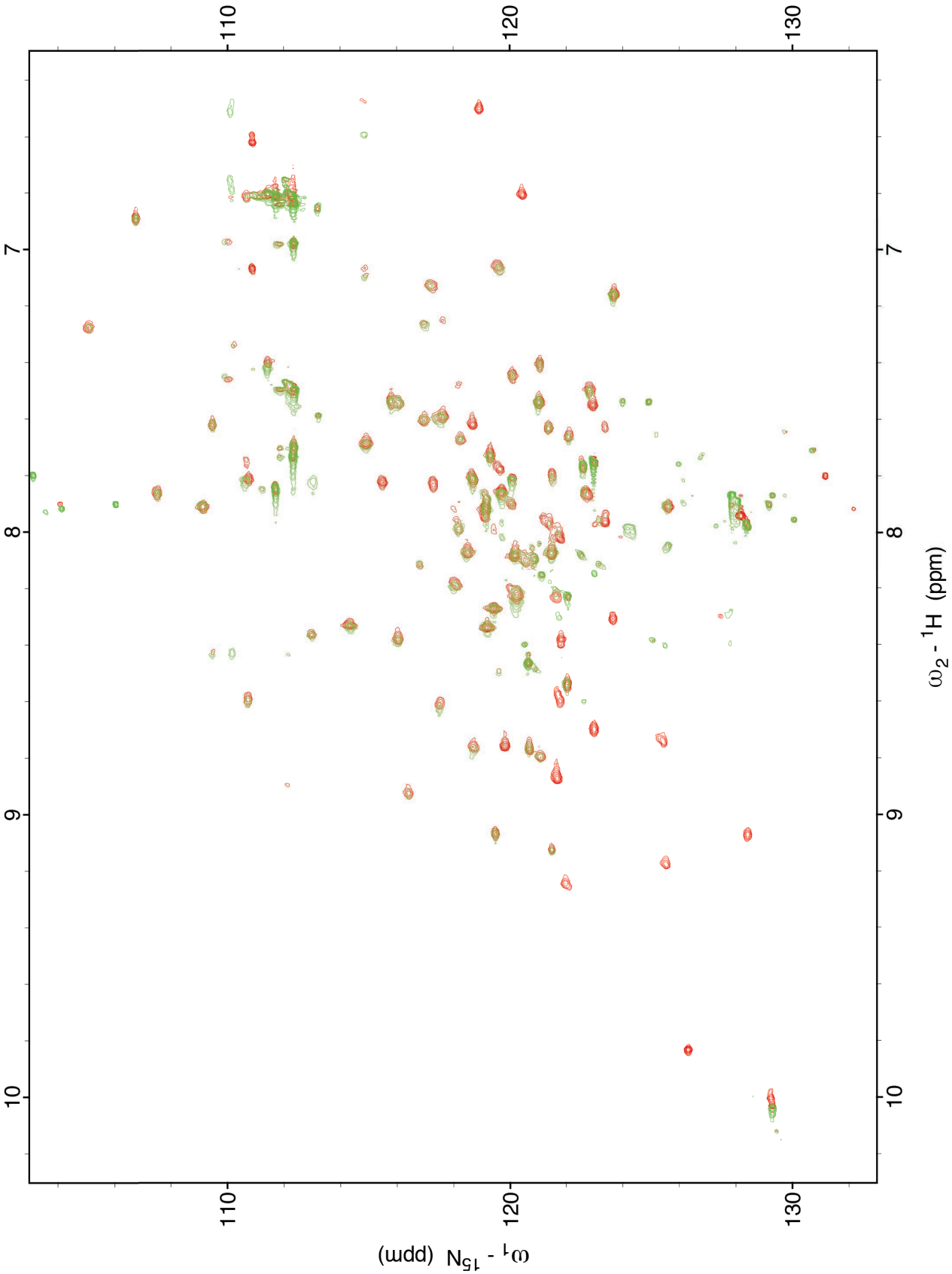

Figure 78: Overlay of HSQC of 0.25 mM ^{15}N KorA and 0.25 mM ^{15}N KorA with 1.25 mM CΔ100 KorB. The red coloured spectrum represents the HSQC of 0.25 mM ^{15}N KorA. The green coloured HSQC spectrum is obtained with the addition of 1.25 mM CΔ100 KorB to the 0.25 mM ^{15}N KorA.

6.6 Discussion

6.6.1 Comparative analysis of techniques to study KorB-DNA interaction

Microscale thermophoresis is an optical detection technique based on the temperature-dependent movement of molecules to analyse biomolecular interactions. MST is sensitive to changes in molecular charge and hydration shell but does not only rely on change in particle size like FA or AUC (Jerabek-Willemsen et al., 2014; Wienken et al., 2010).

MST utilises optical apparatus to detect the fluorophore and the direction of the trace signal is dependent on the choice of fluorophore used in the sample. In this study, two fluorophores were used; KorA was labelled with NHS and the O_B DNA was labelled with fluorescein. The examples for the fluorophore based direction dependence is evident from the traces for fluorescein and NHS-label in figure 70 and figure 75 respectively.

Based on the MST signals it is possible to dissect different binding mechanisms and detect multi-step protein-DNA interactions (Jerabek-Willemsen et al., 2014). The use of fluorescein labelled O_B DNA is presented here as a relatively recent method for performing protein-DNA interaction study for a plasmid partitioning system. The MST data for KorB-DNA interaction contains information that extend single equilibrium dissociation constant and this information is used to infer binding modes for the interaction.

Binding of KorB constructs *viz.* WT, CΔ100 KorB and NΔ31CΔ100 KorB were tested with full-site and half-site O_B DNA with MST. KorB WT binds to dsHL O_B in a single step whereas KorB binds to dsFL O_B in two steps. CΔ100 KorB appears to bind O_B DNA in a two step manner whilst no significant results were obtained for DNA binding of CΔ100 KorB-derived construct, NΔ31CΔ100 KorB.

Circular dichroism spectroscopy is a relatively easy measurement technique because of low sample requirement, less experimentation time and has proven to be useful for studying protein-ligand interactions (Siligardi et al., 2014). In this chapter, CD has been used to quantify dissociation constants for KorB-DNA interactions by identifying changes in the secondary structure of O_B DNA upon binding to KorB constructs.

At wavelengths above 250 nm, minimal contribution from protein is observed whilst DNA dominates the CD signal (Cary and Kneale, 2009). A variety of changes can be observed in CD signal at 280 nm upon addition of protein to DNA. There are examples in the literature where DNA-binding proteins bend the DNA molecule. Proteins containing high-mobility-group (HMG) domain or with a homologous domain bend DNA by intercalation. Example of such proteins are HMG-D and Sox-5 that induces DNA bending by intercalation (Connor et al., 1994; Klass et al., 2003). Intercalators cause a wavelength shift towards blue region in the CD signal but an increase in the signal is not observed (Garbett et al., 2007).

On the other hand, an increase in the CD signal is observed for a controller protein, C.AhdI binding its operator DNA without experiencing any shift in wavelength (Papapanagiotou et al., 2007). Similar to C.AhdI titration with its operator DNA, binding of both CΔ100 KorB and NΔ31CΔ100 KorB to O_B DNA (both dsHL and dsFL) caused an increase in the CD signal without any blue-shift, thereby implying twisting or bending of the O_B DNA without any base intercalation. However, it is not conclusive from the results of this study, whether KorB's interaction with the O_B follows bending/twisting or any specific unwinding of the DNA.

Fluorescence anisotropy involves the excitation of fluorophores in solution and the molecular size of the fluorophore correlates to its rate of rotational diffusion. Anisotropy value of a fluorophore in bound form is more than the value in its free form. Here, KorB binding to fluorescein-tagged O_B DNA was used to measure the change in anisotropy between free form of fluorescein-tagged O_B DNA and tagged-DNA bound to KorB. As

concentration of fluorophore does not impact rate of tumbling, so dilution has no effect on anisotropy. For all the FA experiments, the increase in the protein concentration was the only variable factor, changing the anisotropy and because of the protein addition, the dilution effect was minimal on the intensity (excitation/emission) reduction.

During the anisotropy titration, addition of protein did not quench fluorescence, indicating satisfactory spatial distance between fluorophore tagged to the 17 bp dsFL O_B DNA and the KorB binding site. As quenching was not observed, so it can be inferred that there was no erroneous impact on the anisotropy as a result of protein binding. It can also be assumed that the rotational diffusion of fluorophore was the only factor contributing to the anisotropy.

The overall size (polydispersity) of a biomolecule in the solution has an effect on the anisotropy values because of its tumbling rate. Also the rotational relaxation lifetime affects the intrinsic anisotropy of a molecule (free and bound). Jameson and Ross (2010) defined rotational relaxation lifetime as 'the time for a given orientation to rotate through an angle given by the arccos e^{-1}, which is 68.42°'. An estimate of rotational relaxation lifetime (ns) of a protein assumed to be spherical is equal to MW of protein in kDa. In this study, based on the the MW of the proteins, estimated rotational relaxation lifetime of KorB constructs are - KorB WT: 38.1 ns, CΔ100 KorB: 30.6 ns and NΔ31CΔ100 KorB: 25.3 ns. These effects are expected to reflect in anisotropy maximum for KorB-DNA interactions, but the maximum anisotropy did not seem to be correlated to the size of KorB constructs (table 21). In future anisotropy experiments, KorB-DNA binding studies should correct for this intrinsic difference between different MW proteins binding the same DNA substrate.

6.6.2 Dissociation constants and biological implication

Electrophoretic Mobility Shift Assays (EMSA) were not used to quantify dissociation constants for KorB-DNA and KorB-KorA interactions, as the assays are time consuming, require radioactivity material and the gel matrix discourages a 'true equilibrium'

measurement of dissociation constants (Owen and McMurray, 2009). On the other hand, the gamut of fluorescence techniques including fluorescence anisotropy are more accurate in the range of 10^{-10} to 10^{-3} molar for quantifying dissociation constants (Owen and McMurray, 2009).

Comparison of dissociation constants for KorB constructs binding to dsHL O_B DNA and dsFL O_B DNA are listed in table 31 and table 32 respectively. For WT-dsHL O_B interaction, different dissociation constants were observed from different techniques, with the observed interaction being five-fold tighter from FA (K_d=2.2±0.14 μM) data as compared to MST (K_d=10.2±3.3 μM) data. Anisotropy revealed tight binding between KorB WT with dsFL O_B DNA with dissociation constant of 0.63±0.11 μM. Interaction of WT-dsFL O_B with MST was observed to be a two step process. MST determined first binding event of KorB WT-dsFL (K_{d1}=0.8 μM) was in agreement with results obtained from FA data but the second event, with K_{d2}=9.4±2.3 was not observed in the FA assay. This is possibly because the FA assay was done at lower KorB WT concentrations (0-11 μM) as compared to the MST measurement and the second binding event in the MST was only observed at higher protein concentrations (10-100 μM). Both these results formed a solid foundation for comparison of dissociation constants for truncated KorB constructs with O_B DNA with those of KorB WT.

When compared to wild type KorB, the dissociation constants for CΔ100 KorB and NΔ31CΔ100 KorB were consistently higher in all the CD, FA and MST experiments. Comparing the binding between NΔ31CΔ100 KorB to dsFL and dsHL DNA (from FA and CD data), the protein bound five fold tighter to dsFL DNA, although this is only true if the protein is assumed to bind the DNA as a dimer. On the other hand, the reverse result was seen on comparing CΔ100 KorB binding to dsFL and dsHL DNA. Data from all the techniques (FA, CD and MST), suggest CΔ100 KorB to bind dsHL more tightly than to dsFL. But a consistent correlation was not established between the dissociation constants derived from FA, CD and MST data. It was expected for KorB constructs with truncated dimerisation domain to dimerise indirectly and only in the

presence of the DNA. The reduced affinity of the deletion mutants of KorB to the O_B DNA can be explained with the lack of C-terminal dimerisation domain and absence of pre-formed dimers in solution.

A lack of reproducibility in fluorescence assays could be attributed to the photo-instability of fluorescein (de Jong et al., 2005). Provided the exposure of stock tagged O_B DNA to light was minimal, fluorescence intensity was not observed to decrease significantly even over period of months when used for a variety of experiments. In contrast to FA and CD, MST is more sensitive to binding-induced changes during the protein-DNA complex formation as MST does not only rely on the change in the size (Seidel et al., 2013). For MST and FA experiments, an alternative labelling strategy could be explored (a different fluorophore such as 6-carboxy-fluorescein or 5-carboxy-tetra-methyl-rhodamine) in order to improve the reproducibility and/or reduce the error in the data acquisition.

The CTD of KorA interacts with KorB (shown using NMR spectroscopy), but the exact region of KorB involved in this interaction with the CTD is not known (Bingle et al., 2008). The HSQC titrations of CΔ100 KorB against ^{15}N labelled partner protein KorA confirm that identical KorA peaks in the HSQC are perturbed as by the KorB WT protein. Also the missing peaks in the HSQC with ^{15}N KorA-CΔ100 KorB interaction can be attributed to the overall size (41.9 kDa) of the complex formed. To determine any difference in the binding between KorB WT and CΔ100 KorB, lower concentrations of CΔ100 KorB are needed when titrating against KorA.

Building and expanding on the results presented in this chapter, a number of possible future experiments can be envisaged. KorB constructs can be tested with varying lengths of O_B DNA. Binding profile of KorB constructs and non-specific DNA can be tested with a competition assay. DNA binding of KorB can be tested in the presence of partner protein, KorA and in the presence of O_AO_B DNA. Alternative buffers and salts can be tested for KorB-DNA interactions with FA, CD and MST experiments including testing the effect of magnesium in the buffer as suggested by Taylor et al., 2015.

KorB construct	Equation	K_d (μM)	±Std. Error
	FA		
Wild-type	Hyperbola	2.2	0.14
CΔ100 KorB	Hyperbola	0.44	0.05
NΔ31CΔ100 KorB	Hyperbola	8.4	0.98
	CD		
CΔ100 KorB	Mod. Quad.	6.4	3.6
NΔ31CΔ100 KorB	Mod. Quad.	5.8	0.8
	MST		
Wild-type	Hyperbola	10.2	3.3
CΔ100 KorB	Hyperbola	K_{d1}=5.6	±SE1=1.9
		K_{d2}=2250	±SE2=250

Table 31: Comparison of binding of KorB (monomer) to dsHL O_B DNA using different techniques. Dissociation constants from fluorescence anisotropy (FA), circular dichroism (CD) and microscale thermophoresis (MST) data are compared for DNA-binding KorB constructs. Mod. Quad. stands for modified quadratic equation. K_{d1} and K_{d2} are the dissociation constants and SE1 and SE2 are the standard error for binding event1 and binding event2 respectively.

KorB construct	Protein species	Equation	K_d (μM)	$\pm$Std. Error
		FA		
Wild-type	Dimer	Hyperbola	0.63	0.11
CΔ100 KorB	Monomer	Hyperbola	10.1	1.5
NΔ31CΔ100 KorB	Monomer	Hyperbola	1.8	0.93
		CD		
CΔ100 KorB	Monomer	Mod. Quad.	22.6	10.2
	Dimer	Mod. Quad.	9.6	5.9
NΔ31CΔ100 KorB	Monomer	Mod. Quad.	6.8	2.1
	Dimer	Mod. Quad.	1.1	0.7
		MST		
Wild-type	Monomer	Mod. Hyp.	K_{d1}=0.8	(fixed)
			K_{d2}=9.4	$\pm$SE2=2.3
CΔ100 KorB	Monomer	Mod. Hyp.	K_{d1}=1.1	$\pm$SE1=0.3
			K_{d2}=110	$\pm$SE2=11

Table 32: Comparison of binding of KorB to dsFL O_B DNA using different techniques. Dissociation constants from fluorescence anisotropy (FA), circular dichroism (CD) and microscale thermophoresis (MST) data are compared for DNA-binding KorB constructs. Mod. Quad. and Mod. Hyp. stands for modified quadratic equation and modified hyperbola equation respectively. K_{d1} and K_{d2} are the dissociation constants and SE1 and SE2 are the standard error for binding event1 and binding event2 respectively.

7

Conclusions and Outlook

7 Conclusions and Outlook

7.1 Introduction

Plasmid partitioning is a dynamic process that ensures that each daughter cell gets at least one copy of the plasmid during bacterial cell division. DNA partitioning in bacteria primarily involves two proteins, a motor protein (ParA) and a DNA-binding protein (ParB) and these two proteins interact with each other and mediate faithful partitioning of the plasmid. Being a low-copy number plasmid, RK2 requires an active partitioning system. The partitioning of the RK2 depends on IncC, KorB and centromere-like sites on the plasmid (O_B). IncC belongs to the ParA family of proteins and acts as an ATPase. KorB belongs to the ParB superfamily of proteins and acts as a DNA-binding protein, interacting at twelve O_B sites on the RK2 plasmid (Chikami et al., 1985; Figurski and Helinski, 1979; Lowbury et al., 1969; Macartney et al., 1997; Pansegrau et al., 1994).

KorB is one of the most well-characterised ParB proteins. KorB also acts as a transcription repressor, working co-operatively with another RK2 repressor protein, KorA. Structurally, KorB is a multi-domain protein consisting of a partially structured N-terminal domain, a central helical DNA-binding domain containing the HTH motif, an unstructured linker and a C-terminal dimerisation domain. Of these domains, the crystal structures of DBD and the CTD have been solved previously (Delbruck et al., 2002; Khare et al., 2004). However, there is limited information on the structure and function of the N-terminal domain. The aim of this work was to further our understanding about the structure and function of DNA-binding protein, KorB through *in vitro* studies.

The structure of the N-terminal domain of KorB was determined and structural and biophysical properties of the NTD KorB were investigated using a combination of tech-

niques including solution state NMR spectroscopy, circular dichroism, analytical ultracentrifugation and mass spectrometry. Further, the mobility of the NTD KorB was computationally examined using molecular dynamics simulations.

In addition, multiple deletion mutants of KorB were used to investigate the binding of the KorB to the O_B DNA. KorB-DNA binding studies were undertaken using fluorescence anisotropy, circular dichroism and microscale thermophoresis. Various mutants of KorB (wild type, CΔ100 KorB and NΔ31CΔ100 KorB) capable of binding to the DNA were tested and binding affinities of the KorB constructs were compared to the wild-type KorB protein.

7.2 Insights gained from this work

7.2.1 Structural and biophysical characterisation of N-terminal of KorB

The N-terminal region containing amino acids 31-137 of KorB protein was expressed and purified to homogeneity using affinity chromatography and size exclusion chromatography. From 1 L of the culture, 22 mg of NTD KorB was obtained, sufficient to perform the structural and biophysical experiments. In chapters 3 and 4, structural and biophysical properties of NTD KorB have been investigated.

In the size exclusion chromatogram for the NTD KorB, two peaks were observed, that suggested the presence of more than one species in the solution. It is plausible that there is protein monomer-dimer equilibrium or that the NTD KorB is modified. Size exclusion chromatography and analytical ultracentrifugation (AUC) data confirmed that NTD KorB behaves as a monomer in solution *i.e.* no evidence for dimer formation. The molecular mass of NTD KorB from ESI-MS data almost agreed with the theoretical mass but not exactly. It is plausible that there is a partial covalent modification of the NTD KorB. Circular dichroism spectroscopy data suggested that the NTD KorB is partially folded. In the presence of TFE, a slight increase in α-helical character was observed for NTD KorB (with/without the tag).

In protein crystallisation trials of NTD KorB, precipitation was observed in about one-third of the conditions. During the crystallisation trials of NTD KorB, micro-crystals were observed for the protein, but the crystals failed to grow in size with time. The addition of TFE did increase the secondary structure of the NTD KorB but did not aid in the crystallisation trials of the protein. The difficulty to crystallise the NTD KorB in a stable conformation could be attributed to its partially unfolded character, as indicated by the CD data. Since a structure for the NTD KorB was not elucidated by X-ray crystallography, the structure of the protein was obtained with the help of solution state NMR spectroscopy.

2D HSQC spectra overlays of NTD KorB and KorB WT (data not shown) were used to validate the use of deletion mutant for structural insights. The HSQC data suggested independent and modular domain folding within KorB. Using solution state NMR spectroscopy data, 80% of the backbone of NTD KorB was assigned with confidence. The less than ideal completeness of assignments was due to relatively high number of proline residues (10 Pro residues) in a 130 residue protein and the intrinsic flexibility of the protein which gave highly overlapped spectra. The analysis of the N-terminal domain by NMR proved somewhat harder than anticipated, largely because there were more peaks in the HSQC than NH groups. The additional peaks in the HSQC of NTD KorB seemed to be reproducible between different labelled NTD KorB preparations and so were unlikely due to degradation products. It is not yet clear whether the additional peaks in the NTD KorB's HSQC are because there are protein conformers in very slow exchange or because of partial covalent modification of the protein.

Comparison of secondary structure of NTD KorB from the chemical shift data using TALOS+ and from computationally predicted Jpred and Psipred data showed agreement with all of the methods (TALOS+, Jpred and Psipred) predicting two α-helices towards the C-terminal of the protein. An ensemble of 20 conformers for NTD KorB was calculated compatible with the restraints used for structure refinement. The struc-

ture of NTD KorB was mainly disordered and consisted of two α-helices (α1 and α2) towards the C-terminal region (core region). The N-terminal tag of NTD KorB was highly flexible and did not contribute to any peaks in the HSQC spectrum. It is plausible to suggest that the tag does not contribute to the NTD KorB structure and does not reduce the significance of the ensemble. The NTD KorB ensemble structure suggested that N-terminal of the protein is mostly flexible whilst the two α-helices towards the C-terminal show limited flexibility. The dynamics of the protein was studied using molecular dynamics simulations. The MD data corroborated that N-terminal of the protein exhibits a wide range of motion whereas the same was not observed for the two α-helices towards the C-terminal of the protein.

7.2.2 Biophysical characterisation of DNA-binding domain of KorB

Two KorB deletion mutants (CΔ100 KorB and NΔ31CΔ100 KorB) capable of binding DNA were expressed and purified. From 1 L of the culture, approximately 25 mg of CΔ100 KorB and 20 mg of NΔ31CΔ100 KorB were obtained, sufficient to perform biophysical and DNA-binding experiments.

In the size exclusion chromatogram, CΔ100 KorB was observed as a single species and SDS-PAGE showed that the protein is devoid of any impurities, whilst impurities were seen in the NΔ31CΔ100 KorB purification SDS-PAGE gel (after size exclusion run). Both proteins behaved as monomers on the size exclusion column. Like NTD KorB, crystallisation trials of both deletion mutants were also set to elucidate the structure of DNA-binding domain of KorB. In all of the crystallisation screens, precipitation was observed for almost one-third of the conditions and micro-crystals were observed for both mutants but the crystals failed to grow in size with time. The difficulty to crystallise these deletion mutants might be because of the disorder in the protein possibly due to the flexible N-terminal domain. The NΔ31CΔ100 KorB was observed to be more stable at RT when compared to CΔ100 KorB. CD data suggested an increase in α-helical structure for NΔ31CΔ100 KorB and it appeared to be more ordered when

compared to CΔ100 KorB probably because of the loss of first 53 disordered residues from the N-terminal of the CΔ100 KorB.

7.2.3 KorB and DNA interaction

The DNA-binding properties of KorB deletion mutants (CΔ100 KorB and NΔ31CΔ100 KorB) were compared to KorB WT protein using fluorescence anisotropy (FA), circular dichroism (CD) and microscale thermophoresis (MST).

From FA and MST data, similar dissociation constants were observed for KorB WT-dsFL DNA interaction and those constants were similar to the EMSAs done before this study (data not shown). When compared to KorB WT, the value of dissociation constants for CΔ100 KorB and NΔ31CΔ100 KorB for O_B DNA were consistently higher in all the CD, FA and MST experiments. Comparing the binding between FA and CD data for NΔ31CΔ100 KorB to dsFL and dsHL DNA, the protein bound five fold tighter to dsFL DNA. On the other hand, for CΔ100 KorB, a second binding event with dsFL O_B DNA was observed with weaker affinity than first binding event. Data from all the techniques (FA, CD and MST) suggest CΔ100 KorB binds dsHL more tightly than to dsFL. But a consistent correlation was not established between the binding constants derived from FA, CD and MST data. When comparing the dissociation constants between the truncated KorB mutants and dsFL DNA, binding of NΔ31CΔ100 KorB was found to be tighter than to the CΔ100 KorB. This weaker binding of CΔ100 KorB can be attributed possibly to the steric hindrance posed by the first 53 aa (23 aa from N-terminal tag and 30 aa from KorB). In addition, the biphasic binding was observed from the MST data, for both KorB WT and CΔ100 KorB with dsFL O_B DNA. This biphasic binding indicates a second binding event of KorB binding to the DNA which possibly is a non-specific protein-DNA interaction. It was expected for KorB constructs with deleted dimerisation domains to dimerise indirectly and only in the presence of the DNA. Their reduced affinity to the DNA can be explained by the lack of C-terminal dimerisation domain and absence of pre-formed dimers in solution.

7.3 A spreading model for plasmid partitioning

ParB proteins with spreading and bridging interactions on the DNA can be observed across chromosomal and plasmid partitioning systems and yet the interaction interfaces have remained elusive (Bingle et al., 2005; Graham et al., 2014; Havey et al., 2012; Rodionov et al., 1999). This is likely because of the lack of a full-length structure for a ParB protein and complex domain organisation of the protein (Chen et al., 2015; Fisher et al., 2017). ParB-DNA segregation complexes involve binding of ParB to both DNA and other ParB proteins on the DNA (Sanchez et al., 2015; Taylor et al., 2015). During the course of segregation, the NTD ParB may exhibit different functions and the N-terminal is proposed to be involved in binding both DNA and other ParB proteins (Leonard et al., 2004). The ability of ParB to bind DNA specifically and non-specifically supports ParB spreading on the DNA (Fisher et al., 2017). The flexible nature of the N-terminal domain of ParB is mainly responsible for spreading interactions with a neighbouring ParB (Graham et al., 2014; Havey et al., 2012). The structural study of ParB from *H. pylori* suggested that N-terminal domain in the protein is likely to be important in spreading along the DNA and possibly in bridging the DNA as well (Chen et al., 2015; Song et al., 2017; Taylor et al., 2015). Also, the spreading of the ParB on the DNA is facilitated by *parS*, which act as nucleation points responsible for conformational changes in the protein (Broedersz et al., 2014; Leonard et al., 2004). In case of KorB, the N-terminal of KorB is majorly disordered but the two C-terminal helices show reduced flexibility. As expected, by itself, the N-terminal of KorB does not bind DNA (from FA DNA-binding data) as it lacks the DNA-binding domain. However, based on the DNA partitioning model proposed by Chen et al. (2015) it can be postulated that N-terminal of KorB is likely to facilitate the spreading of the full length protein along the DNA. NTD KorB may interact non-specifically at high DNA or protein concentrations and the flexible N-terminal may provide multiple protein-protein interaction interfaces where KorB can interact with other KorB molecules.

Based on the DNA partitioning model proposed by Chen et al. (2015), a similar spreading model for KorB is proposed. Figure 79 represents the KorB spreading model for plasmid partitioning. KorB contains a partially structured NTD, a central DBD, an unstructured linker, and a C-terminal dimerisation domain, where these three domains of KorB (NTD, DBD and CTD) participate in the DNA binding. The helical DNA-binding domain binds to the plasmid RK2 at a specific O_B site and the C-terminal domain of KorB facilitates the protein to dimerise with a neighbouring KorB monomer and stabilises the KorB-O_B complex (Bingle et al., 2005). Similar to the interaction in the ParB protein from *H. pylori*, the neighbouring KorB dimers are proposed to interact with the adjacent KorB molecule facilitated by the NTD and the protein spreads along the DNA horizontally (Chen et al., 2015). Further, the N-terminal domain helps in the bridging KorB dimers using transverse interactions and vertically recruiting the distal DNA. Considering the specific and non-specific binding of KorB to the DNA, a higher-order nucleo-protein complex can be envisioned. Subsequently, the nucleo-protein complex recruits the motor protein, IncC. Further to the recruitment, the IncC mediates plasmid partitioning using a mechanism that remains elusive.

The caveat to this postulated partitioning model is that there is no direct experimental evidence in this study about the non-specific interaction of KorB and DNA as the N-terminal of KorB neither interacts with itself nor does it bind to the DNA. The experimental data supporting the non-specific interaction of KorB and DNA is the two step binding event (biphasic) from the MST data. The biphasic binding was observed for both KorB WT and CΔ100 KorB with dsFL O_B DNA. The biphasic event indicates a second but weak KorB binding to the DNA possibly because of non-specific protein-DNA interaction. Considering the lack of experimental evidence of non-specific KorB-DNA binding, the spreading model proposed is based on KorB homologue, ParB protein from *H. pylori* as proposed by Chen et al., 2015.

While this study investigates the structure of N-terminal of KorB and DNA-binding properties of KorB, further research is needed to examine the role of the KorB in active

plasmid partitioning. The KorB working along with IncC (the ATPase involved in RK2 partitioning) makes the partitioning system more complicated to understand. Further structural understanding of KorB and its interaction with DNA will be required to underpin KorB's role in orchestrating efficient RK2 plasmid partitioning. Future work would be beneficial to understand the plasmid partitioning at atomic-level and the structural and biophysical principles used in this study can be extended to investigate chromosomal segregation in bacteria.

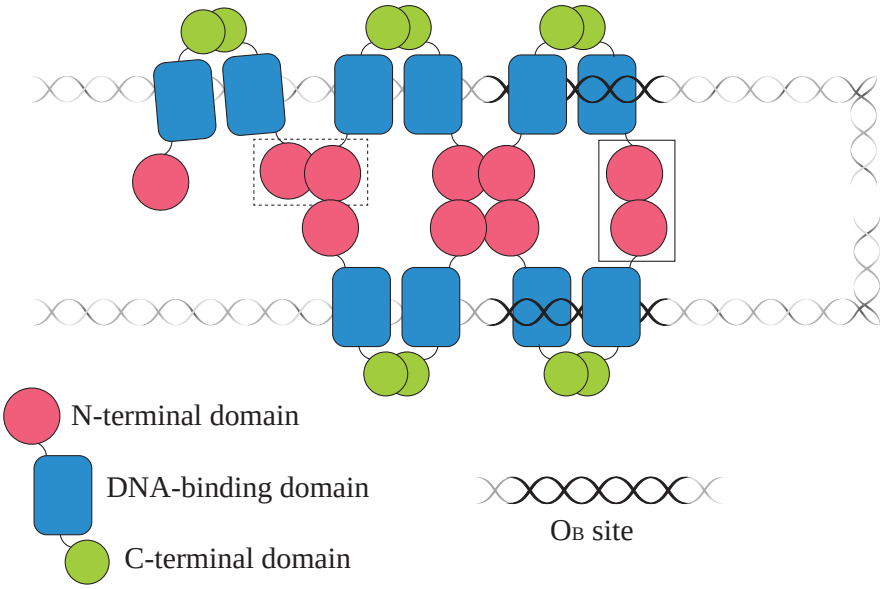

Figure 79: A spreading model for plasmid partitioning. The C-terminal dimerisation (in green), the DNA-binding (in blue) and the flexible N-terminal domains (in red) of KorB are shown. KorB binds plasmid DNA as dimer at the specific O_B site (in black) or non-specific sites (in grey). There is no direct experimental evidence in this study about the non-specific interaction of KorB and DNA. Using the N-terminal domain, multiple KorB molecules spread along the DNA by adjacent interactions (dashed box) horizontally and bridge the DNA through transverse interactions (solid box) vertically. A higher-order nucleo-protein complex recruits the motor protein, IncC mediating the plasmid partitioning using a mechanism that remains elusive. The model is based on the partitioning mechanism proposed by Chen et al., 2015.

Bibliography

Ah-Seng, Y., Lopez, F., Pasta, F., Lane, D., and Bouet, J. Y. (2009). Dual role of DNA in regulating ATP hydrolysis by the SopA partition protein. *Journal of Biological Chemistry*, 284(44):30067–30075.

Ángyán, A. F., Szappanos, B., Perczel, A., and Gáspári, Z. (2010). CoNSEnsX: an ensemble view of protein structures and NMR-derived experimental data. *BMC Structural Biology*, 10(1):39.

Atkins, P. W. and De Paula, J. (2006). *Elements of physical chemistry*. Oxford University Press.

Aylett, C. H. and Lowe, J. (2012). Superstructure of the centromeric complex of TubZRC plasmid partitioning systems. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41):16522–16527.

Aylett, C. H., Wang, Q., Michie, K. A., Amos, L. A., and Lowe, J. (2010). Filament structure of bacterial tubulin homologue TubZ. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46):19766–19771.

Balzer, D., Ziegelin, G., Pansegrau, W., Kruft, V., and Lanka, E. (1992). KorB protein of promiscuous plasmid RP4 recognizes inverted sequence repetitions in regions essential for conjugative plasmid transfer. *Nucleic Acids Research*, 20(8):1851–1858.

Batt, S. M., Bingle, L. E., Dafforn, T. R., and 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. *Journal of Molecular Biology*, 385(5):1361–1374.

Bax, A., Clore, G., and Gronenborn, A. M. (1990). 1H-1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *Journal of Magnetic Resonance (1969)*, 88(2):425–431.

Baxter, J. C. and Funnell, B. E. (2014). Plasmid Partition Mechanisms. *Microbiology Spectrum*, 2(6):1–20, DOI: 10.1128/microbiolspec.PLAS–0023–2014.

Bechhofer, D. H. and Figurski, D. H. (1983). Map location and nucleotide sequence of *korA*, a key regulatory gene of promiscuous plasmid RK2. *Nucleic Acids Research*, 11(21):7453–7469.

Berjanskii, M., Liang, Y., Zhou, J., Tang, P., Stothard, P., Zhou, Y., Cruz, J., MacDonell, C., Lin, G., Lu, P., and Wishart, D. S. (2010). PROSESS: a protein structure evaluation suite and server. *Nucleic Acids Research*, 38(Web Server issue):W633–40.

Bhattacharya, A., Tejero, R., and Montelione, G. T. (2006). Evaluating protein structures determined by structural genomics consortia. *Proteins: Structure, Function, and Bioinformatics*, 66(4):778–795.

Bignell, C. and Thomas, C. M. (2001). The bacterial ParA-ParB partitioning proteins. *Journal of Biotechnology*, 91(1):1–34.

Bingle, L. E., Macartney, D. P., Fantozzi, A., Manzoor, S. E., and Thomas, C. M. (2005). Flexibility in repression and cooperativity by KorB of broad host range IncP-1 plasmid RK2. *Journal of Molecular Biology*, 349(2):302–316.

Bingle, L. E. and Thomas, C. M. (2001). Regulatory circuits for plasmid survival. *Current Opinion in Microbiology*, 4(2):194–200.

Bingle, L. E. H., Rajasekar, K. V., tul Muntaha, S., Nadella, V., Hyde, E. I., and Thomas, C. M. (2008). A single aromatic residue in transcriptional repressor protein KorA is critical for cooperativity with its co-regulator KorB. *Molecular Microbiology*, 70(6):1502–1514.

Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016):164–70.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1-2):248–254.

Breier, A. M. and Grossman, A. D. (2007). Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the Bacillus subtilis chromosome. *Molecular Microbiology*, 64(3):703–718.

Broedersz, C. P., Wang, X., Meir, Y., Loparo, J. J., Rudner, D. Z., and Wingreen, N. S. (2014). Condensation and localization of the partitioning protein ParB on the bacterial chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24):8809–8814.

Bu, Z. and Callaway, D. J. (2011). Proteins MOVE! Protein dynamics and long-range allostery in cell signaling. *Advances in Protein Chemistry and Structural Biology*, 83:163–221.

Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., and Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, 41(Web Server issue):W349–57.

Cary, P. and Kneale, G. G. (2009). Circular Dichroism for the Analysis of Protein-DNA Interactions. pages 613–624. Humana Press.

Chen, B. W., Lin, M. H., Chu, C. H., Hsu, C. E., and Sun, Y. J. (2015). Insights into ParB spreading from the complex structure of Spo0J and parS. *Proceedings of the National Academy of Sciences of the United States of America*.

Chen, J. W., Romero, P., Uversky, V. N., and Dunker, A. K. (2006a). Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *Journal of proteome research*, 5(4):879–87.

Chen, J. W., Romero, P., Uversky, V. N., and Dunker, A. K. (2006b). Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *Journal of proteome research*, 5(4):888–98.

Chikami, G. K., Guiney, D. G., Schmidhauser, T. J., and Helinski, D. R. (1985). Comparison of 10 IncP plasmids: homology in the regions involved in plasmid replication. *Journal of Bacteriology*, 162(2):656–660.

Connor, F., Cary, P. D., Read, C. M., Preston, N. S., Driscoll, P. C., Denny, P., Crane-Robinson, C., and Ashworth, A. (1994). DNA binding and bending properties of the postmeiotically expressed Sry-related protein Sox-5. *Nucleic Acids Research*, 22(16):3339–3346.

Datta, N. and Hedges, R. W. (1972). Host ranges of R factors. *Journal of General Microbiology*, 70(3):453–460.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., Richardson, D. C., and Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35(Web Server issue):W375–83.

Davtyan, A., Schafer, N. P., Zheng, W., Clementi, C., Wolynes, P. G., and Papoian, G. A. (2012). AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *The Journal of Physical Chemistry B*, 116(29):8494–8503.

de Jong, L. A., Uges, D. R., Franke, J. P., and Bischoff, R. (2005). Receptor–ligand binding assays:

Technologies and Applications. *Journal of Chromatography B*, 829(1-2):1–25.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293.

Delbruck, H., Ziegelin, G., Lanka, E., and 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. *Journal of Biological Chemistry*, 277(6):4191–4198.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015). JPred4: A protein secondary structure prediction server. *Nucleic Acids Research*, 43(W1):W389–W394.

Duhr, S. and Braun, D. (2006). Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences of the United States of America*, 103(52):19678–82.

Dunham, T. D., Xu, W., Funnell, B. E., and Schumacher, M. A. (2009). Structural basis for ADP-mediated transcriptional regulation by P1 and P7 ParA. *EMBO J*, 28(12):1792–1802.

Dunker, A., Lawson, J., 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., and Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1):26–59.

Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, 18(6):756–764.

Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3):197–208.

Ebersbach, G. and Gerdes, K. (2001). The double par locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA. *Proceedings of the National Academy of Sciences of the United States of America*, 98(26):15078–15083.

Figurski, D. H. and Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences of the United States of America*, 76(4):1648–1652.

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., and Dillingham, M. S. (2017). The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. *eLife*, 6.

Funnell, B. and Slavcev, R. (2004). Partition systems of bacterial plasmids. In *Plasmid Biology*, chapter 5, pages 81–104.

Galkin, V. E., Orlova, A., Rivera, C., Mullins, R. D., and Egelman, E. H. (2009). Structural polymorphism of the ParM filament and dynamic instability. *Structure*, 17(9):1253–1264.

Garbett, N. C., Ragazzon, P. A., and Chaires, J. B. (2007). Circular dichroism to determine binding mode and affinity of ligand–DNA interactions. *Nature Protocols*, 2(12):3166–3172.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, pages 571–607. Humana Press, Totowa, NJ.

Gayathri, P., Fujii, T., Moller-Jensen, J., van den Ent, F., Namba, K., and Lowe, J. (2012). A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science*, 338(6112):1334–1337.

Gerdes, K., Howard, M., and Szardenings, F. (2010). Pushing and pulling in prokaryotic DNA segregation. *Cell*, 141(6):927–942.

Gerdes, K. and Molin, S. (1986). Partitioning of plasmid R1. Structural and functional analysis

of the parA locus. *Journal of Molecular Biology*, 190(3):269–279.

Gerdes, K., Moller-Jensen, J., and Bugge Jensen, R. (2000). Plasmid and chromosome partitioning: surprises from phylogeny. *Molecular Microbiology*, 37(3):455–466.

Gianni, S., Dogan, J., and Jemth, P. (2014). Distinguishing induced fit from conformational selection. *Biophysical Chemistry*, 189:33–39.

Graham, T. G. W., Wang, X., Song, D., Etson, C. M., van Oijen, A. M., Rudner, D. Z., and Loparo, J. J. (2014). ParB spreading requires DNA bridging. *Genes and Development*, 28(11):1228–38.

Greenfield, N. J. (1996). Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data. *Analytical Biochemistry*, 235(1):1–10.

Grzesiek, S. and Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society*, 114(16):6291–6293.

Grzesiek, S. and Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *Journal of Biomolecular NMR*, 3(2):185–204.

Gsponer, J., Futschik, M. E., Teichmann, S. A., and Babu, M. M. (2008). Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science*, 322(5906):1365–1368.

Hanai, R., Liu, R., Benedetti, P., Caron, P. R., Lynch, A. S., and Wang, J. C. (1996). Molecular dissection of a protein SopB essential for *Escherichia coli* F plasmid partition. *Journal of Biological Chemistry*, 271(29):17469–17475.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature*, 381(6583):571–580.

Havey, J. C., Vecchiarelli, A. G., and Funnell, B. E. (2012). ATP-regulated interactions between P1 ParA, ParB and non-specific DNA that are stabilized by the plasmid partition site, parS. *Nucleic Acids Research*, 40(2):801–812.

Heald, R. (2000). Motor Function in the Mitotic Spindle Minireview. *Cell*, 102(4):399–402.

Heald, R. and Khodjakov, A. (2015). Thirty years of search and capture: The complex simplicity of mitotic spindle assembly. *The Journal of Cell Biology*, 211(6):1103–11.

Higman, V. A. (2017). Triple Resonance Backbone Assignment; www.protein-nmr.org.uk/solution-nmr/assignment-theory/triple-resonance-backbone-assignment.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38.

Hwang, L. C., Vecchiarelli, A. G., Han, Y. W., Mizuuchi, M., Harada, Y., Funnell, B. E., and Mizuuchi, K. (2013). ParA-mediated plasmid partition driven by protein pattern self-organization. *EMBO J*, 32(9):1238–1249.

Hyde, E. I., Callow, P., Rajasekar, K. V., Timmins, P., Patel, T. R., Siligardi, G., Hussain, R., White, S. A., Thomas, C. M., and Scott, D. J. (2017). Intrinsic disorder in the partitioning protein KorB persists after co-operative complex formation with operator DNA and KorA. *Biochemical Journal*, 474(18):3121–3135.

Iakoucheva, L. M., Brown, C. J., Lawson, J., Obradović, Z., and Dunker, A. (2002). Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology*, 323(3):573–584.

Ikura, M., Kay, L. E., and Bax, A. (1990). A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, 29(19):4659–67.

Ishima, R. and Torchia, D. A. (2000). Protein dynamics from NMR. *Nature Structural Biology*,

7(9):740–743.

Jagura-Burdzy, G., Kostelidou, K., Pole, J., Khare, D., Jones, A., Williams, D. R., and Thomas, C. M. (1999). IncC of broad-host-range plasmid RK2 modulates KorB transcriptional repressor activity In vivo and operator binding in vitro. *Journal of Bacteriology*, 181(9):2807–2815.

Jagura-Burdzy, G. and Thomas, C. M. (1995). Purification of KorA Protein from Broad Host Range Plasmid RK2: Definition of a Hierarchy of KorA Operators. *Journal of Molecular Biology*, 253(1):39–50.

Jameson, D. M. and Ross, J. A. (2010). Fluorescence Polarization/Anisotropy in Diagnostics and Imaging. *Chemical Reviews*, 110(5):2685–2708.

Jensen, R. B., Lurz, R., and Gerdes, K. (1998). Mechanism of DNA segregation in prokaryotes: replicon pairing by parC of plasmid R1. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8550–8555.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, 1077:101–113.

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., and Duhr, S. (2011). Molecular interaction studies using microscale thermophoresis. *Assay and Drug Development Technologies*, 9(4):342–53.

Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, 89(3):496–514.

Kelly, S. M., Jess, T. J., and Price, N. C. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta*, 1751(2):119–139.

Khare, D., Ziegelin, G., Lanka, E., and Heinemann, U. (2004). Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nature Structural and Molecular Biology*, 11(7):656–663.

Klass, J., Murphy IV, F. V., Fouts, S., Serenil, M., Changela, A., Siple, J., and Churchill, M. E. A. (2003). The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *Nucleic Acids Research*, 31(11):2852–2864.

Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-Grained Protein Models and Their Applications. *Chemical Reviews*, 116(14):7898–7936.

Kostelidou, K., Jones, A. C., and Thomas, C. M. (1999). Conserved C-terminal region of global repressor KorA of broad-host-range plasmid RK2 is required for co-operativity between KorA and a second RK2 global regulator, KorB. *Journal of Molecular Biology*, 289(2):211–21.

Kostelidou, K. and Thomas, C. M. (2000). The hierarchy of KorB binding at its 12 binding sites on the broad-host-range plasmid RK2 and modulation of this binding by IncC1 protein. *Journal of Molecular Biology*, 295(3):411–22.

Lakowicz, J. (2006). Fluorescence Anisotropy. In *Principles of Fluorescence Spectroscopy*, pages 353–382. Springer US, Boston, MA.

Larsen, R. A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J. (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis. *Genes and Development*, 21(11):1340–1352.

Lebowitz, J., Lewis, M. S., and Schuck, P. (2009). Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Science*, 11(9):2067–2079.

Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009). Discovery Through the Computational Microscope. *Structure*, 17(10):1295–1306.

Lee, W., Cornilescu, G., Dashti, H., Eghbalnia, H. R., Tonelli, M., Westler, W. M., Butcher, S. E.,

Henzler-Wildman, K. A., and Markley, J. L. (2016a). Integrative NMR for biomolecular research. *Journal of Biomolecular NMR*, 64(4):307–32.

Lee, W., Kim, J. H., Westler, W. M., and Markley, J. L. (2011). PONDEROSA, an automated 3D-NOESY peak picking program, enables automated protein structure determination. *Bioinformatics*, 27(12):1727–1728.

Lee, W., Petit, C. M., Cornilescu, G., Stark, J. L., and Markley, J. L. (2016b). The AUDANA algorithm for automated protein 3D structure determination from NMR NOE data. *Journal of Biomolecular NMR*, 65(2):51–57.

Lee, W., Stark, J. L., and Markley, J. L. (2014). PONDEROSA-C/S: client–server based software package for automated protein 3D structure determination. *Journal of Biomolecular NMR*, 60(2-3):73–75.

Lee, W., Tonelli, M., and Markley, J. L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, 31(8):1325–7.

Leonard, T. A., Butler, P. J., and Lowe, J. (2004). Structural analysis of the chromosome segregation protein Spo0J from Thermus thermophilus. *Molecular Microbiology*, 53(2):419–432.

Lobley, A., Whitmore, L., and Wallace, B. A. (2002). DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, 18(1):211–212.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003). Structure validation by Cα geometry: φ,ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3):437–450.

Lowbury, E., Lilly, H., Kidson, A., Ayliffe, G., and Jones, R. (1969). Sensitivity Of *Pseudomonas aeruginosa* To Antibiotics: Emergence Of Strains Highly Resistant To Carbenicillin. *The Lancet*, 294(7618):448–452.

Lukaszewicz, M., Kostelidou, K., Bartosik, A. A., Cooke, G. D., Thomas, C. M., and Jagura-Burdzy, G. (2002). Functional dissection of the ParB homologue (KorB) from IncP-1 plasmid RK2. *Nucleic Acids Research*, 30(4):1046–1055.

Lüthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, 356(6364):83–85.

Macartney, D. P., Williams, D. R., Stafford, T., and Thomas, C. M. (1997). Divergence and conservation of the partitioning and global regulation functions in the central control region of the IncP plasmids RK2 and R751. *Microbiology*, 143(7):2167–2177.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989a). Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy:. *Biochemistry*, 28(15):6150–6156.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989b). Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American Chemical Society*, 111(4):1515–1517.

Moller-Jensen, J., Borch, J., Dam, M., Jensen, R. B., Roepstorff, P., and Gerdes, K. (2003). Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Molecular Cell*, 12(6):1477–1487.

Moller-Jensen, J., Ringgaard, S., Mercogliano, C. P., Gerdes, K., and Lowe, J. (2007). Structural analysis of the ParR/parC plasmid partition complex. *EMBO J*, 26(20):4413–4422.

Motallebi-Veshareh, M., Rouch, D. A., and Thomas, C. M. (1990). A family of ATPases involved in active partitioning of diverse bacterial plasmids. *Molecular Microbiology*, 4(9):1455–1463.

Murray, H., Ferreira, H., and Errington, J. (2006). The bacterial chromosome segregation protein Spo0J spreads along DNA from parS nucleation sites. *Molecular Microbiology*, 61(5):1352–1361.

Musacchio, A., Wilmanns, M., and Saraste, M. (1994). Structure and function of the SH3 domain. *Progress in Biophysics and Molecular Biology*, 61(3):283–297.

Ni, L., Xu, W., Kumaraswami, M., and 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. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11763–11768.

Ogura, T. and Hiraga, S. (1983). Partition mechanism of F plasmid: two plasmid gene-encoded products and a cis-acting region are involved in partition. *Cell*, 32(2):351–360.

Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005). Comparing and Combining Predictors of Mostly Disordered Proteins. *Biochemistry*, 44(6):1989–2000.

Olejniczak, E. T., Xu, R. X., and Fesik, S. W. (1992). A 4D HCCH-TOCSY experiment for assigning the side chain1H and13C resonances of proteins. *Journal of Biomolecular NMR*, 2(6):655–659.

Owen, B. and McMurray, C. (2009). Rapid method for measuring DNA binding to protein using fluorescence anisotropy. *Protocol Exchange*.

Panchal, S. C., Bhavesh, N. S., and 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. *Journal of Biomolecular NMR*, 20(2):135–147.

Pancsa, R. and Tompa, P. (2012). Structural disorder in eukaryotes. *PLoS One*, 7(4):e34687.

Pansegrau, W., Lanka, E., Barth, P. T., Figurski, D. H., Guiney, D. G., Haas, D., Helinski, D. R., Schwab, H., Stanisich, V. A., and Thomas, C. M. (1994). Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *Journal of Molecular Biology*, 239(5):623–663.

Papapanagiotou, I., Streeter, S. D., Cary, P. D., and Kneale, G. G. (2007). DNA structural deformations in the interaction of the controller protein C.AhdI with its operator sequence. *Nucleic Acids Research*, 35(8):2643–50.

Pawson, T. and Schlessingert, J. (1993). SH2 and SH3 domains. *Current Biology*, 3(7):434–442.

Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., Yu, H., Wu, Z., and Schulten, K. (2015). Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, 31:64–74.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

Pillet, F., Sanchez, A., Lane, D., Anton Leberre, V., and Bouet, J. Y. (2011). Centromere binding specificity in assembly of the F plasmid partition complex. *Nucleic Acids Research*, 39(17):7477–7486.

Plimpton, S. (1995). Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117(1):1–19.

Popp, D., Narita, A., Oda, T., Fujisawa, T., Matsuo, H., Nitanai, Y., Iwasa, M., Maeda, K., Onishi, H., and Maeda, Y. (2008). Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability. *EMBO J*, 27(3):570–579.

Povey, J. F., Smales, C. M., Hassard, S. J., and Howard, M. J. (2007). Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function. *Journal of Structural Biology*, 157(2):329–338.

Rabi, I. I., Zacharias, J. R., Millman, S., and Kusch, P. (1992). Milestones in magnetic resonance: 'a new method of measuring nuclear magnetic moment' . 1938. *Journal of Magnetic Resonance Imaging : JMRI*, 2(2):131–133.

Radnedge, L., Youngren, B., Davis, M., and Austin, S. (1998). Probing the structure of complex macromolecular interactions by homolog specificity scanning: the P1 and P7 plasmid partition systems. *EMBO J*, 17(20):6076–6085.

Rajasekar, K., Muntaha, S. T., Tame, J. R., Kommareddy, S., Morris, G., Wharton, C. W., Thomas, C. M., White, S. A., Hyde, E. I., and Scott, D. J. (2010). Order and disorder in the domain organization of the plasmid partition protein KorB. *Journal of Biological Chemistry*, 285(20):15440–15449.

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., and White, S. A. (2016). Flexibility of KorA, a plasmid-encoded, global transcription regulator, in the presence and the absence of its operator. *Nucleic Acids Research*, 44(10):4947–4956.

Ravin, N. V., Rech, J., and Lane, D. (2003). Mapping of functional domains in F plasmid partition proteins reveals a bipartite SopB-recognition domain in SopA. *Journal of Molecular Biology*, 329(5):875–889.

Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1):W320–W324.

Rodionov, O., Lobocka, M., and Yarmolinsky, M. (1999). Silencing of genes flanking the P1 plasmid centromere. *Science*, 283(5401):546–9.

Rosenberg, A. H., Lade, B. N., Chui, D. S., Lin, S. W., Dunn, J. J., and Studier, F. W. (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, 56(1):125–35.

Salje, J. and Lowe, J. (2008). Bacterial actin: architecture of the ParMRC plasmid DNA partitioning complex. *EMBO J*, 27(16):2230–2238.

Sanchez, A., Cattoni, D. I., Walter, J.-C., Rech, J., Parmeggiani, A., Nollmann, M., and Bouet, J.-Y. (2015). Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. *Cell Systems*, 1(2):163–173.

Sanchez, A., Rech, J., Gasc, C., and Bouet, J. Y. (2013). Insight into centromere-binding properties of ParB proteins: a secondary binding motif is essential for bacterial genome maintenance. *Nucleic Acids Research*, 41(5):3094–3103.

Sandhu, K. S. (2009). Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *Journal of Molecular Recognition*, 22(1):1–8.

Scholefield, G., Whiting, R., Errington, J., and Murray, H. (2011). Spo0J regulates the oligomeric state of Soj to trigger its switch from an activator to an inhibitor of DNA replication initiation. *Molecular Microbiology*, 79(4):1089–1100.

Schuck, P., Lary, J., Stafford, W., Liu, S., Olsen, P., Hayes, D., Moody, T., Ridgeway, T., Lyons, D., and Laue, T. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical Journal*, 78(3):1606–19.

Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002). Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems. *Biophysical Journal*, 82(2):1096–1111.

Schumacher, M. A. (2008). Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation. *Biochemical Journal*, 412(1):1–18.

Schumacher, M. A. (2012). Bacterial plasmid partition machinery: a minimalist approach to survival. *Current Opinion in Structural Biology*, 22(1):72–79.

Schumacher, M. A. and Funnell, B. E. (2005). Structures of ParB bound to DNA reveal mecha-

nism of partition complex formation. *Nature*, 438(7067):516–519.

Schumacher, M. A., Glover, T. C., Brzoska, A. J., Jensen, S. O., Dunham, T. D., Skurray, R. A., and Firth, N. (2007a). Segrosome structure revealed by a complex of ParR with centromere DNA. *Nature*, 450(7173):1268–1271.

Schumacher, M. A., Mansoor, A., and Funnell, B. E. (2007b). Structure of a four-way bridged ParB-DNA complex provides insight into P1 segrosome assembly. *Journal of Biological Chemistry*, 282(14):10456–10464.

Schumacher, M. A., Piro, K. M., and Xu, W. (2010). Insight into F plasmid DNA segregation revealed by structures of SopB and SopB-DNA complexes. *Nucleic Acids Research*, 38(13):4514–4526.

Seidel, S. A. I., Dijkman, P. M., Lea, W. A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J. S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F. A., Ladbury, J. E., Schreiber, G., Watts, A., Braun, D., and Duhr, S. (2013). Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*, 59(3):301–15.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4):213–223.

Shiraki, K., Nishikawa, K., and Goto, Y. (1995). Trifluoroethanol-induced Stabilization of the alpha-Helical Structure of beta-Lactoglobulin: Implication for Non-hierarchical Protein Folding. *Journal of Molecular Biology*, 245(2):180–194.

Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., and Dunker, A. K. (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Research*, 35(Database issue):D786–93.

Siligardi, G., Hussain, R., Patching, S. G., and Phillips-Jones, M. K. (2014). Ligand- and drug-binding studies of membrane proteins revealed through circular dichroism spectroscopy. *Biochimica et Biophysica Acta*, 1838(1):34–42.

Simpson, A. E., Skurray, R. A., and Firth, N. (2003). A single gene on the staphylococcal multiresistance plasmid pSK1 encodes a novel partitioning system. *Journal of Bacteriology*, 185(7):2143–2152.

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Genetics*, 17(4):355–362.

Song, D. and Loparo, J. J. (2015). Building bridges within the bacterial chromosome. *Trends in Genetics*, 31(3):164–173.

Song, D., Rodrigues, K., Graham, T. G., and Loparo, J. J. (2017). A network of cis and trans interactions is required for ParB spreading. *Nucleic Acids Research*, 45(12):7106–7117.

Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Analytical Biochemistry*, 287(2):243–51.

Sreerama, N. and 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. *Analytical Biochemistry*, 287(2):252–260.

Studier, F. W. and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*, 189(1):113–30.

Surtees, J. A. and Funnell, B. E. (1999). P1 ParB domain structure includes two independent multimerization domains. *Journal of Bacteriology*, 181(19):5898–5908.

Surtees, J. A. and Funnell, B. E. (2001). The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex. *Journal of Biological Chemistry*, 276(15):12385–12394.

Tamiola, K., Acar, B., and Mulder, F. A. A. (2010). Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society*, 132(51):18000–18003.

Tamiola, K. and Mulder, F. A. A. (2011). ncIDP-assign: a SPARKY extension for the effective NMR assignment of intrinsically disordered proteins. *Bioinformatics*, 27(7):1039–40.

Taylor, J. A., Pastrana, C. L., Butterer, A., Pernstich, C., Gwynn, E. J., Sobott, F., Moreno-Herrero, F., and Dillingham, M. S. (2015). Specific and non-specific interactions of ParB with DNA: implications for chromosome segregation. *Nucleic Acids Research*, 43(2):719–31.

Teng, Q. (2007). *Structural Biology: Practical NMR Applications*. Springer US.

The Nobel Prize (1926). The Nobel Prize in Chemistry; bit.ly/2LzVB6H.

The Nobel Prize (1944). The Nobel Prize in Physics; bit.ly/2LD9m4w.

The Nobel Prize (1952). The Nobel Prize in Physics; bit.ly/2v3BDa3.

Theophilus, B. D. and 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 Research*, 15(18):7443–7450.

Thermofisher. Amine-reactive crosslinker chemistry. *bit.ly/2LUJOAN*, 2017-11-28.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 579(15):3346–3354.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends in Biochemical Sciences*, 37(12):509–516.

van den Ent, F., Moller-Jensen, J., Amos, L. A., Gerdes, K., and Lowe, J. (2002). F-actin-like filaments formed by plasmid segregation protein ParM. *EMBO J*, 21(24):6935–6943.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chemical reviews*, 114(13):6589–631.

Vecchiarelli, A. G., Han, Y. W., Tan, X., Mizuuchi, M., Ghirlando, R., Biertumpfel, C., Funnell, B. E., and Mizuuchi, K. (2010). ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition. *Molecular Microbiology*, 78(1):78–91.

Vecchiarelli, A. G., Hwang, L. C., and Mizuuchi, K. (2013). Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15):E1390–7.

Vecchiarelli, A. G., Mizuuchi, K., and Funnell, B. E. (2012). Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Molecular Microbiology*, 86(3):513–523.

Vecchiarelli, A. G., Neuman, K. C., and Mizuuchi, K. (2014). A propagating ATPase gradient drives transport of surface-confined cellular cargo. *Proceedings of the National Academy of Sciences of the United States of America*, 111(13):4880–4885.

Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., and Schuck, P. (2004). Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Analytical Biochemistry*, 326(2):234–256.

Weitao, T., Dasgupta, S., and Nordstrom, K. (2000). Role of the mukB gene in chromosome and plasmid partition in *Escherichia coli*. *Molecular Microbiology*, 38(2):392–400.

Whitmore, L. and Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32(Web Server issue):W668–73.

Whitmore, L. and Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, 89(5):392–400.

Wiederstein, M. and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server):W407–W410.

Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D., and Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, 1(7):100.

Williams, D. R., Macartney, D. P., and 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 1:3369–3378.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994). A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *Journal of the American Chemical Society*, 116(26):11655–11666.

Appendix

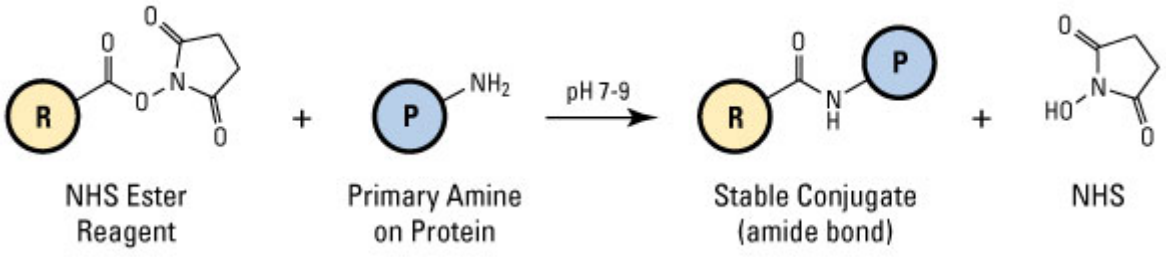

Figure 80: The NHS-ester chemistry reaction. Image reproduced from Thermofisher.

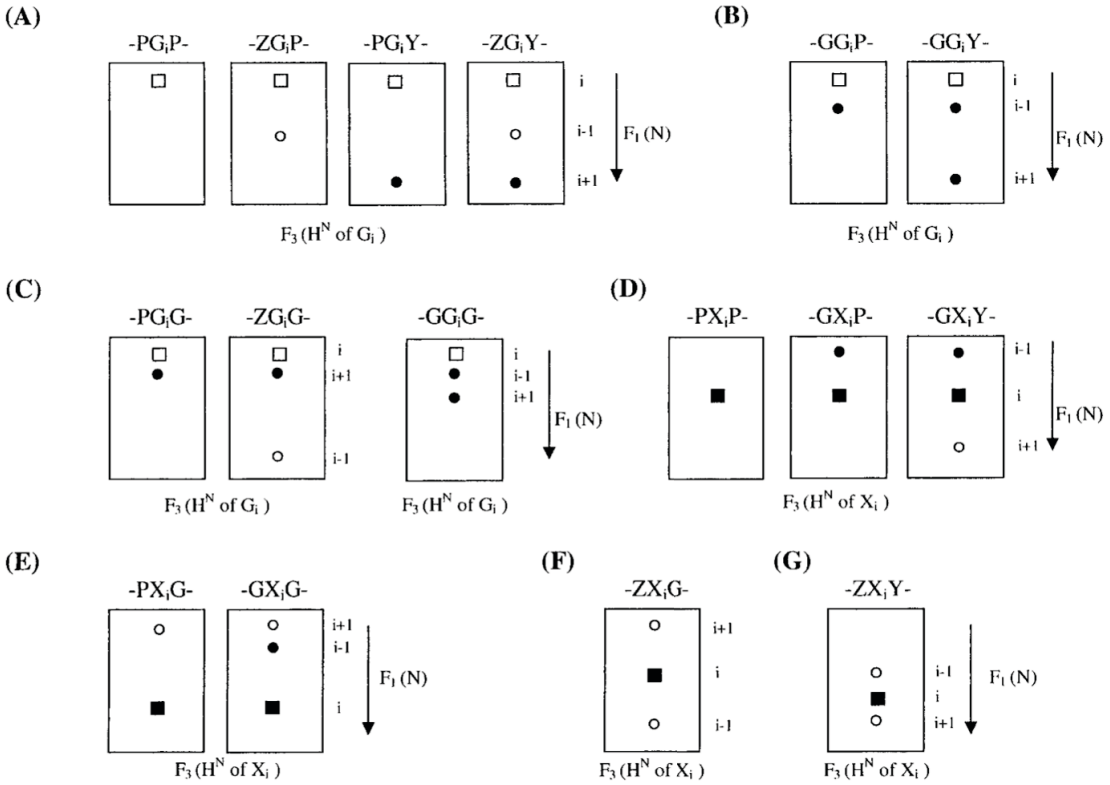

Figure 81: The expected peak patterns in the F1-F3 planes of the HNN spectrum for several triplets of residues. Image taken from Panchal et al. (2001).

KorB wild type

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH SEFMTAAQAK TTKKNTAAAA QEAAGAAQPS GLGLDSIGDL

        70         80         90        100        110        120
SSLLDAPAAS QGGSGPIELD LDLIDEDPHQ PRTADNPGFS PESIAEIGAT IKERGVKSPI

       130        140        150        160        170        180
SVRENQEQPG RYIINHGARR YRGSKWAGKK SIPAFIDNDY NEADQVIENL QRNELTPREI

       190        200        210        220        230        240
ADFIGRELAK GKKKGDIAKE IGKSPAFITQ HVTLLDLPEK IADAFNTGRV RDVTVVNELV

       250        260        270        280        290        300
TAFKKRPEEV EAWLDDDTQE ITRGTVKLLR EFLDEKGRDP NTVDAFNGQT DAERDAEAGD

       310        320        330        340        350        360
GQDGEDGDQD GKDAKEKGAK EPDPDKLKKA IVQVEHDERP ARLILNRRPP AEGYAWLKYE

       370        380
DDGQEFEANL ADVKLVALIE G
```

Number of amino acids: 381
Molecular weight: 41537.8
Theoretical pI: 5.01

Amino acid composition:

Ala (A)	42	11.0%
Arg (R)	22	5.8%
Asn (N)	14	3.7%
Asp (D)	36	9.4%
Cys (C)	0	0.0%
Gln (Q)	16	4.2%
Glu (E)	35	9.2%
Gly (G)	34	8.9%
His (H)	11	2.9%
Ile (I)	22	5.8%
Leu (L)	27	7.1%
Lys (K)	28	7.3%
Met (M)	2	0.5%
Phe (F)	10	2.6%
Pro (P)	21	5.5%
Ser (S)	19	5.0%
Thr (T)	17	4.5%
Trp (W)	3	0.8%
Tyr (Y)	5	1.3%
Val (V)	17	4.5%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 71
Total number of positively charged residues (Arg + Lys): 50

Atomic composition:

Carbon	C	1802
Hydrogen	H	2864
Nitrogen	N	530
Oxygen	O	595
Sulfur	S	2

Formula: $C_{1802}H_{2864}N_{530}O_{595}S_2$
Total number of atoms: 5793

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	23950
Abs 0.1% (=1 g/l)	0.577

Figure 82: Parameters of KorB wild type theoretically calculated from the protein sequence with ProtParam webserver.

CΔ100 KorB

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH SEFMTAAQAK TTKKNTAAAA QEAAGAAQPS GLGLDSIGDL

        70         80         90        100        110        120
SSLLDAPAAS QGGSGPIELD LDLIDEDPHQ PRTADNPGFS PESIAEIGAT IKERGVKSPI

       130        140        150        160        170        180
SVRENQEQPG RYIINHGARR YRGSKWAGKK SIPAFIDNDY NEADQVIENL QRNELTPREI

       190        200        210        220        230        240
ADFIGRELAK GKKKGDIAKE IGKSPAFITQ HVTLLDLPEK IADAFNTGRV RDVTVVNELV

       250        260        270        280
TAFKKRPEEV EAWLDDDTQE ITRGTVKLLR EFLDEKGRDP N
```

Number of amino acids: 281

Molecular weight: 30591.04

Theoretical pI: 5.72

Amino acid composition:

Ala (A)	29	10.3%
Arg (R)	17	6.0%
Asn (N)	11	3.9%
Asp (D)	21	7.5%
Cys (C)	0	0.0%
Gln (Q)	11	3.9%
Glu (E)	23	8.2%
Gly (G)	24	8.5%
His (H)	10	3.6%
Ile (I)	19	6.8%
Leu (L)	20	7.1%
Lys (K)	19	6.8%
Met (M)	2	0.7%
Phe (F)	8	2.8%
Pro (P)	16	5.7%
Ser (S)	19	6.8%
Thr (T)	15	5.3%
Trp (W)	2	0.7%
Tyr (Y)	3	1.1%
Val (V)	12	4.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 44
Total number of positively charged residues (Arg + Lys): 36

Atomic composition:

Carbon	C	1331
Hydrogen	H	2127
Nitrogen	N	395
Oxygen	O	429
Sulfur	S	2

Formula: $C_{1331}H_{2127}N_{395}O_{429}S_2$

Total number of atoms: 4284

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	15470
Abs 0.1% (=1 g/l)	0.506

Figure 83: Parameters of CΔ100 KorB theoretically calculated from the protein sequence with ProtParam webserver.

NΔ31CΔ100 KorB

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MDSIGDLSSL LDAPAASQGG SGPIELDLDL IDEDPHQPRT ADNPGFSPES IAEIGATIKE

        70         80         90        100        110        120
RGVKSPISVR ENQEQPGRYI INHGARRYRG SKWAGKKSIP AFIDNDYNEA DQVIENLQRN

       130        140        150        160        170        180
ELTPREIADF IGRELAKGKK KGDIAKEIGK SPAFITQHVT LLDLPEKIAD AFNTGRVRDV

       190        200        210        220
TVVNELVTAF KKRPEEVEAW LDDDTQEITR GTVKLLREFL DEKGRDPN
```

Number of amino acids: 228

Molecular weight: 25284.25

Theoretical pI: 4.96

Amino acid composition:

Ala (A)	18	7.9%
Arg (R)	16	7.0%
Asn (N)	10	4.4%
Asp (D)	21	9.2%
Cys (C)	0	0.0%
Gln (Q)	8	3.5%
Glu (E)	21	9.2%
Gly (G)	18	7.9%
His (H)	3	1.3%
Ile (I)	19	8.3%
Leu (L)	17	7.5%
Lys (K)	16	7.0%
Met (M)	1	0.4%
Phe (F)	7	3.1%
Pro (P)	14	6.1%
Ser (S)	12	5.3%
Thr (T)	11	4.8%
Trp (W)	2	0.9%
Tyr (Y)	3	1.3%
Val (V)	11	4.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 42
Total number of positively charged residues (Arg + Lys): 32

Atomic composition:

Carbon	C	1107
Hydrogen	H	1776
Nitrogen	N	318
Oxygen	O	357
Sulfur	S	1

Formula: $C_{1107}H_{1776}N_{318}O_{357}S_1$
Total number of atoms: 3559

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	15470
Abs 0.1% (=1 g/l)	0.612

Figure 84: Parameters of NΔ31CΔ100 KorB theoretically calculated from the protein sequence with ProtParam webserver.

NTD KorB

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH SEFLDSIGDL SSLLDAPAAS QGGSGPIELD LDLIDEDPHQ

        70         80         90        100        110        120
PRTADNPGFS PESIAEIGAT IKERGVKSPI SVRENQEQPG RYIINHGARR YRGSKWAGKK

       130
SIPAFIDNDY
```

Number of amino acids: 130

Molecular weight: 14083.44

Theoretical pI: 5.94

Amino acid composition:

Ala (A)	9	6.9%
Arg (R)	8	6.2%
Asn (N)	4	3.1%
Asp (D)	10	7.7%
Cys (C)	0	0.0%
Gln (Q)	4	3.1%
Glu (E)	8	6.2%
Gly (G)	14	10.8%
His (H)	9	6.9%
Ile (I)	11	8.5%
Leu (L)	8	6.2%
Lys (K)	5	3.8%
Met (M)	1	0.8%
Phe (F)	3	2.3%
Pro (P)	10	7.7%
Ser (S)	17	13.1%
Thr (T)	2	1.5%
Trp (W)	1	0.8%
Tyr (Y)	3	2.3%
Val (V)	3	2.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 18
Total number of positively charged residues (Arg + Lys): 13

Atomic composition:

Carbon	C	611
Hydrogen	H	948
Nitrogen	N	186
Oxygen	O	197
Sulfur	S	1

Formula: $C_{611}H_{948}N_{186}O_{197}S_1$
Total number of atoms: 1943

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	9970
Abs 0.1% (=1 g/l)	0.708

Figure 85: Parameters of NTD KorB theoretically calculated from the protein sequence with ProtParam webserver.

KorA

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MKKRLTESQF QEAIQGLEVG QQTIEIARGV LVDGKPQATF ATSLGLTRGA VSQAVHRVWA

        70         80         90        100
AFEDKNLPEG YARVTAVLPE HQAYIVRKWE ADAKKKQETK R
```

Number of amino acids: 101

Molecular weight: 11305.92

Theoretical pI: 9.57

Amino acid composition:

Ala (A)	13	12.9%
Arg (R)	7	6.9%
Asn (N)	1	1.0%
Asp (D)	3	3.0%
Cys (C)	0	0.0%
Gln (Q)	9	8.9%
Glu (E)	9	8.9%
Gly (G)	7	6.9%
His (H)	2	2.0%
Ile (I)	4	4.0%
Leu (L)	7	6.9%
Lys (K)	9	8.9%
Met (M)	1	1.0%
Phe (F)	3	3.0%
Pro (P)	3	3.0%
Ser (S)	3	3.0%
Thr (T)	7	6.9%
Trp (W)	2	2.0%
Tyr (Y)	2	2.0%
Val (V)	9	8.9%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 12
Total number of positively charged residues (Arg + Lys): 16

Atomic composition:

Carbon	C	502
Hydrogen	H	811
Nitrogen	N	147
Oxygen	O	148
Sulfur	S	1

Formula: $C_{502}H_{811}N_{147}O_{148}S_1$
Total number of atoms: 1609

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	13980
Abs 0.1% (=1 g/l)	1.237

Figure 86: Parameters of KorA theoretically calculated from the protein sequence with ProtParam webserver.

[model] residue (phi, psi)

[1] 40 SER (52.3, 88.8)
[1] 42 GLY (-166.7, -61.9)
[1] 104 ILE (-55.9, 0.8)
[1] 117 ALA (-162.3, -39.2)
[2] 54 ILE (-34.5, 147.9)
[3] 57 ASP (69.8, 89.6)
[3] 73 SER (64.9, -83.5)
[3] 74 ILE (70.6, -16.5)
[3] 119 LYS (49.1, 86.3)
[4] 70 SER (59.5, 157.2)
[4] 87 LYS (-179.9, -54.7)
[4] 101 ARG (66.9, 147.2)
[5] 54 ILE (43.4, 111.4)
[5] 112 ARG (178.4, -46.2)
[6] 73 SER (56.8, -89.0)
[6] 74 ILE (72.6, -11.7)
[7] 126 ILE (73.7, 101.1)
[8] 41 GLN (-171.8, -51.5)
[8] 70 SER (-146.7, -61.3)
[8] 101 ARG (70.3, 116.7)
[9] 63 THR (-79.3, -128.8)
[9] 87 LYS (175.5, 105.9)
[9] 101 ARG (44.9, 101.9)
[9] 121 SER (67.2, 137.9)
[10] 51 LEU (178.4, -39.2)
[10] 57 ASP (72.6, 110.8)
[10] 74 ILE (68.6, -21.7)
[10] 104 ILE (70.9, 23.3)
[11] 68 GLY (-167.2, -65.7)
[11] 73 SER (60.1, -83.5)
[11] 74 ILE (64.0, -12.0)
[11] 104 ILE (59.0, 79.4)
[12] 50 ASP (-61.5, 93.4)
[12] 57 ASP (68.9, 115.7)
[12] 99 PRO (-101.6, 109.6)
[12] 101 ARG (63.0, 97.4)
[13] 70 SER (-169.1, -55.0)
[13] 99 PRO (-103.0, 113.9)
[13] 103 ILE (77.4, 108.0)
[13] 122 ILE (172.1, 111.5)
[14] 34 LEU (65.1, 97.3)
[15] 71 PRO (-75.3, -146.0)
[15] 73 SER (60.8, -71.3)
[15] 85 GLY (136.2, 72.4)
[15] 126 ILE (69.9, 99.1)
[16] 74 ILE (74.4, -17.1)
[16] 101 ARG (64.7, -89.7)
[17] 51 LEU (-168.0, -28.2)
[17] 73 SER (62.9, -72.9)
[17] 114 SER (-85.3, -119.5)
[18] 69 PHE (67.1, 132.6)
[18] 73 SER (49.6, -95.8)
[18] 74 ILE (73.8, -17.1)
[18] 103 ILE (71.5, 109.2)
[19] 43 GLY (147.6, 78.4)
[19] 54 ILE (73.0, 119.5)
[19] 73 SER (41.8, 21.8)
[19] 100 GLY (-72.7, -111.5)
[19] 126 ILE (76.4, 143.1)
[20] 68 GLY (161.9, 62.7)

Figure 88: Outliers in the Ramachandran plot for the NTD KorB structure ensemble.

Table 33: NTD KorB ensemble outliers based on Ramachandran plot.

Residue	Occurence	Model number
34 LEU	1	14
40 SER	1	1
41 GLN	1	8
42 GLY	1	1
43 GLY	1	19
50 ASP	1	12
51 LEU	2	10, 17
54 ILE	3	2, 5, 19
57 ASP	3	3, 10, 12
63 THR	1	9
68 GLY	2	11, 20
69 PHE	1	18
70 SER	3	4, 8, 13
71 PRO	1	15
73 SER	7	3, 6, 11, 15, 17, 18, 19
74 ILE	6	3, 6, 10, 11, 16, 18
85 GLY	1	15
87 LYS	2	4, 9
99 PRO	2	12, 13
100 GLY	1	19
101 ARG	5	4, 8, 9, 12, 16
103 ILE	2	13, 18
104 ILE	3	1, 10, 11
112 ARG	2	5, 20

Residue	**Occurence**	**Model number** (cont...)
114 SER	1	17
117 ALA	1	1
119 LYS	1	3
121 SER	1	9
122 ILE	1	13
126 ILE	3	7, 15, 19

Table 33: NTD KorB ensemble outliers based on Ramachandran plot continued.

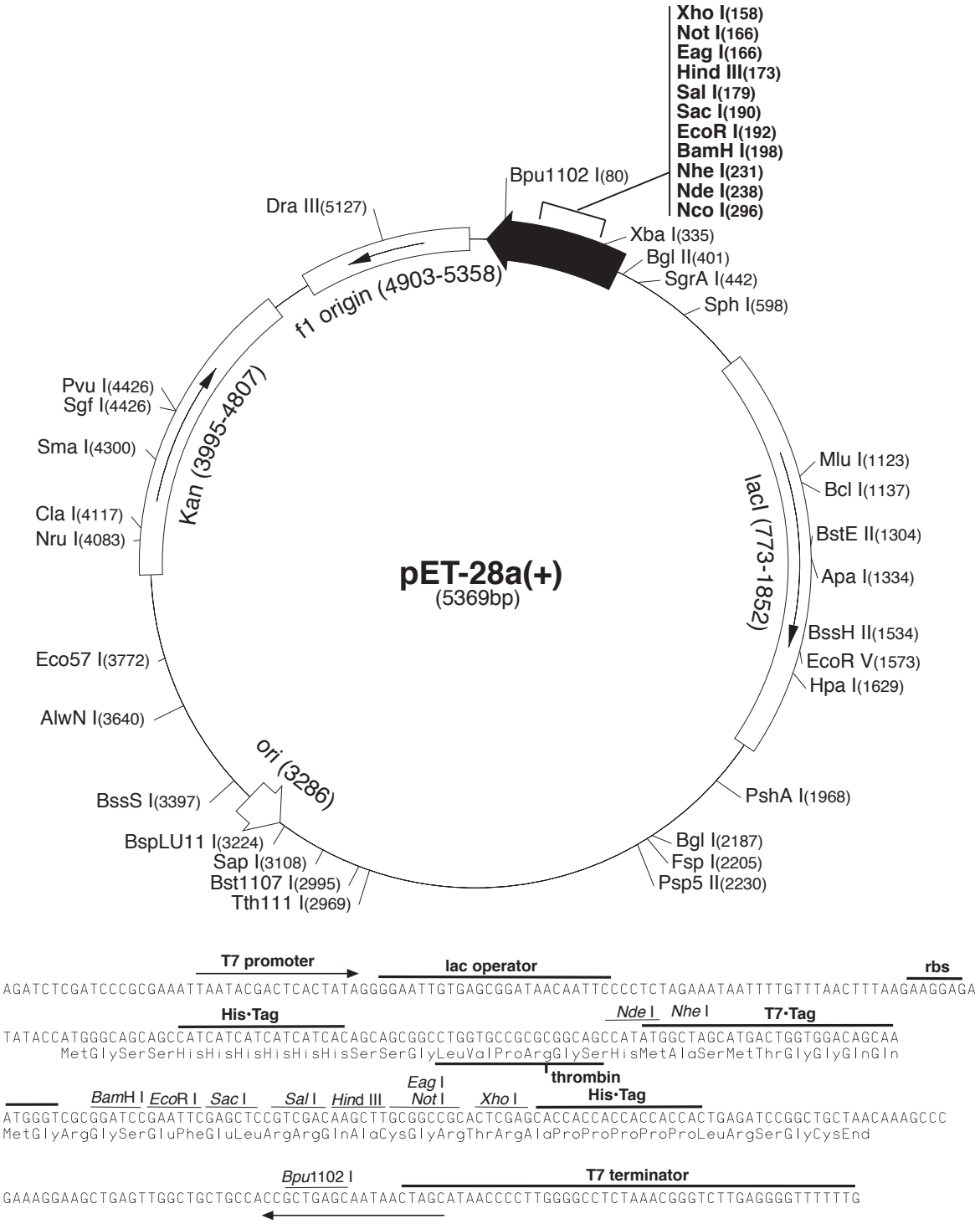

Figure 89: Schematic representation of pET28a(+) plasmid showing the positions of all restriction sites. Ori-start site of plasmid replication. Kan-Kanamycin resistance gene. lacI-Lac repressor gene.

Chain	Atom	Res	Seq		Chain	Atom	Res	Seq	Mol_ID	Distance
A	1HG	GLU	72	-	A	2HH1	ARG	109	1	Dist = 1.52
A	OD1	ASP	52	-	A	1HZ	LYS	115	5	Dist = 1.54
A	1HZ	LYS	120	-	A	O	TYR	130	1	Dist = 1.56
A	HG	SER	26	-	A	OD1	ASP	29	16	Dist = 1.56
A	HG1	THR	80	-	A	O	ARG	112	11	Dist = 1.57
A	HG	SER	73	-	A	OE1	GLU	76	16	Dist = 1.58
A	2HG1	ILE	104	-	A	H	GLY	107	18	Dist = 1.58
A	2HG1	ILE	104	-	A	H	GLY	107	4	Dist = 1.58
A	HG	SER	26	-	A	OD2	ASP	29	10	Dist = 1.58
A	3HZ	LYS	82	-	A	OD2	ASP	127	4	Dist = 1.59
A	OE2	GLU	94	-	A	2HH1	ARG	110	2	Dist = 1.59
A	O	ILE	77	-	A	HG1	THR	80	9	Dist = 1.60
A	OE1	GLU	56	-	A	3HZ	LYS	120	1	Dist = 1.60
A	OD2	ASP	50	-	A	3HZ	LYS	82	10	Dist = 1.60

Figure 90: Close contacts in the NTD KorB ensemble.

Deviation	Residue Name	Chain ID	Sequence Number	Model	AT1	-	AT2	-	AT3	Bond Angle	Dictionary Value
-4.3	GLU	A	72	1	CB	-	CA	-	C	105.8	110.1
4.3	ALA	A	75	1	N	-	CA	-	C	115.5	111.2
-4.9	LEU	A	24	4	N	-	CA	-	C	106.3	111.2
-4.5	SER	A	121	4	N	-	CA	-	C	106.7	111.2
-4.7	ILE	A	103	5	N	-	CA	-	C	106.5	111.2
-4.9	LYS	A	120	6	N	-	CA	-	C	106.3	111.2
-5.0	GLU	A	94	11	N	-	CA	-	C	106.2	111.2
4.5	SER	A	31	16	N	-	CA	-	C	115.7	111.2

Figure 91: Covalent contacts in the NTD KorB ensemble.

Amino Acid	Three letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

JCSG+ tube number	Salt	Buffer	pH	Precipitant
33	–	0.1 M Na/K phosphate	6.2	25 % (v/v) 1,2 propanediol; 10 % (v/v) Glycerol
34	–	0.1 M Bicine	9	10 % (w/v) PEG 20000; 2 % (v/v) 1,4-Dioxane
43	0.2 M Lithium sulfate	0.1 M Tris	8.5	40 % (v/v) PEG 400
44	–	0.1 M Tris	8	40 % (v/v) MPD
84	3 M Sodium chloride	0.1 M Bis-Tris	5.5	–
88	0.2 M Calcium chloride	0.1 M Bis-Tris	5.5	45 % (v/v) MPD
89	0.2 M Ammonium acetate	0.1 M Bis-Tris	5.5	45 % (v/v) MPD
96	0.2 M Ammonium acetate	0.1 M HEPES	7.5	45 % (v/v) MPD

Table 34: Crystallisation conditions of NTD KorB.

Abbreviations

aa	Amino acid
ASU	Asymmetric unit
AUC	Analytical ultracentrifugation
AUDANA	Automated database-assisted NOE assignment
AWSEM	Associative memory water mediated structure and energy model
BE	Binding event
BSA	Bovine serum albumin
CBP	Centromere binding protein
CCR	Central control region
CD	Circular dichroism
CPL	Circularly polarised light
CTD	C-terminal domain
CV	Column volume
DBD	DNA-binding domain
DNA	Deoxyribonucleic acid
dsFL	double stranded full-length
dsHL	double stranded half-length
EDTA	Ethylene diamine tetra-acetic acid
EMSA	Electrophoretic mobility shift assay
ESI-MS	Electrospray ionisation mass spectroscopy
FA	Fluorescence anisotropy
FP	Fluorescence polarisation
FRET	Förster resonance energy transfer
HMG	High mobility group domain
HTH	Helix-turn-helix
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
INT	Intensity

IPTG	Isopropyl β-D-1-thiogalactopyranoside
JCSG	Joint center for structural genomics
LB	Lysogeny broth
MCS	Multiple cloning site
MD	Molecular dynamics
mdeg	millidegree
MS	Mass spectrometry
MST	Microscale thermophoresis
MWCO	Molecular weight cut-off
ncIDP-assign	Neighbour-corrected IDP chemical shift assignment
NHS	N-hydroxysuccinimide esters
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser effect spectroscopy
NTD	N-terminal domain
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB ID	Protein data bank identifier
PMSF	Phenyl methane sulfonyl fluoride
PROMALS3D	Profile multiple alignment with local structure
ProSA	Protein structure analysis
PSVS	Protein structure validation suite
rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
SH3	Src homology 3
TFE	Trifluoroethanol
Tris	Tris-(hydroxymethyl)-aminomethane
WT	Wild type
XRC	X-ray crystallography