^{19}F NMR Studies on DnaB helicase

by

Kekini Vahini Kuppan

A thesis submitted for the degree of
Masters of Philosophy

Research School of Chemistry
Australian National University

December 2013

Australian
National
University

Declaration

The work described herein is the author's own work, unless otherwise stated, and was carried out within the Research School of Chemistry, Australian National University, from June 2011 – Nov 2013. None of the material has been submitted in support of an application for any other degree.

Kekini Vahini Kuppan

26th May 2014

To my family

ACKNOWLEDGEMENTS

I sincerely thank my supervisor Prof Thomas Huber for his academic inspiration, brilliant supervision and constant encouragement throughout my MPhil program. His guidance has helped me through my research and for writing this thesis.

I want to thank Prof Gottfried Otting for his guidance, help and teaching NMR. I would also like to thank Prof Nick Dixon and his group at The University of Wollongong for insights into DNA replication system and DnaB helicase.

I would like to extend my gratitude to the Research School of Chemistry, the Australian National University for giving me an opportunity to work in the School and providing the ANU Research Scholarship and the Tuition Fee waiver.

I offer my many thanks to Dr Kiyoshi Ozawa and Dr Karin Loscha for teaching me everything that I know in cell-free protein synthesis and protein labelling. I want to thank all the members from Otting and Huber group with whom I have made friends for life, for their support and help in research and life.

I am most grateful to my parents, and my brother, who listens to my research even though he does not understand anything. A special thanks goes to my mom for always being there for me, guiding, helping and sharing my experiences in Australia even though she lives outside Australia. I could not have finished my program without her.

LIST OF PUBLICATION

Journal articles:

Ozawa K, Loscha KV, **Kuppan KV**, Loh CT, Dixon NE, Otting G. High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. Biochemical and Biophysical Research Communications. **2012** Feb 24;418(4):652-6.

Contribution: I expressed and purified the tfmF labelled DnaB by cell-free protein synthesis and optimised the expression conditions.

Conference presentations:

Kuppan KV, Ozawa K and Huber T. In-vitro incorporation of unnatural amino acid in DnaB helicase for structural study using ^{19}F-NMR. **2013** Feb 38th Lorne Conference on Protein Structure and function.

ABBREVIATIONS

A_{280}	Absorbance at 280 nm
aaRS	aminoacylated tRNA synthetase
ADP	adenosine diphosphate
ATP	adenosine triphosphate
Gst	*Geobacillus stearothermophilus*
G.stearothermophilus	*Geobacillus stearothermophilus*
B. subtilis	*Bacillus subtilis*
DnaGC	truncated C-terminus of *Gst* DnaG primase (residues 449–581)
DTT	dithiothreitol
dNTP	deoxyribonucleoside triphosphate
E. coli	*Escherichia coli*
G. kaustophilus	*Geobacillus kaustophilus*
EDTA	ethylenediamine-N,N,N',N'-tetraacetate
FPLC	fast protein liquid chromatography
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
H. pylori	*Helicobacter pylori*
IPTG	isopropyl-β,D-thiogalactoside
kDa	Kilo Dalton
LB	Luria-Bertani (medium)
NMR	nuclear magnetic resonance
*p*CNF-RS	*para*-cyanophenylalanine tRNA synthetase
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
RBS	ribosome binding site
S. aureues	*Staphyloccocus aureues*
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis

ssDNA	single-stranded DNA
TEMED	N,N,N',N'-tetramethylethylenediamine
tfmF	Trifluoromethylphenylalanine
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet

ABSTRACT

NMR (nuclear magnetic resonance) spectroscopy is an important technique used for structural characterisation of proteins under near-physiological condition. It is used for determining the structure of protein at atomic resolution and to study protein-protein interactions. However, its major limitation is the size of protein. The limitation can be overcome by solution state ^{19}F NMR, a promising technique which is used for studying the structural dynamics of large proteins by selective fluorine labelling.

Cellular processes such as bacterial DNA replication is carried out by large protein complexes such as DnaB helicase, DnaG primase and DNA polymerase enzyme. It is necessary to study these proteins under physiological conditions to gain insights into the process. DnaB is primary DNA helicase in the replisome and its main function is unwinding the duplex DNA. It interacts with various other proteins and performs different functions during the process. Its structural characterisation by electron microscopy (EM) has established that it adopts two different rotational symmetry states (C3 and C6) and the conformational interchange occurs in the N-terminal domain only. The symmetrical states were determined under non-physiological conditions. The factors triggering the conformational change were determined under physiological conditions however; the functional significance of each conformation adopted under these conditions is elusive. To gain insights into structure, dynamics and interactions of DnaB under physiological conditions, I have studied the DnaB helicase and partner proteins by ^{19}F NMR spectroscopy.

In this thesis, I report my studies that employed fluorine one-dimensional NMR spectroscopy to study conformational changes of hexameric DnaB helicase with mass of 315 kDa in solution under physiological conditions. Trifluoromethylphenylalanine (tfmF) was the chosen fluorine label, incorporated into DnaB site-specifically. DnaB helicase and their partner proteins from *Escherichia coli* (*E. coli*) and *Geoacillus stearothermophilus (Gst)* were studied to observe the significant features in both systems.

TfmF labelled *E. coli* DnaB was expressed by cell-free protein synthesis and examination at different pH established that its N-terminus is flexible in solution under near

physiological conditions rather than adopting rigid conformation as reported by EM studies. Furthermore, the study revealed that its interaction with helicase loader DnaC induced more flexibility into N-terminus of *E. coli* DnaB. The complex of *E. coli* DnaB and DnaC is a 480 kDa protein and the presented fluorine NMR data are first attempts to study the conformational changes in such large protein systems.

Gst DnaB helicase was studied with its primase, helicase loader and Mg^{2+}. TfmF was incorporated into DnaB by *in vivo* method. The fluorine data showed that N-terminus of *Gst* DnaB is flexible similar to *E. coli* DnaB. However, its interaction with primase induces rigidity to the N-terminus and arranges it to form C3 symmetry, which is in concurrence with previous work. Moreover, the NMR and gel filtration data showed that magnesium ion rendered the integrity of hexamers by forming unstable monomers. We report here the initial studies on *Gst* helicase loader, DnaI. The fluorine NMR and gel filtration data suggested that it interacts with DnaB monomer instead of hexamers.

The presented data displays ^{19}F NMR as a useful tool in determining the structural dynamics of large protein systems in solutions and its data can supplement previous structural information.

Table of Contents

CHAPTER 1

Introduction

Medical science is faced with challenge of discovering novel antibiotics due to rapid increase in antibiotic-resistant pathogens. Recent developments in antibiotics have been the chemical modification of existing antibiotics which targeted the same mechanism as its parent compounds and are not effective against antibiotic-resistant bacteria such as methicillin-resistant strains of Gram positive *Staphylococcus aureues* (MRSA), multidrug resistant strains of gram negative *Acinetobacter* and *Pseudomonas spp.*, which pose a great treat[1]. Therefore, the novel antibiotic to be developed must target key mechanisms required for survival of organisms, must be conserved across pathogenic prokaryotes and distinct from processes in humans.

DNA replication, repair and recombination processes are vital in achieving the accurate transmission of genetic material from generation to generation, and thus are essential for survival of organisms. These processes are carried out by nucleoprotein complexes involving 30 or so different replication proteins[2]. These processes operate with high fidelity mediated by well-regulated and coordinated protein-protein and protein-nucleotide interactions[2-6], which have transient and stable protein-protein interactions among them[7,8]. It has been previously reviewed that these proteins involved are underexploited as drug targets[1].

DNA replication is carried out by macromolecular complexes called primosomes and replisomes, which are nucleoprotein complexes classified as motor proteins due to their ability to couple movement with work and act as macromolecular machinery [2]. There is need for deep understanding of their structures, functions and interactions as complexes. Unfortunately, intact replisome and primosome cannot be isolated from cells due to their labile nature, but functional complexes can be reassembled *in vitro*. *Escherichia coli* replication proteins were chosen for *in vitro* studies as they can be isolated in large quantities and reconstituted faithfully as complexes[5]. Other replication mechanisms that have been studied are T4 and T7 bacteriophages, in which host enzymes have been used

for DNA replication. In the past decade, extensive work has been done in understanding DNA replication mechanism and in establishing the protein interaction network in the dynamic macromolecular assembly [3-6]. Combining the knowledge and tools developed over the years may be utilized for drug discovery and in understanding other multiprotein complexes.

1.1. Bacterial DNA Replication

DNA exists in its stable form as duplex DNA and is intertwined in antiparallel conformation. The mode of DNA replication is semiconservative; this means that a new duplex DNA has one parent and one daughter strand. During the replication, the replisome, moves only in one direction from 5' to 3' along the single stranded DNA. The DNA strand continuously synthesized from 5' to 3' is called leading strand and the strand which is synthesized in discontinuous manner from 3' to 5' is called the lagging strand. *E. coli* has a circular chromosome and two replisomes are recruited to replicate circular DNA while moving in opposite direction. Each replisome recruited synthesises both leading and lagging strand of DNA [2,3]. There are three phases in DNA replication: initiation, elongation and termination.

In the initiation phase the specific sites on the chromosome triggers the initiation of the DNA replication coordinated with the cell cycle. The initiation phase involves binding of DnaA, an initiator protein to bind to chromosome at specific sites at origin of replication (in *E. coli* called *oriC*). Multiple DnaA molecules bind to sequence specific sites called DnaA box sequence; in *E. coli* there are five DnaA boxes. The *oriC* also has AT rich regions which aids in destabilizing the duplex DNA at the *oriC* when the complex of DnaA, histone-like proteins and integration host factor stretches the DNA. This allows the loading of DnaB helicase by the helicase loader [5].

DnaB helicase unwinds the duplex DNA for replication. It is a homohexameric protein and is loaded onto DNA bound by DnaA with the aid of the helicase loader in complex with ATP. The complex of DnaA, DnaB helicase and helicase loader is called prepriming complex. The unwound helix is prevented from reannealing by single stranded binding protein (SSB)[3].

This is followed by recruitment of DnaG primase, which catalyses the formation of short RNA primers complementary to the DNA strand being replicated [2,3]. The recruitment is observed in prokaryotes and not in bacteriophages because they have a single protein which performs both unwinding of DNA and synthesis of primers.

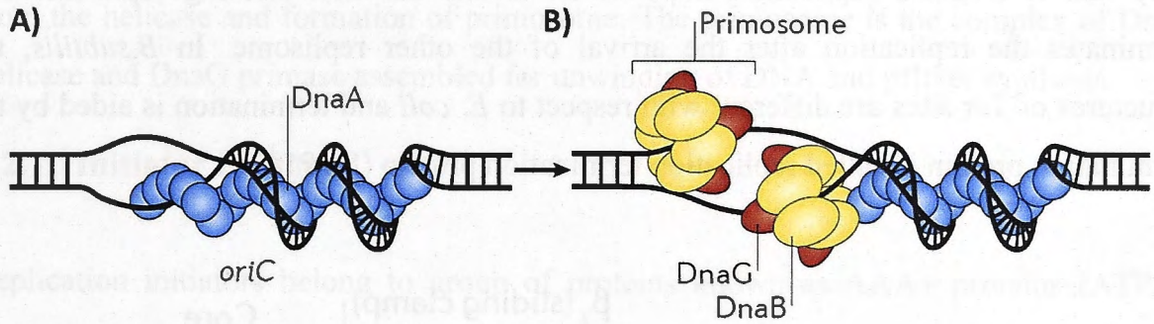

***Figure 1.1**: A) The stretching of duplex DNA by initiator protein DnaA which leads to formation of bubble. B) It shows the formation of the primosome (DnaB/DnaG) in the bubble at the replication fork. (The picture is reproduced from Robinson and coworkers[3]).*

The elongation phase occurs with multiprotein assembly called replisome, which carries out the gene duplication. The main component of the bacterial replisome is DNA polymerase III, a holoenzyme with multiple subunits having three major functional fragments: core polymerase for synthesising new strand and proof reading, the sliding clamps for continuous movement with synthesis and clamp loader complex for loading and removing polymerase from DNA[2,3]. The enzyme is formed by the association of ten different subunits to carry out the process with high fidelity.

The synthesis of the leading strand is straightforward, with unidirectional movement of DNA polymerase adding new nucleotide to the 3' end of the RNA primer synthesised by primase. The lagging strand has a different mode of replication, which involves synthesis of multiple primers by DnaG primase and primers are extended by DNA polymerase as

short fragments of 1000-2000 nucleotides called Okazaki fragments [4]. The gaps between the fragments are then filled by DNA polymerase I and ligated by DNA ligase.

The replication is terminated approximately opposite to the *oriC* site on the circular chromosome. It is accomplished by the strong binding of Tus protein in the DNA polymerase to 23 bp *Ter* sites [13]. In *E. coli*, there are ten Ter sites orientated in such a way that it forms a replication trap, which arrests the first arriving replisome and terminates the replication after the arrival of the other replisome. In *B.subtilis*, the structures of Ter sites are different with respect to *E. coli* and termination is aided by the termination protein is called replication termination protein (RTP) [14].

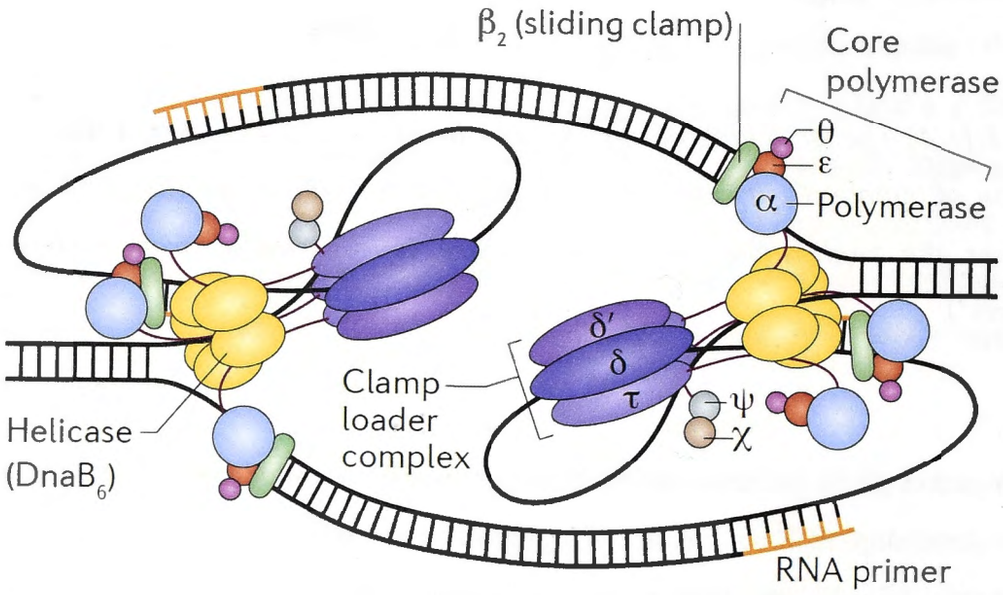

Figure 1.2: *The elongation phase of DNA replication depicting the interactions between various proteins in the replication fork. The picture is the reproduced from Robinson et al., 2013*

In DNA replication, the most critical stage is the initiation stage [5]. The nucleoprotein complexes involved are prepriming complex and primosome. Gaining in-depth knowledge of the replication proteins involved in the initiation stage will lay the foundation for building the protein interaction network. Compare and contrast of the structure and function of homologous proteins from different species will demonstrate diversity of the mechanism.

1.2. Proteins involved in Initiation stage

Prepriming complex is the first nucleoprotein complex formed during DNA replication[2]. It is the complex of DnaA, DnaB helicase and helicase loader. The complex forms bubble in the duplex DNA resulting in single stranded DNA (ssDNA) and the DnaB helicase is loaded onto the ssDNA by the loader. This is followed by dissociation of helicase loader from the helicase and formation of primosome. The primosome is the complex of DnaB helicase and DnaG primase assembled for unwinding of DNA and primer synthesis.

1.2.1. Initiator Protein

Replication initiators belong to group of proteins known as AAA+ proteins (ATPase Associated with various cellular Activities) [7]. The protein uses energy derived from ATP hydrolysis for DNA binding, melting of the duplex DNA, its oligomerisation, and for helicase loading activities[7,5].

DnaA is a prokaryotic initiator protein in *E. coli* and other Gram negative bacterial species such as *Bacillus subtilis,* and *Pseudomonas putida.* There are other important proteins that interact with DnaA for initiation such as DnaN, SeqA and DiaA. *B. subtilis* homolog of DnaN also plays an important role in initiation of replication [7,8,9].There are subtle differences in the DnaA interacting factors between *E. coli and B. subtilis* [9]. *B.subtilis* lacks homolog for SeqA, *hda*, however, it has Soj that controls the initiation of the replication by directly interacting with DnaA and availability of DnaA and access to oriC are controlled differently in *B. subtilis*[9].

DnaA has four functional domains important for replication initiation. The main function of Domain I is interacting with neighbouring DnaA and DnaB-DnaC complex for loading the helicase onto DNA. Domain III is involved binding and hydrolysis of ATP as it contains the distinctive AAA+ motif and also arginine finger motif required for initiation activity. These two domains are connected by domain II of 5kDa. Domain IV is involved in recognizing the DNA box motifs of *oriC* [7,8] . In *E. coli*, DiaA interacts with domain I and II of N-terminal of DnaA to stimulate timely initiation of replication. DiaA was studied as a stable homodimer and shown to initiate replication of minichromosome in in

vitro system when DnaA levels are limited. DNA replication commences only with the binding of DnaA and stretching of the duplex DNA[10].

There are five DNA box motif at *oriC* identified by DnaA. The three motifs R1, R2, and R4 bind to DnaA with high affinities; the other motifs bind to DnaA-ATP complex. It has been shown that *E. coli* DnaA forms stable extended right-handed helical filaments on DNA that aids in melting the duplex and loading the DnaB helicase with help of its loading partner (Fig 1.1). The crystal structure of *Aquifex aeolicus* DnaA has shown that domain III and IV are adequate for forming the right handed filament [11]. This step initiates the replication process.

1.2.2. Helicase Loading Partner

Helicase loaders generally form tight complex with helicases and load them onto duplex DNA. They comprise of two domains, one is the N-terminal domain that binds to the helicase and another domain is the C-terminal domain, which has the AAA+ motif for ATP binding and hydrolysis [1,15].

A functionally well-studied helicase loader is *E. coli* DnaC. Only partial structural information of DnaC is available, which is the crystal structure of C-terminal domain of *Aquifex aeolicus* DnaC. It showed that DnaC shares structural similarity with DnaA by retaining the Box VII and arginine fingers [6,16]. It is also been reported that DnaC complexes with DnaB in the presence of ATP and dissociates after ATP has been hydrolysed to ATP. Hence DnaC is a ATP/ADP switch [17,18]. Six monomers of DnaC interact with DnaB by binding to six of its monomers forming a dodecamer complex [15]. Another EM study indicated that DnaB hexamer binds to three *E. coli* DnaC monomers rather than six molecules [20].

E. coli DnaB-DnaC has been recently studied by electron microscopy (EM) and showed that there is gap running along one side of the complex, forming a righted handed helical structure to load DnaB by ring opening mechanism [16].

In Gram negative bacteria the helicase loader is encoded as DnaC and in Gram positive bacteria, the helicase loader encoded is called DnaI. The homology between DnaC and

DnaI is limited to AAA+ motifs and BoxVII in the C-terminus [7,15]. There is a novel Zn^{2+} binding fold in DnaI, which is not present in *E. coli* DnaC, but both the DnaC/I loaders interact with the C-terminal domain helicase through the N-terminus [21].

In *B.subtilis,* a two-step loading mechanism is proposed for helicase loading onto duplex DNA. DnaI and DnaB are the helicase loaders (where as in *E. coli* DnaB is the helicase) that coordinate with the helicase to be loaded by ring formation mechanism. DnaI and DnaB helicase loader delivers the helicase by assembling the hexamer from monomers around the ssDNA because preformed helicase rings are inactive in the presence of its loaders [22]. In *G.kaustophilus*, another well-studied Gram positive bacterium, the DnaI is proposed to interact with helicase and loaded onto ssDNA by ring opening mechanism similar to *E. coli* system[16, 23].

The structural and biochemical studies suggested that mechanism of helicase delivery onto ssDNA is different among the organisms. Further studies are required to understand precise mechanism of helicase delivery, as the above stated mechanism is speculated based on the protein-protein interactions [14-18].

The biochemical studies showed that excess of DnaC reversibly suppresses the helicase activity by closing the central channel of the complex. The helicase activity is restored by the release of DnaC from DnaB-DnaC complex when in contact with DnaG primase by opening the central channel [3, 24].

1.2.3. Primase

Primase is a DNA dependent RNA polymerase that synthesises the primers. DNA polymerase III cannot initiate the DNA synthesis, because it can only add nucleotide to the 3' end of the strand. Primase initiates the primer synthesis from the start codon ATG by recognizing specific sequence [25]. Primase is called DnaG across all bacteria. DnaG alone has low catalytic activity and is inactive on SSB protein coated ssDNA. It has been shown that the complex of DnaG-DnaB helicase enhances the activity of DnaG by many folds [26]. In complex with DnaB, DnaG recognises sequence 5'-CAG-3' to catalyse the primer synthesis [27].

Structural studies have shown that DnaG exists as monomer in solution [28]. DnaG compromises three domains; the N-terminal domain has a novel zinc binding domain, central domain is core subunit involved in catalysing the primer synthesis and third domain is the C-terminal domain [29].

The catalytic activity is metal dependent process and requires Mg^{2+} to catalyse the RNA synthesis. The main function of C-terminal domain is to interact and complex with DnaB helicase and aids in recruiting primase onto the replication fork. This has been shown by mutational studies involving removal of last eight amino acids in C-terminal resulted in hindering the interaction between DnaG and DnaB [30].

In *E. coli*, the DnaB helicase and DnaG primase has transient interaction because of which it was a challenge to study the interactions. However, in *Geobacillus stearothermophilus,* helicase and primase formed stable complex and was studied the key interacting residues and structural arrangement of DnaB in complex with primase by X-ray crystallography [31].

1.2.4. DNA replicative helicases

DNA helicases are ubiquitous group of enzymes involved in unwinding of the duplex DNA during DNA replication and repair [19]. There are several bacterial DNA replicative helicases such as RuvB, RepA, PcrA and DnaB responsible for binding and unwinding of DNA. The helicases are classified into six different superfamilies (SF) based on their structure and biochemical properties [32,33]. Many replicative helicases exists as monomer, dimers and hexamers. Among all the helicases in the bacterial cells, DnaB helicase is the principle replicative helicases.

Duplex DNA is unwound by breaking the hydrogen bonds between the nucleotides specific polarity of Duplex DNA. The 5' lagging strand of DNA passes through the center of the helicase ring while the 3' leading strand is excluded from the hexamer ring resulting in two single strands of opposite polarity accessible for replication[34]. The DNA binding stimulates ATP hydrolysis and energises the movement of replication fork from 5'to 3' until the replication fork reaches the termination site [35].

The main function of hexameric helicase is unwinding of duplex, however it is also involved in other significant biological activities such as ATP binding and its hydrolysis as energy sourse, binds to both single and/or double stranded DNA [36], and interacts with DnaC helicase loader for delivering the DnaB onto the single stranded DNA[37], recruiting DnaG primase onto replication fork and increasing its primase activity [35], SSB protein (single stranded DNA binding protein) [38], and τ subunit of DNA polymerase III during replication[39]. The various functions of helicase are represented in Figure 1.3.

DnaB helicase in *E. coli* and DnaC in *B. subtilis* forms a ring like structure by oligomerising six monomers into a hexamer. *E. coli* DnaB helicase is one of the well-characterised proteins in the hexameric helicase protein family. The hexameric *E. coli* DnaB helicase belongs to superfamily 4 (SF4) consisting six RecA like folds in the core of the helicase responsible for nucleoside binding [32-36].

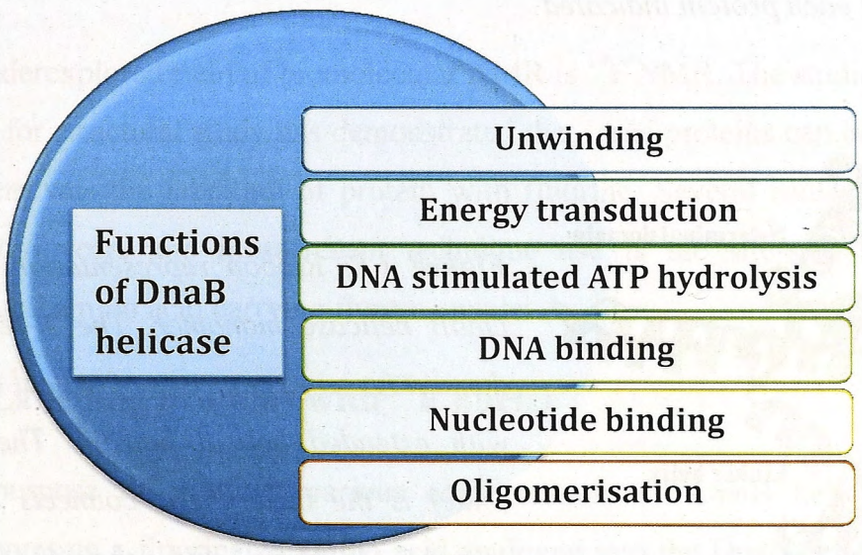

Figure 1.3: *A pictorial representation of various functions of DnaB helicase displaying the versatility of the protein*

The proteolytic assays have shown that each subunit can be broken down to two domains; one is smaller domain N-terminal domain and a larger domain C-terminal domain. N-terminal domain plays a vital role in recruiting and interacting with DnaG primase during primer synthesis and defines the direction of movement [30]. C-terminal domain has the RecA like fold containing nucleoside triphosphate binding pockets with

high affinity towards ATP, five motifs H1, H1a, H2, H3 and H4 for helicase activity and binds to ssDNA. The C-terminal domain also interacts with DnaC [32-36, 40, 41]. These two domains are connected by a flexible helical linker [40].

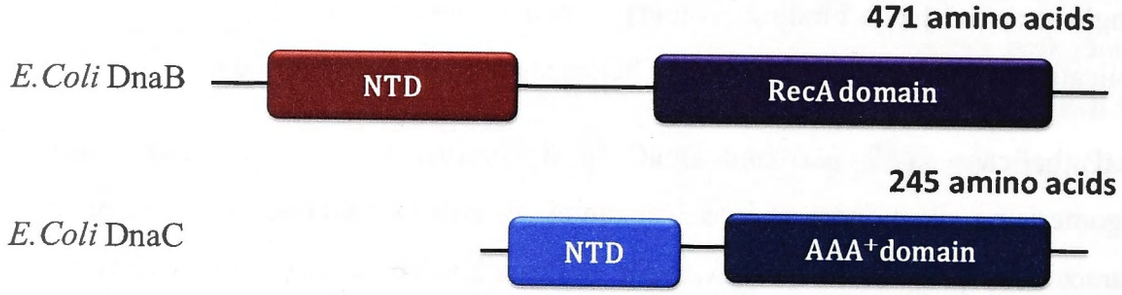

***Figure 1.4**: Pictorial representation displaying the primary structure of E. coli DnaB and DnaC with length of each protein indicated*

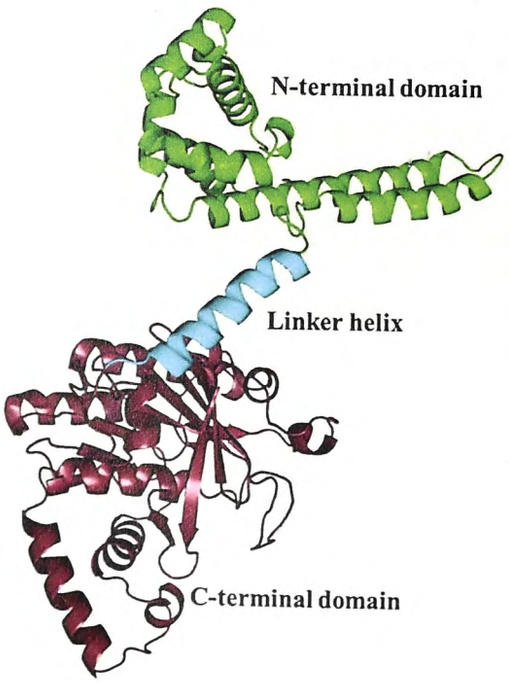

***Figure 1.5**: Ribbon representation of the DnaB helicase monomer. The N-terminal domain (green) comprises of helical bundle with extended helical hairpin. The helix (blue) is the linker helix connects the N-terminal domain and the C-terminal domain. C-terminal domain (raspberry) has a RecA like fold and is comprised of both helical bundle and beta sheets.*

The structural arrangement of the subunits has always been under questioning. So far, there is no atomic resolution structure of the hexameric form of *E. coli* DnaB. Recently,

structure of the full-length *Gst* DnaB helicase of was solved by X-ray crystallography giving insight into the structural arrangement and symmetry of the molecule [31]. EM study of *E. coli* DnaB showed that it exists in two different symmetrical forms [42,43].

So far the structural information of DnaB helicase in solution determined under physiological conditions gave the insights on factors inducing conformational change while the EM studies displayed various conformations under non-physioslogical conditions. One way of obtaining structural information of protein under physiological conditions is studying the protein by NMR spectroscopy. It is a promising field to investigate the symmetry of the helicase under different conditions and to complement the information of EM studies. However, conventional biomolecular NMR usually carried out by isotopes such as ^{15}N, ^{13}C and ^{1}H is not convenient because of the large size of the helicase resulting in non-analysable spectrum with too many resonances crowding the spectrum.

An underexplored field of biomolecular NMR is ^{19}F NMR. The studies using ^{19}F NMR as a tool for structural study has demonstrated that large proteins can be studied. The major problem was the labelling of protein with fluorine. Several routes have been explored over the years and most recent technique use is the site-specific incorporation of unnatural amino acid carrying fluorine nuclei.

1.3. Labelling proteins with ^{19}F nuclei

Our purpose of studying various conformations of DnaB helicase is achieved by incorporating a fluorinated amino acid analogue into the DnaB helicase at specific sites. The non-natural analogue of an amino acid is referred to as unnatural amino acid (UAA).

1.3.1. Background on the various labelling methods

For structural, functional and binding characterisation of a protein, mutations can be introduced in the gene to replace a amino acid of functional importance with other amino acids of different nature. Subsequent loss of functions or change in structure of the protein may than be attributed to the respective residues. Although, these mutagenesis studies are well-established and are used in research for long time providing substantial

information on the function and residues involved in interaction with other protein, it often does not provide information on subtle changes in amino acid side chain due to acidity or hydrogen bonding changes and its effect on the structure of the protein[40].

To assess more subtle changes, it is necessary to modify the side chains of certain amino acids or replace the amino acid with close analogues to observe the resultant changes. These modifications are done either by tagging or labelling a protein. Over the years, number of methods have been developed for labelling and tagging the protein some of the example are semi-synthetic incorporation, chemical modifications, labelling the protein with tags, and incorporating UAA [45,46].

A semi-synthetic method is one of the primary approaches used in labelling a protein. This involves synthesising a part of protein as a peptide carrying unnatural amino acid and fusing it biosynthetically to the remaining protein to generate an active protein for study. It is used for site-specific incorporation of unnatural amino acids into proteins. The drawback is that it is impractical to produce large peptides or proteins in quantities required for NMR quantities, and if done so, protein products tend to easily precipitate and contain high ammounts of byproducts [45,46].

The second method is global replacement of particular amino acid with its analogue. Generally, the protein expression system used is an auxotroph, which is unable to produce the particular amino acid, to be replaced by its analogue. Alternatively, the protein can be expressed under auxotrophic conditions, wherein 19 amino acids are provided in medium along with the unnatural amino acid as supplements. Glyphosate is added in the medium in case of replacing aromatic amino acids with UAA because it inhibits the biosynthesis of aromatic amino acids. An achievement of greater than 95% global replacement was reported by these studies [47,48].

The next system is the addition of desired label to the side chain of an amino acid. The amino acids used for tagging are cysteines (Cys), lysine (Lys), and occassionally threonines (Thr) especially for membrane proteins and soluble large proteins. Cysteines are commonly used for tagging because of its highly reactive thiol side chain that can selectively react with label reagents to form a disulphide bridge. In most cases, single free cysteine is not readily available for reaction, so many proteins have to be

substantially engineered for tagging via cystein mutagenesis. The limitations of the system are the requirement of large quantities of reaction quencher and tags of approximately 100-200 folds excess, mutations may render protein inactive or insolube, especially if the cysteines are required at the active site or for folding. In case of fluorine tags, they are less soluble in aqueous solution and requires co-solvent such as acetonitrile, not a protein friendly solvent to solubilise the tags [45,46]. Adriaensens and coworkers demonstrated that fluorine tagging resulted in disruption of native disuphide bonds, aggregation and precipitation of egg-white lyzosyme [49]. This problem can be bypassed by incorporating unnatural amino acid site specifically.

Conventional biomolecular NMR deciphers conformational change or binding of the protein with atomic resolution. As the size of the protein is a challenge in case of large systems, the spectra can be simplified by specific isotope labelling of the protein. In synthetic and semisynthetic methods yields are low and cannot be applied for large systems. Use of cellular machinery to incorporate UAA is easy to upscale. Recently, incorporation of single UAA site-specifically into protein using *E. coli* expression system has been established [50] and even incorporation of single UAA at multiple sites and multiple UAA at various sites into single protein are new system being developed[51,52].

1.3.2. Principle of incorporation of unnatural amino acids

In all organism, the proteins are biosynthesised from 20 canonical amino acids encoded by 61 degenerate codons. For site-specific incoporation using bacteria as expression system, it is necessary to have a new codon should be introduced for UAA incorporation and identified by expression system as a codon. For this purpose, one of the three nonsense or stop codons (TAG-amber, TAA–ochre, and TGA-opal) can be used, which is recognised by orthogonal tRNA/aminoacyl tRNA synthetase and in response the ribosomes incorporate UAA. The choice of the nonsense is dependent on the expression system because , if the nonsense codon selected is the most common stop codon for host genes, the translation will not be terminated and host proteins will contain the UAA or vice versa, UAA is not incorporated into protein because the protein is terminated at nonsense codon[50].

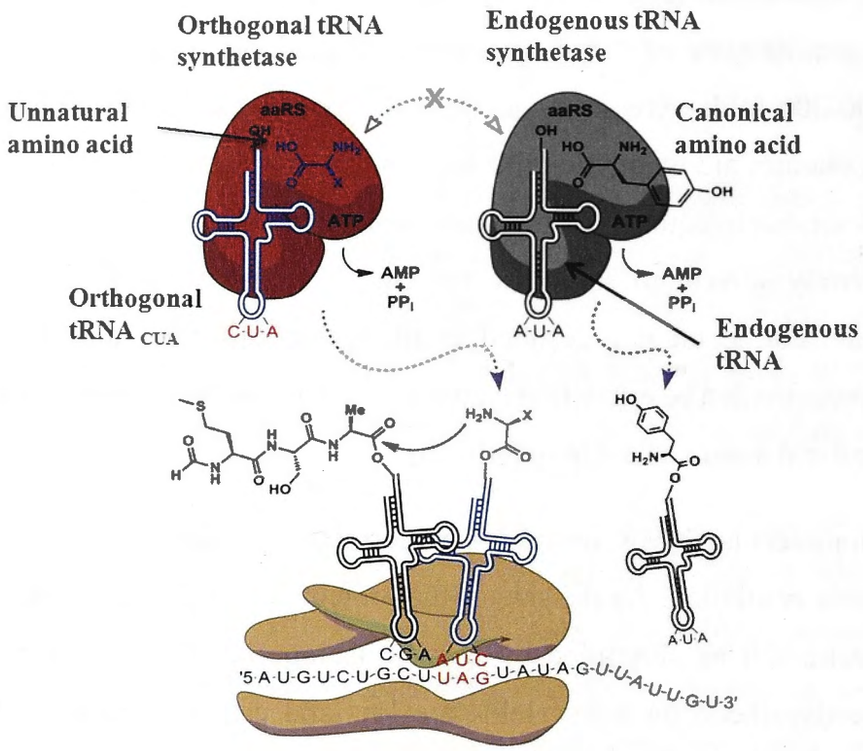

***Figure 1.6**: Schematic diagram of incorporation of unnatural amino acid into protein by the orthogonal tRNA/ aaRS pair by using the host ribosomal machinery. The orthogonal aaRS (red) aminoacylates the orthogonal tRNA (blue) with an unnatural amino acid (blue) and does not cross-react with any of the endogenous synthetases (grey) and tRNA (black). The aminoacylated orthogonal tRNA then travels to the ribosome (brown) and incorporates the unnatural amino acid in response to amber codon. The picture is the modified version from Young.T.S.et al*[50].

E. coli is commonly used expression system and TAG is often chosen as the codon for incorporating a UAA. In *E. coli* system, the amber codon is a rarely used stop codon and is introduced into site of interest. The process involves recognition of codon by aminoacylated amber supressor transfer RNA($tRNA_{CUA}$) and introduce UAA into peptide. The supressor tRNA is acylated with the desired UAA by aminoacylated tRNA synthetase (aaRS) specific for chosen UAA. To achieve this, an orthogonal supressor tRNA/ aaRS is established for the host. The conservation of orthogonality is important that is the supressor tRNA should not be recognised be endogenous tRNA synthetase and the orthogonal tRNA synthetase must not aminoacylate endogenous tRNA. In either case,

the incorporation of UAA is unsuccessful or a mixture of proteins with UAA and without UAA will be synthesised in the system. Furthermore, the orthogonal system must be compatible with the host transcription and translational components [50].

An orthogonal system for *E. coli* system is the archaean *Methanocaldococcus jannaschii* tyrosyl tRNA synthetase and tyrosyl tRNA (*Mj*TyrRS/*Mj*tRNA^{Tyr}) pair. When the tRNA anti-codon was modified to CUA to recognise the amber codon (*Mj*tRNACUA^{Tyr}) in the Prof. Schultz group [53,54], it was shown that the orthogonal pair could be engineered to incorporated more than 30 UAA into proteins [50,53].

The aminoacylated tRNA synthetase are usually developed for the particular unnatural amino acid to be used. There are evolved and selected by positive screening and negative screening

The *in vivo* procedure involves coexpressing distinct plasmid which encodes the orthogonal system and additional plasmid inserted with gene of interest [56].

The *in vitro* system of expression also referred to called as cell-free protein synthesis uses the cellular transcription and translation machinery to produce the protein in a cell-free environment. It is a well-established method and is one of the method employed in this project to study *E. coli* DnaB helicase and its interactions with other important proteins involved in DNA replication machinery.

1.3.3. Merits of the cell-free protein synthesis

The primary advantage of the *in vitro* system is the ability to manipulate the components of the system to acheive desired results. The reaction components such as the concentration of amino acids, tRNA synthetase can be manipulated to attain higher yield of proteins.

The second advantage is that additional components can be added to obtain active protein with desired post-translational modifications or membrane protein embedded in lipid bilayer[57].

In some cases, higher yield of proteins in the order of milligrams can be produced from millilitres of reaction volume and with suppression efficiency (the ratio between the yield of protein with UAA to the yield of native protein) between 25% - 96% [56-58].

Highly toxic proteins and UAA with various side chain could be site-specifically incorporated with ease and without affecting the yields of the protein in contrast to *in vivo* expression.

1.3.4. Cell-free protein synthesis

This tool was developed in 1961 by Nirenberg and Matthaei to understand the translational process [59]. It is strongly based on the discovery that the cellular intergrity is not necessary for protein synthesis and can be synthesized with crude cell extract containing the ribosomal machinery depleted from endogenous mRNA. Nirenberg's group also demonstrated that protein can be synthesised without DNA [59]. The system containing translational machinery, tRNA and enzymes accompanied by additional components such as amino acids, RNA template and energy source is called 'uncoupled' scheme in which only translation of mRNA takes place. In the early experiments, the yield of the protein expressed was very poor restricting the use of the system. The addition of bacteriophage RNA polymerase to the crude extract subsequently improved the protein yields due to increase in the concentration of the mRNA transcipts [60,61]. This makes cell-free protein synthesis more convenient to express proteins from plasmids and PCR product with phage promoter corresponding to the bacteriophage RNA polymerase. This system is called the 'coupled' scheme, it combines transcription and translation process using the DNA template as the source for the protein expression. In this system, usually supplementary components added are T7 RNA polymerase and required NTPs and plasmids used have T7 promoters [62].

The coupled scheme was further improved by the invention of continuous flow cell-free systems (CFCF)[63,64]. The components are seperated by porous membrane with a desired molecular weight cut-off, the membrane contains the higher molecular weight components such as the crude extract while the membrane is immersed in the reservoir containing the lower-molecular weight compounds that act as feeding solution. During

the reaction, there is continuous exchange of components across the membrane to carry out the expression without saturation for longer duration.

In our system, S30 extract from *E. coli* is used as the source of translation/ transcription machinery. Salts such as Mg^{2+} and Ca^{2+}, processed supressor $tRNA_{CUA}$, UAA specific tRNA synthetase, NTPs, cofactors, phosphoenolpyruvate and pyrophosphatases are added along with the S30 cell extract inside the porous membrane. The feeding solution contains the amino acids and UAA for continuous exchange. The range of the volume for preparation of S30 extract and supressor tRNA varies from 2 L to 20 L as desired [63]. The system was further optimised to increase the protein yield to milligrams per mL of reaction volume for studies. The protocols established by Apponyi and coworkers was followed for cell-free extract preparation and for the cell-free reaction set up [64]. CFCF is a simple and efficient mode of protein expression for incorporation of UAA.

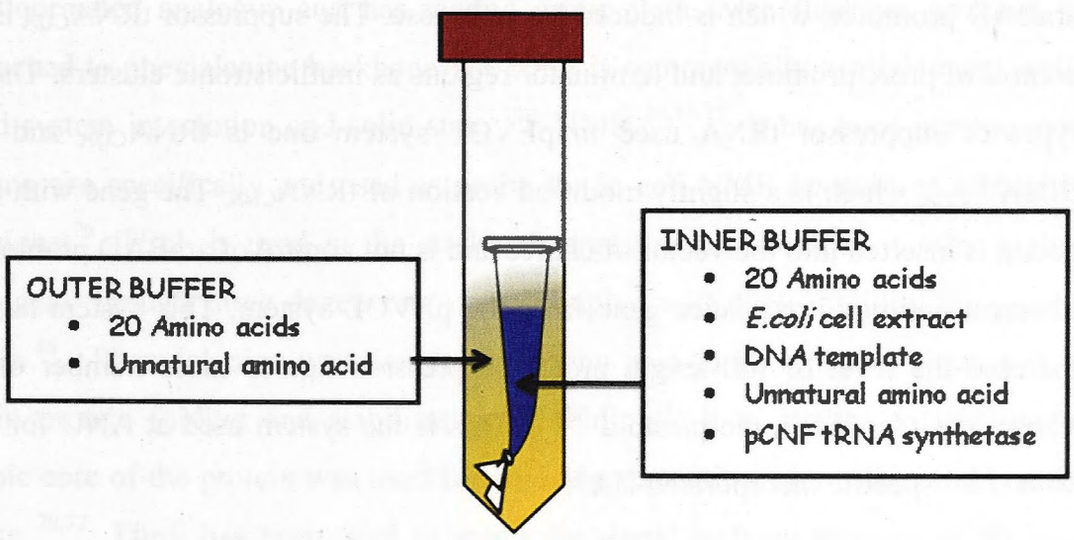

***Figure 1.6**: Schematic representation of the continuous flow cell-free protein synthesis set up in our lab. The bottom end of the membrane (spectrapor dialysis tubing) is closed by a knot and top of the membrane is closed by inserting closed 0.5 mL Eppendorf tube*

The CFCF has been applied to various structural genomics and proteomics studies [65,67]. The system has been used for NMR studies such as selectively labelling the proteins with

^{15}N isotopes [67], optimising the conditions for suppression of isotope scrambling by addition of Pyridoxal-5'-phosphate [68]. UAA with reactive side chains were successuly incorporated into proteins for lanthanide tagging using CFCF [69].

1.3.5. UAA incorporation by *in vivo* expression system

Fluorinated amino acids such as tfmF have been incorporated into proteins successfully using *in vivo* mode. The vector system used is called pDule-tfmFRS (tRNA synthetase for tfmF) along with orthogonal supressor $tRNA_{CUA}$. This vector is transformed into cells along with the vector pBAD containing the gene with amber stop codon, which is induced by arabinose. The induction of both vectors resulted in protein with tfmF [54, 55].

The system was later evolved to pEVOL system, which has a copy of suppressor tRNA along with aaRS gene for the particular UAA. The gene is under the influence of the araBAD promoter, which is induced by arabinose. The suppressor $tRNA_{CUA}$ is under the control of proK promoter and terminator regions as mutlicistronic clusters. There are two types of suppressor tRNA used in pEVOL system one is $tRNA_{CUA}$ and another is $tRNA^{opt}_{CUA}$, which is a slightly modified version of $tRNA_{CUA}$. The gene with amber stop codon is inserted into the vector of choice and is not control of araBAD promoter and has different antibiotic resistance gene from the pEVOL system. This system has shown to increase the level of full-length protein expression due to more number of $tRNA_{CUA}$ transcripts encoded in the plasmid [50, 70]. This is the system used at ANU for expression with site - specific incorporated UAA.

The choice of aaRS selective to the UAA is very important because supressor tRNA must be acylated by aaRS with UAA only and not with endogenous amino acids. In case of, UAA, it must be non-toxic, efficiently transported across the cells and stable to the endogenous enzymes.

1.3.6. Unnatural Amino acid and aminoacyl tRNA

The choice of unnatural amino acid is important because it affects the structure of the protein. Mono-fluorinated variants are generally considered to cause less structural perturbations. Fluorine label present in the side chains of aromatic amino acids have

shown not to affect the structure or the function of the protein. This was proved when 80 % of tryptophans were replaced by 5-fluorotryptophan (5-F-Trp) in *E. coli*; the growth rate of *E. coli* was not affected[71]. On the contrary, pentafluorinated phenylalanines [72], monoflourinated variants of tyrosine[73] and histidines [74] have shown to significantly affect the structure. The perturbation caused by fluorinated histidines, is due to the fluorine in the ring structure affecting the pKa value [74].

The choice of UAA for ^{19}F labelling the protein was based on the purpose of the project, which is to determine the conformational changes in 312 kDa and 480 kDa protein systems. In order to observe the signal from large systems, the spin label should have strong signals trifluorinated analogue was chosen over the monofluorinated to achieve signals with higher intensity in the large proteins [71,72].

4-trifluormethylphenylalanine (tfmF) was chosen as the UAA for the study. It is an aromatic fluorinated analogue and has methyl group with three fluorines attached to carbon attached to pheylalanine backbone.The tfmF is commercially available and well-established system in solution and solid state ^{19}F NMR [55,75-77]. It has been incorporated into proteins site specifically and used as probe for in cell NMR because of its higher intensity signal[76]. TfmF is used in the study of large proteins because it offer higher intensity signal through three degenerate non-*J*-coupled with high mobility due to the methyl axis [46]. Phenylalanine was the chosen amino acid to be replaced because it is involved in protein folding and stabilisation of folding[78]. It is mostly present in the hydrophobic core of the protein was used for studying the native, non-native and protein folded state [70,77]. TfmF has been used to study the signal in large proteins of 98 kDa homodimer histidinol dehydrogenase [77] and in studying the solvent exposure of membrane protein[76]. So far, tfmF has been incorporated into proteins by *in vivo* expression using rich media [55, 75-77].

In this project, the aaRS used for incorporating the UAA was *para*- cyanophenylalanine tRNA synthetase (*p*CNF-RS), which is developed by Prof Schultz group. *p*CNF-RS appears to be highly tolerant of parasubstituted phenylalanine derivatives. It was tested to incorporate various phenylalanine analogues into green fluorescent protein (GFP). It was reported that *p*CNF-RS incorporated the UAA with great efficiency[79]. *p*CNF-RS has not

been previously used as aaRS to incorporate tfmF into GFP. In this project, it is the first attempt to incorporate tfmF using *p*CNF-RS/*Mj*tRNACUA^{Tyr} orthogonal system in *E. coli.*

1.4. Fluorine NMR

The structural information of the proteins are generally determined using biophysical techniques such as NMR, circular dichroism (CD), fluorescence, and X-ray crystallography. In recent times, the conventional ^{1}H, ^{13}C and ^{15}N (HCN) NMR and X-ray crystallography has been widely used for structural determination. X-ray crystallography has advanced to great lenghts, providing an increasing number of high resolution protein crystals. In X-ray crystallography, though the structures are solved, we do lack the knowledge on protein dynamics under biological conditions. Multi-dimensional solution NMR spectroscopy can provide data about the dynamism of proteins and their interactions with other molecules, in addition to structure data. The technique commonly used as alternate approach to x-ray crystallography to studying the structure of smaller proteins (Mw < 30kDa), however its application to proteins with Mw > 40kDa is rare due to broad signals and spectral crowding with larger systems, which only can be circumvented by expensive deuteration of the system and long experimental measurement times in high dimensional experiments [45,46].

This emphasizes the necessity for an alternate approach or technique for studying the protein dynamics under physiological conditions to complement the known structural information [43]. This issue is addressed by ^{19}F NMR, which can be used to study local interactions of larger proteins (Mw > 100kDa) with known structural information[41]. ^{19}F NMR gives a unique insight into topology, residues involved in conformational changes, change in intramolecular distances under biological conditions. It is been well-demonstrated over the years that ^{19}F NMR is a good technique for studying large and unstable proteins. The applications are primarily based on the hypersensitivity of the flourine nucleus, which aids in observing the conformational changes and solvent exposed residues that are undetectable in HCN NMR (2D) [46].

Over the span of two decades from 1970s to 1990s, ^{19}F NMR has been used to study the significant biological processes such as protein folding, protein-protein, protein-ligand, protein-lipid binding, protein aggregation and fibrillation [46]. ^{19}F NMR is underexplored in biomolecular NMR. Eventhough, there are other NMR isotopic nucleus such as ^{31}P, ^{23}Na, which are relatively commonly used; but the ease of observing the ^{19}F chemical shifts in biomolecular NMR due to its unique properties makes it an excellent tool for monitoring the protein dynamics.

1.4.1. Properties of ^{19}F nucleus

Fluorine has magnetogyric ratio which is 83% that of the proton, has a ½ spin nucleus similar to the proton and it exists in 100% natural abundance. In proteins and DNA fluorine is absent. Because of this absence of background it is possible to use one-dimensional fluorine NMR to study protein complexes, conformational changes and monitor weak binding [45,46].

The spectral width of fluorine is 100 times larger than of proton, results in well dispersed fluorine resonances over the range of 1000 ppm. The increase in the range of chemical shifts aids in studying the weak protein-protein, protein-ligand, folding, and enzyme kinetics without the necessity to use 2D NMR for resonance separation [45,46]. It was demonstrated by Li et al., that they observed wide range of chemical shifts on measuring the intestinal fatty acid protein under acidic condition to study the tertiary structural changes using ^{19}F NMR [44].

The attribute of utmost importance is the hypersensitivity of ^{19}F chemical shifts to its local environment. Any change in the conformation due to protein-protein interactions, ligand binding can be detected from the change in chemical shifts or the intensity of resonances[80-82].

Fluorine has lone-pair electrons that generate paramagnetic shielding and in the shielding equation, the large paramagnetic term of the fluorine makes the chemical shifts hypersensitive to local environment of folded protein. The fluorine shielding is contributed by four forces namely, hydrogen bond formation, electrostatic force due to dipoles or charges, local magnetic fields and van der Waal's interactions [83]. In two

independent studies on the forces influencing the chemical shifts involving fluorinated - alkaline phosphatase [83] and ribonuclease-S [84] demonstrated that van der Waal's interaction as the main factor that affects the chemical shift of fluorine nucleus. It was proposed that along with van der Waal's interaction, local magnetic and electric fields also contributed to the shielding in smaller proportions [83,84]. In recent study on galactose binding protein using ^{19}F NMR, it was reported that electric field interactions contributed largely to the chemical shielding term [84]. On the whole, it appears that both van der Waals and electrostatic interactions contribute to the fluorine shielding term.

The high senstivity and strong dipolar coupling of fluorine nucleus allowed for measuring distance restraints by ^{19}F-^{19}F and ^{19}F-^{1}H NOEs [46]. ^{19}F-^{1}H NOEs is used in for establishing solvent contacts, intramolecular contacts for distance restraints and intermolecular interactions between protein and ligand. The difference in cross relaxation rates between fluorine nuclei-water and fluorine nuclei-interior of the protein is used for identifying the solvent exposure or burial of the fluorinated site. This scheme was used in studying exposure of tyrosine to solvent by replacing all the tyrosine with 3-fluorotyrosine in calmodulin in both calcium bound and calcium free state [48].

The distinctive physical property of fluorinated compunds is due to its strong electronegativity. On comparing the C-F and C-H bonds, the former is more stable and less polarised than the C-H bond. The polarisation is in opposite direction to that of C-H bond. The vander waals radius of fluorine is slightly greater than that of hydrogen. However, hydrogen can be often substituted by fluorine in compounds as amino acid analogues or other chemical compounds for incorporation[85].

Fluorine incorporation causes little or no perturbations to the structure and functions of the proteins. This is generally the case with fluorinated aromatic acids which have fluorine in their side chains and especially if fluorine is substituted for hydrogen in the side chain. The covalent radii of fluorine is 1.35 Å and hydrogen is 1.2 Å, due to the similar size, the fluorine replacement causes little steric effect. The effect of fluorine incorporation on the structure and functions has been tested in several proteins and there was no resultant change in the properties of the protein using the UAA with fluorine in the side chain [46,47].

1.4.2. Application of ^{19}F NMR

^{19}F NMR has been applied to detect the binding of the small chemical compounds to the large protein system for drug discovery studies. One of the recent work in this field was by Liu and coworkers where they measured the binding of various small compounds to beta-Adrenergic receptor to detect the activity of the protein by attaching the fluorinated label to cysteines. The change in the peak intensity and shift of the fluorine resonance showed the binding of the agonist and antagonists to the protein [81].

^{19}F NMR has also been applied in the detecting the solvent exposure of important residues in membrane proteins and the solvent induced isotope shifts strategy was used to detect the important residues. The position of residues are identified by observing the change in the intensity of the chemical shifts of the particular fluorinated site by altering the percentage of deuterium oxide in the buffer. The exchange of deuterium oxide with water causes an upfield shifts of the ^{19}F resonance indicating that the residue is solvent exposed[80].

Structural mobility has been examined by ^{19}F NMR, by observing the fluorine relaxation. The fluorinated proteins were studied in intact cells, the peak broadening and narrowing were then assessed when the proteins were present inside cells and after cell lysis. The broader peak showed that the protein was interacting with other proteins inside the cells, when compared to peaks when the cells were disrupted [86].

These applications have demonstrated the fluorine nucleus is highly sensitive to its local environment even a minor change in the interaction with fluorine, it can be detected by change in the chemical shift, line width or intensity of the peak. The combination of mentioned attributes makes ^{19}F NMR it useful tool to study large proteins in solution.Its incorporation has also been proved not to affect the structure or the function of the protein. Some of the largest protein systems investigated using ^{19}F NMR had mass in the range of 100 kDa. Here we are showing that we can observe ^{19}F signals from protein and its complexes whose masses are in the range of 300 to 480 kDa. We are also demonstrating that the conformational changes are well detected by fluorine nucleus due to its sensitivity to change in their interactions with neighbouring residues.

Overview of the thesis

The main aim of this work is to obtain structural information about the replication proteins using NMR. NMR studies give the insight information on the structure of the protein and its interaction in biologically native state of the protein. Though in ^{19}F NMR it is easy to analyse the data, because of high sensitivity of the fluorine nucleus and absence of any background signals and makes it a good tool to study large protein complexes, it is underexpored in biomolecular NMR. It is primarily due to the difficulty in site-specific incorporation of fluorine into the protein. The incorporation of fluorinated amino acids into a protein using orthogonal tRNA/tRNA synthetase system developed by Prof. Schultz group using both in vitro and in vivo modes has become the solution. The strategy aids to site-specifically label the protein with non-toxic fluorinated aromatic amino acids such as tfmF into large protein complexes without affecting structure and functions of the protein. DnaB helicase involved in initiation phase of the replication was chosen to be studied by ^{19}F NMR by incorporating tfmF site-specifically to label the protein with fluorine. The conformational changes of DnaB helicase under different conditions and in complex with other replication proteins were studied using ^{19}F NMR. DnaB helicases from different organisms were studied to compare and contrast the importance of the conformational changes resulting in the better understanding of the replication system.

Chapter 1: Introduces about the replication system and important proteins involved in initiation stage of DNA replication. This is followed by introducing the biosynthetic system of expression of labelled DnaB helicase for ^{19}F NMR studies. A detailed account on the strategy of incorporation of fluorinated amino acid into DnaB using orthogonal system is outlined. The importance of ^{19}F NMR and its properties are explained as the last part of chapter 1

Chapter 2: General methods and protocols are outlined.

Chapter 3: Various conformations of *E. coli* DnaB helicase are studied by ^{19}F NMR. The tfmF is site-specifically incorporated into DnaB by cell-free protein synthesis. TfmF was incorporated into N-terminal domain and helicase was studied under different pHs and in

complex with helicase loader. The distinctive chemical shifts for each mutant and flexibility of the N-terminal domain was observed.

Chapter 4: The tfmF is site-specifically incorporated into *Gst* DnaB by *in vivo* mode of expression. TfmF was incorporated into N-terminal domain of the helicase and was examined under different pHs, different temperature, with Mg^{2+}, primase and helicase loader. The successful formation of hexamers and complex with primase was confirmed by gel filtration under physiological conditions and the conformational changes of DnaB helicase observed by ^{19}F NMR are reported in this chapter.

Chapter 5: Future directions

Reference

1. Robinson, A., Causer, R. J. & Dixon, N. E. Architecture and Conservation of the Bacterial DNA Replication Machinery, an Underexploited Drug Target. *Current Drug Targets* **13**, 352-372 (2012).
2. Schaeffer, P. M., Headlam, M. J. & Dixon, N. E. Protein--protein interactions in the eubacterial replisome. *IUBMB Life* **57**, 5-12, (2005).
3. Robinson, A. & van Oijen, A. M. Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat Rev Microbiol* **11**, 303-315, (2013)
4. Hamdan, S. M., Loparo, J. J., Takahashi, M., Richardson, C. C. & van Oijen, A. M. Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* **457**, 336-U339, (2009).
5. Kaguni, J. M. Replication initiation at the *Escherichia coli* chromosomal origin. *Current Opinion in Chemical Biology* **15**, 606-613, (2011).
6. Pomerantz, R. T. & O'Donnell, M. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends in Microbiology* **15**, 156-164, (2007).
7. Soultanas, P. Loading mechanisms of ring helicases at replication origins. *Molecular Microbiology* **84**, 6-16 (2012).

8. Moriya, S., Kato, K., Yoshikawa, H. & Ogasawara, N. Isolation of a dnaA mutant of Bacillus subtilis defective in initiation of replication: amount of DnaA protein determines cells' initiation potential. *EMBO J* **9**, 2905-2910 (1990).
9. Hill, N. S., Kadoya, R., Chattoraj, D. K. & Levin, P. A. Cell size and the initiation of DNA replication in bacteria. *PLoS Genet* **8**,(2012)
10. Keyamura, K. *et al.* The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev* **21**, 2083-2099,(2007)
11. Duderstadt, K. E., Chuang, K. & Berger, J. M. DNA stretching by bacterial initiators promotes replication origin opening. *Nature* **478**, 209-213(2011)
12. Ozaki, S. & Katayama, T. DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* **62**, 71-82, (2009).
13. Neylon, C. *et al.* Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: A model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* **39**, 11989-11999, (2000).
14. Wilce, J. A. *et al.* Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. *Nature Structural Biology* **8**, 206-210, (2001).
15. Li, Y. & Araki, H. Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication. *Genes to Cells* **18**, 266-277, (2013).
16. Arias-Palomo, E., O'Shea, V. L., Hood, I. V. & Berger, J. M. The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. *Cell* **153**, 438-448, (2013).
17. Gupta, M. K., Atkinson, J. & McGlynn, P. DNA Structure Specificity Conferred on a Replicative Helicase by Its Loader. *Journal of Biological Chemistry* **285**, 979-987, (2010).
18. Davey, M. J., Fang, L. H., McInerney, P., Georgescu, R. E. & O'Donnell, M. The DnaC helicase loader is a dual ATP/ADP switch protein. *Embo Journal* **21**, 3148-3159, (2002).
19. San Martin, C. *et al.* Three-dimensional reconstructions from cryoelectron microscopy images reveal an intimate complex between helicase DnaB and its loading partner DnaC. *Structure with Folding & Design* **6**, 501-509 (1998).
20. Makowska-Grzyska, M. & Kaguni, J. M. Primase Directs the Release of DnaC from DnaB. *Mol Cell* **37**, 90-101, (2010).

21. Loscha, K. V. *et al.* A novel zinc-binding fold in the helicase interaction domain of the *Bacillus subtilis* DnaI helicase loader. *Nucleic Acids Research* **37**, 2395-2404, (2009).
22. Velten, M. *et al.* A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Molecular Cell* **11**, 1009-1020, (2003).
23. Tsai, K. L., Lo, Y. H., Sun, Y. J. & Hsiao, C. D. Molecular Interplay between the Replicative Helicase DnaC and Its Loader Protein DnaI from *Geobacillus kaustophilus*. *Journal of Molecular Biology* **393**, 1056-1069, (2009).
24. Seitz, H., Weigel, C. & Messer, W. The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*. *Molecular Microbiology* **37**, 1270-1279, (2000).
25. Yoda, K. & Okazaki, T. Specificity of Recognition Sequence for *Escherichia-Coli* Primase. *Molecular & General Genetics* **227**, 1-8, (1991).
26. Johnson, S. K., Bhattacharyya, S. & Griep, M. A. DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates. *Biochemistry* **39**, 736-744, (2000).
27. Bhattacharyya, S. & Griep, M. A. DnaB helicase affects the initiation specifity of *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry* **39**, 745-752, (2000).
28. Stamford, N. P. J., Lilley, P. E. & Dixon, N. E. Enriched Sources of *Escherichia-Coli* Replication Proteins - the Dnag Primase Is a Zinc Metalloprotein. *Biochimica Et Biophysica Acta* **1132**, 17-25 (1992).
29. Pan, H. & Wigley, D. B. Structure of the zinc-binding domain of Bacillus stearothermophilus DNA primase. *Structure with Folding & Design* **8**, 231-239, (2000).
30. Tougu, K. & Marians, K. J. The interaction between helicase and primase sets the replication fork clock. *Journal of Biological Chemistry* **271**, 21398-21405 (1996).
31. Bailey, S., Eliason, W. K. & Steitz, T. A. Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* **318**, 459-463, (2007).
32. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry* **76**, 23-50, (2007).

33. Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annual Review of Biochemistry* **69**, 651-697, (2000).

34. Jezewska, M. J., Rajendran, S., Bujalowska, D. & Bujalowski, W. Does single-stranded DNA pass through the inner channel of the protein hexamer in the complex with the *Escherichia coli* DnaB helicase? Fluorescence energy transfer studies. *Journal of Biological Chemistry* **273**, 10515-10529, (1998).

35. Lebowitz, J. H. & Mcmacken, R. The Escherichia-Coli Dnab Replication Protein Is a DNA Helicase. *Journal of Biological Chemistry* **261**, 4738-4748 (1986).

36. Arai, K. & Kornberg, A. Mechanism of dnaB protein action. II. ATP hydrolysis by dnaB protein dependent on single- or double-stranded DNA. *J Biol Chem* **256**, 5253-5259 (1981)

37. Wickner, S. & Hurwitz, J. Interaction of Escherichia-Coli Dnab and Dnac(D) Gene Products Invitro. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 921-925, (1975).

38. Biswas, E. E., Chen, P. H. & Biswas, S. B. Modulation of enzymatic activities of Escherichia coli DnaB helicase by single-stranded DNA-binding proteins. *Nucleic Acids Research* **30**, 2809-2816, (2002).

39. Dallmann, H. G., Kim, S., Pritchard, A. E., Marians, K. J. & McHenry, C. S. Characterization of the unique C terminus of the Escherichia coli tau DnaX protein - Monomeric C-tau binds alpha and DnaB and can partially replace tau in reconstituted replication forks. *Journal of Biological Chemistry* **275**, 15512-15519, (2000).

40. Nakayama, N., Arai, N., Kaziro, Y. & Arai, K. Structural and functional studies of the dnaB protein using limited proteolysis. Characterization of domains for DNA-dependent ATP hydrolysis and for protein association in the primosome. *J Biol Chem* **259**, 88-96 (1984).

41. Biswas, E. E. & Biswas, S. B. Mechanism of DnaB helicase of Escherichia coli: Structural domains involved in ATP hydrolysis, DNA binding, and oligomerization. *Biochemistry* **38**, 10919-10928, (1999).

42. Barcena, M. *et al.* The DnaB center dot DnaC complex: a structure based on dimers assembled around an occluded channel. *Embo Journal* **20**, 1462-1468, (2001).

43. Yu, X., Jezewska, M. J., Bujalowski, W. & Egelman, E. H. The hexameric E. coli DnaB helicase can exist in different Quaternary states. *J Mol Biol* **259**, 7-14, (1996).
44. Li, H. & Frieden, C. Fluorine-19 NMR studies on the acid state of the intestinal fatty acid binding protein. *Biochemistry* **45**, 6272-6278, (2006).
45. Danielson, M. A. & Falke, J. J. Use of F-19 NMR to probe protein structure and conformational changes. *Annual Review of Biophysics and Biomolecular Structure* **25**, 163-195,(1996).
46. Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of 19F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* **62**, 1-33,(2012)
47. Lee, H. W. *et al.* F-19 NMR investigation of F-1-ATPase of Escherichia coli using fluorotryptophan labeling. *Journal of Biochemistry* **127**, 1053-1056 (2000).
48. Kitevski-LeBlanc, J. L., Al-Abdul-Wahid, M. S. & Prosser, R. S. A Mutagenesis-Free Approach to Assignment of F-19 NMR Resonances in Biosynthetically Labeled Proteins. *Journal of the American Chemical Society* **131**, 2054 (2009).
49. Adriaensens, P. *et al.* Investigation of Protein-Structure by Means of F-19-Nmr - a Study of Hen Egg-White Lysozyme. *European Journal of Biochemistry* **177**, 383-394, (1988).
50. Young, T. S. & Schultz, P. G. Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry* **285**, 11039-11044, (2010).
51. Loscha, K. V. *et al.* Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angewandte Chemie-International Edition* **51**, 2243-2246, (2012).
52. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* **52**, 1828-1837, (2013).
53. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annual Review of Biophysics and Biomolecular Structure* **35**, 225-249,(2006).
54. Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *Journal of the American Chemical Society* **122**, 5010-5011, (2000).

55. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C. & Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat Protoc* **2**, 2601-2607, (2007).
56. Albayrak, C. & Swartz, J. R. Using E. coli-based cell-free protein synthesis to evaluate the kinetic performance of an orthogonal tRNA and aminoacyl-tRNA synthetase pair. *Biochemical and Biophysical Research Communications* **431**, 291-295, (2013).
57. Bundy, B. C. & Swartz, J. R. Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein-Protein Click Conjugation. *Bioconjugate Chemistry* **21**, 255-263, (2010).
58. Goerke, A. R. & Swartz, J. R. High-Level Cell-Free Synthesis Yields of Proteins Containing Site-Specific Non-Natural Amino Acids. *Biotechnology and Bioengineering* **102**, 400-416, (2009).
59. Nirenberg, M. & Matthaei, J. H. Dependence of Cell-Free Protein Synthesis in E Coli Upon Naturally Occurring or Synthetic Polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 1588, (1961)
60. Nevin, D. E. & Pratt, J. M. A coupled in vitro transcription-translation system for the exclusive synthesis of polypeptides expressed from the T7 promoter. *FEBS Lett* **291**, 259-263, (1991).
61. Craig, D., Howell, M. T., Gibbs, C. L., Hunt, T. & Jackson, R. J. Plasmid cDNA-directed protein synthesis in a coupled eukaryotic *in vitro* transcription-translation system. *Nucleic Acids Res* **20**, 4987-4995 (1992)
62. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol* **23**, 150-156,(2005).
63. Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. & Alakhov, Y. B. A Continuous Cell-Free Translation System Capable of Producing Polypeptides in High-Yield. *Science* **242**, 1162-1164, (1988).
64. Apponyi, M. A., Ozawa, K., Dixon, N. E. & Otting, G. Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods Mol Biol* **426**, 257-268, (2008)
65. Yokoyama, S. Protein expression systems for structural genomics and proteomics. *Current Opinion in Chemical Biology* **7**, 39-43, (2003).

66. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14652-14657, (2002).
67. Wu, P. S. *et al.* Amino-acid type identification in 15N-HSQC spectra by combinatorial selective 15N-labelling. *J Biomol NMR* **34**, 13-21, (2006).
68. Su, X. C., Loh, C. T., Qi, R. H. & Otting, G. Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enymes for selective N-15-labelling and production of perdeuterated proteins in H2O (vol 50, pg 35, 2011). *Journal of Biomolecular Nmr* **51**, 409-409,(2011).
69. Loh, C. T. *et al.* Lanthanide Tags for Site-Specific Ligation to an Unnatural Amino Acid and Generation of Pseudocontact Shifts in Proteins. *Bioconjugate Chemistry* **24**, 260-268, (2013).
70. Heidary, D. K. & Jennings, P. A. Three topologically equivalent core residues affect the transition state ensemble in a protein folding reaction. *Journal of Molecular Biology* **316**, 789-798, (2002).
71. Kim, H. W., Perez, J. A., Ferguson, S. J. & Campbell, I. D. The Specific Incorporation of Labeled Aromatic-Amino-Acids into Proteins through Growth of Bacteria in the Presence of Glyphosate - Application to Fluorotryptophan Labeling to the H+-Atpase of *Escherichia-Coli* and Nmr-Studies. *Febs Letters* **272**, 34-36, (1990).
72. Bovy, P. R., Getman, D. P., Matsoukas, J. M. & Moore, G. J. Influence of Polyfluorination of the Phenylalanine Ring of Angiotensin-Ii on Conformation and Biological-Activity. *Biochimica Et Biophysica Acta* **1079**, 23-28, (1991).
73. Labroo, V. M. *et al.* Direct Electrophilic Fluorination of Tyrosine in Dermorphin Analogs and Its Effect on Biological-Activity, Receptor Affinity and Selectivity. *International Journal of Peptide and Protein Research* **37**, 430-439 (1991).
74. Taylor, H. C. *et al.* Active Conformation of an Inactive Semi-Synthetic Ribonuclease-S. *Journal of Molecular Biology* **149**, 313-317, (1981).
75. Shi, P. *et al.* In situ ^{19}F NMR studies of an E. coli membrane protein. *Protein Science* **21**, 596-600,(2012).

76. Jackson, J. C., Hammill, J. T. & Mehl, R. A. Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J Am Chem Soc* **129**, 1160-1166, doi:10.1021/ja064661t (2007).

77. Shi, P., Li, D., Chen, H. W., Xiong, Y. & Tian, C. L. Site-specific solvent exposure analysis of a membrane protein using unnatural amino acids and F-19 nuclear magnetic resonance. *Biochemical and Biophysical Research Communications* **414**, 379-383, doi:DOI 10.1016/j.bbrc.2011.09.082 (2011).

78. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing (vol 323, pg 927, 2002). *Journal of Mol Bio* **324**, 1051-1051,(2002)

79. Young, D. D. *et al.* An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **50**, 1894-1900, doi:Doi 10.1021/Bi101929e (2011).

80. Li, C. *et al.* ^{19}F NMR studies of alpha-synuclein conformation and fibrillation. *Biochemistry* **48**, 8578-8584, (2009)

81. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wuthrich, K. Biased Signaling Pathways in beta(2)-Adrenergic Receptor Characterized by F-19-NMR. *Science* **335**, 1106-1110,(2012)

82. Millett, F. & Raftery, M. A. An NMR method for characterizing conformation changes in proteins. *Biochem Biophys Res Commun* **47**, 625-632, (1972).

83. Hull, W. E. & Sykes, B. D. Fluorine-19 nuclear magnetic resonance study of fluorotyrosine alkaline phosphatase: the influence of zinc on protein structure and a conformational change induced by phosphate binding. *Biochemistry* **15**, 1535-1546 (1976).

84. Chaiken, I. M., Freedman, M. H., Lyerla, J. R., Jr. & Cohen, J. S. Preparation and studies of ^{19} F-labeled and enriched 13 C-labeled semisynthetic ribonuclease-S' analogues. *J Biol Chem* **248**, 884-891 (1973).

85. Buer, B. C. & Marsh, E. N. Fluorine: a new element in protein design. *Protein Sci* **21**, 453-462 (2012).

86. Williams, S. P., Haggie, P. M. & Brindle, K. M. ^{19}F NMR measurements of the rotational mobility of proteins in vivo. *Biophys J* **72**, 490-498,(1997).

CHAPTER 2

General Methods and Materials

2.1. Bacterial strains and Plasmid Vectors

2.1.1. Bacterial Strains

E. coli strain BL21(DE3)pLysS: genotype (*F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3) pLysS(cmR)*)[1] was used for protein expression system and *E. coli* DH5α : genotype *(F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–)* was used for cloning purposes[2,3].

2.1.2. Plasmid Vectors

The vector pETMCSI is a T7 expression vector. According to Studier, the vector has phage T7 φ10 promoter for gene expression and requires to be transformed into *E. coli* expression strain, which has T7 RNA polymerase for transcription, leading to overexpression of protein of interest upon induction[3,4].

The vector pETMCSIII had same basic components as pETMCSI. The codon for Met-His_6 is placed between the ribosome binding motif and restriction endonuclease NdeI site of multi-cloning site (MCS) on the vector [4]. A protein is expressed with an additional Met-His_6 on the N-terminus of the protein.

2.1.3. Growth Media

E. coli strains were grown in LBT media[5]. LBT media was prepared according to the recipe published by Luria and Burrous in 1957 with the addition of 25mg/L of thymine as a supplement (LBT) and respective antibiotic. Ampicillin was added to the media with a concentration of 100 μg /ml (LBTA) and/or chloramphenicol is added at the

concentration of 50 μg/mL (LBT(A/C)), if the bacterial strain had ampicillin and/or chloramphenicol resistance. *E. coli* was grown at 37 °C.

2.2. Molecular Cloning Techniques*

2.2.1. Small-scale plasmid extraction by alkaline-lysis (Miniprep)

DH5α carrying the plasmid with the desired gene was grown in LBT medium with the required antibiotic at 37 °C, overnight. The following day, the cells were harvested and lysed using the buffers from the Qiagen Miniprep Kit (Qiagen) and following the protocol of the kit.

2.2.2. Restriction Digestion of DNA by endonucleases

The restriction digestion endonucleases used for cloning were *Eco*RI and *Nde*1, supplied by New England BioLabs. The protocol followed was incubating 100 ng DNA with 1 unit of restriction digestion enzyme in 50 μl at 37 °C, for 4 h. The samples with digested DNA was stored at 4 °C and were isolated by running the samples on agarose gel (section 2.2.3) and extracted from it (2.2.4).

2.2.3. Agarose gel electrophoresis

Both analytical and preparative gels were made of 1% - 1.5% agarose in 1x TAE (40 mM Tris acetate, pH 7.8, 2mM EDTA) [6] containing 0.5 μl/ml of RedSafe DNA stain (ChemBio Ltd, England). Electrophoresis tank (Thermo Fisher Scientific Inc.) was used for running agarose gels. The electrophoresis was carried out at 110 V for 40 min for well resolved DNA fragments. The DNA fragments were then observed under UV light with wavelength of 254nm using UVP TM-15 ChromatoVue UV transilluminator with camera to capture the picture.

---

*Water used in the process is distilled water, which is further purified using MilliQ system (Millipore) for all the methods in section 2.2 and 2.3

2.2.4. DNA extraction from Agarose gel

The DNA fragments were isolated and purified from the gel using the NucleoSpin Extract kit (Machery-Nagel GmbH). The desired DNA fragments were cut out from the gel and dissolved in the buffers from the kit and protocol was followed accordingly.

2.2.5. Ligation

T4 DNA ligase (Fermentas, Thermo Fisher Scientific Inc.) was used to set up ligation reactions with volume of 20 µl using purified DNA fragments (Section 2.2.4) as described by Sambrook and coworker[7]. Reaction mixtures containing 100 ng of DNA fragments, 100 ng of vector, 10x DNA ligase buffer and T4 DNA ligase enzyme were incubated at 16 °C for 16 h.

2.2.6. Electro competent cells and transformation of the plasmids

E. coli electrocompetent cells were made following the procedure described in Miller and Nickoloff [8]. The transformation of the cells is carried out by mixing 2 µl of the ligation mix with 50 µl competent cells and incubated on ice. After the incubation of 10 min, the cells are transferred into electroporation cuvette and pulsed for a second in the electroporator (Bio-Rad, Laboratories Pty., Ltd.). Immediately after electroporation, 1 ml of LB media was added and incubated at 37 °C for 1 h. The transformed cells are plated on the LBT media plates with desired antibiotic.

2.2.7. Colony polymerase chain reactions (Colony PCR)

The reaction mix of 20 µl volume contains 0.2 µM of primers, 0.25 mM of dNTPs, 1.25 units of Taq DNA polymerase, 1X Thermopol buffer (20 mM Tris-HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X-100, pH 8.8) (New England Biolabs, Inc.). The cycle conditions of PCR were : 10 min at 98 °C, single step denaturation followed by 30 s at 94 °C, short denaturation; 15 s annealing at 52 °C and

elongation step a 72 °C for 2 min 30 s (35 cycles) and single final step at 72 °C for 10 min.

2.2.8. Sequencing polymerase chain reactions

The reaction mixture was prepared by mixing 1 µl of big dye v.3.1 and 4 µl big dye dilution buffer supplied by The John Curtin School of Medical Research, ANU, Australia. The mix was further supplemented with 3.2 pmole of primers, 100 ng of DNA template and water to make upto 20 µl. The cycle conditions includes an initial single step of denaturation at 96 °C, which is followed by 10 s denaturation at same temperature, a 15 s at 50 °C and at 60 °C for 4 min. The above mentioned last three steps are repeated for 30 times. The samples were then stored at -20 °C till further processing.

2.2.9. Ethanol Precipitation of DNA for determination of nucleotide sequence

After sequencing PCR, the DNA was ethanol precipitated for DNA sequencing. According to the protocol 20 µl of PCR product was added to 70 µl of 100% ethanol along with 1 µl of 125 mM EDTA and 9 µl of 1M sodium acetate (pH4.8). The mix was incubated at room temperature for 15 minutes and centrifuged for 20 min at 13200 rpm. The supernatant was decanted and the pellet was washed with 250 µl of 70 % ethanol and again was centrifuged as previous step. The pellet was dried using desiccator with vacuum pump.

2.2.10. Nucleotide sequences

The samples were analyzed on ABI 3730 DNA sequencer at the Biomolecular Resource Facility (John Curtin School of Medical Research, ANU).

2.3. Protein Tools*

2.3.1 Purification of proteins using FPLC (Fast Protein Liquid Chromatography)

FPLC is equipment for protein purification from GE Healthcare. The Äkta Explorer and FPLC is controlled by software named UNICORN, it controls the equipment, evaluates

the process and reports the results in the form of intensity of the UV absorption. All the chromatography done with FPLC was performed at 4°C - 6 °C.

2.3.2. Denaturing SDS-polyacrylamide gel electrophoresis

Sodium dodecylsulphate polyacrylamide gel electrophoresis is often referred as SDS-PAGE is a protein analytical tool. SDS is an anionic detergent which denatures the protein and the denatured protein is allowed to run through the polyacrylamide gel and depending on the molecular weight of the protein, it is divided as bands from higher molecular weight to the lower mass.

The SDS PAGE gel comprises two layers; one is the separating gel containing 12-15%, acrylamide (30:2.7 acrylamide:bisacrylamide), 375 mM Tris.HCl (pH 8.8), 0.1% (w/v) SDS, 0.033% (w/v) ammonium persulphate and 0.033% (v/v) TEMED and second is the sacking gel containing 4.5% acrylamide (30:2.7 acrylamide:bisacrylamide), 125mM Tris.HCl (pH 6.8), 0.1% (w/v) SDS, 0.08 % ammonium persulphate and 0.08% (v/v) TEMED. The SDS-PAGE was run in SDS containing buffer with 51 mM Tris base, 384 mM glycine and 0.1% (w/v) SDS. The protein samples loaded onto gel (1.5 × 200 ×150 mm) is prepared by mixing equal volumes of protein sample with loading buffer (300 mM Tris base, 15% (v/v) glycerol, 0.6% (w/v) bromophenol blue, 50 mM DTT and 1–2% SDS). The samples were heated at 90 °C for 10 min. The electrophoresis was carried out at 20 mA for 50 min till the dye front reaches the end of the gel. For reference molecular marker (GE healthcare) was loaded onto gel along with other samples. The gel was washed in water* to remove the SDS for staining purposes. The stain used was Coomassie Blue stain from Bio-Rad Laboratories Ltd) and destained using water*. The whole procedure was followed according to Laemmli [9].

2.3.3. Concentration of Proteins

The proteins were concentrated by centrifugation for NMR spectroscopy. The protein concentrators called centricons (Millipore Corporation) were centrifugal filter units with molecular weight cut-off (MWCO) of 10,000- 30,000 were used for concentrating protein with molecular weight of 52 kDa. The filter units were washed with water by spinning at 5000g for 30 min, followed by equilibration with buffer as same as the protein. The process was repeated till the protein was concentrated to the desired concentration levels.

In case, if the volume of the protein is larger than the capacity of the filter unit, the protein concentrating step was performed stepwise by adding the remaining dilute protein solution to the concentrated solution. The protein solution was mixed well by inverting the unit before the centrifugation to avert protein aggregation.

2.3.4. Determination of protein concentration

The protein concentration was determined using Nanodrop (ND-1000) Spectrophotometer from Thermo Fisher Scientific. The operation is carried out by placing 2 µl of protein sample on the pedestal of the instrument and measure the UV absorption a 280 nm (A_{280}) using nanodrop 1000 3.7 version, a software program which controls and reports the results. The molar extinction coefficients at 280 nm of pure proteins were determined from their amino acid compositions as described by Gill and von Hippel [10].

Protein concentrations determined spectrophotometrically from the absorbance values at 280 nm. All specific absorbance values were calculated using the formula

$$M = A_{280}\ (1\ mg/ml)/(5690W + 1280Y + 120C)$$

where W, Y and C are the numbers of Trp, Tyr and Cys residues in a polypeptide of mass M, and 5690, 1280 and 120 are the respective extinction coefficients for these residues.

2.4. ^{19}F NMR measurements

The NMR experiments were performed on a 500 MHz Varian Inova spectrometer (Varian Inc., Palo Alto, CA). The proton coil was tuned to fluorine frequency (470.114 MHz). For 1D ^{19}F NMR experiments, the samples contained 10% D_2O for locking he sample during the run. ^{19}F NMR spectra were obtained at 25 °C and recording of the transients depended on the protein stability and concentration. 10 Hz line broadening was applied. The NMR was measured with standard spectral parameters being a 10,000 Hz

*Water used in the process is distilled water, which is further purified using MilliQ system (Millipore) for all the methods in section 2.2 and 2.3

spectral width, 90° pulse width, 1.0 s relaxation delay. The spectra were externally referenced to free trifluoromethylphenylalanine amino acid, whose chemical shift is at -62.1 ppm for each spectrum measured.

In Chapter 3 - After purification using Ni-NTA columns, the sample was concentrated. The concentrated samples were then transferred into stable NMR buffer containing 50 mM Tris·HCl pH 8, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM ATP, 10% glycerol. The pH of the buffer was varied for measurement of its effect on the structure. The ^{19}F NMR of DnaB helicase was measured at three different pHs of 8.1, 7.2 and 6.5. The DnaB and $DnaB_6$-$DnaC_6$ complex was measured by ^{19}F NMR with Mg^{2+} and ATP. The concentration of the DnaB helicase was between 10 – 30 μM as a monomer depending on the mutation site.

In Chapter 4- The *Gst DnaB* hexamer samples were dialysed into the NMR buffer (50 mM Tris pH 7.4, 1 mM EDTA, 1 mM DTT, 200 mM NaCl) and concentrated to 70 μM.

2.5. Mutation sites for tfmF incorporation

The mutation sites were introduced into *E. coli* and *Gst* DnaB helicase by site directed mutagenesis. The hydrophobic residues such as phenylalanine and tyrosines were mutated to amber codon for tfmF incorporation. These sites were chosen to minimise the disturbances due to charge of the tfmF and its size. TfmF is hydrophobic in nature due to fluorine atoms and main chain is phenylalanine ring, therefore, replacing phenylalanine and tyrosine with tfmF was expected to cause minimum perturbation to the structure of DnaB helicase.

Table 2.1: Mutation sites in the corresponding gene

Organism and Gene	Mutation sites
E. coli DnaB	F69tfmF, F102tfmF, F147tfmF and F166tfmF
Gst DnaB	F52tfmF, Y104tfmF and Y130tfmF

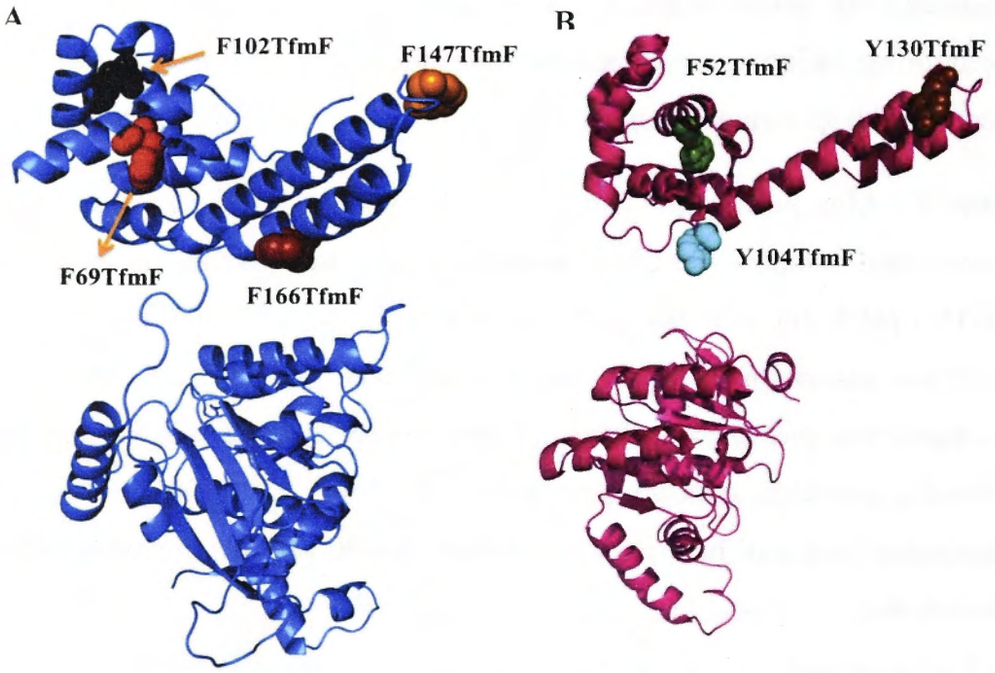

***Figure 2.1**: A) Ribbon representation of E. coli DnaB helicase (blue). Residues of the chosen mutation sites for tfmF incorporation are highlighted and represented with spheres. B) Ribbon representation of Gst DnaB helicase (PDB ID:2R6C) (hotpink) with chosen mutation sites for tfmF incorporation displayed as in A)*

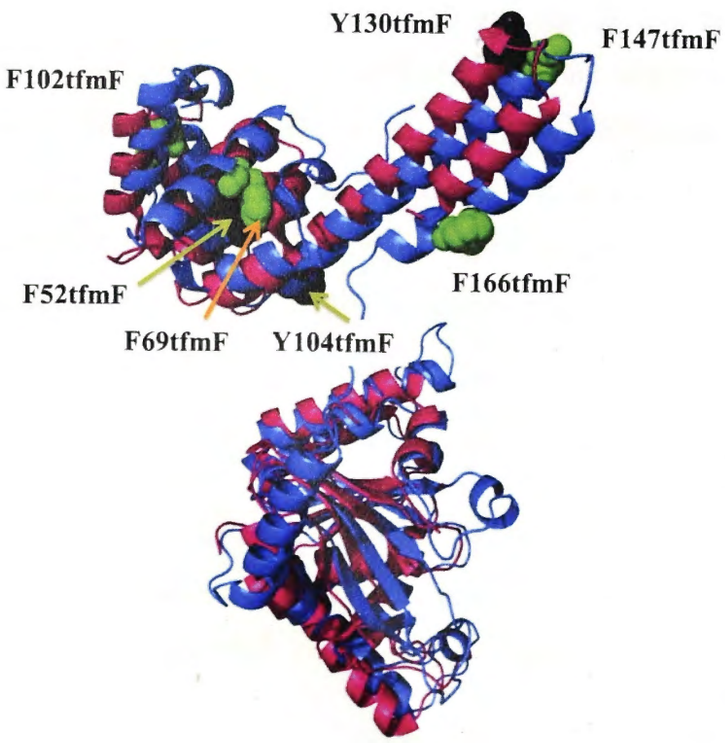

***Figure 2.2**: Structue based sequence alignment of E.coli DnaB (blue) with Gst DnaB (hotpink). The mutation sites in E.coli DnaB (green) and Gst DnaB (black) are represented with spheres.*

Reference

1. Thony, B. *et al.* Hyperphenylalaninemia with high levels of 7-biopterin is associated with mutations in the PCBD gene encoding the bifunctional protein pterin-4a-carbinolamine dehydratase and transcriptional coactivator (DCoH). *American Journal of Human Genetics* **62**, 1302-1311, (1998).
2. Taylor, R. G., Walker, D. C. & Mcinnes, R. R. *Escherichia-Coli* Host Strains Significantly Affect the Quality of Small-Scale Plasmid DNA Preparations Used for Sequencing. *Nucleic Acids Research* **21**, 1677-1678, (1993).
3. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 Rna-Polymerase to Direct Expression of Cloned Genes. *Methods in Enzymology* **185**, 60-89 (1990).
4. Neylon, C. *et al.* Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: A model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* **39**, 11989-11999, (2000).
5. Luria, S. E. & Burrous, J. W. Hybridization between Escherichia-Coli and Shigella. *Journal of Bacteriology* **74**, 461-476 (1957).
6. Reynolds, W. F. & Gottesfeld, J. M. Torsional Stress Induces an S1 Nuclease-Hypersensitive Site within the Promoter of the *Xenopus-Laevis* Oocyte-Type 5s Rna Gene. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 4018-4022, (1985).
7. Sambrook, J., Fritsch, E.F. & Maniatis, T. Molecular Cloning. A Laboratory Manual. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, New York, N.Y. (1989)
8. Miller, E.M. & Nickoloff, J.A. *Escherichia coli* electrotransformation. *Methods Mol. Biol.* 47, 105–113, (1995).
9. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685, (1970).
10. Gill, S.C. & von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326, (1989).

CHAPTER 3

^{19}F NMR study on the symmetry of the *E. coli* DnaB helicase

3.1. Introduction

In the bacterial DNA replication process, DNA polymerase requires single stranded DNA (ssDNA) as template for replication. Unwinding of the duplex DNA into ssDNA is performed by DnaB helicase, which is the primary helicase of the process[1,2]. DnaB helicase is involved in initiation and elongation stages of bacterial DNA replication[2]. The multifunctional DnaB helicase has demonstrated to increase the activity of DnaG primase, to hydrolyse ATP as the energy source for unwinding and translocation, to bind nucleotide triphosphate and its hydrolysis for unwinding [3-7]. Its interaction with other proteins extends to τ subunit of the DNA polymerase III holoenzyme and the replication terminator protein RTP [8,9]. Due to its involvement with other proteins and functional versatility, DnaB helicase from various organisms has been structurally and functionally characterised for the past two decades [3-9]. As it is practically impossible to study replication helicase from each and every bacterium, model organisms were chosen to study replication proteins and their functions in the DNA replication. One of the well-studied organisms is *Escherichia coli (E. coli)* because it's a well-characterised Gram negative bacterium. Its replication proteins were easy to express and could be assembled *in vitro*.

E. coli DnaB is a 52 kDa protein as a monomer containing two subunits of N-terminal domain with molecular weight of 12 kDa and a 33 kDa C-terminal domain connected by a 5 kDa helix linker [10]. The functional *E. coli* DnaB helicase was reported to be a 315 kDa hexamer which has been confirmed by electron microscopy (EM) and mass spectrometry (MS) studies [1, 11-15]. To date there is no well-resolved crystal or NMR structure of the full-length *E. coli* DnaB helicase. However, the well-resolved structure of isolated N-terminal domain was determined by both NMR [16] and X-ray crystallography

methods giving insight into some structural characteristics [17]. The studies have shown that the N-terminal domain has the tendency to dimerise [16], which led to hypothesise that DnaB could be arranged as trimer of dimers with C3 symmetry and N-terminal domain is essential for oligomerisation.

The structural information on full-length DnaB oligomers and $DnaB_6$-$DnaC_6$ complex is limited to low resolution EM images. Initial EM images revealed that DnaB helicase adopts threefold (C3) symmetry [13]. In contrast, later EM study carried out under same conditions by Yu and co-worker reported that DnaB helicase existed in both C3 symmetry and six fold (C6) symmetry in equal population. The latter study showed evidence of inter-conversion between two symmetrical states and reported that N-terminal domain may act like a switch between the two symmetrical states [12]. The existence of two symmetries was further supported by EM study conducted by Donate LE, which revealed pH to be the controlling factor of different quaternary states of DnaB helicase. It demonstrated that at basic pH DnaB helicase was predominantly in C3 symmetry, at neutral pH it existed in both symmetries of equal proportion [11]. Similar phenomenon of quaternary polymorphism was also observed in other helicases such as helicase of *B. subtilis* bacteriophage SPP1 G40P [19]. Although many studies have reported the factors responsible for quaternary polymorphism of *E.coli* DnaB are reported, the functional significance of each quaternary state, are still to be explored.

Some EM studies have demonstrated that interaction with other proteins influence the conformational changes [14,18,19]. DnaC, is a loading partner for *E.coli* DnaB induces conformational change[18]. The studies established that DnaC locks $DnaB_6$ into C3 symmetry by binding to C-terminal domain of the DnaB helicase [14,18,19]. However, it was evident DnaC and the C-terminal domain of DnaB adopted C6 symmetry eventhough DnaB adopted C3 symmetry. These studies indicated two important aspects of *E.coli* DnaB structure, one is that N-terminal domain defines the symmetry of DnaB and second is the fuctional significance of different quaternary states of DnaB necessary at various stages of DNA replication.

These aspects could be observed from DnaB helicases of other organisms, because recently, the high–resolution crystal structure of full-length *Geobacillus*

stearothermophilus DnaB helicase in complex with primase showed that DnaB adopts C3 symmetry in complex with primase with N-terminal domain of the helicase in a triangular arrangement and the C-terminal domain assuming a C6 symmetry [20]. This is further supported by another crystal structure of *Gst* DnaB in complex with DNA, which showed that the N-terminal domain is arranged in right handed helical shape rather than flat as observed in crystal of *Gst* DnaB-DnaG complex. From these researches, it is quite evident that the N-terminal domain defines the symmetry of the helicase as the C-terminal domain always assumes C6 symmetry [21].

The *E. coli* DnaB samples for EM studies were prepared by negative staining of the protein and by frozen hydration[1,12-15], in each preparation the protein was not in its physiological state in solution. To date there is very limited information about conformations of DnaB helicase in solution, which only reveals the conditions under the conformational change is triggered but not the specific conformation it assumes [1,15,23,24].

One of the promising paths to study DnaB helicase in solution is NMR spectroscopy. However, conventional NMR methods are not feasible because of the large size of DnaB and labelling the protein with traditional isotope spin nuclei, which is usually ^{13}C, ^{15}N and ^{1}H will result in spectral over-crowding because amino acids are composed of these elements. This limitation can be overcome by decreasing the density of NMR active nuclei in the protein for easier analysis. One of the isotopic spin nuclei, which occurs rarely in biological systems is ^{19}F [25].

It has been previously shown that ^{19}F NMR is a good tool to study large proteins because of good signals observed in the absence of background signals in 1D spectrum. The fluorine nucleus is very sensitive to any change in the surrounding environment [25,26]. From the EM images indicating two conformations, we know that there is shift in the orientation of the subdomain when the quaternary states of DnaB helicase are interchanged. Exploiting the hypersensitivity of fluorine nucleus we hypothesised that conformational changes in DnaB can be observed. We hypothesised that when DnaB adopts C3 symmetry, a particular amino acid may be surrounded by two different chemical environments and surrounded by only one chemical environment when it

assumes C6 symmetry. This study would be one of the first studies to structurally characterise such a large protein and its complex using ^{19}F NMR.

In this chapter it is described how DnaB helicase was site-specifically labelled with trifluoromethyl phenylalanine (tfmF), a fluorine probe. The labelling involves replacing a particular amino acid with tfmF by site-directed mutagenesis. The main objective is to observe the changes in ^{19}F chemical shift and line width under different pH and in complex with DnaC. ^{19}F NMR proved to be a very effective tool to study the conformational changes in DnaB helicase under different conditions and interaction with other proteins.

3.2. Materials and Method

3.2.1. Site- Directed Mutagenesis

Amber codons were introduced into DnaB helicase at the site of interest by site–directed mutagenesis using PCR. Phenylalanine residues were targeted for mutagenesis. Site-directed mutagenesis was performed by following the principle of PCR overlap extension. The first PCR set up comprises two different PCR reactions with different primer sets for obtaining products with overlapping fragments. The first PCR reaction uses the sequence of T7 promoter as forward primer and the reverse primer is a 30 nucleotide sequence of *E. coli* DnaB with TAG codon replacing the codon of amino acid of interest. The second PCR reaction has the 30 nucleotide sequence complementary of *E. coli* DnaB with TAG codon as the forward primer and T7 termination sequence as reverse primer. It gives two fragments with TAG and 10 overlapping nucleotides. The fragments of correct size were gel purified and further PCR amplified using T7 promoter and terminator primers [27]. The final PCR product with the appropriate mutations was restricted digested using *Eco*RI and *Nde*I as restriction endonucleases. The fragments with correct size were ligated into vector pETMCSIII [28]. The ligated mix was transformed into DH5α by electroporation and confirmed by sequencing. The mutants and site of incorporation is given in table 1.

Table 3.1: Primer sequences

Mutation	Primers with 'TAG' codon at site
F69TfmF	Forward – 5'CACACCGTCATATC**TAG**ACTGAAATGGCG3'
	Reverse – 5'CGCGCCATTTCAGT**CTA**GATATGACGGTGTG3'
F102TfmF	Forward – 5'GCGTCGGTGGT**TAG**GCTTATCTGCAG3'
	Reverse – 5'CTCTGCCAGATAAGC**CTA**ACCACCGACGC
F147TfmF	Forward – 5'GATTGCCGAAGCTGGT**TAG**GATCCGCAGGG3'
	Reverse – 5'CGCCCCTGCGGATC**CTA**ACCAGCTTCGGC3'
F166TfmF	Forward – 5'GAATCCCGCGTC**TAG**AAATTGCCG3'
	Reverse – 5'CTTTCGGCAATTTT**CTA**GACGCGGG3'

The mutations are indicated in **bold**

3.2.2. Expression of DnaB helicase mutants - Cell-free reactions

DnaB helicase mutants were expressed by using *E. coli* cell-free continuous exchange protein synthesis, as described by Ozawa and Kigawa [27,29]. The cell-free reaction is set up in volume of 4 ml. The DNA template which was either linear PCR product or the plasmid, the orthogonal system of suppressor tRNA/purified tRNA synthetase and DNA template was added in the inner buffer along with 20 natural amino acids. The inner buffer contains 1 mM of tfmF, 15 mM of 20 amino acids, 375 µg/ml of $tRNA_{CUA}$ and 40 µM of pCNF-RS along with ATP, folinic acid, rNTPs and DTT. The linear PCR amplified DNA was used as template, cell-free reactions require 10 ng/µl of the DNA template and the concentration of plasmid DNA used as DNA template in cell-free reactions is 16 ng/µl [27,29,30].

The outer buffer is 10x the volume of the inner buffer and contains the 20 amino acids and tfmF of 1 mM concentration. The inner buffer is added inside a membrane with molecular weight cut-off 12000-14000 Da (Spectrapor dialysis tubing).The system is set-up either in 50 mL falcon tubes for 4 mL and in 10 mL tubes for 200 µl reaction volumes. The reactions were carried out at 30 °C from 7 h to 14 h in a water-bath shaker shaking at 150 rpm [27,29,30].

3.2.3. Preparation of DNA Template

The preparation of DNA template was based on the scale of cell-free reaction. For small scale reactions of 200 µl, linear PCR amplified DNA was used, for 4 ml reactions plasmid DNA was used.

3.2.3.1. Linear PCR amplified DNA

The linear PCR amplified DNA involves no restriction digestion, ligation into vector and transformation is required. The method involves using the second round PCR product as the DNA template [31]. The concentration of 30 ng of DNA template is added to the PCR mix using 1131 and 1134 as one set of primers and another set of primers is 1132 and 1133. The PCR product was combined and denatured at 98 °C for 5 min and reannealing at room temperature for 30 min. The DNA has two single-stranded 8 nucleotide overhangs, complimentary and aids in cyclisation of the strands during cell-free reaction using ligase of the cell extract [31,32]. The yield is calculated using nanodrop ND-1000 spectrophotometer. Table 2: Primer sequences used in linear PCR amplified DNA

Primer	Sequences
1131	5'-PO_4-TTAGCTGGTCGATCCCGCGAAATTAATACG-3' (30-mer)
1132	5'-PO_4-CCAGCTAACAAAAAACCCCTCAAGACCCG-3' (29-mer)
1133	5'-PO_4-TCGATCCCGCGAAATTAATACG-3' (22-mer)
1134	5'-PO_4-CAAAAAACCCCTCAAGACCCG-3' (21-mer)

3.2.3.2. Plasmid DNA Extraction (Maxiprep)

For plasmid DNA, Qiagen Maxiprep kit (QIAGEN N.V., Netherlands) was used. The cells were grown overnight at 37 °C in LB broth. The cells were harvested by centrifugation for 20 min at 5000*g*. The harvested cells were lysed using buffers provided in the kit. The plasmid DNA was then purified and eluted using the kit and plasmid DNA was ethanol precipitated. The concentration of the DNA is measured.

3.2.3. Overexpression and Purification of tRNA synthetase

tRNA synthetase used in this project was *p*-cyanophenylalanine tRNA synthetase (*p*CNF-RS), which is a polyspecific tRNA synthetase capable of incorporating various UAA with phenylalanine main chain [33,34].The vector containing *p*CNF-RS/pEVOL was given by Prof. Peter G. Schultz (Scripps Research Institute, CA, USA) [34,35]. The mutant D286R mutant of *p*CNF-RS gene was transferred in to pETMCSIII vector, [28] which introduces a His_6 tag on the N-terminal side of the *p*CNF-RS. The tag aids in faster purification by metal affinity chromatography using Ni-NTA column.

Buffers for Protein Purification

The buffers used for purification are **Buffer A** : 50 mM HEPES, pH 7.5, 300 mM NaCl, 5% glycerol, 20 mM Imidazole and **Buffer B** : 50 mM HEPES, pH 7.5, 300 mM NaCl, 5% glycerol, 300 mM Imidazole. The buffers were prepared in MilliQ water and pH was adjusted accordingly using 1 M HCl. The buffers were filtered using filters 0.22 μm (Sartorius) and vacuum pump from Millipore. The filtered buffers were then stored at 4 °C. The storage buffer **Buffer C**: 50 mM HEPES, pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM DTT.

Overexpression of pCNF-RS

pET vector containing *p*CNF-RS gene was transformed into *E. coli* BL21 (DE3)*recA*$^-$ /pLysS strain. A 10 ml starter culture is prepared by inoculating the sterilised LB medium with single colony from fresh plate and incubated at 37 °C, overnight. The medium contained 100 μg/mL ampicillin and 33 μg/mL chloramphenicol. The following day, the sterilised LB broth of 1L supplemented with antibiotics, was inoculated with 10 ml starter culture and incubated at 37 °C. The growth rate was monitored by measuring optical density of the medium. The cells were grown on shaking till $OD_{600} = 0.6$ was reached. The culture was induced with 1 mM IPTG (final concentration) and cells were incubated at 37 °C for further 3 h. The cells were harvested by centrifugation (5000*g* for 20 min). The cells were freeze-dried using liquid nitrogen and were stored at -70 °C.

The subsequent day the weight of the cells was measured and cells (8 g) were thawed on ice. The thawed cells were resuspended in 40 ml Buffer A with 1 mM PMSF as protease inhibitor. The resuspended cells were lysed using French Press by passing twice through

the French press cell operating at 1200 psi. The lysed cells were collected on ice and for the purpose of separation of cell debris from the protein solution; the lysate was spun down at speed of 34,220*g* for 45 min. After centrifugation, the lysate is collected and stored on ice for purification.

Purification of pCNF-RS

FPLC system was used for purifying *p*CNF-RS using gradient elution method (section 2.4.1). The pump of the FPLC was washed with Buffer A and 5 ml capacity Ni-NTA column (GE Healthcare Life Sciences, USA) was attached to FPLC. The column was equilibrated with 40 ml of Buffer A. The column was then loaded with the lysate containing *p*CNF-RS at flow rate of 1 ml/min using peristaltic pump. After loading the column with protein, the column was washed with 20 ml of Buffer A. Taking the advantage of the FPLC automated system, the column was eluted with mixture of Buffer A and Buffer B, imidazole concentration increases constantly (20-300 mM) over time. At the appropriate concentration of imidazole, the *p*CNF-RS was eluted using 100 ml of buffers. The elutions were collected as fractions of 5 ml in fraction collector. *p*CNF-RS eluted with 150 mM concentration of imidazole.

The purity of *p*CNF-RS was observed and assessed by SDS-PAGE. The SDS-PAGE showed that the eluted fractions of *p*CNF-RS were highly pure and subsequently the fractions were collected (30 ml) and pooled together. The pooled fractions were concentrated as mentioned in section 2.4.3. The concentration was carried out until final volume of 6 ml was achieved. The 6 ml of the protein solution was dialysed against the storage Buffer C. The purification, concentration and dialysis were carried out at 4 °C. After dialysis for 2h, the concentration of protein was measured as mentioned in section 2.4.4. The yield of 191 mgs was achieved with 1.09 mM of concentration. The *p*CNF-RS was aliquoted as 1 ml fractions and stored in -70 °C until use.

3.2.4. Expression and purification of total tRNA containing suppressor $tRNA_{CUA}$

The $tRNA_{CUA}$ used in cell-free reactions is obtained as purified total tRNA. The plasmid contained optimised version of $tRNA_{CUA}$ derived from *M. jannaschii*, [12] with chloramphenicol resistance gene, termed as pKO1474 was used for expression. It has

single copy of amber suppressor tyrosyl-$tRNA_{CUA}$. The vector with $tRNA_{CUA}^{opt}$ was generated in our lab [27]. The expression and purification methodology followed is according to description [27,35,36].

3.2.5. Preparation of *E. coli* S30 extracts

E. coli BL21 star (DE3) strain has T7 RNA polymerase gene for expression, is used as the source of the cell lysate for cell-free reactions. For S30 extract, the cells were grown according to the published procedure [37]. A 20 L Z-medium (pH 7.4) supplemented with glucose, thiamine and chloramphenicol was prepared and sterilized in the fermenter. The cells were grown at 37 °C till an OD_{600}=1 was reached and was induced with 1 mM IPTG for expressing the T7 RNA polymerase. Post-induction, the cells were allowed to grow at 37 °C till the OD_{600}=3. The cells were harvested by centrifugation (10,000*g* for 12 min at 4 °C). Approximately 120 g cells were washed with 1x S30 buffer α twice and were resuspended in the 1.3 ml of the same buffer per gram of cells. The resuspended cells were then passed through French Press once at 6000 psi. The cell suspension was centrifuged at 30,000*g* for 1 h at 4 °C [32,37].

Subsequently, the lysate was dialysed against S30 buffer β and concentrated by dialysing against 50% PEG 8000 in S30 buffer α containing 1 mM DTT. After concentrating the lysate to 1/3 of its initial volume, it was dialysed against S30 buffer α with 1 mM DTT and stored in -70 °C as aliquots of 500 µl, after freeze dried. The extract is thawed just before the cell-free reactions were set up [38].

3.2.6. Co-expression of *E. coli* DnaB mutants and DnaC in cell-free protein system

Cell-free reaction has also proven to be a good system for co-expression of two different proteins. The setup was as same as above. For co-expressing DnaB mutants and DnaC, the DNA template concentration was crucial. The concentration of DNA template of DnaB mutants was 16 ng/µl and concentration of DnaC template was optimised to10 ng/µl for co-expression. For expressing the DnaB mutants, the inner buffer contains 1

mM of tfmF, 15 mM of 20 amino acids, 375 µg/ml of $tRNA_{CUA}$, 40 µM of *p*CNF-RS and 175 µg/ml of tRNA for expressing wild type DnaC and tfmF-labelled DnaB. The expression was carried out at 30 °C for 14 h.

3.3. Result and Discussion

3.3.1. Incorporation of tfmF in DnaB helicase by cell-free protein synthesis

The tfmF was successfully incorporated site-specifically into 315 kDa DnaB by *in vitro* protein synthesis. The ^{19}F resonance was observed for tfmF incorporated at different sites. The chemical shifts observed were distinctly different for all the four mutants. This ascertained that tfmF is good ^{19}F label to study the conformations in large proteins systems.

3.3.1.1. Cell-free protein synthesis

The *in vitro* system was initially tested for tfmF incorporation in 200 µl reactions. The linear PCR amplified DNA was used as template for test reactions. The SDS PAGE analysis of the tfmF labelled DnaB helicase revealed that the yield of each mutant was dependent on the site of tfmF incorporation. The fluorine labelled DnaB expressed and tested by ^{19}F NMR were mutants F69tfmF, F102tfmF, F147tfmF and F166tfmF.

The gel showed two prominent bands, one was the full length DnaB helicase with tfmF with molecular weight of 52 kDa and another band was of smaller molecular weight depending on the site of incorporation. During tfmF incorporation DnaB, the translation process is terminated at the amber stop codon producing a smaller truncated product. For example the molecular weight of truncation product of mutant F69tfmF is 9 kDa. The band intensity of full length DnaB and the truncated DnaB observed were same; suggesting that the yield of DnaB may be reduced due to formation of truncation product. The reason for the truncation is due to Release Factor 1 (RF1) in *E. coli* protein translation machinery competing with tfmF amino-acylated suppressor tRNA in identifying the TAG codon as the stop codon and results in termination of the protein

synthesis [39]. This competitive binding of RF1 results in lower yield of the full-length tfmF labelled DnaB helicase.

On comparison of the protein yields of various mutants, mutants F69tfmF and F166tfmF had relatively higher and similar levels of expression whereas the yield of mutant F102tfmF was lesser with mutant F147tfmF having the least expression. Although the level of expression varied among the mutants, there was 95% incorporation of tfmF in full length DnaB. This was confirmed by the test expression of mutants in the absence of tfmF in the cell-free reaction, which had no expression of full-length DnaB.

tfmF labelled DnaB protein was prone to precipitation. The most stable protein was mutant F69tfmF however it precipitated during concentration of the protein to 50 μM. Mutants F166tfmF and F147tfmF had higher tendency to precipitate and majority of the protein precipitated after purification. For mutant, F102tfmF, most of the protein precipitated during cell-free expression. To obtain maximum amount of protein in solution the duration of the cell-free protein synthesis was optimised.

3.3.1.2. Optimisation of the duration of expression

The reactions were carried out at 30 °C for 14 h and proteins of all the mutants except for mutant F69tfmF showed significant level of precipitation after 14 h cell-free reaction. To minimize precipitation during the expression and obtain higher yields of soluble protein, the duration of the cell-free reaction was optimised. A time course experiment was conducted on mutant F102tfmF to access the level of expression and precipitation after every 2 h. The small volume of reaction mix was collected and separated into precipitate and the soluble protein, to be later analysed by SDS-PAGE.

The presence of maximum amount of the soluble protein was considered as the main criterion to select the optimal duration of the reaction. The SDS-PAGE analysis of the samples showed that mutant F102tfmF expression increased over 12 h of expression; nevertheless there was also considerable increase in the amount of precipitation. The samples collected on and after 6 hours had the same amount of soluble protein, but the amount of precipitate increased with time. Hence, 6 h was chosen as the optimum

duration for cell free reactions because the amount of soluble protein did not increase in time and it had relatively lesser amount of precipitate when compared to samples collected after 6 h. For expressing tfmF labelled DnaB for ^{19}F NMR measurements, the volumes of cell-free reactions were up scaled to 4 ml.

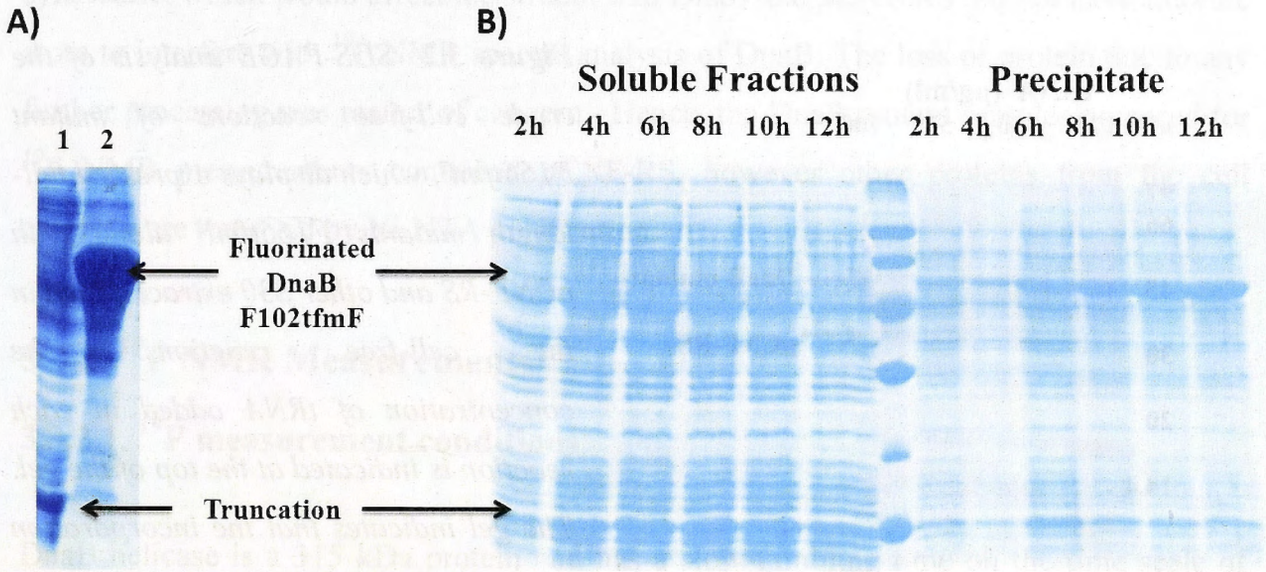

***Figure 3.1**: A) SDS PAGE gel displaying the expression of mutant F102tfmF crude cell-free reaction after being separated into supernatant and precipitate. Lane 1: Supernatant of crude showing the soluble fraction of the truncation; Lane 2: Precipitate of the crude showing the thick band of mutant F102tfmF. B) The gel displays the soluble and insoluble fractions of the crude cell-free reaction collected after every two hours for 12 h. In both figures the full length fluorinated DnaB and truncated product are indicated.*

3.3.1.3. Optimisation of suppressor tRNA

According to previous protocols used in our lab, the optimal concentration of suppressor tRNA was 350 µg/ml. However, the suppressor tRNA concentration was optimized by varying the concentration of tRNA between 175 to 700 µg/ml for tfmF labelled DnaB. F166tfmF was the mutant used for optimising the concentration of tRNA and protein was soluble after expression. The SDS-PAGE analysis showed that the expression increased with increasing concentration of supressor tRNA in the reactions. So far, 350 µg/ml was used because there was no further increase in the expression levels of protein with other

unnatural amino acids but in case of tfmF labelled DnaB, the expression levels considerably increased. 700 µg/ml of suppressor tRNA was chosen as optimum concentration for further cell-free reactions as a compromise to obtain good protein expression yield but not consuming large amounts of suppressor tRNA, which is time-consuming to produce.

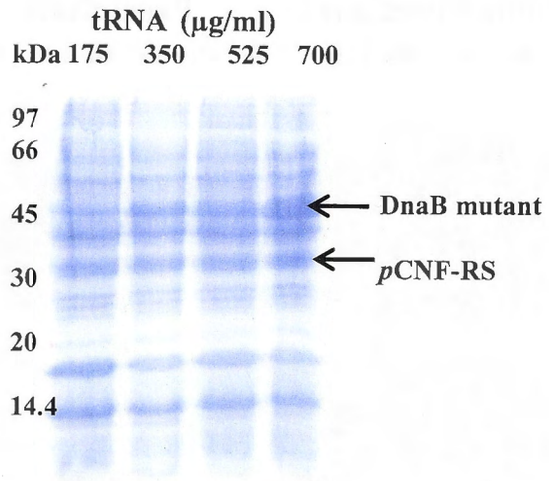

***Figure 3.2**: SDS-PAGE analysis of the crude cell-free reactions of mutant F166tfmF, which displays expressed full-length mutants F166tfmF along with pCNF-RS and other S30 extract added in the cell-free reactions. The concentration of tRNA added in each reaction is indicated at the top of the gel. The gel indicates that the incorporation of tfmF in DnaB helicase depends on the concentration of tRNA.*

3.3.1.4. Purification of DnaB and DnaB/DnaC complex

Fluorine labelled DnaB mutants were purified using 1 ml Ni-NTA columns with buffer A and buffer B. The purified samples had full length fluorine labelled DnaB, *p*CNF-RS and the truncated product, which had N-terminal His_6 tag. As a result of low molecular weight of the product, most of the truncated product was removed during the protein concentration using concentrators and the sample measured by ^{19}F NMR contained both fluorine labelled DnaB, *p*CNF-RS and relatively lower concentration of truncation product.

DnaB/DnaC complex were purified in a similar manner. In case of DnaC, the gene was not tagged with purification tag. Hence, it was purified as a complex with DnaB. For the formation of the complex, the purification buffer A and B were supplemented with 10 mM $MgCl_2$ and 1 mM of ATP to stimulate and enhance the complex formation. It was

very evident from SDSPAGE, the complex was stable and DnaC eluted along with DnaB and *p*CNF-RS.

Methods such as gel filtration, sepharose or DEAE column were not employed to remove the *p*CNF-RS because there was no evidence of interactions between DnaB and synthetase, which would affect the structure of DnaB and *p*CNF-RS did not have fluorine atom to interfere with ^{19}F NMR spectra analysis of DnaB. The loss of protein due to any further processing was matter of concern. Hence, the DnaB mutant samples prepared for ^{19}F NMR measurements contained *p*CNF-RS, however other proteins from the cell extract were removed by Ni-NTA purification.

3.3.2. ^{19}F NMR Measurements

3.3.2.1. ^{19}F measurement conditions

DnaB helicase is a 315 kDa protein and has a slow tumbling time on the time scale of NMR experiments, which results in broad and un-detectable signals. The problem can be alleviated by incorporating an unnatural amino acid with fast intrinsic spin motion and increasing the temperature of measurement. TfmF was chosen as the unnatural amino acid because it has a trifluoromethyl group with fast rotation around the methyl axis to give relatively sharper signals. The NMR measurements were carried out at 25 °C for allowing the reasonable tumbling and obtain resonances with good resolution. The temperature could not be raised any further because *E. coli* DnaB helicase precipitated at higher temperature. Other alternatives to increase signal to noise of NMR measurement were to increase concentration of the protein or increase the total number of measurements.

The concentrations of the DnaB were in range of 10 – 30 μM depending on the site of tfmF incorporation. The concentration mentioned above falls in the range considered as lower end of the concentrations required for ^{19}F NMR measurements, and to obtain good signals relative to background noise samples were measured for 12 h at 25 °C. TfmF labelled DnaB completely precipitated after the measurement. The samples were collected before and after NMR measurement to check the solubility of the protein by

SDS-PAGE. SDS-PAGE analysis of the precipitate showed the presence of fragments of lesser molecular weight indicating degradation of the DnaB, and consequently. Hence, the duration of measurements was reduced to 4 h.

As the ^{19}F resonance reflects the local interactions of fluorine atoms with the neighbouring residues, any significant change in the environment can be observed in ^{19}F NMR. The first objective of the project to successfully incorporation of tfmF by cell-free reactions using *p*CNF-RS was achieved.

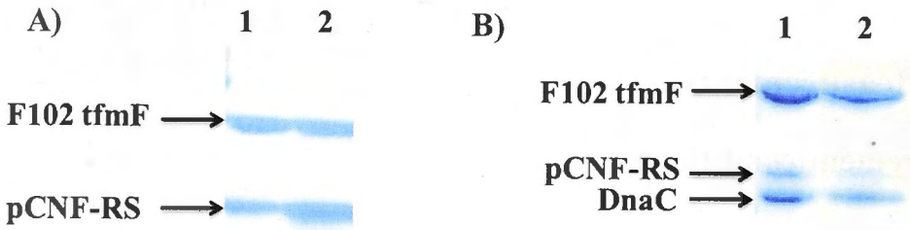

Figure 3.3: A) SDS-PAGE depicting the sample of F102tfmF that was measured by ^{19}F NMR. Lane 1 and 2 shows the bands of the F102tfmF DnaB helicase before and after the 4 h of ^{19}F measurement respectively. The second band of pCNF-RS is present due to the presence of His_6 tag. B) SDS-PAGE depicting the samples of F102tfmFDnaB in complex with DnaC. Lane 1 and 2 shows the bands of the complex before and after ^{19}F NMR measurement.

3.3.2.2 Effect of pH on ^{19}F chemical shift of free tfmF

To eliminate possible pH dependent effects on tfmF amino acid on the ^{19}F chemical shifts of tfmF labelled DnaB, control experiments were conducted on the free tfmF amino acid. Any pH dependent changes in chemical shifts of free tfmF were to be used for re-calibration of the chemical shift changes observed for fluorinated DnaB mutants. ^{19}F NMR of the free tfmF amino acid was measured at different pHs, which are between 6.4 and 8.1. ^{19}F NMR of tfmF amino acid was recorded in 128 scans. The chemical shifts of tfmF amino acid were between -61.12 and -61.14 ppm and did not depend on pH of the

solution. The average of the chemical shift was -62.13 ppm and this value was used as the reference for all further experiments conducted on fluorinated DnaB mutants. Figure 3.4 shows that ^{19}F chemical shift changes less than 0.02 ppm over the pH range from 6.5 to 8.1, a not significant effect on the chemical shifts of ^{19}F spectra and no further re-calibration was necessary.

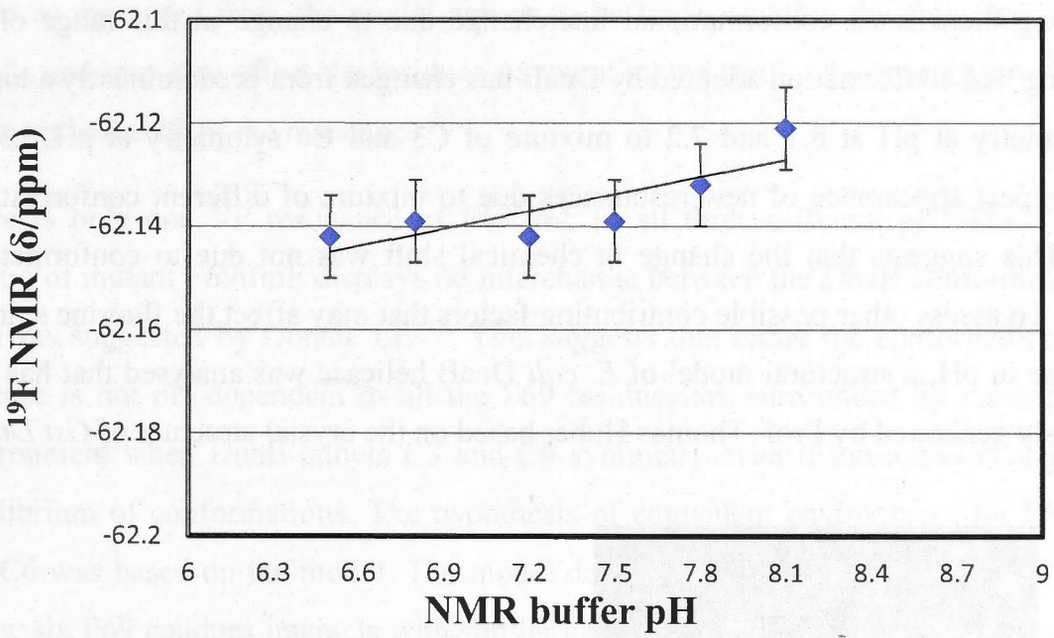

***Figure 3.4**: Comparison of ^{19}F chemical shifts observed for free-tfmF amino acid at different pH of NMR buffer. The vertical axes displays the ^{19}F chemical shifts in the units of ppm and the horizontal axes displays the pH values. The error is 0.02 ppm is depicted by error bar and display that there is pH does not affect the tfmF amino acid significantly*

3.3.3. ^{19}F NMR to study the effect of pH on fluorinated DnaB mutants

3.3.3.1. Mutant F69tfmF

The residue F69 is at the beginning of N-terminus of the DnaB helicase and mutant F69tfmF protein was the most stable protein when compared to other mutant proteins. The ^{19}F resonance of mutant F69tfmF was measured at pH 8.1, 7.2 and 6.5 with concentration of 30 μM as a monomer.

The line width of the ^{19}F resonances was 20 Hz for all pHs measured. The ^{19}F chemical shift was identical at pH 8.1 and 7.2 with δ = -61.5 ppm. At pH 6.5, there was upfield shift of 0.1 ppm in the ^{19}F resonance (Fig 3.6). According to Donate LE and coworkers at pH 6.5 and 7.2, DnaB helicase adopts C6 symmetry and C3 symmetry in equal proportion and at pH 8, it assumes predominantly C3 symmetry[13]. There is no apparent change in chemical shift or line width of the ^{19}F resonances between pH 8.1 and pH 7.2, suggesting there is no conformational interchange due to change in this range of pH. Assuming that conformation adopted by DnaB has changed from predominantly adopting C3 symmetry at pH at 8.1 and 7.2 to mixture of C3 and C6 symmetry at pH 6.5 one would expect appearance of new resonances due to mixture of different conformational states. This suggests that the change in chemical shift was not due to conformational change. To assess other possible contributing factors that may affect the fluorine spin due to change in pH, a structural model of *E. coli* DnaB helicase was analysed that has been previously generated by Prof. Thomas Huber based on the crystal structure of *Gst DnaB*.

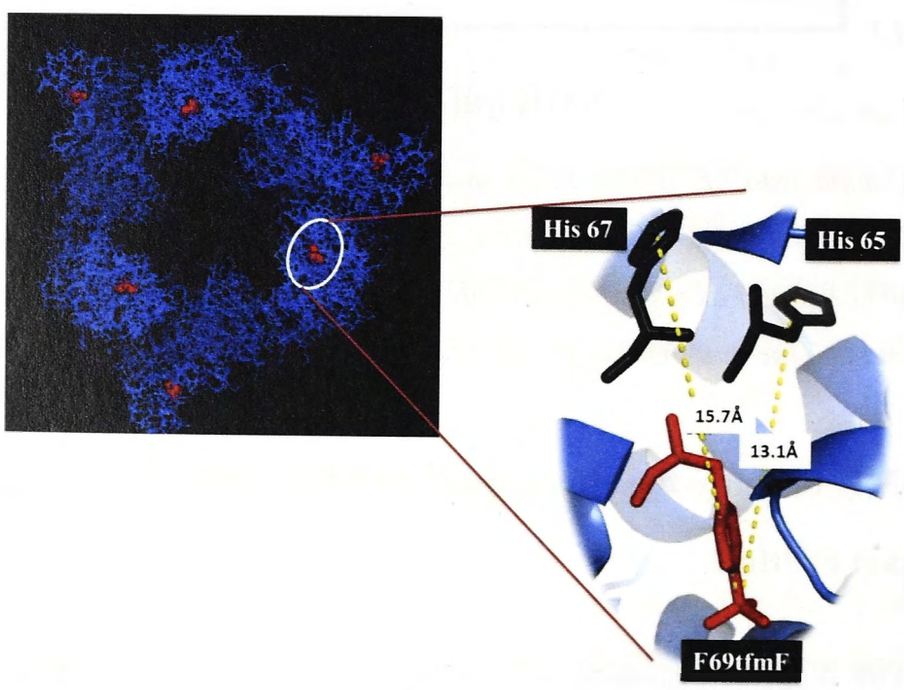

***Figure 3.5**: Line representation of DnaB helicase with red spheres highlighting the position of F69in C3 symmetry. The position of tfmF at F69 (red sticks) and surrounding Histidines 65 (black stick) and 67 (black stick) are shown. The distance of imidazole ring of histidine residue to the trifluromethyl group is specified to show the proximity of the residues in helix.*

The molecular model revealed two histidine residues near the residue F69. The change in protonation states of the histidine residues may be responsible for the change in ^{19}F resonance. As the pKa value of histidine is 6.8, when the pH was reduced below pH 6.8, the changes in protonation states of His65 and His67 which are at distance of 13.1 Å and 15.7 Å respectively, would have affected the electrostatic environment of tfmF (Fig 3.5). Although the distance between side chains of the histidine residues and trifluoromethyl group as measured from the model appear to be large, solution the side chains can be mobile and may also affect the residues surrounding the tfmF, which could have resulted in the upfield shift of the resonance [26].

There is only one ^{19}F resonance of F69tfmF at all three different pH. The ^{19}F NMR spectra of mutant F69tfmF displays no interchange between the DnaB conformations due to pH as suggested by Donate LE [13]. This suggests that either the conformations of the helicase is not pH dependent or all the F69 residues are surrounded by same chemical environment when DnaB adopts C3 and C6 symmetry even if there was change in the equilibrium of conformations. The hypothesis of equivalent environment for F69 in C3 and C6 was based on the model. The model depicts DnaB in its C3 symmetry state and all the six F69 residues interacts with similar neighbouring residues (Histidines) resulting in same chemical environment and in C6 symmetry the six F69 have equivalent environment.

The line width of resonances was considered quite narrow for 315 kDa protein. The possible reasons for narrower resonances were either due to the position of residue F69 or the protein is present as monomer or dimer. The sample was prepared in the presence of $MgCl_2$ and ATP and the further experiments confirmed that F69tfmF formed a stable complex with DnaC. DnaC forms complex with DnaB only when it exists as a hexamer. In literature, it is been established that the DnaB helicase form stable hexamer at lower concentration and in the presence of $MgCl_2$ and ATP [10-14]. Subsequently, the location of F69 was examined and it is revealed that it is placed towards the N-terminus and buried inside helical bundle. The narrower signal suggests flexibility of the structure holding it, which may be more flexible than the remaining part of the protein resulting in sharper resonance due to faster tumbling.

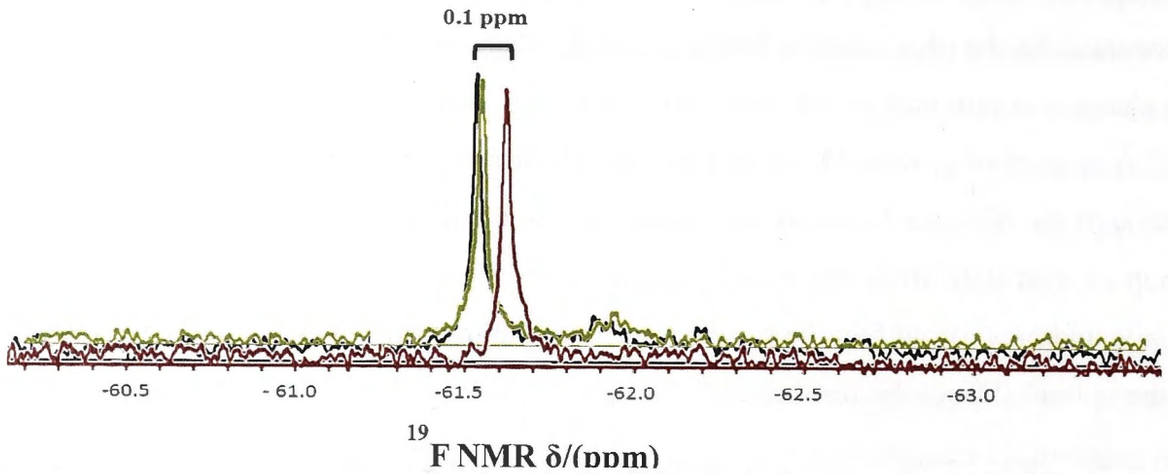

***Figure 3.6**: [19]F NMR spectra of trifuorophenylalanine (tfmF) labeled DnaB helicase placed at F69. The line width of all [19]F spectra measured at pH8.0 (black), pH7.2 (green) and pH6.5 (red) is 20 Hz. There is 0.1 ppm upfield shift of [19]F resonance, when measured at pH6.5.*

3.3.3.2. Mutant F102tfmF

It has been reported that residue F102 is located near the dimer interface of N-terminal domain [17] and is important for dimerisation of the N-terminal domains [40]. It is of particular interest because the molecular model of *E. coli* DnaB helicase displays that F102 has two different chemical environments, when DnaB adopts C3 symmetry. We hypothesised that subsequently there should be two distinct ^{19}F resonances representing each chemical environment. Mutant F102tfmF was unstable in buffer with pH 7.2 and 6.5 and could be measured by ^{19}F NMR only at pH 8.1.

The concentration of mutant F102tfmF was 17 μM. The ^{19}F NMR spectrum displayed presence of two resonances of different intensities. One was a sharper resonance at -62.24 ppm with line width 36 Hz and another was a minor peak at -61.9 ppm with line width of 52 Hz. The two resonances either represent different local environment or presence of native and denatured state, since F102tfmF is prone to precipitation. However, the latter hypothesis is rejected because the denatured state precipitates immediately and denatured soluble F102tfmF in solution would be low effective concentration to produce a signal.

Therefore, the minor resonance cannot be attributed by the denatured state of F102tfmF rather represents the different local environments.

The hypothesis of two different chemical environments surrounding residue F102 in concurrence with the molecular model suggesting that DnaB is predominantly adopting C3 symmetry and consistent with previous work displaying the DnaB is predominantly in C3 symmetry at pH 8 [13].

The ^{19}F resonance with different intensities may suggest that at pH 8.0, there might be small proportion of DnaB adopting C6 symmetry, in which the mutant F102tfmF has same chemical shift as the ^{19}F resonance at -62.24 ppm resulting in a sharper signal with higher intensity. The sharpness of the signal indicates that the solvent exposed F102 residue has faster tumbling.

The line width of the two ^{19}F resonances are comparatively larger than the ^{19}F resonance of mutant F69tfmF, however the broadening of the resonances confirms the hypothesis that the N-terminus is flexible.

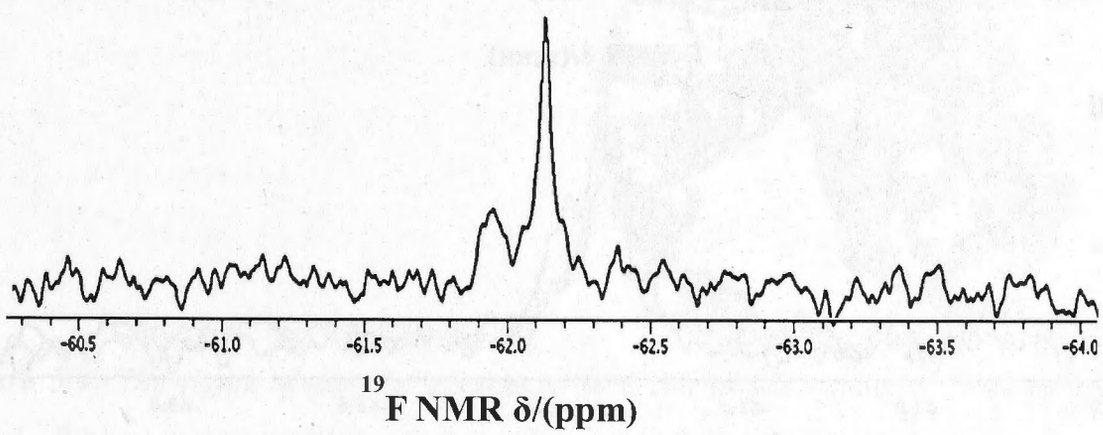

***Figure 3.7**: ^{19}F NMR of mutant F102tfmF in buffer pH 8.1. Line broadening was set as 10 Hz and measured with 10% D_2O as the solvent. The spectrum displays two peaks of different intensities. The chemical shift of narrower resonance is -62.24 ppm and broader resonance is -61.9 ppm.*

3.3.3.3. Mutants F147tfmF and F166tfmF

The residues F147 and F166 are in the helix turn helix motif of the N-terminal domain of DnaB. Both mutants were measured at pH 8.1. The protein of mutant F147tfmF precipitated during ^{19}F NMR measurement and could not be measured at pH 7.2 and 6.5. The mutant F166tfmF protein was relatively stable and was measured at all the three pHs.

The concentrations of mutants F147tfmF and F166tfmF were 10 μM and 30 μM, respectively. The chemical shifts of the ^{19}F resonances of both the mutants were similar at pH 8.1. However, the line width was different. For mutant F147tfmF, there was one resonance at -61.84 ppm with line width of 60 Hz at pH 8.1.

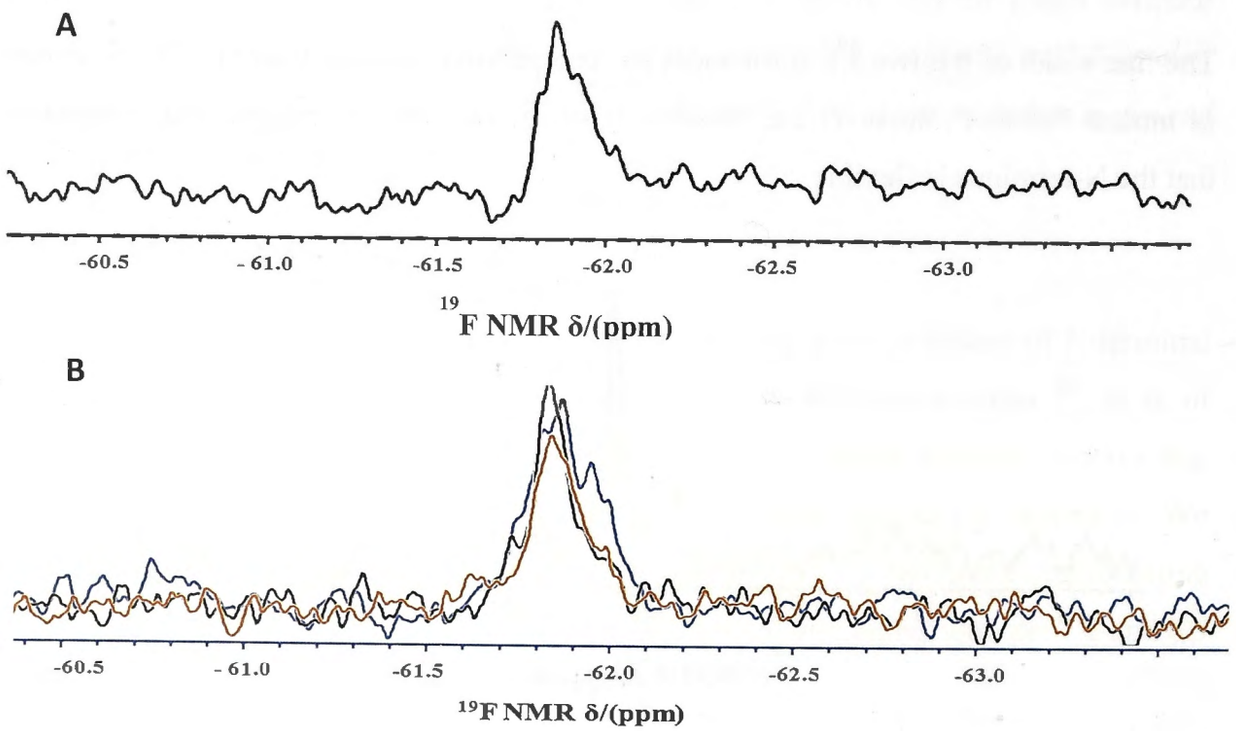

***Figure 3.8**: ^{19}F NMR spectra of F147tfmF and F166tfmF DnaB. A) 1D ^{19}F spectra of F147tfmF in pH 8.1 depicts the single peak observed at -61.8 ppm. The line width of the peak is 70 Hz. B) 1D ^{19}F spectra of F166tfmF displaying the overlay of the spectrum acquired in different pHs. The ^{19}F NMR spectra at pH 8 (blue) depicts the ^{19}F resonance split into two peaks and line width of the peak is 100 Hz, whereas the ^{19}F spectra measured in pH 7.2 and 6.5 (black and yellow), the line width of the peak is 70 Hz.*

For mutant F166tfmF, at pH 7.2 and 6.5, there was a resonance with chemical shift at -61.8 ppm, with line width of 70 Hz suggesting that structure of helix turn helix is well ordered. At pH 8.1, the chemical shift of resonance was at δ = -61.82 ppm similar to other spectra at pH 7.2 and pH 6.5, but the resonance was broader with split and had line width of 100 Hz. The smaller split resonance is at -62 ppm (Fig 3.8).

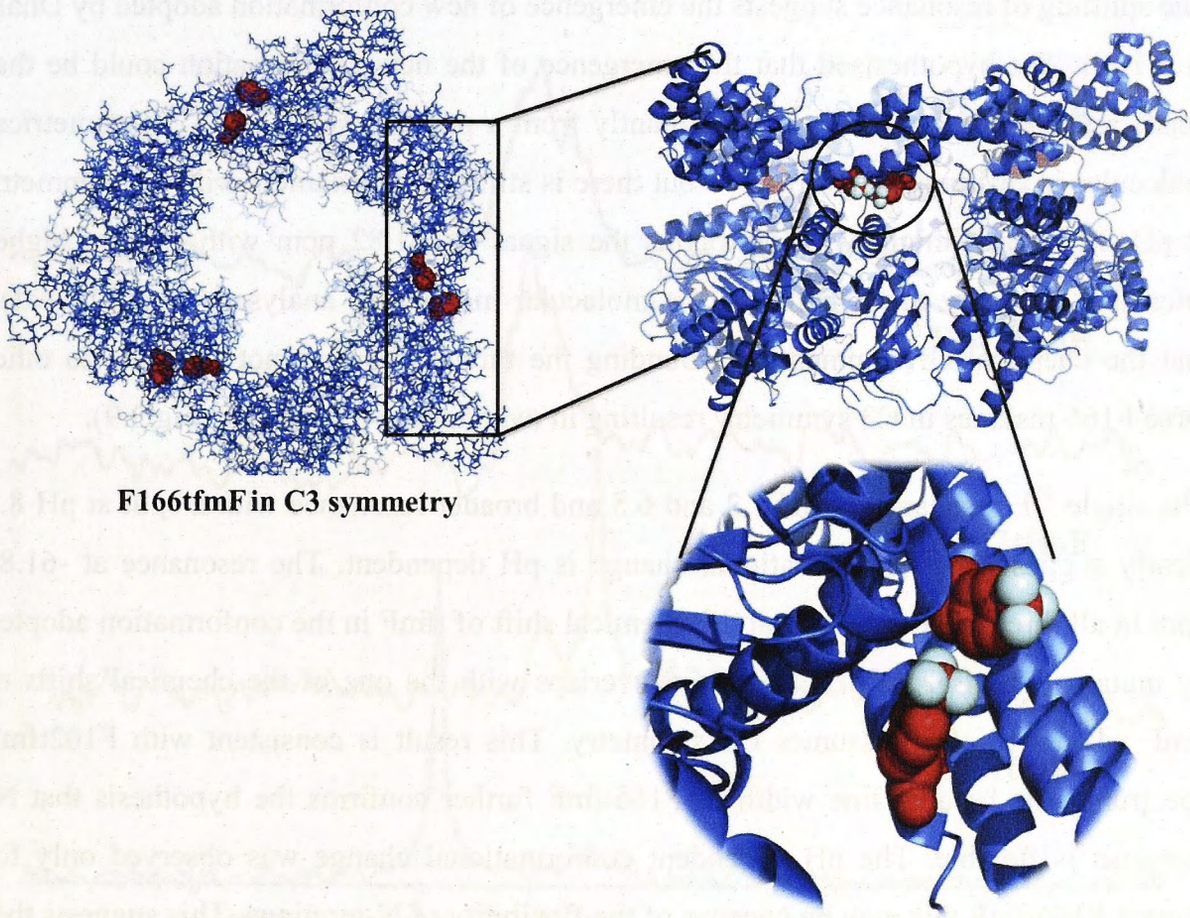

***Figure 3.9**: The figure shows the placement of F166 in C3 symmetry from molecular model. Ribbon representation of the side view of DnaB helicase shows the tfmF incorporated at 166 as red spheres with trifluoromethyl group. The zoomed version is displaying the close proximity between the residues and nearly similar chemical environment surrounding each residue*

The split of resonance was considered as two signals rather than single broad peak because at lower pHs, the line shape of resonances clearly shows single distinct peak whereas at pH 8, there is small peak with peak maxima at -62 ppm. However, the peak is considered as single broad peak, then the signal must have broadened on either sides with

peak maxima at -61.8 ppm rather than having two peak maxima separated with resolution of 0.2 ppm. Even though, the intensity of the smaller resonance is less than the major resonance; the peak maxima are well resolved. Hence, it was considered as two overlapping resonances.

The split of resonance displays two broad resonances having nearly similar chemical shifts overlapping on one another resulting in one broad signal with two peak maxima. The splitting of resonance suggests the emergence of new conformation adopted by DnaB at pH 8.1. We hypothesised that the emergence of the new conformation could be that DnaB assumes C3 symmetry predominantly from a population C6 and C3 symmetrical molecules in different proportions [13] but there is still DnaB hexamers with C6 symmetry at pH 8.1, which might which explains the signal at -61.82 ppm with slightly higher intensity. To verify the hypothesis, the molecular model was analysed, which revealed that the chemical environments surrounding the three F166 was not identical to other three F166 residues in C3 symmetry resulting in overlapping resonances (Fig 3.9).

The single ^{19}F resonance at pH 7.2 and 6.5 and broader resonance with a split at pH 8.1 clearly suggests that conformational change is pH dependent. The resonance at -61.82 ppm in all three pH indicates that the chemical shift of tfmF in the conformation adopted by mutant F166tfmF at pH7.2 and 6.5 overlaps with the one of the chemical shifts of tfmF when F166tfmF assumes C3 symmetry. This result is consistent with F102tfmF spectrum. The broader line width of F166tfmF further confirms the hypothesis that N-terminus is flexible. The pH dependent conformational change was observed only for mutant F166tfmF this may be because of the flexibility of N-terminus. This suggests that structure of helix turn helix is well ordered.

3.3.3.4. Comparing the ^{19}F resonance of different mutants

^{19}F resonance was observed for all the mutants displaying the successful incorporation of tfmF into DnaB. The ^{19}F resonance incorporated at each position is different from the other demonstrates the successful incorporation of tfmF site-specifically.

By comparing the chemical shifts of resonance of each mutant with free tfmF amino acid (δ= -62.13 ppm) and approximate location of the residue could be identified. The ^{19}F

resonance of mutant F69tfmF is at -61.5 ppm, suggesting that the residue is buried because the chemical shift is not closer to that of free tfmF. The resonances of F102tfmF (δ= -61.9 ppm and -62.2 ppm), resonances of F147tfmF (δ= -61.78 ppm), and F166tfmF (δ= -61.82 ppm and -62 ppm), suggests that these residues are closer to solvent exposed surface. The resonances of mutant F102tfmF indicates that some of F102 residues is more solvent exposed than the other in a hexamer and the deduction concurs with placement F102 residue in the model.

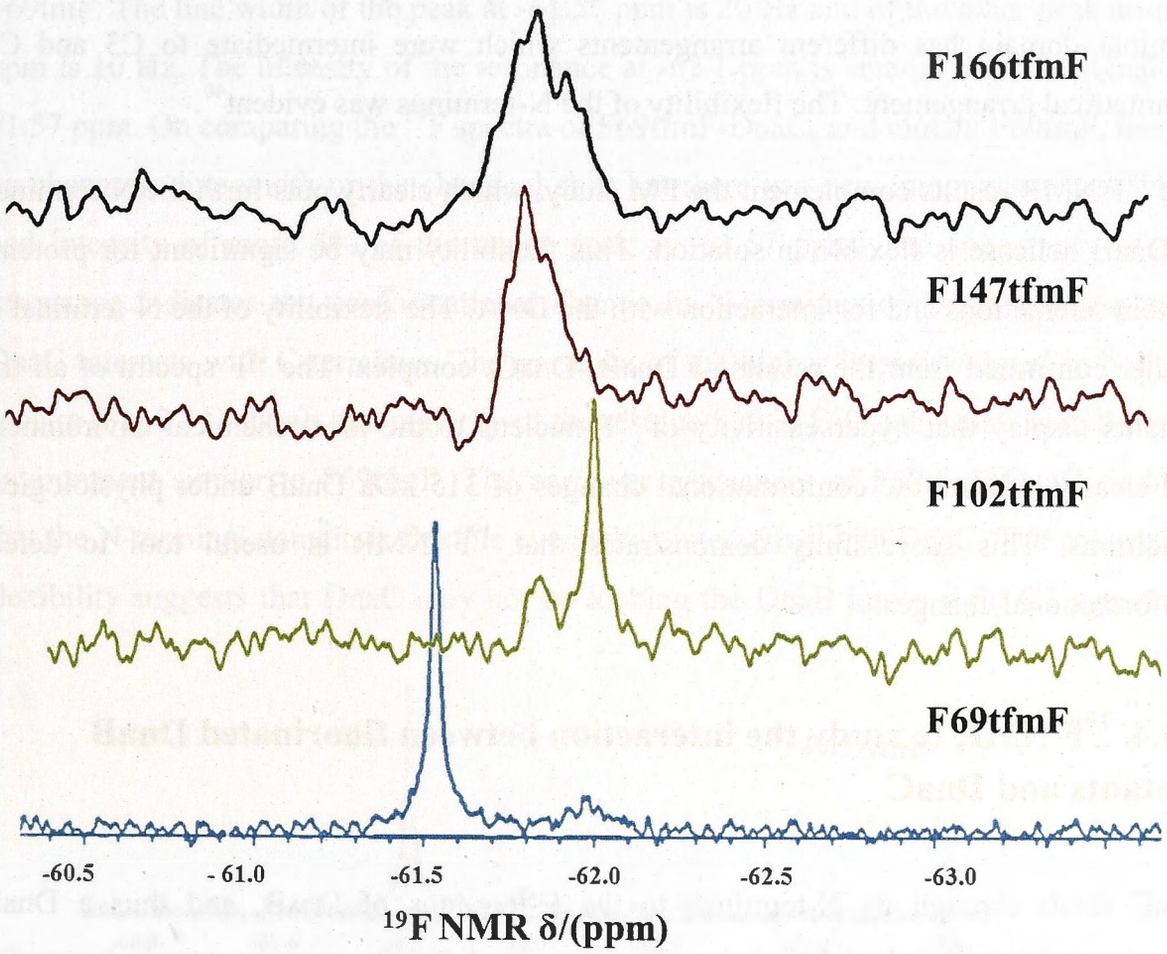

Figure 3.10: *^{19}F NMR spectra of tfmF labelled DnaB measured at pH 8. From the bottom of the spectra, F69tfmF – blue spectrum, F102tfmF – green spectrum, F147tfmF – red spectrum and F166tfmF – black spectrum. The broadening of the resonances is evident from the ^{19}F spectra.*

The number of peaks was an important factor for observing the conformational change. In the mutants F69tfmF and F147tfmF there was one ^{19}F resonances in pH 8. In case of mutants F102tfmF and F166tfmF, at pH 8, there were two ^{19}F resonances.

The line width of the ^{19}F resonances observed for each mutant played a major role in inferring the structural information of DnaB helicase in solution. The line width of higher intensity ^{19}F resonances observed for the mutants F69tfmF is 20 Hz, F102tfmF is 30 Hz, F147tfmF is 60 Hz and F166tfmF is 70Hz. The gradual decrease in the line width of ^{19}F resonances as the position of tfmF is shifted towards the N-terminal domain indicates that N-terminus is less-ordered and flexible compared to the C-terminus. The flexibility of N-terminal domain have previously reported by Yang and coworkers in which EM images show that though overall DnaB is adopting C3 symmetry and C6 symmetry, the N-terminal domain has different arrangements which were intermediate to C3 and C6 symmetrical arrangement. The flexibility of the N-terminus was evident[40].

The ^{19}F NMR results complement the EM study, which clearly puts forth the N-terminus of DnaB helicase is flexible in solution. This flexibility may be significant for protein-protein interactions and for interaction with the DNA. The flexibility of the N-terminal is further confirmed from the results of $DnaB_6$-$DnaC_6$ complex. The ^{19}F spectra of all the mutants display that hypersensitivity of ^{19}F nucleus to the local chemical environment and clearly indicate the conformational changes of 315 kDa DnaB under physiological conditions. This successfully demonstrates that ^{19}F NMR is useful tool to detect conformational changes.

3.3.4. ^{19}F NMR to study the interaction between fluorinated DnaB mutants and DnaC

DnaC binds through its N-terminus to the C-terminus of DnaB, and thus a DnaB construct with N-terminal His_6 tag and an untagged DnaC constructs were chosen for structural analysis. The main aim of the project was to observe the conformational changes in DnaB and these changes were only evident in the N-terminal domain and C-terminal domain always assumed C6 symmetry. Therefore, the mutation sites were selected in N-terminal domain and not in the C-terminal domain, eventhough C-terminal domain interacts with DnaC.

It was observed that co-expression of DnaB and DnaC helped in increasing the stability and reducing the precipitation of DnaB mutants. The complex of 480 kDa of molecular

weight was formed in the presence of ATP and Mg^{2+}. The ^{19}F NMR spectra of complex of fluorine labelled DnaB mutants- $DnaC_6$ was compared with the ^{19}F NMR spectra of fluorine labelled DnaB mutants in the presence of ATP and Mg^{2+}. Three mutants were studied in complex with DnaC by ^{19}F NMR at pH 8.

The ^{19}F NMR spectrum of F69tfmF- $DnaC_6$ showed two ^{19}F resonances. One narrow signal was at -61.57 ppm and the second well resolved resonance was at -62.1 ppm. The resonance at -61.57 ppm which is the same as the single resonance observed in mutant F69tmF. The line width of the peak at -61.57 ppm is 20 Hz and of the other peak at -62.1 ppm is 10 Hz. The intensity of the resonance at -62.1 ppm is smaller than the signal at -61.57 ppm. On comparing the ^{19}F spectra of F69tfmF-$DnaC_6$ and mutant F69tmF, there is no change in line width or the chemical shift but there is a new minor resonance which had intensity of more 5% of the major peak at -61.57 ppm. The appearance of new resonance indicates the conformational change in N-terminus of DnaB helicase when DnaC interacts with C-terminus. The presence of the higher intensity signal in both the spectra (Fig 3.11) shows that majority of the residue F69 in F69tfmF-DnaC is in the same environment similar to F69tfmF. The narrower resonances of F69tfmF-$DnaC_6$ indicate that the N-terminal domain is flexible even when interacting with DnaC. The presence of flexibility suggests that DnaC may not be locking the DnaB into a rigid C3 symmetry.

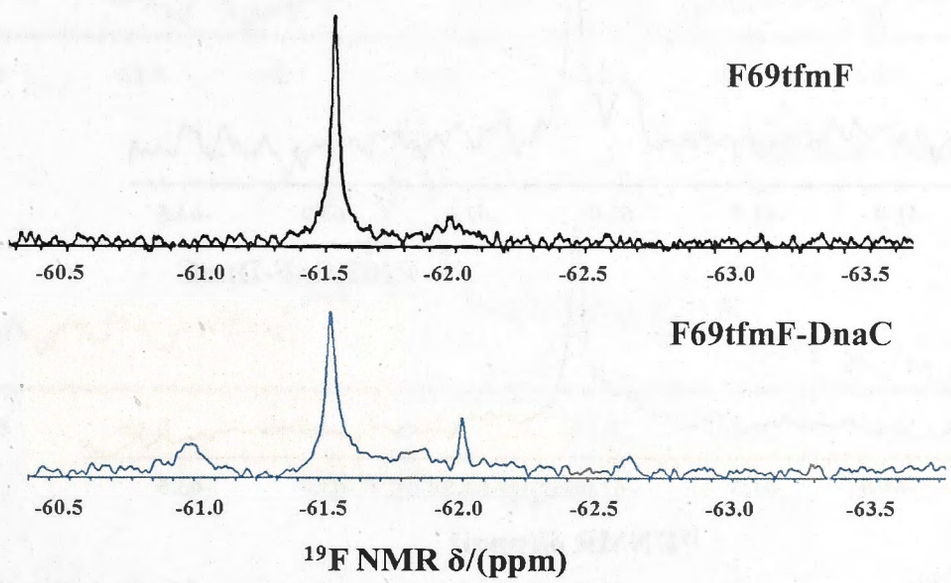

Figure 3.11: *^{19}F spectra of mutant F69tfmF (black) and in complex with DnaC (blue). The mutant F69tfmF has ^{19}F resonance at -61.55 ppm and two ^{19}F resonances observed for mutant F69tfmF-DnaC are at -61.55 and -62.1 ppm.*

The appearance of new signal indicates that complex might assume different conformations with one dominant form. Since, residue F69 has same chemical environment in C3 symmetry; the new signal represents a different conformation apart from C3 and C6 symmetry. To get deeper understanding of the appearance on new signals, other mutants were measured by ^{19}F NMR.

In ^{19}F NMR spectrum of mutant F102tfmF-$DnaC_6$ only one resonance was observed at -62.2 ppm instead of two resonances present in F102tfmF. The resonance at -61.9 ppm with smaller intensity observed in ^{19}F NMR spectrum of mutant F102tfmF has disappeared. There was no apparent change in line width of ^{19}F resonances at -62.2 ppm between F102tfmF and complex of F102tfmF- $DnaC_6$ (Fig 3.12). The disappearance of the signal when the complex is formed shows strong evidence that DnaB helicase has undergone major conformational change upon binding to DnaC. However it is evident that DnaB is not assuming rigid C3 symmetry when in complex with DnaC, which varies from results reported by EM studies [13,14,18]. The single resonance with the line width of 36 Hz suggests all the residues F102 have equivalent environment, which may be due to increased flexibility of N-terminus of mutant F102tfmF.

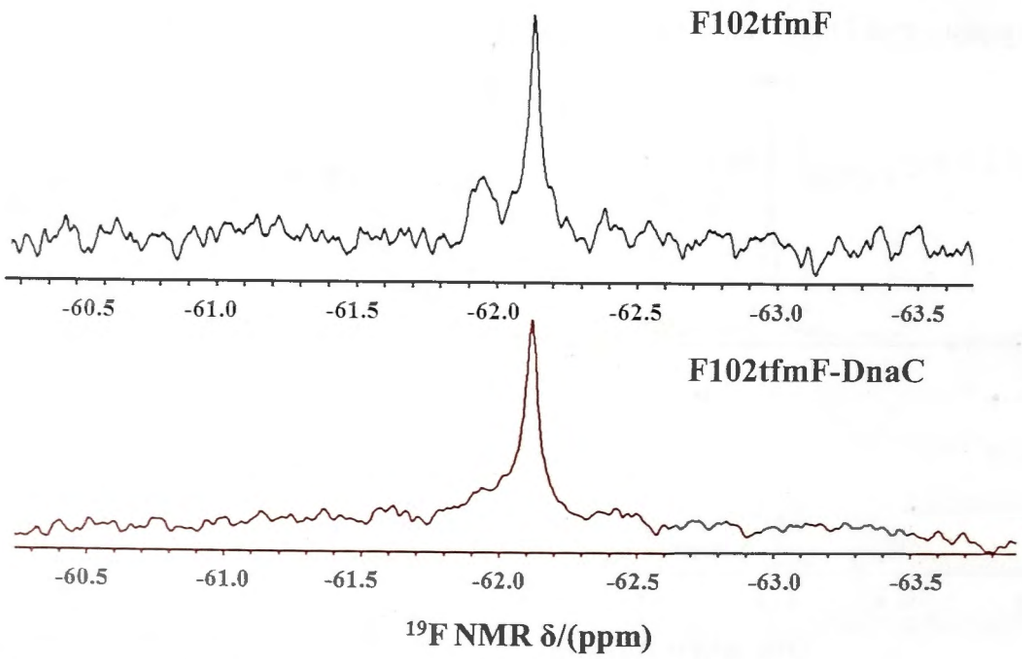

Figure 3.12: *^{19}F spectra of mutant F102tfmF (black) and in complex with DnaC (red). The mutant F102tfmF has two ^{19}F resonances at -62.2 ppm and -61.9 ppm. Mutant F102tfmF in complex with DnaC has only one signal at -62.2 ppm.*

The yield of mutant F147tfmF was very low when compared to other mutants. Due to lower concentration and precipitation, mutant F147tfmF - $DnaC_6$ could not be measured by ^{19}F NMR.

For mutant F166tfmF-DnaC complex, there is only single resonance at -61.87 ppm, which coincides with the peak maxima of one of the signals observed in the ^{19}F spectrum of F166tfmF (Fig 3.13). The line width of the signal has decreased to 70 Hz, which is similar to the resonance observed for F166tfmF at lower pHs. The signal at -61.82 ppm in spectrum F166tfmF has disappeared similar to mutant F102tfmF. The disappearance of signal indicates apparent change in conformation, which is consistent with ^{19}F NMR spectrum of mutant F102tfmF- $DnaC_6$. The single resonance shows that all the tfmF at F166 has similar environment in complex may be due to the increase in the flexibility of N-terminus induced by DnaC binding and suggesting that helix turn helix becomes less ordered when DnaB complexes with DnaC.

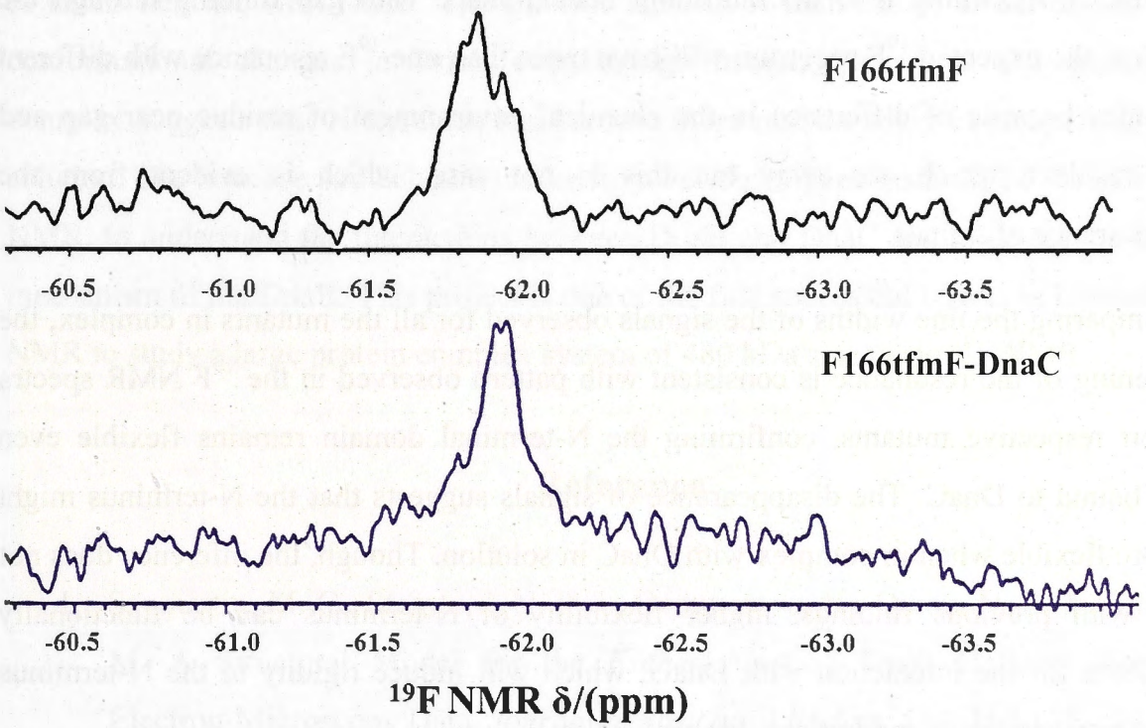

Figure 3.13: *^{19}F spectra of mutant F166tfmF (black) and in complex with DnaC (purple). The mutant F166tfmF has two ^{19}F resonances at -61.84 ppm and -61.95 ppm. Mutant F166tfmF in complex with DnaC has only one signal at -61.94 ppm.*

3.3.4.1. Comparing the ^{19}F spectra of DnaB-DnaC complex

^{19}F NMR study, there was only one resonance for mutants F102tfmF$_6$-DnaC$_6$ and F166tfmF$_6$-DnaC$_6$ and two resonances of different intensity in F69tfmF$_6$-DnaC$_6$ complex. The speculated reason for disappearances of resonance in DnaB$_6$-DnaC$_6$ complex could be because of increased flexibility of N-terminus due to DnaC binding at C-terminus.

The ^{19}F spectra clearly shows that the DnaC does not lock DnaB into C3 symmetry in solution as suggested by EM studies [13,14,18]. If DnaB helicase did confer rigid C3 symmetry when bound to DnaC, the ^{19}F NMR spectrum will display two signals as observed in DnaB instead the spectra revealed disappearance of signals. Although, the previous EM studies have reported that DnaC induced C3 symmetry in DnaB, recent studies have stated otherwise [42].

Recently, EM study by Arias-Palomo E and co-workers showed that DnaC changes the DnaB flat hexamer into right handed spiral helical molecule with gap running along the complex [42]. Assuming if DnaB did adopt helical shape with gap running through the complex, the expected ^{19}F spectrum will have more than one ^{19}F resonance with different intensities because of difference in the chemical environment of residue near gap and other residues which are away but this is not case, which is evident from the disappearance of signals.

On comparing the line widths of the signals observed for all the mutants in complex, the broadening of the resonance is consistent with pattern observed in the ^{19}F NMR spectra of their respective mutants, confirming the N-terminal domain remains flexible even when bound to DnaC. The disappearance of signals suggests that the N-terminus might be more flexible when in complex with DnaC in solution. Though, the inference does not agree with previous findings, higher flexibility of N-terminus can be functionally significant for the interaction with DnaG, which will induce rigidity to the N-terminus and activates helicase and initiates the replication.

All the ^{19}F spectra clearly shows that fluorine nucleus is sensitive to local environment because small changes in conformation due to binding of DnaC to the C-terminus of DnaB helicase is evident in the ^{19}F resonances of tfmF placed in the N-terminus.

3.5. Conclusion

In biomolecular NMR, it is always a challenge to study large protein systems in solution because of non-resolvable and un-assignable resonances. Introduction of exogenous nuclei which is different from the nuclear spins that exist naturally will solve the problem. The exogenous nuclei spin label introduced in this project is fluorine. The main aim of the project was to incorporate fluorine label into large macromolecular system by cell-free protein synthesis and to successfully observe the ^{19}F signals from the system. The fluorine signals for all the mutants were distinct for their position and well- resolved. The second major objective was to determine the symmetry of DnaB helicase under various conditions in solution. The factor varied was pH and results showed that N-terminus is mobile when compared to the C-terminus of the helicase in native biological conditions. This has been previously established by EM studies that highlighted the flexibility in the N-terminus of the protein[41].

The sensitivity of fluorine spin was further explored to study the structural arrangement of DnaB helicase with DnaC. The fluorine label in N-terminus of DnaB displayed the conformational changes in DnaB even though interactions are at C-terminus. The complex suggests that N-terminus has become more mobile when in complex DnaC in solution. The helicase and helicase loader from various organisms must be studied by ^{19}F NMR, to understand the interactions between DnaB and DnaC and to detect the loading mechanism of the DnaB. This project is one of the first successful efforts in biomolecular NMR to study a large protein complex system of 480 kDa size using ^{19}F NMR.

Reference

1. Sanmartin, M. C., Stamford, N. P. J., Dammerova, N., Dixon, N. E. & Carazo, J. M. A Structural Model for the *Escherichia-Coli* Dnab Helicase Based on Electron-Microscopy Data. *Journal of Structural Biology* **114**, 167-176, (1995).
2. Lebowitz, J. H. & Mcmacken, R. The *Escherichia-Coli* Dnab Replication Protein Is a DNA Helicase. *Journal of Biological Chemistry* **261**, 4738-4748 (1986).

3. Kaguni, J. M. Replication initiation at the *Escherichia coli* chromosomal origin. *Current Opinion in Chemical Biology* **15**, 606-613, (2011).
4. Jezewska, M. J., Kim, U. S. & Bujalowski, W. Interactions of Escherichia coli primary replicative helicase DnaB protein with nucleotide cofactors. *Biophys J* **71**, 2075-2086,(1996).
5. Yoda, K. & Okazaki, T. Specificity of Recognition Sequence for Escherichia-Coli Primase. *Molecular & General Genetics* **227**, 1-8, (1991).
6. Lu, Y. B., Ratnakar, P. V. A. L., Mohanty, B. K. & Bastia, D. Direct physical interaction between DnaG primase and DnaB helicase of Escherichia coli is necessary for optimal synthesis of primer RNA. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 12902-12907, (1996).
7. Chang, P. & Marians, K. J. Identification of a region of Escherichia coli DnaB required for functional interaction with DnaG at the replication fork. *Journal of Biological Chemistry* **275**, 26187-26195,(2000).
8. Kim, S., Dallmann, H. G., McHenry, C. S. & Marians, K. J. Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell* **84**, 643-650,(1996).
9. Manna, A. C. *et al.* Helicase-contrahelicase interaction and the mechanism of termination of DNA replication. *Cell* **87**, 881-891,(1996).
10. Bujalowski, W., Klonowska, M. M. & Jezewska, M. J. Oligomeric Structure of Escherichia-Coli Primary Replicative Helicase Dnab Protein. *Journal of Biological Chemistry* **269**, 31350-31358 (1994).
11. Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annual Review of Biochemistry* **69**, 651-697, (2000).
12. Yu, X., Jezewska, M. J., Bujalowski, W. & Egelman, E. H. The hexameric E. coli DnaB helicase can exist in different Quaternary states. *J Mol Biol* **259**, 7-14, (1996).
13. Donate, L. E. *et al.* pH-controlled quaternary states of hexameric DnaB helicase. *J Mol Biol* **303**, 383-393, doi:10.1006/jmbi.2000.4132 (2000).
14. San Martin, C. *et al.* Three-dimensional reconstructions from cryoelectron microscopy images reveal an intimate complex between helicase DnaB and its loading partner DnaC. *Structure with Folding & Design* **6**, 501-509 (1998).

15. Watt, S. J. *et al.* Multiple oligomeric forms of Escherichia coli DnaB helicase revealed by electrospray ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry* **21**, 132-140, (2007).
16. Weigelt, J., Miles, C. S., Dixon, N. E. & Otting, G. Backbone NMR assignments and secondary structure of the N-terminal domain of DnaB helicase from E. coli. *J Biomol NMR* **11**, 233-234 (1998).
17. Fass, D., Bogden, C. E. & Berger, J. M. Crystal structure of the N-terminal domain of the DnaB hexameric helicase. *Structure* **7**, 691-698 (1999).
18. Barcena, M. *et al.* The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel. *EMBO J* **20**, 1462-1468, (2001).
19. Wang, G. *et al.* The structure of a DnaB-family replicative helicase and its interactions with primase. *Nature Structural & Molecular Biology* **15**, 94-100, (2008).
20. Bailey, S., Eliason, W. K. & Steitz, T. A. Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* **318**, 459-463, (2007).
21. Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J. M. & Steitz, T. A. The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation. *Cell* **151**, 267-277, (2012).
22. Bujalowski, W. & Jezewska, M. J. Kinetic mechanism of nucleotide cofactor binding to Escherichia coli replicative helicase DnaB protein. Stopped-flow kinetic studies using fluorescent, ribose-, and base-modified nucleotide analogues. *Biochemistry* **39**, 2106-2122, (2000).
23. Roychowdhury, A., Szymanski, M. R., Jezewska, M. J. & Bujalowski, W. Interactions of the Escherichia coli DnaB-DnaC Protein Complex with Nucleotide Cofactors. 1. Allosteric Conformational Transitions of the Complex. *Biochemistry* **48**, 6712-6729, (2009).
24. Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of 19F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* **62**, (2012)
25. Danielson, M. A. & Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct* **25**, 163-195,(1996).

26. Ozawa, K. *et al.* High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochemical and Biophysical Research Communications* **418**, 652-656, (2012).
27. Neylon, C. *et al.* Interaction of the Escherichia coli replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* **39**, 11989-11999, (2000).
28. Kigawa, T. *et al.* Cell-free production and stable-isotope labeling of milligram quantities of proteins. *Febs Letters* **442**, 15-19,(1999).
29. Wu, P. S. *et al.* Amino-acid type identification in 15N-HSQC spectra by combinatorial selective 15N-labelling. *J Biomol NMR* **34**, 13-21, (2006).
30. Wu, P. S. C. *et al.* Cell-free transcription/translation from PCR-amplified DNA for high-throughput NMR studies. *Angewandte Chemie-International Edition* **46**, 3356-3358, (2007).
31. Ozawa, K., Jergic, S., Park, A. Y., Dixon, N. E. & Otting, G. The proofreading exonuclease subunit of Escherichia coli DNA polymerase III is tethered to the polymerase subunit via a flexible linker. *Nucleic Acids Research* **36**, 5074-5082, (2008).
32. Young, D. D., Young, T. S., Ahmad, I. & Schultz, P. G. Evolved aminoacyl-tRNA synthetases with unusual polyspecificities. *Abstracts of Papers of the American Chemical Society* **240** (2010).
33. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annual Review of Biochemistry, Vol 79* **79**, 413-444, (2010).
34. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *Journal of Molecular Biology* **395**, 361-374, (2010).
35. Ugwumba, I. N. *et al.* Using a Genetically Encoded Fluorescent Amino Acid as a Site-Specific Probe to Detect Binding of Low-Molecular-Weight Compounds. *Assay and Drug Development Technologies* **9**, 50-57, (2011).
36. Apponyi, M. A., Ozawa, K., Dixon, N. E. & Otting, G. Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods Mol Biol* **426**, 257-268, (2008).
37. Loscha, K. V. *et al.* Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angewandte Chemie-International Edition* **51**, 2243-2246, (2012).

38. Korostelev, A., Zhu, J., Asahara, H. & Noller, H. F. Recognition of the amber UAG stop codon by release factor RF1. *EMBO J* **29**, 2577-2585, (2010)
39. Williams, N. K. *et al.* Stabilization of native protein fold by intein-mediated covalent cyclization. *Journal of Molecular Biology* **346**, 1095-1108, (2005).
40. Yang, S. X. *et al.* Flexibility of the rings: Structural asymmetry in the DnaB hexameric helicase. *Journal of Molecular Biology* **321**, 839-849, (2002).
41. Arias-Palomo, E., O'Shea, V. L., Hood, I. V. & Berger, J. M. The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. *Cell* **153**, 438-448, (2013).

Chapter 4

^{19}F NMR Studies on *Geobacillus Stearothermophilus* DnaB and its protein-protein interactions

4.1. Introduction

Geobacillus stearothermophilus is a thermophilic and Gram-positive bacterium. It is used as an expression system for thermostable proteins because of its capacity to grow at temperatures as high as 65 °C [1]. *G. stearothermophilus* is also common contaminant in diary industry. It has been shown to form biofilms and poses a problem in milk powder manufacture plants [2]. *G. stearothermophilus* is both a saviour and problem, which makes it an interesting organism to be studied. The insights into its DNA replication process might aid in both enhancing its thermostability and decontamination. In order to understand a process, it has always been best to begin with critical components. In case of DNA replication, it is the DnaB helicase. *G. stearothermophilus* DnaB helicase (*Gst DnaB*) was first cloned and expressed in *E. coli* by Bird and co-workers. *Gst DnaB* is a 454 amino acids sequence with 45% identity and 69% similarity to *E. coli* DnaB helicase [3]. The trypsin proteolysis showed that it had two domains, the N-terminal domain with molecular weight of 12 kDa and C-terminal domain with 33 kDa, similar to *E. coli* DnaB helicase [3].

E. coli and *Gst DnaB* belong to same family as SF4 sharing some of the important characteristic such as possessing a RecA domain in C-terminal domain, and forming a hexamer [4]. Since, *E. coli* DnaB has been studied extensively; some of its structural characteristics can be assigned to *Gst DnaB* to gain insights into the protein and vice versa.

Several EM studies have examined and established that *E. coli* DnaB adopts different conformations under different conditions [5-8]. The deductions from these studies are still under debate. However, it can be inferred that the N-terminal domain is dynamic in solution under physiological conditions. A possible way to induce rigidity into the N-

terminus is by interacting with DnaG primase. In case of *E. coli* DnaB helicase, the interactions between DnaB and DnaG are transient and studies have identified that DnaG associates and dissociates rapidly from *E. coli* DnaB during primer synthesis on lagging strand [9]. In contrast, *Gst DnaB*-DnaG forms a stable complex that can be isolated through gel filtration, suggesting that *Gst*DnaG is associated permanently with DnaB during DNA replication [10].

Gst DnaB-DnaG complex was crystallised and a high resolution structure was determined by X-ray crystallography [11]. The crystal structure shows that three monomers of *Gst*DnaG forms complex with one molecule of hexameric *Gst DnaB* helicase, which supports the atomic force microscopy images (AFM) of *Gst DnaB*-DnaG and gel filtration data reported by Thirlway and coworkers [12]. The AFM images showed that the *Gst*DnaG binds to the linker helix of *Gst DnaB* helicase [12], being consistent with interaction studies on *E. coli* DnaB, which reported that interacting residues reside on both the N-terminal and C-terminal domains [13, 14]. However, the crystal structure resolved by Steitz group showed that *Gst*DnaG interacts with the residues in the N-terminal domain of *Gst DnaB*. All the studies showed that interaction is primarily between C-terminal domain of *Gst*DnaG and the N-terminal domain and linker helix of *Gst DnaB* [11-14]. The domain interacting with helicase is called helicase binding domain. The domain with residues 454-597 is known as p16 domain having a mass of 16 kDa. The p16 domain is necessary for the complex formation and DnaG primase activity. The interaction with DnaG stimulates ATPase and helicase activity and in turn DnaB increases rate of RNA primer synthesis [15]. The interaction studies and crystal structure have demonstrated that DnaG induces triangular arrangement of N-terminus to adopt C3 symmetry from mixture of C6 and C3 symmetry [11,12].

In *E. coli*, the association of DnaG to the N-terminal domain of DnaB leads to dissociation of the helicase loader DnaC [16]. However, in Gram positive bacteria, the studies demonstrated that the whole complex of helicase-helicase loader (DnaI) and primase can be isolated [17] and *in vitro* synthesis of RNA primers in presence of DnaI [18,19] has proven otherwise. The studies on *Gst DnaB* have been with the helicase loader of *B. subtilis* DnaI (*Bsu* DnaI) [17, 20]. The complex of DnaB-I- DnaGC was isolated by gel filtration and crystallised by two different research groups. The crystal structure

displayed that six subunits of *Bsu*DnaI interacts with six *Gst DnaB* subunits. The paper also suggested that during complex formation and the release of *Bsu*DnaI, there is a significant conformational change in the helicase. They suggested that helicase transforms from open-ring to open spiral to closed spiral eventually releasing the loader form the complex [20]. *Bsu* DnaI has AAA+ fold for ATPase activity and nucleotide binding in the C-terminal domain. The N-terminal domain contains zinc binding fold and it is homologous to the N-terminal domain of *E. coli* DnaC [21]. To this date there is no structural information on the *G. Stearothermophilus* DnaI helicase loader (*Gst*DnaI) and its interaction with *Gst DnaB*. This will be early stages of research reporting on the behaviour of *Gst*DnaI and its interaction with *Gst DnaB*. In *E. coli*, DnaB-DnaC has shown to adopt different quartenary structures [7,22, 23]. EM images have shown that DnaB adopts C3 symmetry when bound to helicase loader [7,22]. It would be interesting to observe any conformational changes induced in *Gst DnaB* by *Gst*DnaI interaction as it was observed for *Bsu*DnaI and *Gst DnaB*.

The studies indicate that *Gst DnaB*-DnaG complex is a promising system to validate the ^{19}F NMR as a potential tool to investigate the conformational changes in DnaB helicase. *Gst DnaB* is well-behaved protein because it is stable and has well-resolved crystal structure and there is ample information from its well–resolved structure in complex with DnaG, which *E. coli* DnaB lacks.

Here, we incorporate trifluoromethylphenylalanine (tfmF) into *Gst DnaB* by *in vivo* protein expression using pEVOL system with *p*-cyano-L-phenylalanyl-tRNA synthetase (*p*CNF-RS) developed by Prof. Schultz and coworkers [24,25]. The tfmF labelled *Gst DnaB* and its interactions with Mg^{2+}, nucleotides, DnaG and DnaI are studied by ^{19}F NMR. We report a new aspect that Mg^{2+} induces the formation of monomers, the N-terminus of *Gst DnaB* is flexible and ^{19}F NMR spectra confirm the conformational change induced by DnaG in solution.

4.2. Materials and Methods

4.2.1. Plasmids and Genes

The *Gst DnaB* and p16 domain of *Gst*DnaG (DnaGC) was codon optimized using the OPTIMISER software [26], for expressing the proteins in *E. coli* system. The optimized genes were synthesized by GenScript USA Inc., NJ, USA. The genes were synthesised with a solubility tag comprising of MASMTG, His_6, GB1 domain is the soluble domain of IgG [31], and Tobacco Etch Virus (TEV) protease cleavage site attached to N-terminus of both the genes as described in the figure 1. The solubility tag was attached to the second amino acid of the *Gst* DnaB/DnaGC, the first methionine in both the genes were removed. *Gst* DnaI with C-terminal His_6 tag inserted in pETMCSI in the bacterial strain of BL21 (DE3) was a kind gift from Prof. Nick Dixon (University of Wollongong, Wollongong, Australia). *Gst DnaB* and DnaGC were inserted into pET20b vector by GenScript. The gene sequences of Gst DnaB and DnaGC is given in table 4.1.

***Figure 4.1**: The pictorial representation of the Gst DnaB and DnaGC with the tag attached to N-terminus of both the genes.*

***Table 4.1**: The sequence of the genes with the attached tags*

Genes	Sequence with the tag
Gst DnaB	ATGGCTTCTATGACCGGTCACCATCACCATCACCATCAGTACAAACTGATCCTGAACGGTAAAACCC TGAAAGGTGAAACCACCACCGAAGCTGTTGACGCTGCTACCGCGGAAAAAGTTTTCAAACAGTACG CTAACGACAACGGTGTTGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCG AAAACCTGTACTTCCAGTCTTCTGAACTGTTCTCTGAACGTATCCCGCCGCAGTCTATCGAAGCGGA ACAGGCGGTTCTGGGTGCGGTTTTCCTGGACCCGGCGGCGCTGGTTCCGGCGTCTGAAATCCTGATC CCGGAAGACTTCTACCGTGCGGCGCACCAGAAAATCTTCCACGCGATGCTGCGTGTTGCGGACCGTG GTGAACCGGTTGACCTGGTTACCGTTACCGCGGAACTGGCGGCGTCTGAACAGCTGGAAGAAATCG GTGGTGTTTCTTACCTGTCTGAACTGGCGGACGCGGTTCCGACCGCGGCGAACGTTGAATACTACGC GCGTATCGTTGAAGAAAAATCTGTTCTGCGTCGTCTGATCCGTACCGCGACCTCTATCGCGCAGGAC GGTTACACCCGTGAAGACGAAATCGACGTTCTGCTGGACGAAGCGGACCGTAAAATCATGGAAGTT TCTCAGCGTAAACACTCTGGTGCGTTCAAAAACATCAAAGACATCCTGGTTCAGACCTACGACAACA TCGAAATGCTGCACAACCGTGACGGTGAAATCACCGGTATCCCGACCGGTTTCACCGAACTGGACCG TATGACCTCTGGTTTCCAGCGTTCTGACCTGATCATCGTTGCGGCGCGTCCGTCTGTTGGTAAAACCG CGTTCGCGCTGAACATCGCGCAGAACGTTGCGACCAAAACCAACGAAAACGTTGCGATCTTCTCTCT GGAAATGTCTGCGCAGCAGCTGGTTATGCGTATGCTGTGCGCGGAAGGTAACATCAACGCGCAGAA

	CCTGCGTACCGGTAAACTGACCCCGGAAGACTGGGGTAAACTGACCATGGCGATGGGTTCTCTGTCT AACGCGGGTATCTACATCGACGACACCCCGTCTATCCGTGTTTCTGACATCCGTGCGAAATGCCGTC GTCTGAAACAGGAATCTGGTCTGGGTATGATCGTTATCGACTACCTGCAGCTGATCCAGGGTTCTGG TCGTTCTAAAGAAAACCGTCAGCAGGAAGTTTCTGAAATCTCTCGTTCTCTGAAAGCGCTGGCGCGT GAACTGGAAGTTCCGGTTATCGCGCTGTCTCAGCTGTCTCGTTCTGTTGAACAGCGTCAGGACAAAC GTCCGATGATGTCTGACATCCGTGAATCTGGTTCTATCGAACAGGACGCGGACATCGTTGCGTTCCT GTACCGTGACGACTACTACAACAAAGACTCTGAAAACAAAAACATCATCGAAATCATCATCGCGAA ACAGCGTAACGGTCCGGTTGGTACCGTTCAGCTGGCGTTCATCAAAGAATACAACAAATTCGTTAAC CTGGAACGTCGTTTCGACGAAGCGCAGATCCCGCCGGGTGCGTAAGAATTC
*Gst*DnaGC	ATGGCTTCTATGACCGGTCACCATCACCATCACCATCAGTACAAACTGATCCTGAACGGTAAAACCC TGAAAGGTGAAACCACCACCGAAGCTGTTGACGCTGCTACCGCGGAAAAAGTTTTCAAACAGTACG CTAACGACAACGGTGTTGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCG AAAACCTGTACTTCCAGTCTCTGGCGAAAAAACTGCTGCCGGCGTTCCAGAACGCGGAACGTCTGCT GCTGGCGCACATGATGCGTTCTCGTGACGTTGCGCTGGTTGTTCAGGAACGTATCGGTGGTCGTTTC AACATCGAAGAACACCGTGCGCTGGCGGCGTACATCTACGCGTTCTACGAAGAAGGTCACGAAGCG GACCCGGGTGCGCTGATCTCTCGTATCCCGGGTGAACTGCAGCCGCTGGCGTCTGACGTTTCTCTGCT GCTGATCGCGGACGACGTTTCTGAACAGGAACTGGAAGACTACATCCGTCACGTTCTGAACCGTCCG AAATGGCTGATGCTGAAAGTTAAAGAACAGGAAAAAACCGAAGCGGAACGTCGTAAAGACTTCCTG ACCGCGGCGCGTATCGCGAAAGAAATGATCGAAATGAAAAAAATGCTGTCTTCTTCTTAAGAATTC
*Gst*DnaI	ATGGAACGAGTAAATCAACTGTTGCAGCGGCTGTTCGGAAACGAAGGGTTCCGGCGGCGCTATGAA CAAATGCGGCGCTATATTTTGACGCATCCGGACGTGCAGCCGTTTTTGCAGGCGCACGAGCAGCAGC TGTCGCGCGATGCGGTGGACCGAAGTTTAATGAAGCTGTACGAATTTATCGAGCAACATGGCCATTG CCGCCAGTGCCCAGGGCTCGAGCAATGCCCAAATATGTTGCCGGGGTATCATCCGAACTTGGTGGTC GCCGGCGGGCGAATTGACGTTGAATACGACCGCTGCCCGAAAAAAGTGCAAGATGATGAACGGAGA AGGCAGGAAGCGCTCATTCAAAGCATGTTCGTGCCGCGGGAAATTTTGCAAGCTTCGCTGTCGGATG TGGATcTTAGCGACGATGGGCGCATTAAAGCCATCCAGTTTGCAGAGAAGTTCGTGACGGAGTACGA GCCGGGGAAAAAAATGAAAGGATTGTACTTGTACGGGTCGTTCGGCGTCGGCAAAACGTATTTGCT CGGGGCGATCGCCAATGAACTGGCGAAACGGAACATTCCGTCGCTCATCGTCTATGTGCCGGAGCTG TTTCGCGAGCTGAAGCATTCATTGCAAGATCAGACGATGAACGAAAAGCTCGATTATGTGAAAAAA GTGCCGGTGCTCATGCTCGATGACCTTGGAGCGGAGGCGATGTCGAGCTGGGTGCGCGACGATGTG CTCGGCCCAATTTTGCAATACCGGATGTTTGAAAATTTGCCGACCTTTTTCACCTCCAACTTTGATAT GAAGCAGCTCGCCCACCATTTGACGTATTCGCAGCGCGGCGAGGAAGAAAAAGTGAAAGCCGCCCG CATTATGGAGCGGATCCGCTCACTCGCGCACCCGGTTGAAATTACCGGGCCAAACCGCCGCGAACA CCATCACCATCACCATTAA

4.2.2. Site-Directed Mutagenesis

The *Gst DnaB* in pET20b vector was transformed into DH5α and plasmid DNA was extracted by miniprep. The extracted plasmid was used as template for the PCR reactions. The first PCR set up comprises two different PCR reactions with different primer sets for obtaining products with overlapping fragments. The primers in first PCR reaction consists of sequence of T7 promoter as forward primer and reverse primer is 30 nucleotide sequence of *Gst DnaB* with TAG codon replacing the codon of amino acid of interest. The primers of second PCR reaction consists of the 30 nucleotide sequence complementary to the *Gst DnaB* with TAG codon as the forward primer and T7 termination sequence as reverse primer. The products from both the reactions have the fragments with 10 nucleotide overlap. The fragments are then gel purified and amplified by the second PCR to obtain the *Gst DnaB* gene with TAG codon at site of interest, using the sequence of T7 promoter and T7 terminator as forward and reverse primer, respectively [27]. The amplified gene with TAG codon is subjected to PCR clean-up to

remove primers contamination. The product is restricted digested using restriction endonucleases – *Eco*RI and *Nde*I and ligated into pETMSCI vector using T4 DNA ligase. The ligated mix was transformed into DH5α by electroporation. The mutants were confirmed by sequencing. The mutants and mutation site is mentioned in table 4.2.

***Table 4.2**: Primer sequence for site-directed mutagenesis*

Mutation	Primers with 'TAG' codon at site
F52TfmF	Forward – 5' CAGCAGAAAATC**TAG**CACGCGATGCT-3'
	Reverse – 5' CAGCATCGCGTG**CTA**GATTTTCTGGTG-3'
Y104TfmF	Forward – 5' GCGAACGTT**TAG**TACGCGCGTAT-3'
	Reverse – 5' CATACGCGCGTA**CTA**TTCAACGTTCGC-3'
Y130TfmF	Forward – 5' GCGCAGGACGCT**TAG**ACCCGTGAAGA-3'
	Reverse – 5' GTCTTCACGGGT**CTA**ACCGTCCTGCGC-3'

The mutation sites are indicated in **bold**

4.2.3. Site-specifically tfmF labelled *Gst DnaB* Expression

For tfmF labelled *Gst DnaB* helicase expression, *in vivo* expression was chosen. The pETMCSI vector carrying *Gst DnaB* helicase gene with amber codon replacing the codons of F52, Y104 or Y130 was transformed into BL21 (DE3)*recA*$^-$. The cells were also carried the pEVOL plasmid containing *p*CNF-RS gene and suppressor tRNA, was a gift from Prof. Schultz (The Scripps Research Institute, La Jolla, CA), was used for expression of fluorine labelled *Gst* DnaB[28].

The cells were grown at 37 °C in LB medium supplemented with 100 μg/ml ampicillin and 33 μg/ml chloramphenicol. 2.5 mL of an overnight culture was used to inoculate 250 mL LB medium supplemented with antibiotics and 1 mM tfmF [29]. The cells were grown till OD_{600} of 1.5 was reached and the expression of *p*CNF-RS was induced with 0.02% Arabinose and *Gst* DnaB with amber codon was induced with 1 mM IPTG. The cells were grown for further 4 h after induction at 37 °C and 42 °C for achieving best condition for overexpression. Only for the culture grown at 25 °C was incubated for 16 h, instead of

4 h. After the 4 h or 16 h of incubation, 1 mL of the culture was set aside and spun down at 5000*g* for analysis by SDS-PAGE.

The cells were harvested by centrifugation at 5000g for 15 min. The cells were dissolved in the lysis buffer (50 mM Tris.HCl pH7.4, 500 mM NaCl, 10 mM Imidazole) and lysed using mechanical pressure by French Press at 12,000 psi. The cell lysates were centrifuged for 1 h at 32,000*g*. The supernatant was collected for purification using Ni-NTA column.

4.2.4. *Gst* DnaCG and DnaI Expression

The T7 system pET20b vector containing the DnaGC was transformed into BL21 (DE3) expression cells. The 2.5 ml of LB medium supplemented with 100 µg/ml ampicillin was inoculated with single colony and incubated overnight at 37 °C. The overnight culture was used the following day to inoculate 500 mL of LB media supplemented with ampicillin. The culture was grown at 37 °C till OD_{600} reached 0.8 and induced with 1 mM IPTG. The cells were incubated for 4 h post-induction at the same temperature for overexpression. The cells were harvested and lysed in the similar method as mentioned above.

*Gst*DnaI with C-terminal His_6 tag inserted in pETMCSI in BL21 (DE3) was a kind gift from Prof. Nick Dixon (University of Wollongong, Wollongong, Australia). 500 ml of LB medium with ampicillin is inoculated with 5 mL of overnight culture in the presence of 100 µg/ml ampicillin. The cells were grown at 37 °C till OD_{600} 0.6 was reached and induced with 1 mM IPTG. After induction, the cells are grown at 25 °C, overnight for overexpression. The cells were harvested by centrifugation at 5000*g* for 15 min. The cells were dissolved in the lysis buffer (50 mM Tris.HCl pH7.4, 500 mM NaCl, 10 mM Imidazole) and lysed using mechanical pressure by French Press at 12,000 psi. The cell lysates were centrifuged for 1 h at 32,000*g*. The supernatant was collected for purification using Ni-NTA column.

4.2.5. Analytical Gel Filtration

The superdex S-200 10/30 (Amersham Pharmacia Biotech) gel filtration column was used for determination of the stability of hexamer formation and *Gst DnaB*-DnaGC and *Gst DnaB*-DnaI complex formation. The column was equilibrated with NMR buffer (50 mM Tris pH 7.4, 1 mM EDTA, 1 mM dithiothreitol (DTT), 200 mM NaCl) and 2 ml of the concentrated *Gst DnaB* or the complex is loaded onto the column. The complex formation is initiated by mixing 60 μM of *Gst DnaB* with 100 μM of DnaGC was incubated at room temperature for 30 min prior to loading onto the column. Gel filtration was carried out at 4 °C and eluted with the NMR buffer at flow rate at 0.5 ml/min and 2 ml of elution fractions were collected using FPLC system. Fractions with UV absorbance fraction containing the protein were pooled together and concentrated and were analysed by SDS-PAGE.

4.3. Results and Discussion

4.3.1. Protein Expression

4.3.1.1. *Gst* DnaB expression

The expression of *Gst DnaB* by cell-free protein synthesis gave lower protein yields when compared to *E. coli* DnaB yields. Therefore, tfmF labelled *Gst DnaB* was expressed by *in vivo method* using pEVOL system with *p*CNF-RS as tRNA synthetase.

Mutants Y104tfmF and Y130tfmF were expressed at 25 °C, 37 °C and 42 °C for optimising the post-induction temperature for overexpression and incorporation of tfmF with higher efficiency. After expressing the protein at different temperatures, 1 ml culture was lysed by dissolving in SDS sample buffer and was analysed by SDS-PAGE. The molecular weight of full-length *Gst* DnaB expressed is 58 kDa including the residues of the solubility tag. The solubility tag (MASMTG + His_6 + GB1 domain + TEV protease site) was attached to facilitate efficient protein production. The specific purpose of

MASMTG, which is a T7 gene 10 tag, was to increase the protein expression [30]. The His_6 tag was attached for purification using Ni-NTA column, GB1 domain has shown to increase the solubility of proteins [31] and TEV protease cleavage site was added between the tag and *Gst DnaB* gene to remove the tag by proteolysis [32].

SDS-PAGE analysis showed that the expression levels of truncation product were similar at all the temperatures, however, the expression of full length tfmF labelled *Gst DnaB* varied. *Gst DnaB* expressed well at 37 °C and 42 °C and the expression levels were similar. However, at 25 °C, the level of expression was relatively lower. The solubility of tfmF labelled DnaB at these temperatures was compared and was evident that *Gst DnaB* expression at 42 °C was less soluble than at 37 °C. The experimented indicated at higher temperatures expression of tfmF labelled *Gst* DnaB is higher however, it is more soluble at 37 °C. Therefore, 37 °C was chosen as optimum temperature for expression of tfmF labelled *Gst DnaB*.

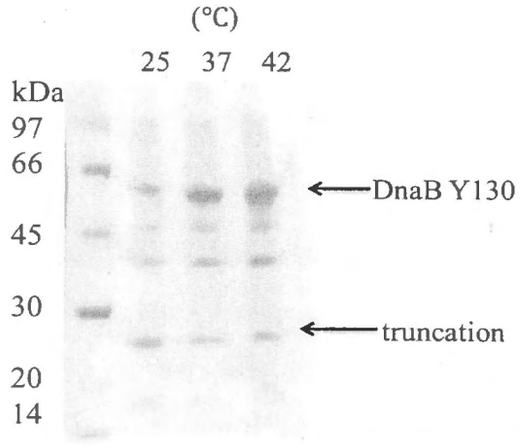

***Figure 4.2**: Optimisation of temperature for the tfmF labelled Gst DnaB expression. A representative coomassie stained gel of crude cell lysate displaying the different expression levels of mutant Y130tfmF at 25 °C, 37 °C and 42 °C. The arrows indicate the full length tfmF incorporated Gst DnaB and truncation product.*

4.3.1.2. *Gst*DnaGC and DnaI Expression

Instead of full-length DnaG, only a construct of solubility tag with the C-terminal p16 DnaB binding domain, referred as DnaGC, was used to study the interaction between *Gst DnaB* and DnaGC and subsequently study the conformation of DnaB by ^{19}F NMR. The expression of *Gst*DnaGC was carried out at 37 °C for 3 h after induction. Analysis of the crude cells and cell lysate by SDS-PAGE gel showed that *Gst*DnaGC overexpressed and was in soluble form. The DnaGC with the solubility tag is a 25 kDa protein.

The expression of *Gst*DnaI was carried out at 16 °C for 14 h because the expression of *Gst*DnaI was relatively lesser than the expression levels of *Gst DnaB* and DnaGC, so the temperature was decreased and post-induction temperature was increased. *Gst*DnaI degraded after purification when stored on ice more than 2 h. *Gst*DnaI showed tendency to precipitate and degraded into lesser molecular weight peptides during gel filtration. The molecular weight of *Gst*DnaI is 37 kDa.

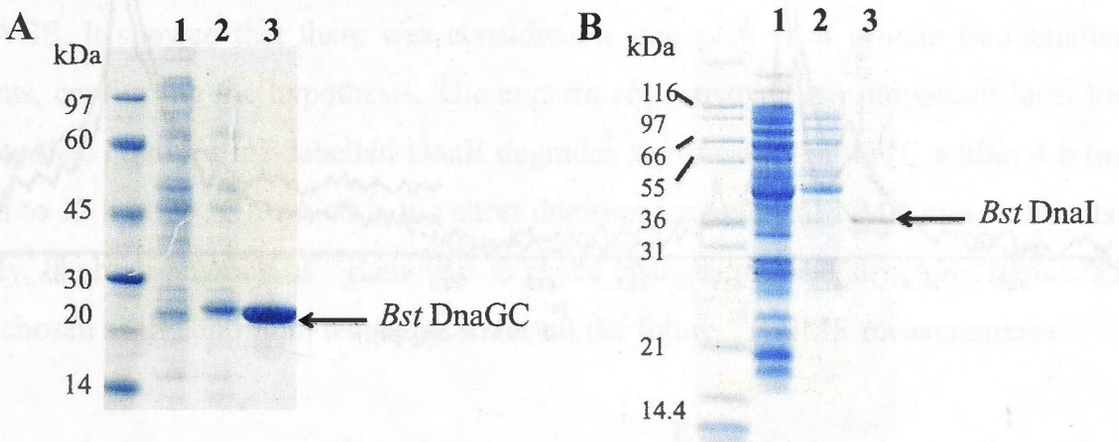

Figure 4.3: *SDS PAGE gels displaying the fractions collected at different stages of purification of GstDnaGC and DnaI. A) GstDnaGC purification Lane1: Flowthrough collected while the DnaGC was passed through the Ni-NTA column, Lane 2: Wash fraction collected and Lane 3: Elution fraction of pure GstDnaGC. B) GstDnaI purification Lane1: Flowthrough collected, Lane 2: Wash fraction collected and Lane 3: Elution fraction of pure GstDnaI*

4.3.2. ^{19}F NMR Spectra of tfmF labelled *Gst DnaB* helicase

4.3.2.1. Effect of solubility tag

All the mutants were expressed with solubility tag on the N-terminus of the protein. The tag was cleaved by TEV protease and removed by second round of Ni-NTA column purification, in which *Gst DnaB* without the tag was eluted in the flow through. The tfmF labelled DnaB was examined by ^{19}F NMR with and without the solubility tag, to study if the tag had any effect on the ^{19}F resonance and/or on the structure of DnaB.

The ^{19}F NMR spectra displayed there was no change in chemical shift or the line width of the resonances observed in the presence and absence of the solubility tag. It was concluded that tag did not affect the ^{19}F resonances and on the structure of DnaB. For further ^{19}F NMR experiments the solubility tag was not removed.

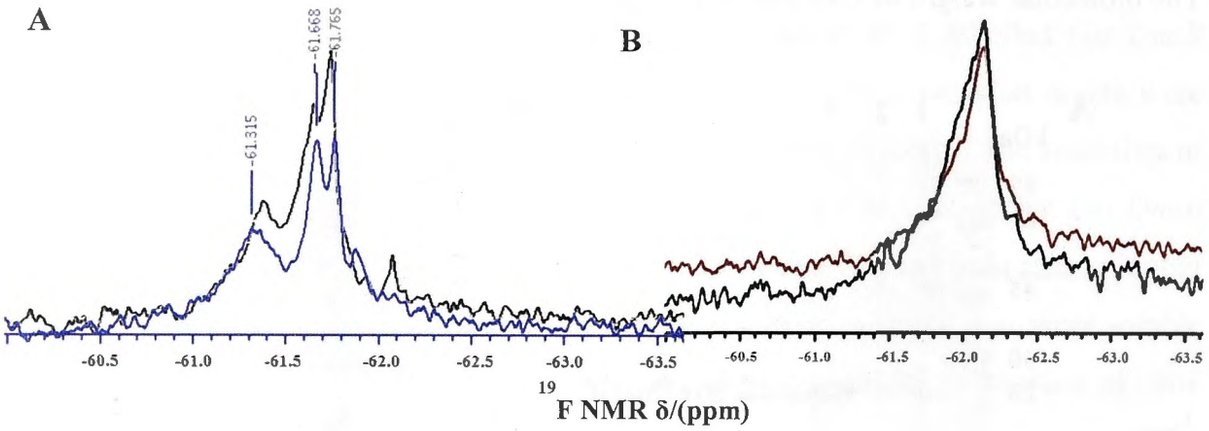

***Figure 4.4**: ^{19}F spectra of Gst DnaB mutants with tfmF incorporated at Y130 and Y104. A) ^{19}F spectra of mutant Y130tfmF with (black spectrum) and without the tag (blue spectrum) have been displayed. B) ^{19}F spectra of mutant Y104tfmF with (black spectrum) and without the tag (red spectrum) displaying the ^{19}F resonance.*

4.3.2.2. Effect of temperature

Liu and co-workers have previously shown that in ^{19}F NMR the fluorine relaxation depends on the measurement temperature. They demonstrated that ^{19}F resonances of 2, 2, 2-trifluoroethanethiol incorporated into G-protein coupled receptors became narrower with the rise in temperature [33]. The reduction in the line width is attributed to faster tumbling of the molecules at higher temperature and couplings between spins are more efficiently averaged out. The optimal growth temperature of *B.stearothermophilus* is 60 - 65 °C, therefore we hypothesized that *Gst DnaB* retains its structure to a temperature ≥50 °C. It was observed that *Gst DnaB* was stable at 25 °C for 12 h of measurement. The ^{19}F NMR measurement temperature was increase to 42 °C for mutant Y104tfmF to observe a sharper resonance.

^{19}F NMR spectra of mutant Y104tfmF had single resonance at -62.2 ppm at 25 °C. At 42°C, there was two well-resolved resonances one at -61.7 ppm and another broad resonance at -61.9 ppm with a split. Apparently, the resonance at -62.2 ppm has shifted downfield to -61.9 ppm at 42 °C. The ^{19}F spectrum at 42 °C was not as expected; instead of narrower resonance the chemical shift changed and new resonances appeared. The change in spectrum was attributed to the protein degradation. The protein samples of mutant Y104tfmF collected before and after the NMR measurement were analysed by SDS-PAGE. It showed that there was considerable degradation of protein into smaller fragments, confirming the hypothesis. The experiment conveyed two important facts for future studies. Firstly, tfmF labelled DnaB degrades significantly at 42 °C within 4 h (as opposed to 12 h at 25 °C), which is too short duration for practical NMR measurements. Secondly, the high potency of ^{19}F nucleus to probe changes in local structure. Hence, 25 °C was chosen as the optimum temperature for all the future ^{19}F NMR measurements.

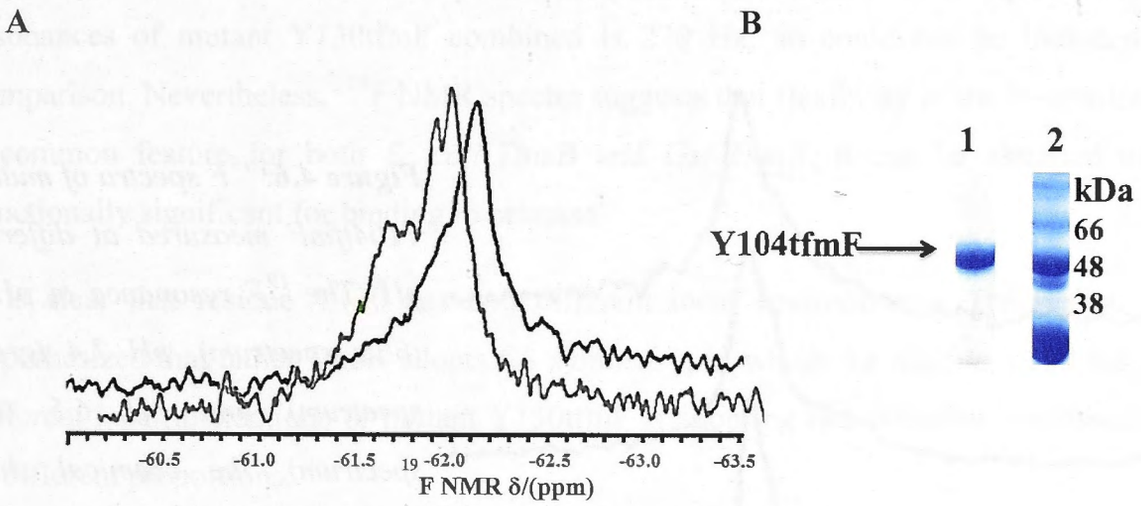

***Figure 4.5**: A) ^{19}F spectra shows comparison of the ^{19}F resonance measured at 25 °C (black spectrum) and 42 °C (green spectrum) for 4 h. B) SDS-PAGE gel showing the samples of Y104tfmF collected before and after the ^{19}F measured at 42 °C. Lane 1 shows the NMR sample Y104tfmF before the measurement and lane 2 shows the degraded sample after the measurement.*

4.3.2.3. Effect of pH

EM studies on *E. coli* DnaB have shown that pH influences the conformational equilibrium of DnaB hexamers. At basic pH, DnaB was predominantly in C3 symmetry and at acidic to neutral pH it assumed C6 symmetry and C3 symmetry in different proportion [7]. The ^{19}F studies carried out on *E. coli* DnaB demonstrated that pH induced conformational change in DnaB but could not reach firm conclusion due to the flexibility of the N-terminus in solution, as indicated by ^{19}F NMR study.

The flexibility of the N-terminus and the effect of pH on the symmetry of *Gst DnaB* were investigated by ^{19}F NMR. The mutants were dissolved in NMR buffer with varying pH (8.0, 7.4 and 6.5) and measured for 8 h at 25 °C. For all the mutants the resonances and line width remained the same with varying pH. The ^{19}F spectra of mutant Y104tfmF at different pH is shown in Figure 4.6 as an example. From the figure, on can clearly observe that pH does not affect the conformational equilibrium.

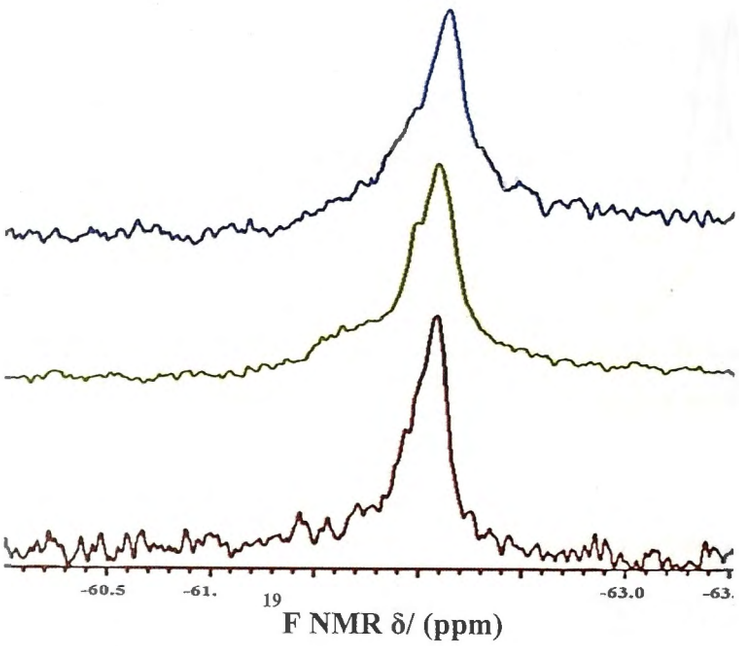

***Figure 4.6**: ^{19}F spectra of mutant Y104tfmF measured at different pH. The ^{19}F resonance at pH 8 (blue spectrum), pH 7.4 (green spectrum) and pH 6.5 (red spectrum), the chemical shifts and line width are the same.*

4.3.2.4.1. Comparison of the ^{19}F spectra of tfmF labelled DnaB measured at pH 8.0

The ^{19}F spectra of all the mutants were compared to deduce the structural characteristic of *Gst DnaB*. It is evident from the distinctive chemical shifts observed for each mutant that

tfmF incorporated at each site has different local environment reassuring that site-specific incorporation of tfmF by *in vivo* expression was success.

Comparison of the ^{19}F chemical shift of mutant F52tfmF (-60.5 ppm) and mutant Y104tfmF (-62.2 ppm) with the ^{19}F chemical shift of free tfmF amino acid (-62.13 ppm), indicates that residue F52 is buried and Y104 is solvent exposed because chemical shift of mutant Y104tfmF is closer to the chemical shift of free tfmF. MutantY130tfmF have two resonances at -61.3 ppm and -61.7 ppm comparing with free tfmF shows that some Y130 residues are buried deeper while others are moderately solvent exposed indicating the two different chemical environments. Among all the residues, F52 is most buried residue in the protein.

The line width of the resonances of mutants F52tfmF (93.5 Hz) and Y104tfmF (150 Hz) were compared. The broadening of resonances indicates that the N-terminus is mobile, consistent with our previous observations on *E. coli* DnaB. The line width of the two resonances of mutant Y130tfmF combined is 270 Hz, so could not be included for comparison. Nevertheless, ^{19}F NMR spectra suggests that flexibility of the N-terminus is a common feature for both *E. coli* DnaB and *Gst DnaB*, it can be assumed to be functionally significant for binding to primase.

It is clear that residue Y130 has two different local environments. The reason was hypothesized that either DnaB adopts C3 symmetry, in which the residue Y130 has two different local interactions or mutant Y130tfmF is adopting two different conformations in different proportions.

The crystal structure (PDB ID: 2R6C) [11] shows that all the residues of Y130 has same local environment in C3 symmetry, so ideally there should be only one fluorine resonance. Another crystal structure of *Gst DnaB* with ssDNA and GDP-AlF_4 during translocation showed that it formed a right handed helical shape arranged in a staircase like structure. In the crystal, the five N-terminal domains are in contact with neighbouring C-terminal domains except the N-terminal domain of first monomer, which is free of contact [35]. Subsequently, mutant Y130tfmF will have two ^{19}F resonances with an intensity difference of 5:1 instead the intensity of the resonances is in ratio of 2:1

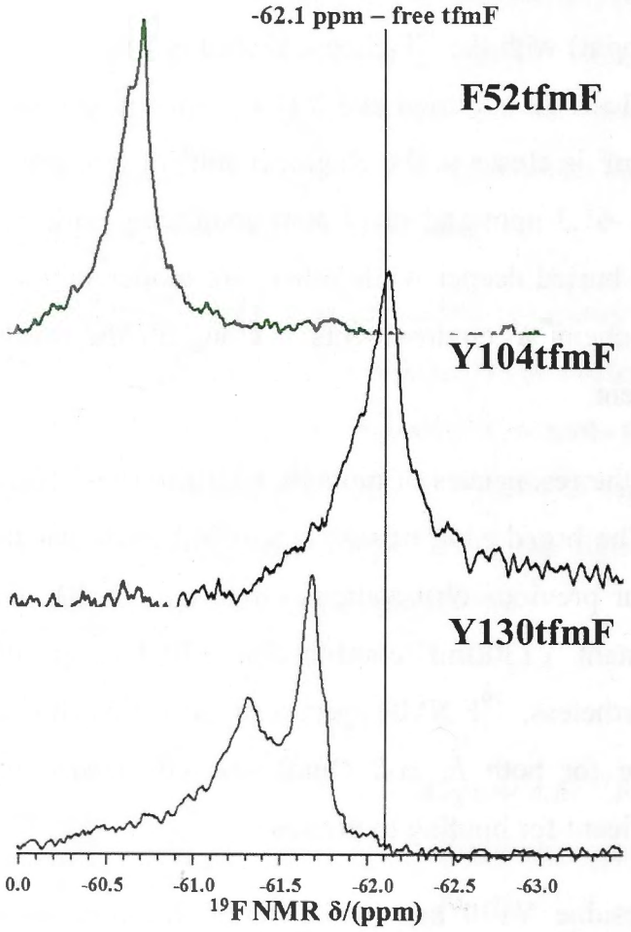

***Figure 4.7**: Comparison of ^{19}F spectra of tfmF labelled GstDnaB at different positions in the N-terminal domain without Mg^{2+} and ATP. The straight line depicts the position of the chemical shift of free tfmF amino acid at -62.13 ppm.*

The AFM images of *Gst DnaB* showed that DnaB adopted C3 and C6 symmetry in different proportions and assumes only C3 symmetry when bound to DnaG [12]. This strongly supports the latter hypothesis of DnaB adopting two different conformations. The two ^{19}F resonances with different intensities suggest that there is one dominant conformation producing higher intensity resonance. The ratio of two conformations could be in the ratio of 2:1 represented by the intensity of the resonances. The reason for the lack of indication of conformational equilibrium in mutants F52tfmF and Y104tfmF ^{19}F spectra could be due to the flexibility of the N-terminus.

4.3.2.5. Effect of Magnesium ion

Mg^{2+} aids in formation of *E. coli* DnaB hexamers, preserves its integrity and is required for complex formation with DnaC along with nucleotide [34,36-38]. However, *Gst DnaB* formed stable hexamers in the absence of Mg^{2+} [10,11,35,39]. So far, there are no studies indicating the Mg^{2+} effect on the structure or/and oligomerisation of *Gst DnaB*. Therefore, Mg^{2+} effect on tfmF labelled *Gst DnaB* was studied by ^{19}F NMR and analytical gel filtration.

5 mM $MgCl_2$ was added to the NMR buffer and all the mutants were measured with each ^{19}F NMR spectrum acquired in 2 h to record the dynamic ^{19}F chemical shift changes. The ^{19}F spectra of the mutants were than compared to the earlier ^{19}F NMR spectra measured without Mg^{2+} to observe the spectral difference.

^{19}F NMR spectra of mutants Y104tfmF and F52tfmF with Mg^{2+} had single resonances at the same chemical shifts as their respective spectra without Mg^{2+}. The line widths of the resonances were similar. Since, there was no spectral difference, it was deduced that Mg^{2+} had no effect on the structure and oligomerisation of the mutants Y104tfmF and F52tfmF.

^{19}F spectrum of mutant Y130tfmF -Mg^{2+} has three prominent resonances at -61.3 ppm, -61.65 ppm and -62.76 ppm. Without Mg^{2+}, mutant Y130tfmF had two ^{19}F resonances at at -61.3 ppm and -61.7 ppm. Though, the resonance at -61.3 ppm is present in both the spectra, the resonance at -61.7 ppm has shifted and split into two resonances in the presence of Mg^{2+}. The time course experiment monitored the appearance of new resonances and their changing intensities for 12 h. The change in chemical shift clearly indicates the structural change in mutant Y130tfmF. The dynamic nature of intensities hypothetically indicates the probable degradation of the protein.

The analysis of before and after NMR protein samples by SDS-PAGE showed that proteins of mutants Y104tfmF and F52tfmF were intact except for mutant Y130tfmF, which showed significant degradation. Mg^{2+} affected the stability of mutant Y130tfmF, which is otherwise the stable mutant. Mg^{2+} was affecting either the structure or the integrity of hexamers of mutant Y130tfmF.

The tfmF labelled DnaB was subjected to gel filtration to investigate the hexamer formation with and without Mg^{2+}. The analytical gel filtration profile in the absence of Mg^{2+} showed that tfmF labelled DnaB eluted at 10 ml corresponding to the elution of hexamers (Fig 4.8) which is supported by previous work using the same column [10,39]. The SDS-PAGE analysis of the elution fractions verified the presence of *Gst DnaB* and demonstrated all the mutants formed hexamers without Mg^{2+}.

In the presence of 5 mM Mg^{2+}, the gel filtration profile of all the mutants showed a relatively high UV absorbance peak at 10 ml elution volume corresponding to hexamer and lower absorbance peak at 15 ml corresponding to the elution of monomer. The well-resolved elution peaks displayed that hexamers were well-separated from monomers. However, the relative intensity of absorbance peak at 15 ml varied among mutants. The SDS-PAGE analysis of the elution fractions showed that *Gst DnaB* was present in both 10 ml and 15 ml fractions indicating that DnaB eluted as hexamers and monomers. The data indicates that either Mg^{2+} affect the integrity of tfmF labelled DnaB hexamers or the protein sample has metalloproteases. However, the absence of the degraded protein after the gel filtration excludes the presence of metalloproteases.

The extent of monomer formation is shown to be dependent on the mutation site. The data showed that mutant Y104tfmF has the least tendency to form monomers whereas mutant Y130tfmF has maximum tendency to form monomers in presence of Mg^{2+}, which is evident from the varying intensity of absorbance of the mutants' monomers. The higher tendency of mutant Y130tfmF to form monomers suggests that residue Y130 might play a significant role in oligomerisation of *Gst DnaB*. The formation of monomer by mutant Y130tfmF and its degradation during ^{19}F NMR indicates that monomers are less stable than the hexamers and susceptible to degradation. The degradation of other mutants was not evident because the proportion of monomers was significantly less.

It is clear that Mg^{2+} does not enhance the integrity of *Gst DnaB* hexamers as expected instead it triggers formation of monomer, which is mutation site dependent. The effect of Mg^{2+} on the integrity of hexamers may be confined to *Gst DnaB* because in case of *Thermus aquaticus* DnaB (*Taq* DnaB), the monomer of the protein were crystallised

instead of hexamer [40] and mass spectrometry study has reported that Mg^{2+} stabilised *E. coli* DnaB oligomers of higher order [41].

The study revealed three new aspects on *Gst DnaB*, one is the Mg^{2+}induces formation of monomers in tfmF labelled *Gst DnaB*, the second is that monomers are less stable and last but not the least is that residue Y130 might be significant in stabilising hexamers or is necessary for interaction between monomers.

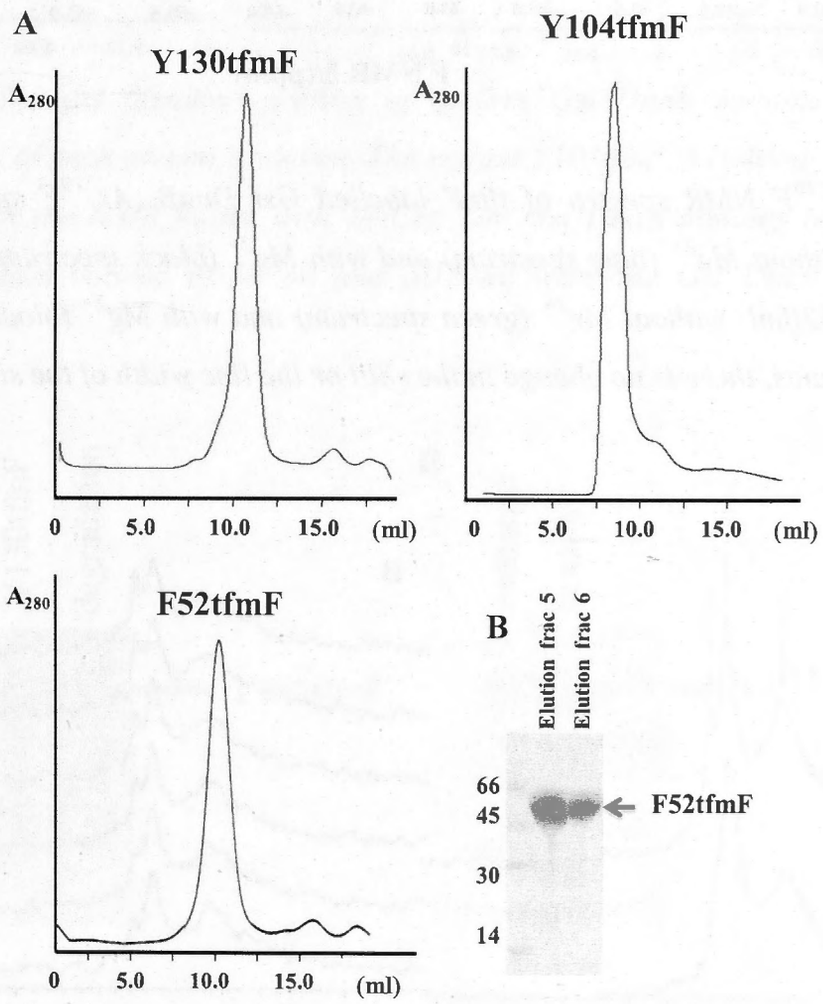

***Figure 4.8**: The gel filtration profiles of various Gst DnaB mutants. A) The gel filtration profile of each mutant is shown. The mutant Y103tfmF, Y104tfmF and F52tfmF were eluted with the NMR buffer without $MgCl_2$. The Gst DnaB mutants have eluted as hexamer at elution volume of 10 ml. B) A representative SDS-PAGE gel analysis of the elution fraction containing hexamer F52tfmF is shown. The arrows indicate the protein on the gel.*

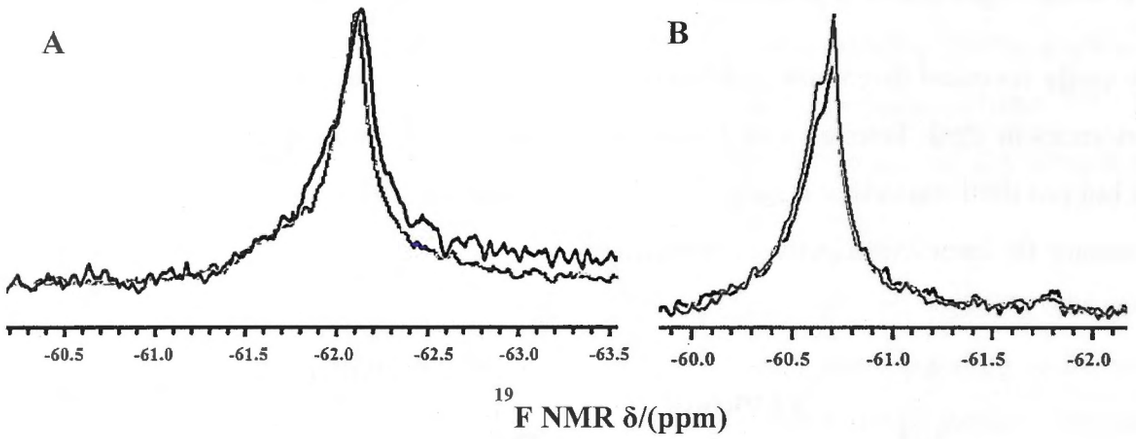

Figure 4.9: *^{19}F NMR spectra of tfmF labelled Gst DnaB. A) ^{19}F spectra of mutant Y104tfmF without Mg^{2+} (blue spectrum) and with Mg^{2+} (black spectrum). B) ^{19}F spectra of mutant F52tfmF without Mg^{2+} (green spectrum) and with Mg^{2+} (black spectrum). For both the mutants, there is no change in the shift or the line width of the signal..*

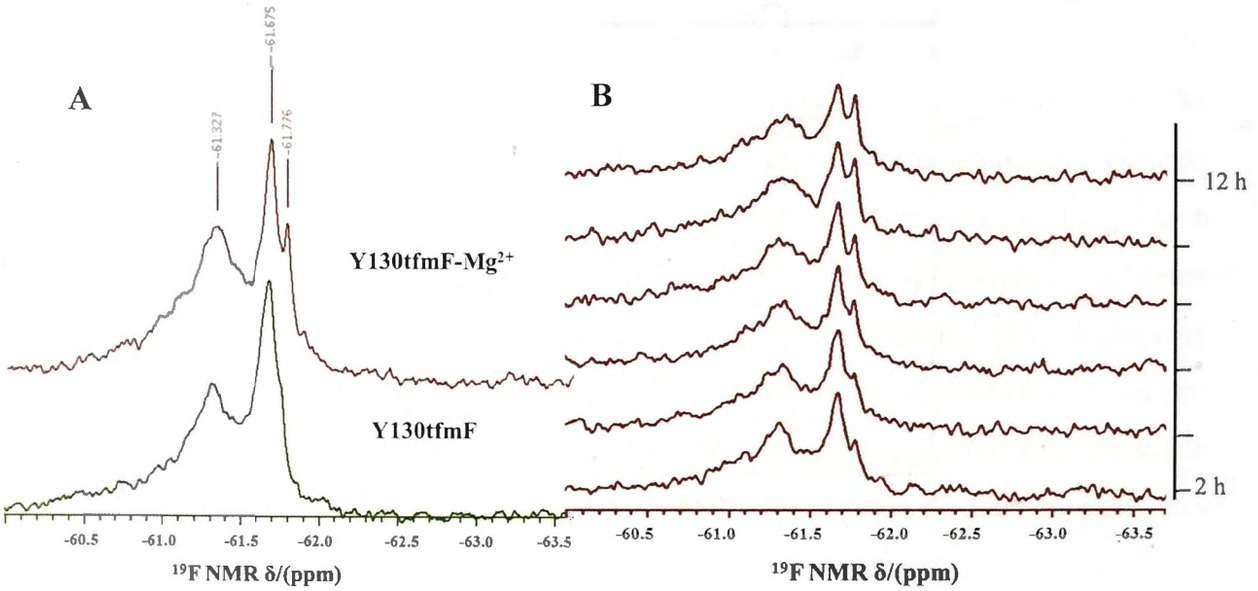

Figure 4.10: *^{19}F NMR spectra of tfmF labelled Gst DnaB mutant Y130tfmF. A) ^{19}F spectra of Y130tfmF measured with (red spectrum) and without Mg^{2+} (green spectrum) that shows the change in chemical shifts with appearance of new signals. The new signals are indicated with the values of chemical shifts. B) The time course measurement of mutant Y130tfmF with Mg^{2+} for 12 h with 4200 transients collected every 2 h. Plot of intensity change in mutant Y130tfmF-Mg^{2+} as function of time.*

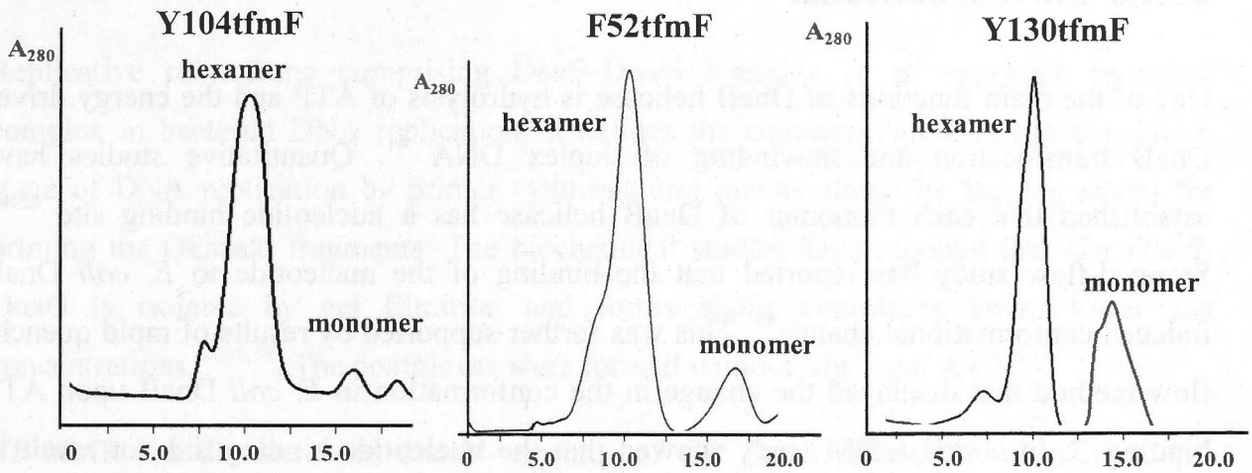

***Figure 4.11**: The gel filtration profiles of various Gst DnaB mutants. A) The gel filtration profile of each mutant is shown. The mutant Y103tfmF, Y104tfmF and F52tfmF were eluted with the NMR buffer with $MgCl_2$. The Gst DnaB mutants have eluted as hexamer at elution volume of 10 ml and at15 ml were the Gst DnaB is eluted as monomer.*

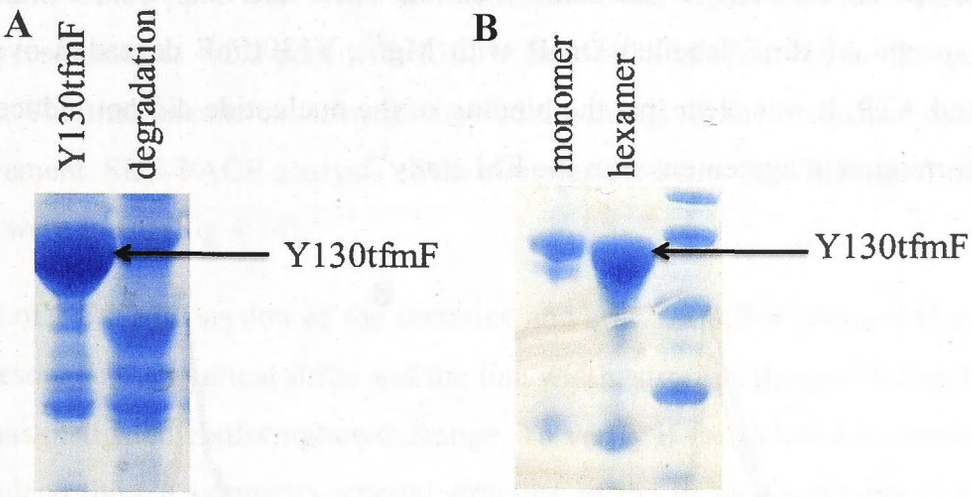

***Figure 4.12**: SDS-PAGE gels showing the mutant Y130tfmF protein samples of NMR and analytical gel filtration. A) SDS-PAGE gel showing the samples before (Y130tfmF) and after (degradation) the ^{19}F NMR measurement. B) SDS-PAGE gels displaying the samples which were collected under elutions fractions of hexamer and monomer.*

4.3.2.6. Effect of nucleotide

One of the main functions of DnaB helicase is hydrolysis of ATP and the energy drives DnaB translocation and unwinding of duplex DNA [41]. Quantitative studies have established that each monomer of DnaB helicase has a nucleotide binding site [42-44]. Stopped-flow study has reported that the binding of the nucleotide to *E. coli* DnaB induced conformational change[43]. This was further supported by results of rapid quench-flow method that displayed the change in the conformation in *E. coli* DnaB upon ATP binding [44]. In contrast, EM study showed that the nucleotide binding did not result in conformational change of *E. coli* DnaB [34]. This would be first time to investigate the conformational change in *Gst DnaB* due to ATP. *Gst DnaB* was dissolved in NMR buffer with 5 mM $MgCl_2$ and 1 mM ATP. The ^{19}F spectrum of each mutant with ATP was compared with their respective ^{19}F spectrum with Mg^{2+}.

The comparison of ^{19}F NMR spectra of unbound and ATP bound state of *Gst DnaB* showed that all the mutants had same chemical shifts and line width as observed in ^{19}F NMR spectra of tfmF labelled DnaB with Mg^{2+}. Y130tfmF degraded over time with Mg^{2+} and ATP. It was clear that the binding of the nucleotide did not induce any change in conformation in agreement with the EM study[34].

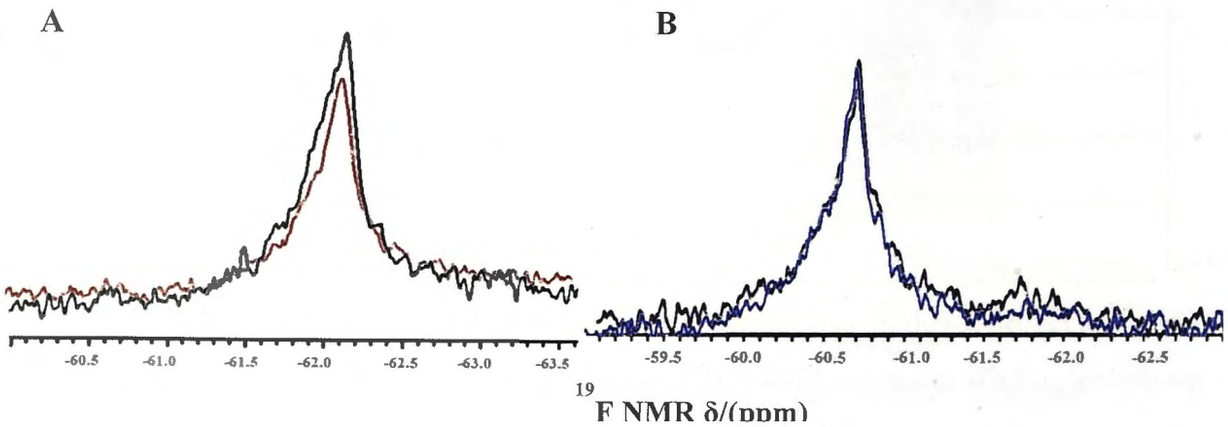

***Figure 4.13**: ^{19}F spectra of mutants Y104tfmF and F52tfmF measured with ATP and without ATP. A) ^{19}F NMR spectra of mutant Y104tfmF with ATP (red) and without ATP (black). B) ^{19}F NMR spectra of mutant F52tfmF with ATP (blue) and without ATP (black).*

4.3.2.4. Effect of DnaGC

Replicative primosome comprising DnaB-DnaG complex is an important initiating complex in bacterial DNA replication. It defines the commencement of the elongation stage of DNA replication by primer synthesis and moves along the lagging strand for priming the Okazaki fragments. The biochemical studies have reported that *Gst DnaB*-DnaG is isolable by gel filtration and forms stable complexes under lower salt concentrations [10,11,39]. The complexes were formed without Mg^{2+} and ATP [9].

^{19}F NMR has demonstrated that N-terminus of *Gst DnaB* is flexible in solution. We speculated that interaction with DnaGC induces rigidity to the N-terminus of DnaB. The *Gst DnaB*-DnaGC complex would be good system to test, whether ^{19}F NMR is a sensitive tool to investigate the conformational changes in proteins as large as 375 kDa complex.

^{19}F NMR spectra of mutant Y104tfmF-DnaGC displayed two well-resolved resonances at -61.52 ppm and -62.06 ppm with same line width of 90 Hz separated by 0.5 ppm in the place of one resonance at -62.2 ppm with 150 Hz line width as observed in the spectra of its apo form. The time course experiment showed no change in the resonance during the 12 h measurement. SDS-PAGE analysis of the complex after the ^{19}F NMR confirmed that the proteins were intact (Fig 4.14).

Comparison of ^{19}F NMR spectra of the complex and apo Y104tfmF showed changes in number of resonances, chemical shifts and the line width, strongly proposing that binding of DnaGC has instigated conformational change. To verify if the induced conformational change is indeed the C3 symmetry, crystal structure of the DnaB-DnaG complex (PDB ID: 2R6C) was analysed to identify the local environment of each Y104.

It revealed that three residues of Y104 are solvent exposed and other three are buried inside *Gst DnaB* hexamer, which clearly is depicted by the two ^{19}F signals. Moreover, the equal numbers of residues interacting with different local environment explains the resonances with same intensity and line width. It clearly depicts that DnaGC instigates conformation change by inducing rigidity to the N-terminus by arranging them to form C3 symmetry (Fig 4.16).

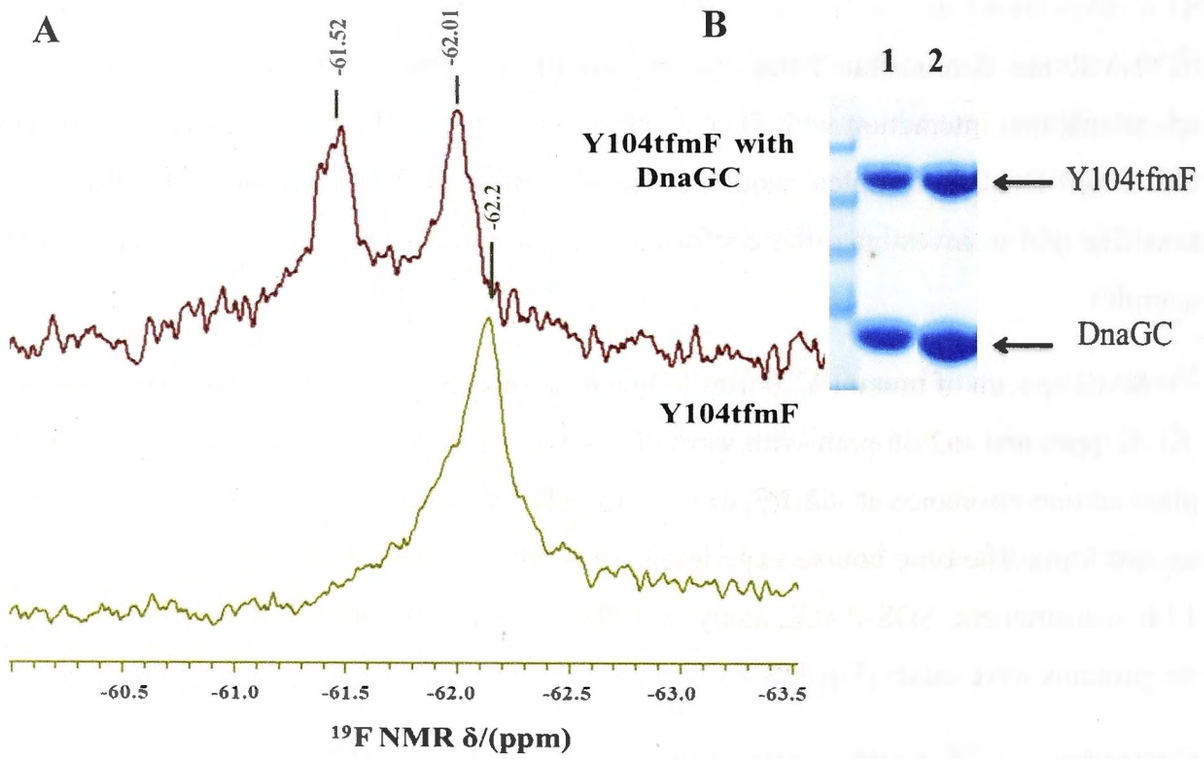

Figure 4.14: *A) ^{19}F spectra of apo mutant Y104tfmF displaying the single resonance of apo form (green) and Y104tfmF-DnaGC having two resonances (red). B) SDS-PAGE gel displaying the complex of mutant Y104tfmF -DnaGC. Lane 1 displays the before ^{19}F NMR sample and lane 2 displays after ^{19}F NMR sample of the intact complex.*

The ^{19}F NMR of mutant Y130tfmF-DnaGC showed single broad resonance at -61.63 ppm with 300 Hz line width which replaced two ^{19}F resonances in spectra of apo Y130tfmF (Fig 4.16). Though the change in number of resonances strongly indicates the structural change, the pattern of spectral change is different from mutant Y104tfmF.

The single resonance suggests that all the residue of mutant Y130tfmF has same local environment and was consistent with the observations from the crystal structure of the complex. All the Y130 residues indeed have same local environment in C3 symmetry, which is depicted by single ^{19}F signal (Fig 4.17).

The broadening of the resonance can be attributed to the DnaB interaction with DnaGC. Thirlway and coworkers reported that DnaGC binds to the linker helix of DnaB[12]. As residue of mutant Y130tfmF is closer to the linker helix, the induced rigidity of the region might have led to broadening of the resonance. The broad signal confirms that DnaG binds to the linker helix.

The ^{19}F NMR structural information being consistent with previous results demonstrates that ^{19}F NMR is capable of depicting the precise structure under physiological conditions.

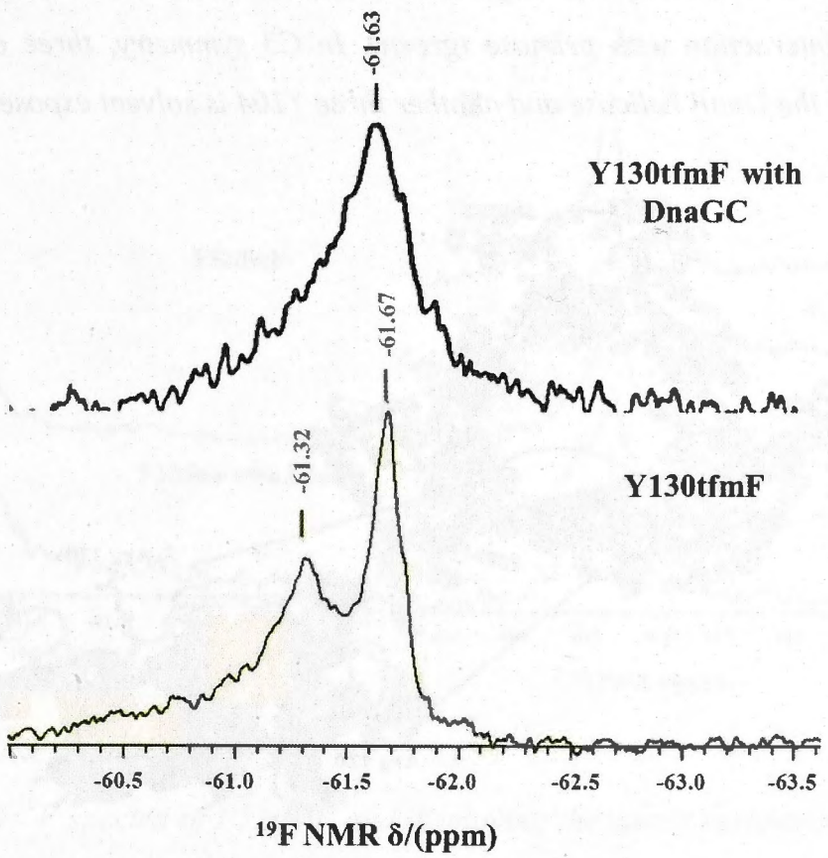

Figure 4.15: *Comparison of ^{19}F spectra of mutant Y130tfmF in apo and in complex with DnaGC. ^{19}F spectra of Y130tfmF displaying the two resonances of apo form (green) and the single resonance in complex with DnaGC (black).The chemical shifts are mentioned on the tip of the signals in ^{19}F NMR spectrum of the complex*

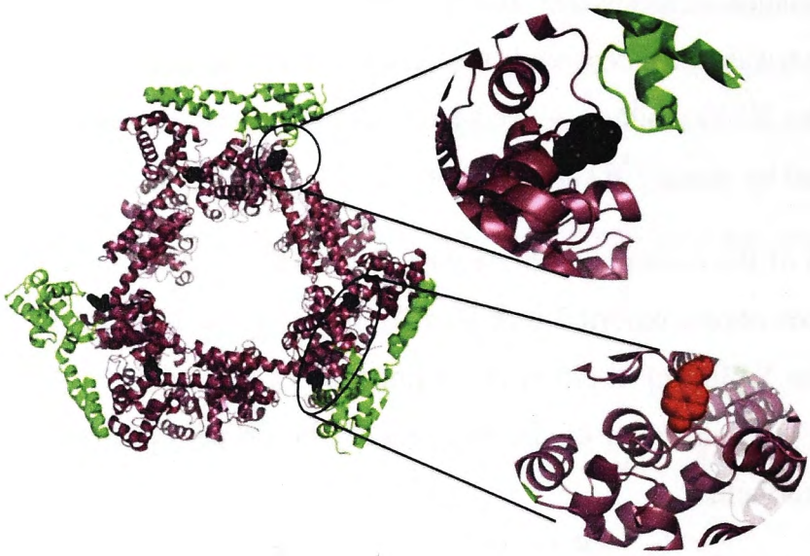

***Figure 4.16**: The structure of DnaB helicase (2R6C) displaying the six residues of Y104 (black) as spheres in C3 symmetry. The placement of one of the Y104 (red) is buried inside the DnaB hexamer (raspberry) and another Y104 (black) is solvent exposed and involved in interaction with primase (green). In C3 symmetry, three of the Y104 are buried inside the DnaB helicase and another three Y104 is solvent exposed.*

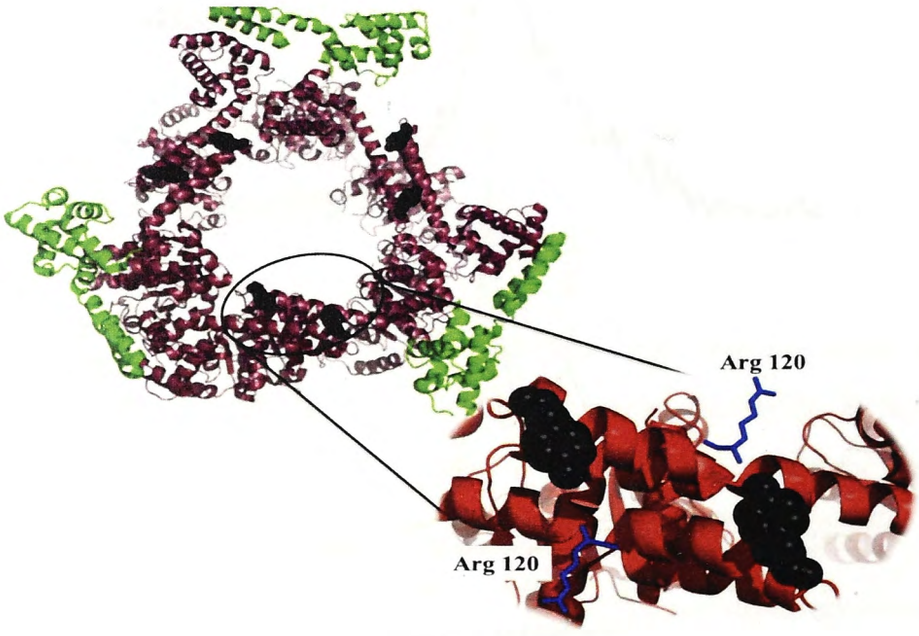

***Figure 4.17**: The structure of DnaB helicase (2R6C) displaying the six residues of Y130 (black) as spheres in C3 symmetry. The placement of all six Y130 (black) is solvent exposed and interacts with same residues. Therefore, all the tfmF have the same surrounding environment.*

^{19}F spectrum of mutant F52tfmF-DnaGC had single resonance at -60.6 ppm. There is no change in the chemical shift between the apo form and the complex. The single resonance suggests that all F52 are surrounded had same local environment in C3 symmetry and was confirmed by the crystal structure analysis, which showed residue F52 is buried inside the protein. Though, the ^{19}F resonance complies with crystal structure in terms of number of signals, the conformational change is not being evident from the spectrum. It could be either because of the flexibility of the N-terminus or mutant F52tfmF-DnaGC complex is not as stable.

However, the ^{19}F resonance of the complex is broad in the beginning of the NMR and narrows with time. The line width reduced from 181 Hz to 93 Hz over 10 h. The change in line width of the ^{19}F resonance is evident from the time course experiment, which displays the sharpening of the line width is gradual. The line width of the resonance in the final spectrum (93 Hz) is same as the line width of the resonance in its apo form.

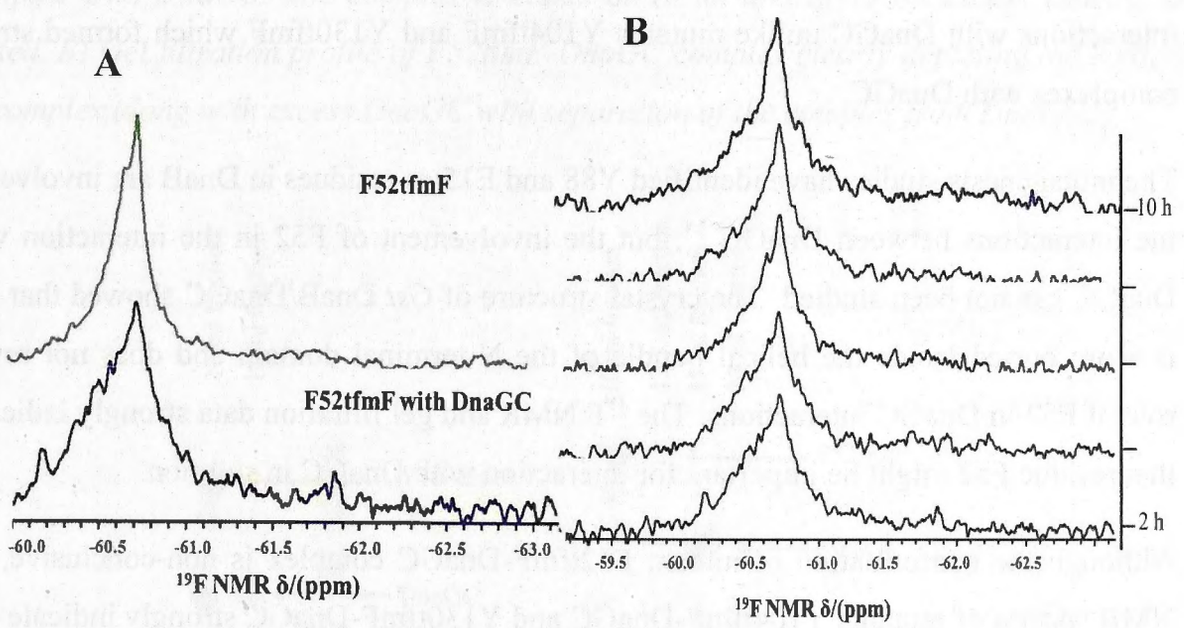

***Figure 4.19**: A) ^{19}F spectra of F52tfmF and displaying the single resonance of apo form (green) and the two resonances in complex with DnaGC (blue). B) Time course of the F52tfmF-DnaGC complex dissociation over time of 10 h.*

This is an indication that DnaGC might be dissociating from mutant F52tfmF due to weak interaction. This was further confirmed with analytical gel filtration.

The mutants Y130tfmF and Y104tfmF formed strong complexes with DnaGC and were isolable by gel filtration. For these complexes, tfmF labelled DnaB-DnaGC complex was well separated from the excess DnaGC verified by SDS-PAGE.

The gel filtration profile of mutant F52tfmF-DnaGC was distinctive; the elution peaks were not separated. The elution of the protein at 20 ml suggested possible elution of degradation product. SDS-PAGE analysis showed the elution fraction at 10 ml had DnaB hexamer predominantly with very small fraction DnaGC. The elution volume at 13 ml had minimal proportions of DnaB and large proportion of DnaGC, suggesting that the proteins were eluted in different fractions in various proportions because of weak interactions.

The gel filtration and SDS-PAGE analysis proved that mutant F52tfmF had weak interactions with DnaGC unlike mutants Y104tfmF and Y130tfmF which formed strong complexes with DnaGC.

The mutagenesis studies have identified Y88 and E15 as residues in DnaB are involved in the interactions between DnaGC [14], but the involvement of F52 in the interaction with DnaGC has not been studied. The crystal structure of *Gst* DnaB/DnaGC showed that F52 is more buried inside the helical bundle of the N-terminal domain and does not reveal role of F52 in DnaGC interactions. The ^{19}F NMR and gel filtration data strongly indicates that residue F52 might be important for interaction with DnaGC in solution.

Although the conformation of mutant F52tfmF-DnaGC complex is non-conclusive, ^{19}F NMR spectra of mutants Y104tfmF-DnaGC and Y130tfmF-DnaGC strongly indicate that DnaGC forms strong complex with DnaB by interacting at the linker region and arranges the otherwise flexible N-terminus to adopt C3 symmetry.

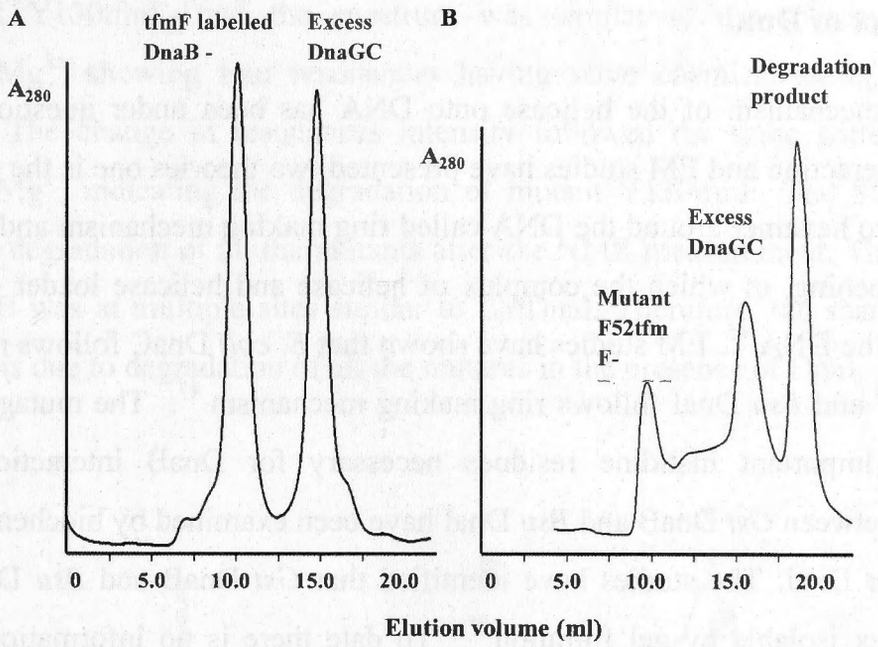

***Figure 4.20**: The gel filtration profiles of various Gst DnaB mutants in complex with DnaGC. A) Representative gel filtration profile for mutant Y130tfmF, Y104tfmF in complex with DnaGC. The complex is eluted at 10 ml and at 15 ml excess DnaGC is eluted. B) Gel filtration profile of F52tfmF-DnaGC complex clearly depicting the elution of complex along with excess DnaGC with separation of the complex from DnaGC.*

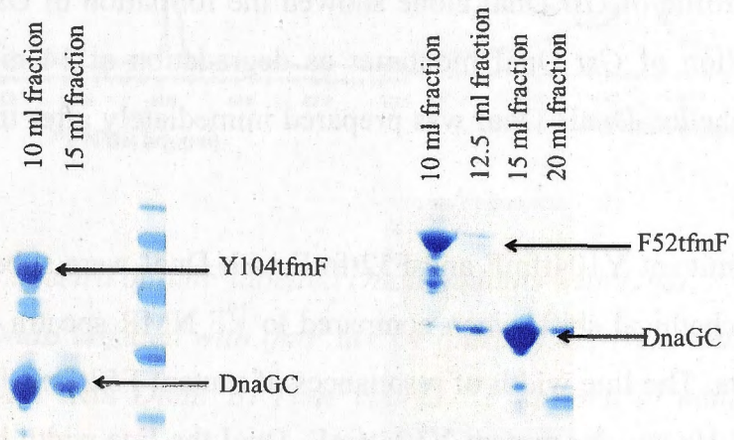

***Figure 4.21**: SDS-PAGE gels showing the samples of analytical gel filtration. A) SDS-PAGE gel showing the samples of mutant Y104tfmF collected as a complex and excess of DnaGC monomer collected. B) SDS-PAGE gels displaying the samples which were collected under elutions fractions of complex 1(F52tfmF), complex 2(F52tfmF+DnaGC) and monomer of DnaGC elution volumes.*

4.3.3.2. Effect of DnaI

The loading mechanism of the helicase onto DNA has been under question for many years. The interaction and EM studies have presented two theories one is the assembly of monomers into hexamer around the DNA called ring making mechanism and the other is called ring opening, in which the complex of helicase and helicase loader opens up to wrap around the DNA [45]. EM studies have shown that *E. coli* DnaC follows ring opening mechanism [46] and *Bsu* DnaI follows ring making mechanism [47]. The mutagenic studies showed the important histidine residues necessary for DnaB interactions [48]. The interactions between *Gst* DnaB and *Bsu* DnaI have been examined by biochemical studies instead of *Gst* DnaI. The studies have identified that *Gst* DnaB and *Bsu* DnaI forms a stable complex isolable by gel filtration[47]. To date there is no information about *Gst* DnaI, these are first attempts to understand and study *Gst* DnaI and its interaction with *Gst* DnaB.

The comparison of primary sequence of *Bsu* DnaI and *Gst* DnaI indicated that they shared the zinc binding fold, cysteines and histidines required for interactions with the C-terminal domain of DnaB.

The gel filtration profile of *Gst* DnaI alone showed the formation of *Gst*DnaI oligomers along with the elution of *Gst* DnaI monomer as degradation at 14 ml. Therefore, the complex of tfmF labelled DnaB-DnaI was prepared immediately after the purification of DnaI.

The ^{19}F signals of mutant Y104tfmF and F52tfmF with DnaI were substantially sharper without change in chemical shift, when compared to ^{19}F NMR spectra of the respective apo form of mutants. The line width of resonances of mutant F52tfmF-DnaI has reduced to 30 Hz from 93.5 Hz and for mutant Y104tfmF- DnaI the line width has reduced from 150Hz to 20 Hz. We speculated that either increase in mobility of the N-terminus or due degradation of DnaB in the presence of DnaI might be the reason for sharper signals.

For mutant Y130tfmF-DnaI the spectrum was similar to the spectrum of mutant Y130tfmF-Mg^{2+} showing four resonances having same chemical shifts with varying intensities. The change in resonances intensity followed the same pattern as mutant Y130tfmF-Mg^{2+} indicating the degradation of mutant Y130tfmF. The SDS-PAGE gel showed the degradation of all the mutants after the NMR measurement. The degradation of the DnaB was at multiple sites similar to *Gst*DnaI. Therefore, the sharpening of the resonances is due to degradation of all the mutants in the presence of DnaI.

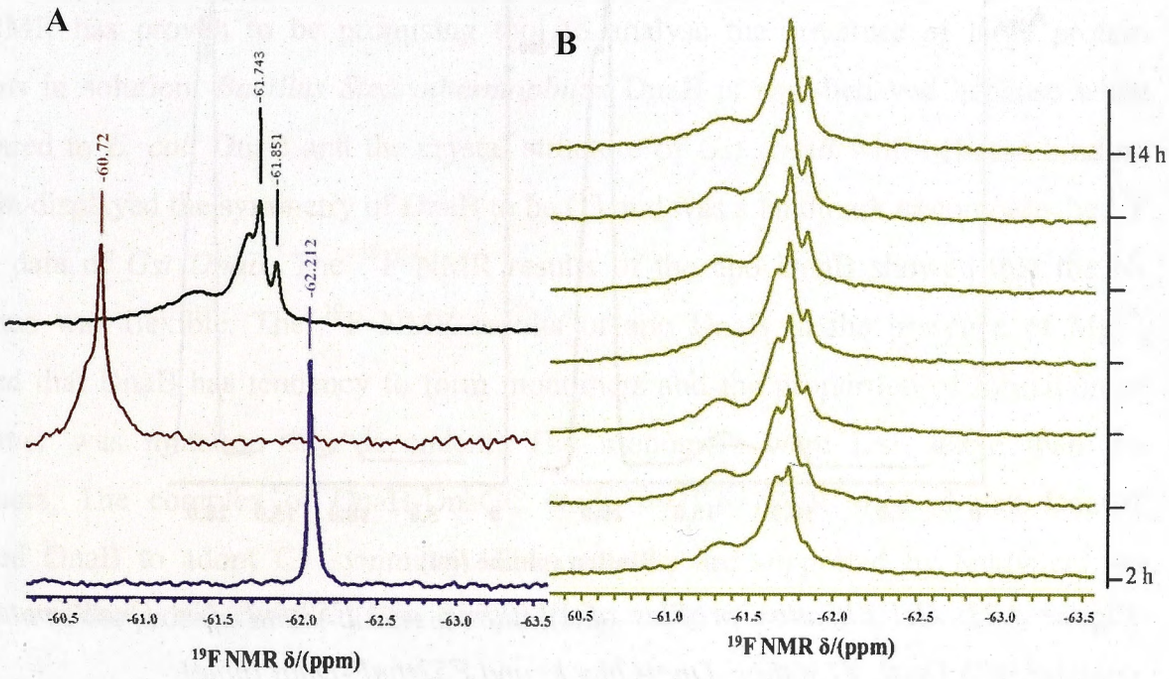

Figure 4.22: *^{19}F spectra of tfmF labelled DnaB mutants with DnaI. A) Comparison of ^{19}F spectra of Gst DnaB labelled with tfmF at 104 (purple), 52 (red) and 130 (green) in the N-terminal domain with DnaI. B) Time course ^{19}F spectra of mutant Y130tfmF-DnaI displaying the change in intensity of signals over the time.*

The complex of mutants F52tfmF and Y130tfmF with DnaI was passed through gel filtration to possibly isolate the complex of DnaB-DnaI without degradation. The absorbance peaks were at 10 ml, 14 ml and 20 ml. The SDS-PAGE analysis of the elution

fractions showed that the degraded products DnaB and DnaI were eluted at 14 ml and 20 ml. A small quantity of DnaB was eluted as hexamer at 10 ml.

We speculate that DnaB and DnaI have interaction which might be influencing the degradation of DnaB. To eliminate this possibility, DnaI sample was mixed with lysozyme and incubated for 30 mins on ice and analysed by SDS-PAGE which showed no degradation of lysozyme. This demonstrates that DnaI is likely to directly influence the structure of *Gst DnaB* and lead to enhanced degradation.

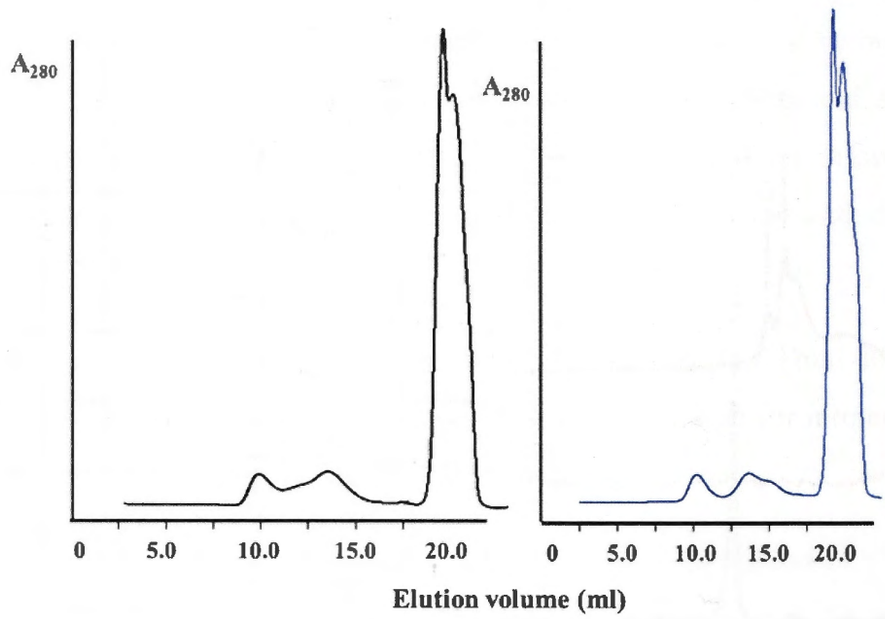

Figure 4.23: *Gel filtration profiles of Y130tfmF and F52tfmF Gst DnaB mutants in complex with DnaI. Y130tfmF-DnaI(black) and F52tfmF-DnaI (blue).*

The previous experiment showed that mutant Y130tfmF degraded in the presence of Mg^{2+} when it exists as monomer. With DnaI, all the mutants degraded rapidly leading to the speculation that probably *Gst*DnaI interacts with DnaB as monomers and as monomers are more susceptible to degradation. This speculation agrees well with the studies on *Bsu* DnaI reporting that helicase is assembled as hexamer from monomers by ring making mechanism[47]. *Gst*DnaI belonging to the same family as *Bsu* DnaI, might interact with monomers of DnaB for loading the helicase. The stability of *Gst*DnaI has to be tested before proceeding any further to study the helicase-helicase loader interactions.

More research has to be put into studying the interactions between *Gst*helicase-helicase loader to understand the loading mechanism in *Bacillus Stearothermophilus*. The interaction could be probed deeper by testing the complex formation in the presence of additives such as Mg^{2+} and ATP or with helicase loading partner such as DnaD and study by site-directed mutagenesis to identify key residues interacting between the helicase and its loader.

4.4. Conclusion

^{19}F NMR has proven to be promising tool to analyse the structure of large protein systems in solution. *Bacillus Stearothermophilus* DnaB is well-behaved helicase when compared to *E. coli* DnaB and the crystal structure of *Gst DnaB* with helicase binding domain displayed the symmetry of DnaB to be C3 and was a landmark to compare the ^{19}F NMR data of *Gst DnaB*. The ^{19}F NMR results of the apo DnaB showed that the N-terminus was flexible. The ^{19}F NMR results of apo DnaB in the presence of Mg^{2+}, showed that DnaB has tendency to form monomers and the proportion of formation of monomer was mutation site dependent. The monomers were less stable than the hexamers. The complex of DnaB-DnaGC study by ^{19}F NMR showed that DnaGC induced DnaB to adopt C3 symmetry. These results are supported by analytical gel filtration of the proteins under different conditions.

The interaction with *Gst*DnaI showed that DnaB is degraded after both ^{19}F NMR measurement and analytical gel filtration, suggesting that DnaB may interact with DnaI as monomer, and as monomers are structurally more dynamic, lead to degradation of the proteins. This system has to be studied more to completely understand the interactions but our results show promising directions on how to study such large, transient protein complexes. As the flexibility of the N-terminus is a common feature in both *E. coli* and *Gst DnaB*, it clearly must be functionally important for DNA replication.

Reference

1. Srivastava, R. A. K. & Baruah, J. N. Culture Conditions for Production of Thermostable Amylase by *Bacillus-Stearothermophilus*. Applied and Environmental Microbiology 52, 179-184 (1986).
2. Flint, S., Palmer, J., Bloemen, K., Brooks, J. & Crawford, R. The growth of *Bacillus stearothermophilus* on stainless steel. *Journal of Applied Microbiology* **90**, 151-157, (2001).
3. Bird, L. E. & Wigley, D. B. The *Bacillus stearothermophilus* replicative helicase: cloning, overexpression and activity. *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1444**, 424-428, (1999).
4. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry* **76**, 23-50, (2007).
5. Sanmartin, M. C., Stamford, N. P. J., Dammerova, N., Dixon, N. E. & Carazo, J. M. A Structural Model for the *Escherichia-Coli* Dnab Helicase Based on Electron-Microscopy Data. *Journal of Structural Biology* **114**, 167-176, (1995).
6. Yu, X., Jezewska, M. J., Bujalowski, W. & Egelman, E. H. The hexameric *E. coli* DnaB helicase can exist in different Quaternary states. *J Mol Biol* **259**, 7-14, (1996).
7. Donate, L. E. *et al.* pH-Controlled quaternary states of hexameric DnaB helicase. *Journal of Molecular Biology* **303**, 383-393, (2000).
8. Yang, S. X. *et al.* Flexibility of the rings: Structural asymmetry in the DnaB hexameric helicase. *Journal of Molecular Biology* **321**, 839-849, (2002).
9. Tougu, K., Peng, H. & Marians, K. J. Identification of a Domain of *Escherichia-Coli* Primase Required for Functional Interaction with the Dnab Helicase at the Replication Fork. *Journal of Biological Chemistry* **269**, 4675-4682 (1994).
10. Bird, L. E., Pan, H., Soultanas, P. & Wigley, D. B. Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from *Bacillus stearothermophilus*. *Biochemistry* **39**, 171-182, (2000).

11. Bailey, S., Eliason, W. K. & Steitz, T. A. Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* **318**, 459-463, (2007).
12. Thirlway, J. *et al.* DnaG interacts with a linker region that joins the N- and C-domains of DnaB and induces the formation of 3-fold symmetric rings. *Nucleic Acids Research* **32**, 2977-2986, (2004).
13. Chang, P. & Marians, K. J. Identification of a region of *Escherichia coli* DnaB required for functional interaction with DnaG at the replication fork. *Journal of Biological Chemistry* **275**, 26187-26195,
14. Lu, Y. B., Ratnakar, P. V. A. L., Mohanty, B. K. & Bastia, D. Direct physical interaction between DnaG primase and DnaB helicase of *Escherichia coli* is necessary for optimal synthesis of primer RNA. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 12902-12907,(1996).
15. Syson, K., Thirlway, J., Hounslow, A. M., Soultanas, P. & Waltho, J. P. Solution structure of the helicase-interaction domain of the primase DnaG: A model for helicase activation. *Structure* **13**, 609-616, (2005).
16. Makowska-Grzyska, M. & Kaguni, J. M. Primase Directs the Release of DnaC from DnaB. *Molecular Cell* **37**, 90-101, (2010).
17. Soultanas, P. A functional interaction between the putative primosomal protein DnaI and the main replicative DNA helicase DnaB in *Bacillus*. *Nucleic Acids Research* **30**, 966-974, doi:Doi 10.1093/Nar/30.4.966 (2002).
18. Rannou, O. *et al.* Functional interplay of DnaE polymerase, DnaG primase and DnaC helicase within a ternary complex, and primase to polymerase hand-off during lagging strand DNA replication in *Bacillus subtilis*. *Nucleic Acids Research* **41**, 5303-5320, (2013).
19. Mitkova, A. V., Khopde, S. M. & Biswas, S. B. Mechanism and stoichiometry of interaction of DnaG primase with DnaB helicase of *Escherichia coli* in RNA primer synthesis. *Journal of Biological Chemistry* **278**, 52253-52261, (2003).
20. Bin, L., Eliason, W. K. & Steitz, T. A. Structure of a helicase-helicase loader complex reveals insights into the mechanism of bacterial primosome assembly. *Nature Communications* **4**,(2013).

21. Loscha, K. V. *et al.* A novel zinc-binding fold in the helicase interaction domain of the *Bacillus subtilis* DnaI helicase loader. *Nucleic Acids Research* **37**, 2395-2404, (2009).
22. Barcena, M. *et al.* The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel. *EMBO J* **20**, 1462-1468,(2001).
23. Arias-Palomo, E., O'Shea, V. L., Hood, I. V. & Berger, J. M. The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. *Cell* **153**, 438-448, (2013).
24. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. coli*. *Journal of Molecular Biology* **395**, 361-374,(2010).
25. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Biochemistry* **52**, 1828-1837, (2013).
26. Puigbo, P., Guzman, E., Romeu, A. & Garcia-Vallve, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research* **35**, W126-W131, (2007).
27. Ozawa, K. *et al.* High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochemical and Biophysical Research Communications* **418**, 652-656, (2012).
28. Young, D. D. *et al.* An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **50**, 1894-1900, (2011).
29. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C. & Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat Protoc* **2**, 2601-2607, (2007).
30. Ozawa, K. *et al.* Proofreading exonuclease on a tether: the complex between the *E-coli* DNA polymerase III subunits alpha, epsilon, theta and beta reveals a highly flexible arrangement of the proofreading domain. *Nucleic Acids Research* **41**, 5354-5367, (2013).
31. Cheng, Y. & Patel, D. J. An efficient system for small protein expression and refolding (vol 317, pg 401, 2004). *Biochemical and Biophysical Research Communications* **334**, 968-968,(2005)

32. Polayes, D. A., Parks, T. D., Johnston, S. A. & Dougherty, W. G. Application of TEV Protease in Protein Production. *Methods Mol Med* **13**, 169-183, doi:10.1385/0-89603-485-2:169 (1998)
33. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wuthrich, K. Biased Signaling Pathways in beta(2)-Adrenergic Receptor Characterized by F-19-NMR. *Science* **335**, 1106-1110, (2012).
34. Biswas, E. E. & Biswas, S. B. Mechanism of DnaB helicase of Escherichia coli: Structural domains involved in ATP hydrolysis, DNA binding, and oligomerization. *Biochemistry* **38**, 10919-10928, doi:Doi 10.1021/Bi990048t (1999).
35. Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J. M. & Steitz, T. A. The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation. *Cell* **151**, 267-277, (2012).
36. Bujalowski, W., Klonowska, M. M. & Jezewska, M. J. Oligomeric structure of *Escherichia coli* primary replicative helicase DnaB protein. *J Biol Chem* **269**, 31350-31358 (1994).
37. Jezewska, M. J. & Bujalowski, W. Global conformational transitions in Escherichia coli primary replicative helicase DnaB protein induced by ATP, ADP, and single-stranded DNA binding. Multiple conformational states of the helicase hexamer. *J Biol Chem* **271**, 4261-4265 (1996).
38. Arai, K. & Kornberg, A. Mechanism of dnaB protein action. II. ATP hydrolysis by dnaB protein dependent on single- or double-stranded DNA. *J Biol Chem* **256**, 5253-5259 (1981).
39. Soultanas, P. & Wigley, D. B. Site-directed mutagenesis reveals roles for conserved amino acid residues in the hexameric DNA helicase DnaB from *Bacillus stearothermophilus*. *Nucleic Acids Res* **30**, 4051-4060 (2002).
40. Bailey, S., Eliason, W. K. & Steitz, T. A. The crystal structure of the Thermus aquaticus DnaB helicase monomer. *Nucleic Acids Res* **35**, 4728-4736,(2007).
41. Watt, S. J. *et al.* Multiple oligomeric forms of Escherichia coli DnaB helicase revealed by electrospray ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry* **21**, 132-140, doi:Doi 10.1002/Rcm.2818 (2007).

42. Bujalowski, W. & Klonowska, M. M. Close proximity of tryptophan residues and ATP-binding site in Escherichia coli primary replicative helicase DnaB protein. Molecular topography of the enzyme. *J Biol Chem* **269**, 31359-31371 (1994).
43. Jezewska, M. J., Kim, U. S. & Bujalowski, W. Interactions of Escherichia coli primary replicative helicase DnaB protein with nucleotide cofactors. *Biophys J* **71**, 2075-2086, (1996).
44. Bujalowski, W. & Jezewska, M. J. Kinetic mechanism of nucleotide cofactor binding to Escherichia coli replicative helicase DnaB protein. Stopped-flow kinetic studies using fluorescent, ribose-, and base-modified nucleotide analogues. *Biochemistry* **39**, 2106-2122, (2000)
45. Rajendran, S., Jezewska, M. J. & Bujalowski, W. Multiple-step kinetic mechanism of DNA-independent ATP binding and hydrolysis by Escherichia coli replicative helicase DnaB protein: Quantitative analysis using the rapid quench-flow method. *Journal of Molecular Biology* **303**, 773-795, (2000).
46. Davey, M. J. & O'Donnell, M. Replicative helicase loaders: Ring breakers and ring makers. *Current Biology* **13**, R594-R596, (2003).
47. Kornberg, A. & Baker T.A., *DNA Replication*, (Second edn)W.H. Freeman and Co, New York (1992)
48. Velten, M. *et al.* A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol Cell* **11**, 1009-1020, (2003).
49. Ioannou, C., Schaeffer, P. M., Dixon, N. E. & Soultanas, P. Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in Bacillus subtilis. *Nucleic Acids Research* **34**, 5247-5258, doi:Doi 10.1093/Nar/Gkl690 (2006).

Chapter 5

Future Directions

The applications of one dimensional fluorine NMR are still being explored in biomolecular NMR. The study on DnaB helicase by ^{19}F NMR has established as a promising tool for studying large protein systems with molecular weight of more than 400 kDa. The fluorine NMR data has brought out new aspects about DnaB that are interesting and requires more research.

Chapter 3: The expression of full-length fluorinated *E. coli* DnaB was rendered difficult due to competing RF1, the protein responsible for identifying TAG stop codon and terminating the translation[1]. Recently,many modified bacterial strains have been developed to improve the yield of unnatural amino acid incorporated proteins. *E. coli* cells were evolved with orthogonal ribosome called ribo-X which cannot interact with RF1[2,3]. Another method was removal or inactivation of the RF1 from the cell extract for cell-free protein synthesis [4,5,6]. More recently, *E. coli* strains were developed without RF1 gene in the genome and demonstrated the labelled protein yields improved[7]. Utilising any one of the technique would aid in increasing the yield of protein labelled with unnatural amino acid.

Similar study can be conducted on DnaB helicase from *Helicobacter pylori (H.pylori)*, a Gram-negative bacterium similar to *E. coli.* The insights into *H.pylori* DnaB helicase will be helpful in pharmaceutical industry to develop new drugs because it is pathogen infecting half the global population [8].*H.pylori* DnaB helicase is a well-behaved and characterised protein [9,10,11]. It assumes dodecamer structure instead of flat hexamer as suggested in *E. coli* DnaB [12]. The binding of DnaG primase is shown to induce conformational change in N-terminal domain of the DnaB helicase and demonstrated to increasing to binding affinity to DNA [10]. It would be interesting to observe fluorine resonances from a dodecamer DnaB helicase with a mass in order of 600 kDa and to study the conformational changes induced by DnaG in solution using ^{19}F NMR.

Chapter 4: The gel filtration data revealed that Mg^{2+} affects the *Gst DnaB* hexamers and triggers the formation of monomers and proportion of monomer formation was dependent on the site where the tfmF was incorporated. It indicates that residue Y130 might play an important role in oligomerisation of the protein. The role of Y130 could be studied in depth by mutagenesis studies by substituting Y130 with other amino acids and study formation of hexamers by either gel filtration or mass spectrometry.

Further research in mutations at site F52 could lead to interesting new insights. Gel filtration and ^{19}F NMR showed that F52tfmF had weak interactions with DnaG suggesting that F52 is important for interacting with DnaG. So far, mutagenesis studies showed Y88 and E15 affected the interaction with DnaG[13]. It would be interesting to conduct similar mutagenic studies on F52 and its neighbouring residues to study the residues important for DnaG binding.

*Gst*DnaI has not been structurally characterised, however, its function can be deduced from the homologous helicase loader - *Bsu* DnaI. The structure of *Gst*DnaI can be determined using conventional NMR spectroscopy and its interaction with *Gst DnaB* can be examined by electron microscopy and these studies will give ample information about helicase loading mechanism in *Geobacillus stearothermophilus.*

Application of ^{19}F NMR: There are other potential ^{19}F-labelling groups that can be used to study large proteins. 4-trifluoromethoxyphenylalanine was used in studying the binding of the small molecules to thioesterase domain of fatty acid synthase [14]. 3-fluorotyrosine was used as fluorine label to study a 22 kDa human manganese superoxide dismutase, which formed a tetramer of 88 kDa protein considered as a large system in NMR[15]. Fluorinated tryptophans are less toxic than other fluoroaromatics and 5-fluorotryptophan has shown to have large dispersions in signals and indicated minor conformational changes in $_D$-lactate dehydrogenase[16].

A well-behaved protein system is necessary to obtain ^{19}F spectrum with highly sensitive signals. By comparing ^{19}F NMR spectra of *E. coli* and *Gst DnaB* helicase, it is evident that *Gst DnaB* produces signals with higher sensitivity than *E. coli* DnaB. The primary factor is the concentration of the protein. The concentration of tfmF labelled *E. coli* DnaB used was in the range of 10- 30 μM; whereas concentrations of tfmF labelled *Gst DnaB*

were 70 μM. *Gst DnaB* is better behaved helicase than *E. coli* DnaB. For future study by ^{19}F NMR, DnaB helicases from *H.pylori* and *Thermus aquaticus* (*Taq*) would be good systems as their partial or whole structure has been determined by X-ray crystallography[17,18]. Moreover, these DnaB helicases are new systems to be characterized, it would be very interesting if they share structural characteristics with *E. coli* or/and *Gst DnaB* helicase.

Effects of ATP on the *Gst* DnaB, DnaB helicases from *H.pylori* and *Thermus aquaticus* (*Taq*) can further studied using ^{19}F NMR in complex with DNA strands. The literature has given potential evidence that DNA is required for ATP hydrolysis and responsible for conformational change[19, 20].

The successful incorporation of fluorine label into the protein using both *in vitro* and *in vivo* system, the study by ^{19}F NMR can be applied to various large protein systems. Understanding these systems would give a detail understanding of the prokaryotic DNA replication mechanism.

Reference

1. Korostelev, A., Zhu, J., Asahara, H. & Noller, H. F. Recognition of the amber UAG stop codon by release factor RF1. *EMBO J* **29**, 2577-2585, (2010)
2. Rackham, O. & Chin, J. W. A network of orthogonal ribosome center dot mRNA pairs. *Nature Chemical Biology* **1**, 159-166, (2005).
3. Wang, K. H., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature Biotechnology* **25**, 770-777, (2007).
4. Agafonov, D. E., Huang, Y., Grote, M. and Sprinzl, M. Efficient suppression of the amber codon in *Escherichia coli in vitro* translation system. *FEBS Lett* (2005).
5. Sando, S., Ogawa, A., Nishi, T., Hayami, M. & Aoyama, Y. In vitro selection of RNA aptamer against Escherichia coli release factor 1. *Bioorg Med Chem Lett* **17**, 1216-1220, (2007).

6. Loscha, K. V. *et al.* Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angewandte Chemie-International Edition* **51**, 2243-2246, (2012).
7. Lajoie, M. J. *et al.* Genomically Recoded Organisms Expand Biological Functions. *Science* **342**, 357-360, (2013).
8. Parsonnet, J. The Incidence of Helicobacter-Pylori Infection. *Alimentary Pharmacology & Therapeutics* **9**, 45-51 (1995).
9. Soni, R. K., Mehra, P., Choudhury, N. R., Mukhopadhyay, G. & Dhar, S. K. Functional characterization of Helicobacter pylori DnaB helicase. Nucleic Acids Research 31, 6828-6840, (2003).
10. Nitharwal, R. G. *et al.* DNA binding activity of Helicobacter pylori DnaB helicase: the role of the N-terminal domain in modulating DNA binding activities. *Febs Journal* **279**, 234-250, (2012).
11. Nitharwal, R. G. *et al.* The domain structure of Helicobacter pylori DnaB helicase: the N-terminal domain can be dispensable for helicase activity whereas the extreme C-terminal region is essential for its function. *Nucleic Acids Research* **35**, 2861-2874, doi:Doi 10.1093/Nar/Gkm167 (2007).
12. Stelter, M. *et al.* Architecture of a Dodecameric Bacterial Replicative Helicase. *Structure* **20**, 554-564, 020 (2012).
13. Thirlway, J. *et al.* DnaG interacts with a linker region that joins the N- and C-domains of DnaB and induces the formation of 3-fold symmetric rings. *Nucleic Acids Research* **32**, 2977-2986, (2004).
14. Cellitti, S. E. *et al.* In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **130**, 9268-9281, (2008).
15. Quint, P. *et al.* Structural mobility in human manganese superoxide dismutase. *Biochemistry* **45**, 8209-8215, (2006)
16. Peersen, O. B., Pratt, E. A., Truong, H. T. N., Ho, C. & Rule, G. S. Site-Specific Incorporation of 5-Fluorotryptophan as a Probe of the Structure and Function of the Membrane-Bound D-Lactate Dehydrogenase of Escherichia-Coli - a F-19 Nuclear-Magnetic-Resonance Study. *Biochemistry* **29**, 3256-3262, (1990).

17. Bailey, S., Eliason, W. K. & Steitz, T. A. The crystal structure of the *Thermus aquaticus* DnaB helicase monomer. *Nucleic Acids Research* **35**, 4728-4736,(2007)
18. Kashav, T. *et al.* Three-Dimensional Structure of N-Terminal Domain of DnaB Helicase and Helicase-Primase Interactions in Helicobacter pylori. *Plos One* **4**, (2009)
19. Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J. M. & Steitz, T. A. The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation. *Cell* **151**, 267-277, (2012).
20. Bujalowski, W. & Jezewska, M. J. Kinetic mechanism of nucleotide cofactor binding to Escherichia coli replicative helicase DnaB protein. Stopped-flow kinetic studies using fluorescent, ribose-, and base-modified nucleotide analogues. *Biochemistry* **39**, 2106-2122, (2000).