The orientation of the Pyrococcus furiosus transcription factor TFB2 in the transcription initiation complex

Arati Bhattarai
Portland State University

Let us know how access to this document benefits you.
Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the Biology Commons

Recommended Citation

10.15760/etd.1937
The orientation of the *Pyrococcus furiosus* transcription factor TFB2 in the transcription initiation complex

by

Arati Bhattarai

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Thesis Committee:

Michael Bartlett, Chair

Jeffrey Singer

Justin Courcelle

Portland State University

2014
ABSTRACT

The hyperthermophile archaeon, *Pyrococcus furiosus* encodes two eukaryotic TFIIB family proteins, TFB1 and TFB2. TFB1 is very similar to TFIIB in terms of sequence homology and function, whereas TFB2 is unusual as it is missing highly conserved sequences in its N-terminal domain that are present in TFIIB and TFB1. Despite this, TFB2 is effective in transcription process, albeit with lower efficiency compared to TFB1. Other archaea also contain multiple TFBs, but unlike *Pyrococcus furiosus* TFB2, these multiple TFBs have higher sequence homology to each other and have similar transcription efficiencies. Photochemical cross-linking experiments have shown that the B-reader of TFB in archaea and TFIIB in eukaryotes is close to transcription start site and is very important in RNAP recruitment to promoter DNA and transcription start site selection. Thus the lack of the highly conserved B reader region in *P. furiosus* TFB2 presents the opportunity to further study the functional importance of this region.

In this study several amino acids in N-terminal domain of TFB2 were mutated with photoactivable unnatural amino acid *p*-benzoyl L-phenylalanine (pBpa) and the proximity of TFB2 relative to DNA was determined by photochemical cross-linking experiments. The results showed that TFB2 interacts with DNA near the TATA box via its C-terminal domain, and interacts with both strands of DNA near the transcription start site via its divergent B-reader and the B-linker sequences. The B-reader loop region is close to transcription start site and
interacts with the transcribed strand of promoter DNA while the B-linker strand cross-links with the non-transcribed strand. Some of the amino acids in between the B-reader loop and the B-linker strand region in TFB2 are seen to cross-link both the transcribed and the non-transcribed strand. Thus, despite the absence of strong homology to conserved B-reader and B-linker sequences, TFB2 is likely to interact with DNA in the transcription bubble and facilitate in transcription initiation.
ACKNOWLEDGEMENT

I would like to dedicate this thesis to my Advisor Dr. Michael Bartlett, my committee members Dr. Jeffrey singer and Dr. Justin Courcelle, my husband Dr. Ishwar Niraula, my son Ishav Niraula, my mother, father, sisters and all family members and friends.
# Table of contents

ABSTRACT ........................................................................................................................................... i

ACKNOWLEDGEMENT ...................................................................................................................... iii

LIST OF FIGURES ................................................................................................................................ vi

CHAPTER I ........................................................................................................................................... 1

1. INTRODUCTION ............................................................................................................................ 1
   1.1 TRANSCRIPTION MACHINERY IN THREE DOMAINS OF LIFE .................................................. 1
   1.2 PROMOTER OPENING IN RNA POLYMERASE II TRANSCRIPTION INITIATION .................... 3
      1.2.1 Promoter opening in archaeal transcription initiation ...................................................... 5
      1.2.2 Promoter opening in bacterial transcription initiation ...................................................... 6
   1.3 ARCHAEA TRANSCRIPTION INITIATION MECHANISM ............................................................ 6
   1.4 TRANSCRIPTION FACTOR TFIIB IN EUKARYOTES AND TFBs IN ARCHAEA ...................... 7
   1.5 TRANSCRIPTION FACTOR TFIIE IN EUKARYOTES AND TFE IN ARCHAEA ......................... 9
   1.6 TFB1 AND TFB2 TRANSCRIPTION FACTORS IN THE ARCHAEO Pyrococcus furiosus........... 11
   1.7 CURRENT UNDERSTANDING OF Pyrococcus furiosus TFB2 STRUCTURE AND FUNCTION...... 12

CHAPTER II .......................................................................................................................................... 19

2. EXPERIMENTAL PROCEDURES ...................................................................................................... 19
   2.1 MUTAGENESIS OF TFB2 BY INSERTION OF A CROSS-LINKABLE UNNATURAL AMINO ACID..... 19
   2.2 PROTEIN OVEREXPRESSION AND PURIFICATION ................................................................. 23
   2.3 STANDARD TRANSCRIPTION ASSAYS ....................................................................................... 25
   2.4 PROTEIN DNA CROSS-LINKING ASSAYS ............................................................................... 26

CHAPTER III ......................................................................................................................................... 30

3. RESULTS ......................................................................................................................................... 30
   3.1 PURIFICATION OF THE TFB2 PROTEINS .................................................................................. 32
   3.2 TFB2 PROTEINS WITH UNNATURAL AMINO ACIDS ARE ACTIVE IN TRANSCRIPTION ASSAYS... 34
   3.3 PROTEIN-DNA CROSS-LINKING CONFIRMS THE PROXIMITY OF TFB2 V175 PROTEIN TO THE -19
       POSITION OF THE TRANSCRIBED STRAND ............................................................................. 38
   3.4 COMPARING TWO TFB2 MUTANTS TFB2 R54 AND TFB2 P62 WITH TFB1 MUTANTS TFB1 W44
       AND TFB1 R52 .......................................................................................................................... 42
   3.5 POSITIONING OF OTHER TFB2 N-TERMINUS AMINO ACIDS IN THE TRANSCRIPTION INITIATION
       COMPLEX .................................................................................................................................. 47

CHAPTER IV ......................................................................................................................................... 53

4. DISCUSSION .................................................................................................................................... 53
   4.1 THE PATH OF TFB2 IN A TRANSCRIPTION COMPLEX ............................................................. 53
   4.2 THE TFB1 AND TFB2 C-TERMINAL DOMAINS ......................................................................... 54
   4.3 CROSS-LINKING BASED ALIGNMENT OF TFB2 TO TFB1 AND TFIIB ..................................... 55
   4.4 STRUCTURAL ORIENTATION OF TFB2 IN TRANSCRIPTION INITIATION COMPLEX ............. 63
List of Figures

Figure 1: Partial alignment of B-finger region of TFBs from various archaea and eukaryotes. ................................................................. 13
Figure 2: Structure of TFIIB ................................................................................................................................. 14
Figure 3: Closed and open initiation complex model of eukaryotic RNAP II and TFIIB with TBP and DNA .................................................................................................................. 15
Figure 4: Structure of TFIIB with DNA showing possible interactions of DNA sequences with different parts of TFIIB. .............................................................................................................. 15
Figure 5: Structure of the initially transcribing RNAP II-TFIIB complex. .............................................................. 16
Figure 6: TFIIB structure (PDB: 4BBS). .................................................................................................................. 17
Figure 7: Partial alignment of Pyrococcus furiosus TFB1 and TFB2 proteins. ............................................................ 19
Figure 8: Unnatural amino acid pBpa incorporation ..................................................................................................... 20
Figure 9: Radioactive incorporations ...................................................................................................................... 27
Figure 10: Partial alignment of N-terminus of Pyrococcus furiosus TFB1 and TFB2 showing part of the N terminal region .................................................................................................................................. 31
Figure 11: Purification of TFB2 R54pBpa protein followed by coomassie blue staining ...................................................... 33
Figure 12: Transcription of wild type TFB1 and wild type TFB2 .................................................................................. 35
Figure 13: Transcription of TFB2 variants, TFB2 V175, TFB2 P62 and TFB2 R54 ........................................................................ 36
Figure 14: Transcription of TFB2 variants, TFB2 K57, TFB2 N59 and TFB2 I68. .......................................................... 38
Figure 15: Protein-DNA cross-linking ....................................................................................................................... 39
Figure 16: TFB2 V175 is close to -19 position on the template strand ........................................................................ 40
Figure 17: Cross-linking of TFB1 F192 and TFB2 V175 with gdh promoter DNA radiolabelled at -6, -4 and -19 positions of transcribed strand ......................................................................................... 41
Figure 18: Cross-linking of TFB2 R54 and TFB2 P62 to the transcribed strand probes ................................................................................................. 44
Figure 19: Cross-linking of TFB1 E74pBpa and TFB2 P62pBpa to the non-transcribed strand ......................................................... 46
Figure 20: Cross-linking of TFB2 K57pBpa to transcribed and non-transcribed DNA strands ................................................................................................................................. 48
Figure 21: Cross-linking of TFB2 N59 with transcribed and non-transcribed DNA strands ................................................................................................................................. 48
Figure 22: Cross-linking of TFB2 I68 with transcribed and non-transcribed DNA strands ................................................................................................................................. 50
Figure 23: New alignment of P. furiosus TFB2 N-terminus .............................................................................................. 51
Figure 24: TFB2 R54 and TFB2 P62 with transcribed and non-transcribed DNA. ............................................................... 60
Figure 25: TFB2 K57, TFB2 N59 and TFB2 I68 with transcribed and non-transcribed DNA. ........................................................................................................ 62
Figure 26: Structure of TFB2 showing the missing structures that are normally present in TFIIB/TFB1. ........................................................................................................ 64
Figure 27: TFB2 N-terminus mutations shown along with transcribed and non-transcribed DNA. ........................................................................................................ 65
Figure 28: TFB2 orientation in transcription initiation complex.....................63
CHAPTER I

1. INTRODUCTION

Archaea constitute a major branch in universal tree of life along with bacteria and eukaryotes (Woese et al, 1977 and 1990). Archaea and bacteria both lack a nucleus, and while metabolic pathways and the gene regulation system of archaea are related to those in bacteria, the information processing systems (DNA replication, transcription and translation) in archaea more closely resemble those in eukaryotes (Soppa J, 1999). Archaea have a simplified information processing system compared to eukaryotes, and thus molecular study of the archaeal system may help in understanding the evolution of complex organisms from simpler life forms.

A hyperthermophile is an organism which thrives in temperatures exceeding 75°C. It has been proposed that the last common ancestor of all life on earth may have been a hyperthermophile (Woese et al, 1990). Most hyperthermophiles belong to the domain Archaea, and this thesis concerns gene regulation in the hyperthermophile euryarchaeon Pyrococcus furiosus (Pfu).

1.1 Transcription machinery in three domains of life

Transcription is the synthesis of RNA from a DNA template, and requires the enzyme RNA polymerase along with transcription factors. It is the first step in gene expression, and is a target for important regulation in all domains of life. The transcription apparatus of archaea resembles the RNA polymerase I, II and III system of eukaryotes (Baumann et al, 1995; Qureshi and Jackson, 1997). In particular, archaeal
RNA polymerase and its basal transcriptional machinery share many properties with the
eukaryotic RNA polymerase II (RNAP II) transcriptional apparatus (Thomm M 1996; Kosa et al, 1997). While eukaryotes have three RNA polymerases that synthesize 3 classes of RNA, archaea like bacteria have only one RNA polymerase. Most of the archaeal RNAP subunits are homologous to subunits in eukaryotic RNA polymerase II (Hirata et al, 2008; Kusser et al, 2008; Kwapisz et al, 2008). The RNA polymerase structure from the archaean Sulfolobus shibatae reveals 13 subunits and most of the subunits share immunological cross-reactivity, sequence and functional similarity with eucaryal RNA polymerase subunits (Langer et al, 1995; Darcy et al, 1999; Werner et al, 2000; Werner F, 2008; Korkhin et al, 2009 and Reich et al, 2009). Additionally, other transcriptional factors such as TATA binding protein (TBP), transcription factor B (TFB) and transcription factor E (TFE) display high levels of structural and functional conservation with their eukaryotic counterparts, TBP, TFIIB, and the TFIIE α subunit (Marsh et al, 1994; Rowlands et al, 1994; Darcy et al, 1999; Kyrpides et al, 1999 and Hanzelka et al, 2001). Eukaryotic RNAP II requires the additional transcription factors TFIIH, TFIIA and TFIIF whose functions have not been found in archaea (Soppa J, 1999; Aravind et al, 1999 and Kyrpides et al, 1999). In general, we can say that the archaenal transcription machinery is a simplified version of eukaryotic transcriptional machinery as it only requires TFB and TBP for transcription in vitro (Langer et al, 1995; Nikolov et al, 1997; Bell et al, 1998 and Werner F, 2007). Therefore archaenal systems are useful for modeling the basic functions of the more complex eucaryal transcription process.
In contrast to the archaeal and eukaryotic RNA polymerases, bacterial RNA polymerase and transcription initiation is simpler. Bacterial transcription is catalyzed by RNAP holoenzyme, which has 6 subunits. The RNAP without sigma is called core enzyme and does not recognize promoters. The core subunits of bacterial RNAP have homologs in both archaeal and eukaryotic RNAPs. However sigma subunit is unique to bacteria and no homologs are seen in archaea or eukaryotic RNAP. The sigma subunit of holoenzyme interacts with the promoter DNA at conserved -35 and -10 elements (Gross et al, 1998 and Campbell et al, 2002). Following recruitment to the promoter by the sigma subunit, RNAP can initiate transcription without the assistance of other factors.

1.2 Promoter opening in RNA polymerase II transcription initiation

In all domains of life, transcription initiation is a multistep process, and requires assembly of transcription factors along with the RNA polymerase (RNAP) on the core promoter region of double stranded DNA, which forms the pre-initiation complex (Zawel et al, 1993; Hampsey 1998; Hahn 2004 and Saecker et al, 2011). The transcription initiation process begins with promoter melting, in which the double stranded DNA is separated into single strands and the transcription start site in the template strand of DNA is placed near the active center of RNAP. Promoter melting involves the separation of 10-12 bp of DNA forming a structure called the transcription bubble (Saecker et al, 2011; Darst and Feklistov, 2011; Kostrewa et al, 2009; Murakami and Darst, 2003).

In eukaryotes the open complex is a catalytically important intermediate in the transcription initiation process by RNAP II. Photochemical cross-linking experiments have shown that promoter melting is initiated by TFIIH in the presence of ATP, through
its XPB and XPD subunits which have DNA helicase and ATPase activity (Kim et al, 2000). The melted region is then stabilized by TFIIE which interacts with DNA downstream of the transcription bubble, and with the clamp of RNAP II (Kim et al, 2000 and Grohmann et al, 2011). TFIIF binds to the non template strand and keeps the bubble open (Bushnell et al, 2004). TFIIF also stabilizes RNAP II and TFIIB binding to the promoter DNA, and can assist TFIIE and TFIIH to join the complex (Flores et al, 1990; Cabart et al, 2011 and Luse DS 2012). Recent cryoelectron microscopy structures of the human transcription initiation complex suggest that TFIIF stabilizes both the closed and open complex (Nogales et al, 2013). TFIIF localizes TFIIH to Pol II where it functions in promoter opening (He et al, 2013). In the minimal open complex from yeast, the core of TFIIB is positioned away from wall of Pol II, but when TFIIF was added it positions back to the wall (Fishburn and Hahn, 2012). Thus TFIIF and TFIIB work cooperatively to stabilize the RNAP II open complex.

**In vitro** experiments have shown that TFIIB is essential for the transition from closed complex to open complex, since the N-terminal domain brings the DNA into proper position relative to the active site of RNAP. The B-reader region and B-linker region of TFIIB are important for open complex formation (Kostrewa et al, 2009). The TFIIB B-reader scans the TSS (transcription start site) to detect the Inr sequences, and then the TFIIB linker region melts the DNA strand near the transcription start site. The template strand then bends and enters the RNAP active site. The fork loop region present in RNAP helps in this movement of template strand. After the template strand moves to the active site more bases separate out, bringing the total transcription bubble to almost 20 bp (Kostrewa et al, 2009 and Sainsbury et al, 2012). Along with these events, the
RNA polymerase clamp also closes during open complex formation (Gries et al, 2010; Chakraborty et al, 2012).

1.2.1 Promoter opening in archaeal transcription initiation

Archaeal transcription can be reconstituted with only TBP, TFB and TFE, homologs of eukaryotic TBP, TFIIB and TFIIE. There are no known functional homologs of TFIH and TFIIF. Archaeal transcription initiation in vitro requires only RNAP, TBP and TFB, with TFB being the major factor responsible for open complex formation and transcription start site selection. Footprinting experiments in highly purified cell free Methanococcus shows that the open complex in tRNA val promoter spans at least the -11 to -1 region (Hausner W and Thomm M, 2001). Open complex formation in vitro does not require TFIH or ATP hydrolysis (Hausner W and Thomm M, 2001 and Bell et al, 1998). In vitro experiments with Pyrococcus furiosus RNAP also show that E’ and F subunits of RNAP, also known as stalk of RNAP, bind to the core region of RNAP and influence the opening and closing of clamp of RNAP during the transition from closed to open complex. Footprinting experiments show that the RNAP E’ subunit stimulates the transcription bubble formation at low temperature and that TFE interacts with upstream edge of transcription bubble making the open complex more stable (Naji et al, 2007). Biochemical experiments and structural analysis of the transcription initiation complex shows that B-linker region of TFB is close to the clamp region of RNAP and may help in the formation and stabilization of open complex (Werner and Weinzierl, 2002; Kostrewa et al, 2009).
1.2.2 Promoter opening in bacterial transcription initiation

In bacteria, the holoenzyme, formed by core RNAP subunits with sigma, can accomplish promoter opening without hydrolyzing ATP (Murakami and Darst, 2003). RNAP recognizes downstream sequences (positions -4 to +2) and σ recognizes upstream sequences (positions -12 to -4) of the non-template strand of the transcription bubble, and helps in maintaining the melted state (Zhang Yu et al, 2012). The σ factor contains several domains, of which region 2.4 has sequence specific interaction with non template strand at -10 promoter region in closed complex (Roberts and Roberts, 1996; Darst and Feklistov, 2011). This -10 region is particularly important for melting the DNA template, which starts with a T-A base pair and involves torsional strain on the double helical DNA at this site (de Haseth et al, 1995, Artsimovitch et al, 1996, Young et al, 2001 and 2004). The adenine residue at -11 position on promoter DNA initiates melting by interaction with sigma, loosening the contacts between the two DNA strands, and allowing the template strand to enter into the active channel (Darst and Feklistov, 2011). The melted non-template strand interacts with 2.3 region of the sigma and helps in maintaining the melted state (Schroeder et al, 2009). In summary, initially a few bases are melted, and later up to 12 base pairs are melted to form the transcription bubble, with sigma being involved at all stages of melting.

1.3 Archaeal transcription initiation mechanism

The transcription process starts in archaea starts with recruitment of transcription factors in the promoter region of the gene. Three general transcription factors TBP, TFB and TFE are utilized for transcription in vitro (Hausner et al, 1996 and Bell et al, 2001).
First TBP recognizes and binds an AT rich TATA box sequence located about -10 to 25 bp upstream from the transcription initiation site (+1 site) (Thomm et al, 1988 and Qureshi et al, 1998). The crystal structure of archaeal TBP-TFB-TATA box complex shows that the saddle shaped TBP binds to the minor groove of the DNA and causes conformational change of DNA in the A-T rich sequences, creating a bend to which the TFIIB core domain binds (Littlefield et al, 1999). The TFB C-terminus interacts with B-recognition element (BRE) sequences upstream and downstream of the TATA sequence, and interacts with the TBP-promoter complex (Hausner et al, 1996). The TBP-TFB-DNA complex recruits RNAP to the promoter through TFB N-terminal domain-RNAP interactions, and transcription initiates (Bell et al, 1999). This process of RNAP recruitment to promoter is similar in eukaryotes although eukaryotes require additional transcription factors (Hampsey 1998, Ranish et al, 1999 and Orphanides et al, 1996).

1.4 Transcription factor TFIIB in eukaryotes and TFBs in archaea

TFIIB in eukaryotes and TFB in archaea, collectively termed TF(II)B, are central transcription factors in transcription initiation process. All archaeal and eukaryotic genome sequences have ORFs that encode TF(II)B homologs, which indicates the functional importance of this protein. TFIIB specifically binds to the TBP-DNA complex and helps recruit RNAP II to the promoters via protein-protein interactions (Orphanides et al, 1996 and Hampsey 1998). Sequence analysis shows that TFIIB resembles the Brf protein part of TFIIBB specificity factor of RNAP III system, and Rrn7 protein, a subunit of yeast RNAP I core factor, both of which, like TFIIB, help in recruiting the RNA

Structurally TF(II)B consists of two distinct domains: a highly conserved N-terminal domain (NTD) connected to a highly conserved C-terminal core domain (CTD) by a linker (Hahn, 2004). The CTD of TF(II)B contains a helix-turn-helix motif that interacts with BRE sequences upstream of the TATA box in a sequence specific manner, contributing to the formation of stable pre-initiation complex (Lagrange et al, 1998; Littlefield et al, 1999; Bartlett et al, 2004). This interaction of the TF(II)B core domain with DNA governs the orientation of DNA along the central cleft of RNAP II (Chen and Hahn, 2004) which helps to determine direction of the transcription process (Korkhin et al, 2009).

The TF(II)B N-terminal domain (NTD) has a role in transcription start site selection (Nikolov et al, 1995; Qureshi et al 1995; Pardee et al, 1998 and Bell et al, 1999). The NTD of TF(II)B contains a Zinc ribbon motif (also called a B-ribbon) and a B-finger region that was structurally defined in a medium-resolution co-crystal structure of RNAP II with TFIIB (Zhu et al, 1996; Bushnell et al, 2004; Chen and Hahn, 2004). Sequence analysis of archaeal TFBs and eukaryotic TFIIBs shows high conservation of amino acids in the zinc ribbon and B-finger regions, suggesting the functional importance of these portions of the protein. The zinc ribbon domain of the protein plays a direct role in recruiting RNAP II to the promoter through interactions with the dock domain of RNAP II (Buratowski et al, 1993; Tubon et al, 2004; Elsby and Tubon, 2008). A more recent, higher resolution (3.8Å) structure of RNAP II-TFIIB structure showed an alternative structure in the B-finger region, so this structure was renamed the B-reader
region, containing the B-reader helix and B-reader strand. In addition a B-linker region was identified that contains the B-linker strand and B-linker helix (Liu et al, 2010; Kostrewa et al, 2009). In a subsequent TFIIB-RNAP II structure, a new B-reader loop region was observed which was not seen previously (Sainsbury et al, 2012). Several X-ray crystal structures of TFIIB-RNAP II show that the B-finger/B-reader-linker regions of TFIIB extend from the active center to the RNA exit channel of RNAP II (Bushnell et al, 2004; Kostrewa et al, 2009; Liu et al, 2010 and Sainsbury et al, 2012). Photochemical cross-linking experiments show that the N-terminal region of TFB in archaea and TFIIB in eukaryotes is close to DNA near the transcription start site and is very important in RNAP recruitment to promoter DNA and transcription start site selection (Ranish et al, 1999; Bartlett et al, 2000; Renfrow et al, 2004 and Hahn 2004). In summary, the TF(II)B B-ribbon region interacts with the RNAP II dock region, the B-reader helix with the RNA exit tunnel, the B-reader strand with the lid region of Pol II, the B-reader loop with Pol II rudder and fork loop region, and the B-linker helix with coiled-coil above the rudder (Chen and Hahn, 2003; Bushnell et al, 2004; Kostrewa et al, 2009 and Sainsbury et al, 2012). TF(II)B thus plays a central role in formation of transcription initiation complex and therefore is vital for the transcription mechanism.

1.5 Transcription factor TFIIE in eukaryotes and TFE in archaea

Archaeal TFE shares structural homology with the amino terminal domain of α subunit of eukaryotic TFIIE (Bell et al, 2001; Hanzelka et al, 2001 and Meinhart et al, 2003). TFIIE enhances the ATPase, kinase and DNA helicase activities of TFIIH, helps in the formation of the open complex and can rescue transcription from stalled RNAP II
open complexes (Okhuma and Roeder, 1994; Serizawa et al, 1994 and Lee et al, 2000). The archaeal TFE consists of an N-terminal winged helix domain whose structure is solved, and a C- terminal zinc ribbon (Bell et al, 2001 and Meinhart et al, 2003). It lacks the β subunit of TFIIE present in eukaryotes (Kuldell and Buratowski, 1997). TFE is not required for transcription in archaea in vitro but it has been shown to enhance the transcription initiation process at some promoters under sub-optimal TBP-TATA box binding conditions (Bell et al, 2001). The TFE from the archaeon Methanobacterium thermoautotrophicum stimulates transcription from almost inactive promoters and less active promoters in presence of RNAP, TBP and TFB (Hanzelka et al, 2001). EMSA experiments have shown that TFE can compensate for mutations in the TFB zinc-ribbon or B-finger motifs, assisting in recruitment of RNAP to form a stable initiation complex, which was not possible by the mutated TFBs alone (Werner et al, 2005). In photochemical cross-linking experiments, the winged helix domain of TFE interacts upstream of transcription start site with the non-template strand of DNA (Grunberg et al, 2007). Protein-protein cross-linking experiments shows that both eukaryotic TFIIE and archaeal TFE interact with the RNAP II clamp domain and that archaeal TFE competes with Spt4/5 (an elongation factor) for binding to coiled-coil region in clamp domain of RNAP (Chen et al, 2007 and Grohmann et al, 2011). The TFIIEα winged helix domain and TFIIE β winged helix domain forms a heterodimer and encircles the promoter DNA interacting with single-stranded DNA, suggesting that TFIIE may help in unwinding the double stranded DNA and help in the stabilization of the open complex (Grunberg et al, 2012). Thus TFE has a central role in transcription initiation process, through its influence on formation and stabilization of the open complex.
1.6 TFB1 and TFB2 transcription factors in the archaeon *Pyrococcus furiosus*

Many archaeal genomes encode multiple TFB and TBP homologs. For example, haloarchaeal genomes encode up to 6 TBPs and 7 TFBs (*Baliga NS et al., 2000*). Additional TFBs occasionally lack intact N or C terminal regions (*Werner F, 2007*). The hyperthermophile archaeon *Pyrococcus furiosus* contains two homologs of the highly conserved TFIIB transcription factor family, named TFB1 and TFB2. The TFB1 polypeptide consists of 300 amino acids whereas TFB2 polypeptide has 283. Some regions of the N-terminus of TFB2 are less conserved compared to the same regions of TFB1. The zinc ribbon domain of TFB2 (amino acids 17 to 49) is 45\% identical to the TFB1 zinc ribbon domain (amino acids 7 to 39), and TFB2 also lacks 27 amino acids present in the B finger region of TFB1. In contrast the TFB2 C-terminus (amino acids 73 to 283) is 63\% identical to the TFB1 C-terminus (amino acids 86 to 300) which includes the helix-turn-helix motif that interacts with BRE sequences.

Previous work from our lab has shown that the Zn-ribbon and B-finger of the N-terminal region of TFB1 are critical for promoter opening (*Micorescu et al., 2008*). *In vitro* experiments have shown that TFB2 has low transcription efficiency compared to TFB1, at least partly because of the divergent B-finger region of TFB2. This low transcription efficiency was partly overcome by addition of TFE, which helps to stabilize the open complex (*Micorescu et al., 2008*). The specific function of the B-finger region in promoter opening is still not clear, and the lack of this highly conserved region in *P.furiosus* TFB2 allows further study of the functional importance of this region.
*Thermococcus kodakarensis*, an archaeon closely related to *P. furiosus* also encodes two TFB homologs, TFB1 and TFB2. However, its TFB2 has an intact B finger region. *In vivo* experiments with *Thermococcus kodakarensis* shows that deletion of TFB1 or TFB2 does not affect the viability of the organism although TFB1 supported transcription *in vitro* at a lower salt concentration compared to TFB2 (*Santangelo et al., 2007*). This suggested that these TFBs have redundant functions. A recent study in *Thermococcus kodakarensis* shows that TFB1 is the more abundant and heat-inducible TFB. Transcriptome analysis also showed that the expression of genes for motility and adhesion is under the control of TFB1 whereas TFB2 controls expression of genes for mevalonate/lipid biosynthesis (*Hides et al., 2014*).

The TFB2 mRNA in *Pyrococcus furiosus* has been shown to be upregulated during heat shock conditions *in vivo* (*Shockley et al., 2003*). It was previously shown that the TFB2 from *Haloferax volcanii* and TFBb from *Halobacterium sp. NRC-1* bind to the heat shock responsive promoter his5 in a temperature-dependent manner (*Lu et al., 2008*). We predict that TFB2 in *Pyrococcus furiosus* has functions that contribute to cell survival at high temperature.

**1.7 Current understanding of *Pyrococcus furiosus* TFB2 structure and function**

There is no *P. furiosus* TFB2 structure at this time. Since the topography of the archaeal initiation complex is similar to eukaryotic initiation complex as shown by protein DNA crosslinking experiments (*Bartlett et al., 2000, Bartlett et al., 2004 and Vassylyev et al., 2002*) we can predict the orientation of archaeal TFB by comparison with TFIIB in the initiation complex. Partial alignment of multiple TFBs from various
organisms (Figure 1) below shows that TFB2 from *Pyrococcus furiosus* is largely similar to other TFBs and TF(II)B, except for missing residues in the B-reader and B-linker regions. Since the sequence of TFB2 is similar to TFB1 and eukaryotic TFIIB (Figure 1), I will discuss the structure of TFB2 based on the known TFIIB structure shown in Figure 2.

**Figure 1**: Partial alignment of B-finger region of TFBs from various archaea and eukaryotes. The alignment shows missing residues in *Pyrococcus* TFB2. The alignment was done using Clustal X 2 software, and shows the sequence conservation of amino acids in B finger region. The domain boundaries are based on the *Saccharomyces cerevisiae* TFIIB (Kostrewa et al 2008). The conserved residues are highlighted using default color setting by the software. *Pfu*, *Pyrococcus furiosus*; *Tk*, *Thermococcus kodakarensis*; *Sso*, *Sulfolobus solfataricus*; *Sa*, *Saccharomyces cerevisiae*; *Hs*, *Homo sapiens*

Co-crystal structures of RNAP II with TFIIB provided information that was used to help name the sub-structures in TFIIB shown in Figure 2. One of these structures was also used to model the positioning of TFIIB relative to DNA during closed complex and open complex formation (Figure 3 and Figure 4). In the closed complex DNA runs parallel to the RNAP II cleft, the TFIIB B-ribbon is interacting with the dock domain of RNAP, the B-reader is in the cleft, the B-linker is close to clamp coiled-coil, and the B-core is above the wall of RNAP. In the open complex model, the template strand with transcription start site +1 reaches the active center of RNAP. The initiator sequences near +1 are close to the B-reader helix, as seen more clearly in Figure 4. In this paper
information on the B-reader loop region of TFIIB was missing from the electron density, so it was modeled. A later report describes more details on the positions of the TFIIB B reader helix, the B reader strand and the B reader loop, in the open complex, and with a 6 nucleotide RNA chain (Figure 5).

A

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>36</th>
<th>63</th>
<th>103</th>
<th>205</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B- ribbon</td>
<td>B-reader</td>
<td>B-linker</td>
<td>B-core N terminal Cyclin domain</td>
<td>B-core C terminal Cyclin domain</td>
<td></td>
</tr>
</tbody>
</table>

B

Figure 2: Structure of TFIIB. (A) Primary structure of TFIIB with domain boundary numbering based on *Pyrococcus furiosus* TFB1. (B) TFIIB from *Saccharomyces cerevisiae* showing domain organization in new cartoon model (open promoter complex structure downloaded from [http://www.lmb.uni-muenchen.de/cramer](http://www.lmb.uni-muenchen.de/cramer) website; Kostrewa et al, 2009) drawn using VMD software. The B-reader loop was not seen in actual crystal structure and was later modeled in (PDB 3K1F).
Figure 3: Closed and open initiation complex model of eukaryotic RNAP II and TFIIB with TBP and DNA. The model was made using the coordinates from Kostrewa et al 2009. The RNAP II is color coded silver, TFIIB is green, TBP is purple, template strand DNA blue and non-template strand cyan (A) The closed complex in which DNA is running parallel to the RNAP II cleft. (B) The open complex, in which the template strand is at the active site of RNAP II

Figure 4: Structure of TFIIB with DNA showing possible interactions of DNA sequences with different parts of TFIIB. The view was made using the open complex model (Kostrewa et al 2009) by VMD 9.1 software. The DNA was not in the original crystal structure, and its positioning is only inferred.

From these models the possible functions of B-reader and B-linker regions of TFIIB can be predicted more accurately. The B-linker region helps in DNA melting at
around 14 base pairs away from the transcription start site and stabilizes the open complex with the help of the clamp coiled-coil and rudder region of RNAP. The B-reader region scans the initiator sequences located upstream of the +1 site and helps in transcription start site selection whereas B-reader loop region is needed for initiation of transcription. A more recent paper by *Sainsbury et al, 2012* shows the structure of initially transcribing RNAP II- TFIIB with a 6 nucleotide RNA (Figure 5). This structure agrees with a previous model (Figure 3), with the TFIIB B-linker helix interacting with RNAP coiled coil, the B-linker strand with rudder, the B-ribbon interacting with wall, the B-reader region close to initiator sequences and core domain above the wall outside the RNAP.

**A**

**B**

*Figure 5: Structure of the initially transcribing RNAP II-TFIIB complex. The model was made using VMD 9.1 software using the coordinates provided by *Sainsbury et al, 2012*. The TFIIB is color coded green, DNA is blue, RNA is red and RNAP II is silver in the model. (A) The initially transcribing model with 6 nucleotide RNA chain. The mobile B-reader loop and B-reader strand, which were not seen in previous structures, are seen in this structure. The B-reader loop is interacting with rudder and fork loop 1 region and B-reader strand interacts with lid of RNAP II which might help to stabilize the open complex and proper positioning of DNA in active center (B) Structure of TFIIB with only DNA and growing RNA chain. The growing RNA chain clashes with B-reader loop region indicating that B-reader loop might direct RNA towards the RNA exit tunnel. The RNA is only 6 nucleotides long and the non-template strand is missing in the structure.*
Like TFB1, TFB2 can form transcription initiation complexes but with low efficiency, perhaps because of a deficiency in the promoter opening process (Micorescu et al, 2008). Poor promoter opening could be due to different orientation of TFB2 N-terminus because of missing parts in the B-reader, B-linker, or both (Figure 6).

Figure 6: TFIIB structure (PDB: 4BBS). TFIIB structure in cartoon model (PDB 4BBS; Sainsbury et al, 2012) showing the complete B-reader and the B-linker region. The mobile N-terminus and C-terminus cyclin fold was removed from TFIIB prior to crystallization. Based on initial alignments, TFB2 is likely to be missing some parts of the B-reader and B-linker; indicated by the circled area.

In this thesis, I investigate the structure of TFB2. Since TFB2 lacks 27 amino acids in its N-terminus that are present in TFB1, I tested the hypotheses that i) TFB2 interacts with DNA near the transcription start site, and ii) The promoter opening deficiency of TFB2 is a result of different orientation of TFB2 around promoter DNA. I compared the *Pfu* TFB2 structure with *Saccharomyces* TFIIB structure by Kostrewa et al, 2009 and Sainsbury et al, 2013 to predict exactly which particular region is missing in
TFB2 (Figure 2B and Figure 6). For example, it may be that part of the B-reader and part of the B-linker strand are missing in TFB2 instead of the whole B-reader region.

Protein-DNA cross-linking assays were done to orient TFB2 to DNA in the transcription complex, using the well-characterized glutamate dehydrogenase promoter from *Pfu* (Pf 1602) and observing the interactions between the promoter DNA and B-reader and B-linker regions of TFB2. In an initial control, I mutated an amino acid in the B-core region of TFB2, which aligns with F192 of TFB1 that cross-links to -19 position on the template strand of DNA, to see if the core region of TFB2 behaves similar to TFB1. Previous cross-linking experiments by *Micorescu et al, 2008* show that the amino acids W44, R52, T58 in the B-reader regions of TFB1 are close to transcription start site (since these amino acids were shown crosslinking to -6, -4 and -2 positions) on template strand. Using this as a guide I have mutated amino acids aligning to two of these amino acids W44 and R52 in TFB2 and see if they cross-link at similar positions on DNA as TFB1 as my initial experiment. Later using the information I obtained from the initial experiment, I further mutated 3 other amino acids in TFB2 and tried to determine where they cross-link to DNA. Data from these cross-linking experiments has helped to specify the TFB2 orientation, which could help to further elucidate the role of TFB2 *in vivo*. 
CHAPTER II

2. EXPERIMENTAL PROCEDURES

2.1 Mutagenesis of TFB2 by insertion of a cross-linkable unnatural amino acid.

Sequence alignments of multiple TFBs and TFIIBs from archaeal and eukaryotic organisms were done using Clustal X2 software. Cross-linking experimental results with *Pyrococcus furiosus* TFB1 by Mike Micorescu (unpublished) were used as a guide, and based on the alignment of TFB1 and TFB2 (Figure 7), amino acids R54, K57, N59, P62, I68 and V175 in TFB2 were chosen for mutagenesis.

![Figure 7: Partial alignment of *Pyrococcus furiosus* TFB1 and TFB2 proteins. TFB1 amino acids mutated by Mike Micorescu are color coded red and the TFB2 amino acids mutated are color coded green. The placements of the boxed amino acids of TFB2 were uncertain.](image)

The unnatural amino acid \( p \)-benzoyl L-phenylalanine (pBpa) was inserted at the positions in TFB2 indicated above following a protocol by Ryu and Schultz (2006). This approach utilizes a plasmid expressing an engineered tRNA and aminoacyl tRNA synthetase system as shown in Figure 8. The engineered tRNA reads the stop codon UAG and inserts the unnatural amino acid pBpa at that position. Thus an overexpressed protein product containing the photoactivatable amino acid pBpa is produced, which can be used for cross-linking reactions later on.
Figure 8: Unnatural amino acid pBpa incorporation. A Modified genetic system containing engineered tRNA-aminoacyl tRNA synthetase used for incorporation of unnatural amino acid into protein was described by Ryu and Schultz, 2006. (A) The unnatural amino acid p-benzoyl L-phenylalanine (pBpa) is shown. The exposed oxygen is excited upon UV irradiation and abstracts hydrogen, followed by radical capture by the central carbon causing pBpa to get inserted into C-H bond of DNA non-specifically, resulting in the formation of covalent bond between DNA and protein. This defines the proximity of amino acid with DNA. (B) Illustration of how the system works during protein overexpression.

A site-directed mutagenesis protocol was performed to create the TFB2 variants, as follows:

The oligonucleotides were purchased from Eurofins MWG Operon. Original stock were diluted to 100uM (100pmol/ul) final concentration by addition of double distilled water as recommended in the oligonucleotide data sheet and stored at -20°C freezer. The primers used for mutagenesis in this study are listed below.

R54 Bpa forward

5’ ACAACCTTGTCGATTCAGAGTTAAGTAGAAGACCAAGACTAATGATATCCCAAG 3’

R54 Bpa Reverse

5’ CTTGGGATATCATATTAGTCTTGCTCTTAGAACTAACTCTGAATCGACAAGGTTTGT 3’
K57 Bpa forward

5’ ATTCAGAGTTAAGTAGGAAGACCTAGACTAATGATATCCCAAGATAC 3’

K57 Bpa Reverse

5’ GTATCTTGGGATATCATTAGTCTAGGTCTTTCTACTTAACTCTGAAAT 3’

N59 Bpa Forward

5’ GAGTTAAGTAGAAAGACCAAGACTTAGGATATCCCAAGATACACTAAAAGA 3’

N59 Bpa Reverse

5’ TCTTTTAGTGTATCTTGGGATATCTAGTCTTTGGTCTTTCTACTTAACTC 3’

P62 Bpa Forward

5’ TTAAGTGAAGAAAGACCAAGACTAATGATATCTAGAGATACACTAAAAGAATTGGAGA GTT CAC 3’

P62 Bpa Reverse

5’ GTGAACCTCTCCAATTCTTTTAGTGTATCTAGAGATACATTGTTGGTCTTTGTCTACT TAA 3’

I68 Bpa Forward

5’ CCAAGACTAATGATATCCCAAGATACACTAAAAGAAGTAGGGAGAGTTCACTAGAGAG 3’

I68 Bpa Reverse

5’ CTCTCTAGTGAACTCTCCCTATCTTTTAGTGTATCTTGGGATATCATITAGTCTTGGG 3’
V175 Bpa Forward

5’ AGAAAGAGATTATGAAAGCCTT**TAG**ATAGATAGTTAGGAATCTCAACTTAACAC 3’

V175 Bpa Reverse

5’ GTGTTAAGTTGAGATTCTA**ACT**ATCTATCTAAAGGCTTTCCATAATCTCTTTCT 3’

A PCR reaction was set up with a TFB2 overexpression plasmid (TFB2 inserted in PET 21b) and respective forward and reverse primers for all 6 positions, and efficiency of mutant plasmid generation was estimated by blue/white screening of a pWhitescript 4.5 plasmid (pWS) control. A 50ul PCR reaction contained 5ul of 10 X thermopol buffer, 30ng TFB2 plasmid, 125ng each of forward and reverse primer set, 1ul dNTP mix and 125ng of DNA polymerase. The PCR was set up for 1 initialization cycle at 98°C for 30seconds, and after that there were subsequent steps of denaturation at 95°C for 30 seconds, annealing at 55°C for 60 seconds and extension at 68°C for 12 minutes which was repeated 18 times. The final extension was done at 68°C for 15 minutes. After the PCR reaction was complete, template DNA was digested by adding 2ul of DpnI and incubating at 37°C for 2 hours. To check the efficiency of PCR reactions the products were run on a 8% PAGE gel, stained with ethidium bromide, and viewed with a UV transilluminator. After the formation of products was confirmed, each of them was subjected to phenol/chloroform extraction and ethanol purification step. The DNA pellet obtained was dissolved in ddH2O and to remove excess salt the DNA was dialyzed against 1mM Tris PH 8.0. The dialyzed DNA was transformed into electrocompetent XL1 blue E. coli cells by electroporation. Transformants were sub cultured twice on LB ampicillin
plates and were grown by shaking at 37°C overnight in 10ml LB ampicillin broth medium with 200ug/ml ampicillin. The plasmid was then extracted from the culture using Gene JET plasmid Miniprep kit, quantitated by spectrophotometer and sent to the OHSU core facility for sequencing to confirm the mutation. The resulting mutant TFB2 plasmid DNA was then transformed into BL21 gold cells with plasmid pSup-BpaRS-6TRN (which encodes modified tRNA and aminoacyl tRNA genes) and transformants were grown in LB media supplemented with 100ug/ml ampicillin and 25ug/ml chloramphenicol medium at 37°C overnight. The freezer stocks of the double transformants were placed at -80°C freezer for future use.

2.2 Protein overexpression and Purification

The overexpression strain (BL21 Gold) with required plasmids was cultured on LB plates supplemented with 100ug/ml ampicillin and 25ug/ml chloramphenicol overnight at 37°C. A loopful of culture was taken from the plate and inoculated in 2X YT media with ampicillin and chloramphenicol, and containing 1mM pBpa. The cells were grown to A$_{600}$ of 1.0 by continuous shaking at 37°C. After the cell density reached 1.0, TFB2 was overexpressed by adding IPTG to final concentration of 0.5mM, and cultures were incubated at 37°C overnight for protein overexpression. The cells were then pelleted, resuspended in lysis buffer and sonicated with a microprobe, 3× 30 seconds to release the proteins from cells.

The protein was then purified using standard his-tag purification protocol. The sonicated cells were centrifuged at 14K rpm, and the supernatant was applied to a Ni-
NTA agarose column. Proteins with a His-tag bound to the resin beads on the column. The column was then washed with buffer containing different concentrations of imidazole and urea, and fractions were collected. The column was first washed with buffer containing 10mM imidazole and 8M urea, and fractions was collected as Flow through (FT). Subsequently the column was then washed with buffer containing 10mM imidazole and 8M urea, followed by 20mM imidazole and 8M urea, and then 20mM imidazole and 3M urea, and fractions were collected as W1, W2 and W3. Washing with 20mM imidazole and 3M urea was repeated and collected as W4. The His-tagged protein was then eluted out twice with buffer having 200mM imidazole and 3M urea, and fractions were collected as E1 and E2. The imidazole competes with the 6X His-tagged TFB2 protein for binding to the column and urea is a denaturant that causes protein to unfold. With this washing and elution scheme, the contaminants are eluted first with buffer containing a low concentration of imidazole and a high concentration of urea, and the protein of interest is eluted later in E1 and E2 fractions with a high concentration of imidazole and a lower concentration of urea. The TFB2 proteins were more difficult to purify compared to TFB1 protein as they inefficiently bind the Ni-NTA column under native solution conditions, presumably because the 6x His tag is inaccessible. Thus 3M urea was added along with 200mM imidazole for eluting final fractions of protein (E1 and E2) which is a modification to the previous native purification protocol used in the lab.

The purified protein fraction was subjected to a protein concentration step using Amicon 3000MW cutoff filtration units with TFB storage buffer containing 500mM NaCl, 20mM Tris-Cl pH8 and 1mM EDTA. To remove the excess urea the E1 and E2
fractions were washed until the urea concentration was reduced to about $10^{-7}$ M. The purified proteins were then quantitated by running them on 12% SDS-PAGE gel along with a Bovine serum albumin dilution series, staining with Coomassie blue, and measuring band intensity using Image Quant 5.2 program. Recovery of TFB2 proteins after the concentration step was inefficient, because TFB2 sticks to the filter and does not dissolve completely in the TFB buffer used during concentration step. In summary, much of the TFB2 remained attached to the filter of the protein concentrating devices instead of being collected in the concentrated fraction, and the final concentration of TFB2 proteins was generally low, but enough was recovered to complete the transcription and cross-linking experiments described here. The final concentrations of all proteins were: wild type TFB2 6.83uM, TFB2 R54pBpa 4.832uM, TFB2 P62pBpa 3.63uM, TFB2 K57pBpa 13.84uM, TFB2 N59pBpa 13.11uM, TFB2 I68pBpa 7.13uM, TFB2 V175pBpa 18.44uM and wild type TFB1 87.5uM.

2.3 Standard transcription assays

To confirm the activity of the purified and concentrated proteins standard in vitro transcription assays were performed as described previously. The transcription buffer contains 40mM Na-HEPES pH 7.3, 250mM NaCl, 2.5mM MgCl$_2$, 0.1mM EDTA, 5mM beta-mercaptoethanol, 5% (v/v) glycerol and 0.1 ug/ul BSA solutions in 12.5ul of final reaction volume. The buffers were mixed with 10nM gdhp promoter DNA, 10nM RNAP, 60nM TBP and 30nM to 240nM TFB proteins. The mixture was then overlaid with 30 ul of mineral oil and incubated at 65°C for 40 mins. After 40 mins ribonucleotide triphosphates (0.5mM of ATP, CTP and GTP, and 10uM of [$\alpha^{-32}$P] UTP ([~40Cimmol$^{-1}$])
were added at 30 sec interval. The reaction mixture was further incubated at 65°C for 20 mins for transcription initiation process to occur and stopped using 12.5ul stop buffer containing 8M urea, 0.05M EDTA, 0.09M Tris-borate buffer, 0.02% xylene cyanol and 0.02% bromophenol blue. The transcripts were then run on a 14% polyacrylamide gel and the bands visualized by exposing the gel to a phosphoimager plate as described previously (Bartlett et al, 2000). The plate was scanned by phosphoimager and bands analyzed using Image Quant TL software package.

2.4 Protein DNA cross-linking assays

Protein DNA crosslinking assays were performed to determine the proximity of TFB2 protein to DNA during transcription initiation. First, biotinylated transcribed or non-transcribed strand of the gdhpr promoter DNA was made. To make biotinylated transcribed strand DNA template, 1pmol/ul each of biotinylated lower primer (+37 L bio) and -56 upper primers were used in 100ul of final PCR reaction mixture. Similarly for making biotinylated non-transcribed strand DNA template 1 pmol/ul each of biotinylated upper primer (-60Ubio) and +35 lower primers were used in 100ul of PCR reaction. The PCR products were run on 8% PAGE gel and visualized on UV transilluminator after staining the gel in 0.5ug/ml ethidium bromide solution. The PCR product (DNA) was purified using Gene Jet PCR purification kit and quantitated using a UV spectrophotometer. These DNAs were then used to make bead bound DNA essentially as described (M Bartlett, 2004). 100pmol of DNA was mixed with 100 ul of Streptavidin T1 magnetic beads so that the final concentration of bead bound DNA after binding and washing steps using different buffers following protocol would be 1pmol/ul. The biotin
labelled DNA was attached to the streptavidin magnetic beads and then suspended in 0.1N NaOH to denature the DNA. The non-biotinylated DNA strand was released in the supernatant and washed away, and the end product attached to the beads was left as single stranded DNA. The efficiency of the protocol was checked using 8% PAGE gel where the supernatant and the initial DNA solution was loaded to see if there was decreased amount of DNA in supernatant fraction. Gdh promoter DNA templates with radioactive ATP’s at -4, -6, -19, +6, +11 positions in template strand and -3 and -7 positions of non template strand were prepared following a protocol as described (M. Micorescu, 2008 and M. Bartlett, 2004). The gdh promoter and oligonucleotides that give rise to desired radiolabel incorporations are illustrated in Figure 9.

Figure 9: Radioactive incorporations. Promoter for the glutamate dehydrogenase (gdh) gene of Pyrococcus furiosus and the oligonucleotides used for making radioactive incorporations are shown. The positions where the radioactive α32P dATP were inserted is shown with red arrows. The respective oligos used for making incorporations at required positions in both template and non-template strands are as shown.

For synthesizing promoter DNA strands with radioactive label at specific positions first the bead bound biotinylated upper (U bio) or biotinylated lower (L bio) DNA prepared earlier was mixed with shorter, gel-purified oligonucleotides and allowed
to anneal. Radioactive dATP was added and extended with Klenow exo'. All 4 dNTPs were then added to 500uM, and T4 DNA polymerase allowed complete extension of the strand. To release the DNA from the beads the product was then digested with restriction enzymes DraI and BsrB1 for template strand and non template strand respectively. The final product was double stranded DNA with radioactive dATP at one specific position in one of the DNA strands.

Cross-linking assays were then performed with the labelled DNA. The reaction mixture contained of 0.5nM radioactive gdhp promoter DNA, 10nM RNAP, 60nM TBP, 120 or 240nM TFB (TFB2 R54 pBpa, TFB2 P62 pBpa and TFB2 V175 pBpa were added at 120nM and TFB2 K57 pBpa, TFB2 N59 pBpa and TFB2 I68 pBpa were added at 240nM) and 120nM of TFE where indicated along with buffer containing 5% glycerol, bovine serum albumin 0.1ng/ul, 5mM beta mercaptoethanol and 250mM NaCl. The reaction mixture was overlaid with 10ul of mineral oil and incubated at 65°C for 1 hour under 365 nm UV light generated by a hand held UV lamp. This excites the pBpa molecule and allows it to crosslink with the DNA positions nearby it. Following cross-linking the complex was then subjected to nuclease treatment by DNase I and S1 nuclease as described previously (Bartlett et al, 2004). The nucleases digest the promoter DNA that is not bound by any transcription proteins and the cross-linked DNA is protected near the site of cross-linking. The cross-linking is then visualized by running the complexes in 9% SDS–PAGE gel, exposing the gel to a storage phosphor plate and collecting data with a phosphoimager. The gel analysis was done using Image Quant TL software. If the pBpa molecule was close to the radioactive labelled position in the DNA,
the cross-link formed was visualized, but if pBpa was not close enough, no radiolabel was transferred to the protein, and the cross-link was not detectable. This technique allows identification of proximity of pBpa to different positions on DNA.
CHAPTER III

3. RESULTS

All archaeal genomes encode a TFIIB homolog, which is required for transcription initiation. Many archaea, including *Pyrococcus furiosus*, encode two or more TFIIB homologs. *Pyrococcus furiosus* TFB2 lacks 27 amino acids in its N-terminus those are otherwise present in TFB1 and eukaryotic TFIIB. A partial amino acid alignment (Figure 1) shows that most of the missing amino acids out of 27 are likely in the B-reader and B-linker strand region. However the alignment does not indicate which part is missing, or whether parts of each region in the N-terminus are missing.

To better define the potential structure of the TFB2 N-terminus, I started with a prediction that the TFB2 N-terminus has structural similarity to TFB1, even though the sequence doesn’t align well with TFB1 and TFIIB consensus sequences in places. Since TFB2 has minimal or no defect in transcription start site selection, this suggests that it may have at least a partly functional B-reader helix, so a manual alignment was made between TFB1 and TFB2 in which there is no gap in the B-reader helix region (Figure 10). Based on this alternative alignment between TFB2 and TFB1, two predictions were made:

1. TFB2 V175 lying in C-terminal cyclin domain should cross-link -19 position of the transcribed strand similar to TFB1 F192.

2. TFB2 R54 and P62 are in the B-reader region, so they should be close to -6 and -4 positions upstream of TSS (+1) on the template strand similar to TFB1 W44 and R52.
Testing the first prediction would help show that both TFB1 and TFB2 are structurally and functionally similar in C-terminus, as suggested by the high degree of homology seen there between TFB1 and TFB2 (Figure 7). Testing the second prediction would help determine whether the orientation of the N-terminal region of TFB2 in the transcription initiation complex is similar to or different from TFB1. To test this, the DNA proximity of TFB2 R54 and P62, which align to TFB1 W44 and R52, was investigated by cross-linking. The interactions of TFB1 W44 and R52 were previously shown to cross-link to the promoter T strand at -4 and -6 relative to the transcription start site (Mike Micorescu unpublished results).

This experiment addresses whether TFB2 is lacking the B-linker strand region. The TFB1 B-linker region is important for the promoter opening process in archaea (Werner and Weinzierl, 2002; Naji et al, 2007; Kostrewa et al 2009), and it may be that the poor promoter opening by TFB2 is caused by the lack of the B-linker strand region.

To test predictions 1 and 2, variants of TFB2 were created that contain the unnatural amino acid p-benzyol L- phenylalanine (pBpa). The proximity of specific surfaces of TFB2 to DNA could then be measured by site specific protein-DNA
photochemical cross-linking. The unnatural amino acid was inserted into the different positions of TFB2 as described (Ryu and Schultz, 2006). The desired amino acid sequences in TFB2 were replaced with stop codon “TAG”, and the mutant TFB2 was co-expressed with plasmid containing engineered tRNA amino acyl synthetase and tRNA system that reads the stop codon and inserts the unnatural amino acid at the desired position in the translated product.

3.1 Purification of the TFB2 proteins

The protein was then purified using Ni$^{++}$ ion chromatography method following a denaturing protocol. Due to changes in the N-terminus, TFB2 might fold in such a way that the N-terminal tag is buried so that native TFB2 does not attach well to the Ni$^{++}$ column. The 6×His tagged protein was bound to the Ni$^{++}$ column at a high urea concentration, washed with high urea and low imidazole concentrations to remove other unwanted proteins, and then eluted with buffers containing 200mM Imidazole and 3M urea. In comparison to wild type TFB1, wild type TFB2 and TFB2 mutants were typically obtained in lower concentrations.

The wash fractions and the eluted fractions were all collected separately and run on 9% SDS-PAGE gel along with the marker. Purification of the wild type TFB1 and TFB2 proteins of *Pyrococcus furiosus* and TFB2 pBpa mutants with N-terminal 6× His-tag by NiNTA column, eluted out almost pure protein (Figure 11).
Figure 11: Purification of TFB2 R54pBpa protein followed by coomassie blue staining. 9% SDS gel was used to show the expression of 34 KD TFB2 R54pBpa mutant in presence of pBpa in the media. The eluted fractions are represented as E1 and E2, the four wash fractions are W1, W2, W3 and W4 and the pellet, supernatant and flow through fractions are represented by P, S and FT along with marker M. The TFB2 R54PBpa mutant is represented by 34 KD band in the gel.

Lanes 3, 4, 6 and 7 show little of the 34 KD band corresponding to TFB2, compared to lanes 1 and 2 indicating that the proteins bound well to the Ni\textsuperscript{2+} column and did not wash out along with buffer. The darker 34KD band in E1 and E2 fraction (Lanes 1 and 2) shows that the almost pure protein was eluted out with buffer containing 200mM imidazole and 3M urea. The elution fractions E1 and E2 were mixed and concentrated using a Microcon 3M filter device.

Other TFB2 pBpa mutants were obtained with a similar approach. The purity of the final products were similar to that observed for the R54 mutant (Figure 11 and data not shown). TFB1 W44pBpa, TFB1 R52pBpa and TFB1 F192pBpa used for my experiments were expressed and purified by Ni\textsuperscript{2+} chromatography in the absence of urea by M. Micorescu and M. Bartlett (unpublished results). In the absence of pBpa in the media, full length mutant proteins were not expressed (data not shown). For consistency in the experiment, I also added pBpa to the growth media expressing wild type TFB1 and wild type TFB2.
3.2 TFB2 proteins with unnatural amino acids are active in transcription assays

To test whether the addition of $p$-benzoyl L-phenylalanine affects TFB2 activity, the purified TFB2 variants were tested for transcription activity in standard transcription assays. The transcription reactions contained gdh promoter DNA, TBP, RNAP, radiolabelled UTP, other NTPs along with TFB1 or TFB2. Proteins were incubated with DNA, nucleotides were added, and radioactive transcripts were measured. To determine the working concentrations of TFB1 and TFB2, transcription reaction was done with different concentrations (30nM, 60nM, 120nM and 240nM) of TFBs. After the reaction was completed the run-off transcript (37nt) was visualized in 14% Polyacrylamide urea gel followed by phosphorimaging. All purified proteins were active in transcription (Figure 12, 13 and 14).
Figure 12: Transcription of wild type TFB1 and wild type TFB2. Gdh promoter DNA was used for transcription assay. 14% Polyacrylamide urea gel was used for comparison of the transcription efficiency of the wild type (Wt.) TFB1 and TFB2 followed by phosphorimaging. The black triangle at the top of the gel image shows the concentration of protein used in each reaction from lowest to highest concentration. The higher molecular weight bands represented by asterisk were likely the result of end to end template switching by RNAP. The transcription experiment was done three times and a representative gel is shown.

Figure 12 shows the comparison of the transcription efficiency of the wild type TFB1 and wild type TFB2 at concentrations 30nM, 60nM, 120nM and 240nM. Lanes 1 to 4 shows transcription in the presence of wild type TFB2, and lanes 5-8 show transcription in the presence of wild type TFB1. Lane 9 shows transcription in the absence of TFB and as expected no transcript was seen. Wild type TFB2 has low transcription efficiency compared to TFB1 at all concentrations as previously observed (Micorescu et al., 2008). Wild type TFB2 shows highest efficiency at highest concentration (240nM, lane 4)
compared to the transcription at 30nM (lane1), 60nM (lane 2) and 120nM (lane 3). Since the highest efficiency was seen at 240nM for TFB2, I used 240nM for the cross-linking experiments. The transcription efficiency of TFB1 was almost the same at 30nM (lane 5), 60nM (lane 6) and 120nM (lane 7) and little less at 240nM (lane 8). The reason for lower activity at 240nM is unclear. In the following cross-linking experiments 120nM wild type TFB1 was used.

Transcription assays were carried out for all TFB2 pBpa variants in a similar way as done for wild type TFB1 and TFB2. The results of the experiments are shown in Figures 13 and 14.

Figure 13: Transcription of TFB2 variants, TFB2 V175, TFB2 P62 and TFB2 R54. 14% Polyacrylamide urea gel was used for comparison of the transcription efficiency of the wild type (Wt.) TFB1 and TFB2 followed by phosphorimaging. The black triangle at the top of the gel image shows the concentration of protein used in each reaction from lowest to highest concentration. The higher molecular weight bands represented by asterisk were likely the result of end to end template switching by RNAP. The transcription experiment was done three times and a representative gel is shown.
TFB2 pBpa mutants V175, P62 and R54 were added at 30nM, 60nM, 120nM and 240nM concentrations (Figure 13). The run off transcripts shows that all proteins were active in transcription (lanes 1 to 12). The maximum transcription efficiency was seen with 120nM concentration of proteins (lane 2 for TFB2 V175, lane 6 for TFB2 P62 and Lane 10 for TFB2 R54). Increasing TFB2 pBpa mutants concentration above 120nM did not increase transcription efficiency (Compare lanes 2, 3 and 4 for TFB2 V175, lanes 6, 7 and 8 for TFB2 P62 and lanes 10, 11 and 12 for TFB2 R54). Thus 120nM of proteins were determined as a working concentration and used in cross-linking reactions later on. Lane 13 represents a control transcription reaction, done with all the components except a TFB2 protein. As expected no run off transcript was seen (lane 13).

The transcription assay for other TFB2 pBpa mutants (TFB2 K57, TFB2 N59 and TFB2 I68) was also carried out similarly and the result was visualized in the 14% polyacrylamide urea gel (Figure 14).
Figure 14: Transcription of TFB2 variants, TFB2 K57, TFB2 N59 and TFB2 I68. 14% polyacrylamide urea gel was used for comparison of the transcription efficiency of the wild type (Wt.) TFB1 and TFB2 followed by phosphorimaging. The black triangle at the top of the gel image shows the concentration of protein used in each reaction from lowest to highest concentration. The black dots close to the transcript in lane 3, 6 and lane 12 are background radioactivity which might have come there while transferring the radioactive gel from the plate to the chromatography paper used for drying the gel. The higher molecular weight bands represented by asterisk were likely the result of end to end template switching by RNAP. The transcription experiment was done three times and a representative gel is shown.

Lanes 1 to 4 represent the transcription of TFB2 K57, lanes 5 to 8 represent the transcription of TFB2 N59 and lanes 9 to 12 represent the transcription of TFB2 I68 at different concentrations. It shows that all the proteins had highest transcription efficiency at 240nM concentration (Lane 4, lane 8 and lane 12) and lowest at 30nM concentration (lane 1, lane 5 and lane 9). From this the concentration of TFB2 mutants to be used in cross-linking experiments was determined to be 240nM.

3.3 Protein-DNA cross-linking confirms the proximity of TFB2 V175 protein to the -19 position of the transcribed strand
To determine the surfaces of TFB2 that interact with DNA, a photochemical cross-linking approach was used. The proximity between site-specifically incorporated pBpa and DNA that is site-specifically radiolabelled was determined following cross-linking and nuclease treatment (illustrated in Figure 15 below).

Figure 15: Protein-DNA cross-linking. Cross-linking is position specific and depends on the proximity of unnatural amino acid pBpa and radiolabel on the DNA.

I predicted that the C-terminus of TFB2 is structurally similar to C-terminus of TFB1 and has similar orientation in transcription initiation complex. To test this the V175 amino acid of TFB2, which aligns with F192 amino acid of TFB1 in the N-terminal cyclin fold domain (Figure 7), was mutated to the unnatural amino acid pBpa and used for cross-linking using DNA radiolabelled at the -19 position in transcribed strand. The DNA radiolabelled at -19 position was selected because it was previously established by Mike Micorescu that the TFB1 F192 protein cross-links DNA at the -19T position.
(Littlefield et al, 1999). Using the *P. furiosus* TBP/TFB1/DNA co-crystal structure as a model, TFB2 V175 is predicted to be close to -19 position of DNA (Figure 16).

![TFB2 V175 structure](image)

Figure 16: TFB2 V175 is close to -19 position on the template strand. The structural model was made using the coordinates from the *Pyrococcus woesei* crystal structure of TBP, TFB1 (C-terminal core) and BRE+ TATA-box-containing promoter DNA (PDB: 1D3U). The T strand was represented by Chain D and residue 1430 in D represents -19T position in the PDB file. The V175 amino acid in Pfu TFB2 aligns with F192 in Pfu TFB1 as well as in Pyrococcus woesei which was identified as residue 1192 in chain B in PDB file.

I predicted that TFB2 V175 cross-links -19T DNA and not to any other positions in DNA. If V175 cross-links to -19T position, this would help confirm that TFB2 is similar to TFB1 structurally in the C-terminus.

Cross-linking assays were carried out with TFB2 V175pBpa and TFB1 F192pBpa protein along with TBP, RNAP, gdh promoter DNAs radiolabelled with $^{32}$P at several positions, in the presence or absence of TFE. The proteins were mixed and the reaction tubes were exposed to 365nm UV light for 1 hour at 65°C. The reactions were then subjected to nuclease treatment to cleave off unbound DNA. If the DNA cross-links to protein, only a small portion of the DNA will remain attached to protein after nuclease
treatment. The cross-linking reaction was then visualized in 9% SDS-PAGE gel following phosphorimaging (Figure 17).

![Diagram showing cross-linking of TFB1 F192 and TFB2 V175 with gdh promoter DNA radiolabelled at -6, -4 and -19 positions of transcribed strand. 9% SDS-PAGE gel was used to observe the cross-linking signals followed by phosphorimaging. The marker/ladder is denoted by letter L in lane 5. The lanes which contain TFE and which do not contain TFE can be identified by + and – sign respectively at the top of the gel. The 36KD band for TFB1 and 34 KD band for TFB2 observed in the gel confirms the cross-linking. The background radioactivity is shown by *. The experiment was done three times and a representative gel is shown.]

Lane 2 shows the cross-linking intensity of TFB2 V175 with -19T DNA and lane 9 shows the cross-linking of TFB1 F192 with -19T DNA. The cross-linking signal generated by both of these proteins had almost same intensity (see bands at ~34 KD and ~36KD for TFB2 and TFB1 respectively). Cross-linking of TFB1 F192 and TFB2 V175 was not seen with -6T (lanes 4 and 11) or -4T DNA probes (Lanes 7 and 13). Though no cross-linking was seen some background crosslinking was seen with -4T and -6T which might be due to the cross-linking of aromatic residues of TFBs or RNAP subunits with
the probe. The experiment was repeated 2 more times and it was confirmed that the cross-linking was only with -19T probe and not with other DNA probes.

The transcription factor TFE has been shown to help TFB1 and TFB2 to form a stable open complex (Grunberg et al, 2007 and Naji et al, 2007). To see the effect of TFE on the interactions of the TFB1 and TFB2 C-terminal domains with DNA, TFE was added to the cross-linking reactions. It was found that the addition of TFE had no effect on cross-linking of TFB1 F192 and TFB2 V175 with -19T probe DNA (lanes 1 and 8) when compared with cross-linking reactions without TFE (lanes 2 and 9). In addition, TFE does not cause TFB1 F192 or TFB2 V175 to interact with DNA at -6T or -4T (Figure 17, lanes 3, 6, 10 and 12).

3.4 Comparing two TFB2 mutants TFB2 R54 and TFB2 P62 with TFB1 mutants TFB1 W44 and TFB1 R52

Based on the manual alignment (Figure 10) I proposed that TFB2 R54 and TFB2 P62 are analogous to TFB1 W44 and TFB1 R52. Since TFB1 W44 and TFB1 R52 are close to the -6 and -4 positions of DNA in the template strand (M. Micorescu unpublished), I predicted that TFB2 R54 and TFB2 P62 should also be close to -6 and -4 positions. To answer this I performed a protein-DNA photochemical cross-linking assay with promoter DNA radiolabelled at -6 and -4 positions with or without presence of TFE. First, cross-linking was attempted with wild type TFB1 and TFB2 (see Figure 18A). No cross-linking was seen with the negative controls, wild type TFB2 (lanes 1- 4 in 18A) and wild type TFB1 (lanes 5- 8 in 18A) with both -4T and -6T probe DNA. With wild type TFB1 a slight cross-link was seen with -4T and -6T, a background cross-link which
might have occurred by cross-linking of aromatic residues of TFB1 or RNAP to the DNA.

Next, cross-linking of TFB1 W44 and TFB2 R54 were compared using -19T, -6T and -4T radiolabelled DNA probes (Figure 18B). As predicted TFB2 R54 cross-linked to -6T and -4T probes as shown in Figure 18B.
Figure 18: Cross-linking of TFB2 R54 and TFB2 P62 to the transcribed strand probes. Lanes labelled “L” denote the protein molecular ladder bands and background radioactivity is denoted by * in all gels. Wild type TFB2 and wild type TFB1 (A), TFB2 R54 pBpa and TFB1 W44 pBpa (B) and TFB2 P62pBpa and TFB1 R52pBpa (C) were used to form transcription complexes with gdh promoter DNA radiolabelled at the indicated positions. Where TFE indicated was present at 120nM. Proteins that cross-linked to DNA were identified by 9% SDS-PAGE followed by phosphorimaging. The experiment was done three times and a representative gel is shown.

Cross-linking was confirmed with 34 KD band for TFB2 and 36 KD band for TFB1. TFB1 W44 was compared with TFB2 R54 which was used as the positive control for the reaction since TFB1 W44 was seen cross-linking transcribed DNA at -4 and -6 previously. Both TFB2 R54 and TFB1 W44 cross-linked with -6T (compare lanes 3 and 11 in 18B) although no clear cross-linking of TFB2 R54 with -4T was seen in absence of
TFE (lane 1 in 18B). TFB2 R54 cross-linked to -4T only in presence of TFE (lane 2 in 18B) whereas TFB1 W44 cross-links -4T both in the absence or presence of TFE but the cross-linking intensity was higher in presence of TFE (compare lanes 12 and 13 in 18B). The cross-linking signal intensity with -6T probe was also higher in the presence of TFE for both TFB1 W44 and TFB2 R54 (compare lanes 3, 4 and 10, 11 in 18B). Lanes 5, 6, 7 and 9 in 18B shows that both TFB2 R54 and TFB1 W44 do not interact with -19T probe DNA whether TFE is present or not.

My next prediction was that TFB2 P62 and TFB1 R52 are analogous to each other and interact with the -6T and -4T DNA probes in a similar way. To my surprise TFB2 P62 did not show any cross-linking with -6T and -4T probe (lanes 4 and 6 in 18C), although TFB1 R52 cross-linked to both -6T and -4T probes as previously seen (Micorescu unpublished, lanes 9 and 11 in Figure 18C). TFB1 P62 did not show any cross-link with -6T and -4T even in the presence of TFE (lanes 5 and 7 in 18C) which was unexpected. Addition of TFE increased the cross-linking of TFB1 R52 with both -6T and -4T (lanes 8 and 10 in 18C). Lanes 1 and 13 in 18C shows that TFB2 P62 and TFB1 R52 do not cross-link with -19T probe. -19T probe DNA was used as a control, and no cross-linking with it confirms that the N-terminus of the TFB2/TFB1 protein is confined near the transcription start site. Also in this gel we can see some background cross-linking which again might be due to non-specific cross-linking of aromatic residues present in the proteins.

Since TFB2 P62 did not show any cross-linking with -6T or -4T DNA probes, it seemed likely that my manual alignment was incorrect, and that TFB2 P62 may be part of
B-linker strand, not the B-reader region. In that case TFB2 P62 might interact with the NT strand instead of T strand. It was previously found that the TFB1 E74 amino acid, which is part of the B-linker strand, cross-links to the -3 position of non-transcribed strand of DNA (Micorescu unpublished data). To address this, the -3NT probe DNA was constructed and the crosslinking experiments were done comparing TFB1 E74 with TFB2 P62 on -3NT and -4T DNA probes. Figure 19 below shows the cross-linking results for TFB1 E74 and TFB2 P62.

![Figure 19: Cross-linking of TFB1 E74pBpa and TFB2 P62pBpa to the non-transcribed strand. A) wild type TFB1 and TFB1 E74pBpa and B) wild type TFB2 and TFB2 P62 were used to form transcription complexes, cross-linked with the DNA radiolabelled at the indicated positions and analyzed. The background cross-linking is indicated by * in both gel. The experiment was done three times and a representative gel is shown.](Image)

The results confirm that TFB1 E74 cross-links to the promoter at -3NT (lane 5 in 19A) but not at -4T (lane 7 in 19A). As predicted TFB2 P62 showed crosslinking with -3NT probe (lane 4 in 19B). The cross-linking increased with TFE addition for both TFB1 E74 (lane 6 in 19A and TFB2 P62 (lane 3 in 19B). Wild type TFB1 and wild type TFB2
were used as controls and as before they did not show any cross-link with -4T and -3NT (lanes 1-4 in 19A and lanes 5,7, 8 and 9 in 19B). Taken together, this confirms that P62 is not part of the B-reader region, but instead is likely to be a part of the B-linker strand region, suggesting that the new manual alignment is incorrect (Figure 10).

3.5 Positioning of other TFB2 N-terminus amino acids in the transcription initiation complex

The conclusion from first set of experiments suggested that TFB2 R54 might be a part of B-reader helix/B reader loop region and TFB2 P62 might be a part of B-linker strand region. These results were quite exciting because there are just 6 amino acids in between TFB2 R54 and TFB2 P62 amino acids, suggesting that a transition of TFB2 interactions from the T strand to NT strand could potentially be mapped. These results opened door for further research and new hypotheses were constructed.

The next challenges were to find out whether the amino acids in between R54 and P62 are interacting with the T strand, NT strand, or with both T and NT strands. To answer this, 3 amino acids near R54 and P62 were chosen for mutagenesis to pBpa. Amino acids K57 and N59 are predicted to be part of the B-reader loop, and the third amino acid I68 a part of the B-linker strand according to the initial Clustal X alignment (Figure 1). My new hypotheses were

1. K 57 crosslinks with T strand DNA and not with NT strand
2. N 59 crosslinks with both T strand and NT strand.
3. I68 crosslinks strongly with NT strand and does not crosslink T strand at all.
Thus to answer these questions, I again performed cross-linking assays with different sets of radiolabelled DNA at various positions. I started out with TFB2 K57 and 5 DNA probes radiolabelled at -4 and +6 and +11 positions of the transcribed strand and -7 and -3 positions of non-transcribed strand. According to the structural alignment and protein-DNA cross-linking experiments, the B-reader loop region of TFBs are close to transcription start site (Kostrewa et al, 2009; Sainsbury et al, 2012 and Micorescu unpublished). Since TFB2 K57 is part of B-reader loop, I predicted that TFB2 K57 will cross-link to -4T probe and not with other probes. Figure 20 below describes the cross-linking pattern of TFB2 K57.

![Cross-linking Pattern of TFB2 K57](image)

Figure 20: Cross-linking of TFB2 K57pBpa to transcribed and non-transcribed DNA strands. A) Cross-linking of K57 with +6 and -4 positions of template and -7 and -3 positions of non-template strand DNA. B) Cross-linking of K57 with -19 and +11 positions of template strand DNA. Transcription complexes were formed and cross-linked to several radiolabelled DNAs. 120nM TFE was added to the assay where indicated. The background cross-linking is indicated by *. The 35KD band in the lanes represent the cross-linked product. The laddar is indicated by lane L. The experiment was done three times and a representative gel is shown.
As predicted TFB2 K57 cross-linked strongly with -4T probe (lane 9 in 20A) although the cross-linking was strongest in presence of TFE (lane 8 in 20A). However, K57 also showed some faint crosslinking with NT strand probes; -3NT (lane 7 in 20A) and -7NT (lane 5 in 20A) and also with upstream T strand probes; +6T (lane 3 in 20A) and +11T (lane 5 in 20B). The cross-linking was confirmed by comparing the cross-linking with -19T which is a control since the N-terminus of TFB is confined near transcription start site, and should not cross-link upstream DNA. There was no cross-link product formed with -19T (lane 3 in 20B).

My next step was to explore how the amino acid in between TFB2 K57 and TFB2 P62 would interact with DNA. The amino acid we had planned for using to answer this was TFB2 N59. We predicted that since P62 cross-linked to non-transcribed DNA strand and K57 cross-linked strongly to transcribed DNA strand, N59 would interact with transcribed and non-transcribed strand with almost similar intensity. The same sets of DNA probes were used in cross-linking assays with TFB2 N59. It was found that TFB2 N59 cross-linked with both transcribed and non-transcribed strand of DNA as per our prediction although the pattern of cross-linking was somewhat different as shown in Figure 21 below.
Figure 21: Cross-linking of TFB2 N59 with transcribed and non-transcribed DNA strands. A) Cross-linking of N59 with +6 and -4 positions of template strand and -7 and -3 positions of non-template strand DNA. B) Cross-linking of N59 with -19 and +11 positions of template strand DNA. The marker/ladder is represented by lane L. All bands less than 25 KD in both gels represent background cross-linking which is indicated by *. 120nM TFE was added to the lanes where indicated. The experiment was done three times and a representative gel is shown.

Lanes 1 and 2 in 21A shows the cross-linking of TFB2 N59 with +6T in presence and absence of TFE and lanes 8 and 9 shows the cross-linking with -4T in presence and absence of TFE. Lanes 3, 4, 5 and 6 in 18 shows cross-linking with -7NT and -3NT probes respectively with or without TFE. Cross-linking of N59 with non-transcribed DNA strand is stronger in the absence of TFE (compare lanes 2, 4, 6, 9 in 21A and lane 5 in 21B). However, TFB2 N59 cross-linking was stronger with downstream T strand probes +6T and +11T instead of upstream T DNA strand probe -4T in presence of TFE (compare lanes 1 and 8 in 21A and lane 3 in 21B). In addition, cross-linking with upstream non-transcribed DNA strand probes -3NT and -7NT was almost similar to that of -4T probe in presence of TFE (compare lanes 3, 5 and 8 in 21A). Unexpectedly, the
strongest cross-linking was seen with downstream +11T probe in presence of TFE (Lane 3 in 21B).

To test the hypothesis that P62 and amino acids to its C-terminal side are part of the B-linker, I examined the cross-linking pattern of TFB2 I68. Since TFB2 I68 amino acid would lie in the B-linker helix region according to the alignment in Figure 1, I predicted that it should cross-link with non-transcribed DNA strand but not with transcribed DNA strand. As predicted, TFB2 I68 showed cross-linking only with NT probes (-3NT and -7NT) with strongest being with -7NT probe as shown in Figure 22 below.

A                                                                               B

Figure 22: Cross-linking of TFB2 I68 with transcribed and non-transcribed DNA strands. Lane L represents the ladder. A) Cross-linking of I68 with +6 and -4 positions of template and -7 and -3 positions of non-template strand DNA. B) Cross-linking of I68 with -19 and +11 positions of template strand DNA. The background cross-linking is indicated with *. 120nM TFE is added where indicated. The experiment was done three times and a representative gel is shown.

TFB2 I68 did not cross-link to -4T (lane 1 in 22A), +6T (lane 3 in 22A) and +11T (lane 5 in 22A). A cross-linking assay was performed with -19T DNA probe, as a control, as
for all TFB2 N-terminus variants. As seen with all TFBs variants, no cross-linking was seen with TFB2 I68 as well (lane 3 in 22B). Lane 7 in 22A shows cross-linking with -3NT, but the signal was stronger in presence of TFE (lane 8 in 22A). Also there was cross-linking with -7NT probe (lane 1 in 22B) and the stronger signal was seen in the presence of TFE (lane 2 in 22B). Comparing the cross-linking signal intensity between both non-transcribed DNA probes, the strongest cross-linking was seen with -7NT (compare lane 8 in 22A to lane 1 in 22B).
CHAPTER IV

4. DISCUSSION

Structural and biochemical experiments have shown that TFB plays central role in the transcription initiation process. The C-terminus of TFB interacts with BRE sequences upstream and downstream of TATA box sequences and helps in stabilizing the pre-initiation complex, whereas the N-terminus has a role in RNAP recruitment to the promoter, selection of transcription start site and formation of the open complex (Bushnell et al, 2004; Chen and Hahn, 2004; Kostrewa et al, 2009; Sainsbury et al, 2012; Bartlett et al, 2004).

4.1 The path of TFB2 in a transcription complex

My research focused on the N-terminus of transcription factors TFB1 and TFB2 from the archaeon *Pyrococcus furiosus*. The TFB2 transcription factor protein in *Pyrococcus furiosus* is very unique in that it lacks 27 amino acids in its N-terminus and shows less transcription efficiency *in vitro*, which is thought to be due to a defect in promoter opening process during transcription initiation. Mike Micorescu, 2008 has shown by domain swapping experiments that the promoter opening defect in TFB2 is caused by its defective N-terminal region, but he was unable to show specifically which part of the N-terminal region was responsible for the promoter opening defect. While sequence alignment of TFB2 with other TFBs was done, no experiments were done until now to determine which specific regions in the N-terminus of TFB2 are missing.
To see how the N-terminus of TFB2 tracks relative to DNA, I used a cross-linkable unnatural amino acid, \( p\)-benzoyl L-phenylalanine (pBpa) to find out the proximity of specific DNA and protein surfaces. The protein-DNA cross-linking assay using the unnatural amino acid pBpa is an established and powerful tool for determining the proximity of protein and DNA in the transcription initiation complex (Ryu and Schultz, 2006; Mike Bartlett and Mike Micorescu unpublished). The cross-linking complexes are only formed in the presence of unnatural amino acid pBpa, RNAP, other transcription factors, and UV irradiation (Micorescu et al, 2008). Using this method, I have mapped cross-links between several promoter DNA positions and the C-terminus amino acid TFB2 V175, as well as 5 other N-terminus amino acids TFB2 R54, K57, N59, P62 and I68.

4.2 The TFB1 and TFB2 C-terminal domains

Based on the conservation of C-terminal domain amino acids in \( P.furiosus \) TFB1, TFB2 and \( S.\ ceresvisiae \) TFIIB, I assumed that the C-terminal domain of TFB2 is oriented similarly to the C-terminal domain of TFB1 and TFIIB in transcription initiation complex. To test this, TFB2 V175 amino acid (TFB1 F192 and TFIIB I209) was mutated to the photoactivable unnatural amino acid pBpa and cross-linking experiments were done with the DNA radiolabelled at -19 position of the transcribed strand. TFB1 F192 served as a positive control since it was previously shown that TFB1 F192 cross-links to -19T (Micorescu unpublished). As predicted, TFB2 V175 cross-links to -19T with similar intensity as TFB1 F192 (Figure 17 lanes 2 and 9), in agreement with the structural model where TFB2 V175 is adjacent to the -19 position of the template strand of DNA (Figure
16). In addition, the similar intensity of cross-linking implies that the two proteins TFB1 F192 and TFB2 V175 recognize TBP/BRE equivalently, and that the defect in TFB2 activity is not due to poor promoter/TBP recognition. This implies that the orientation of TFB2 C-terminus is similar to that of TFB1 and TFIIB C-terminus in initiation complex.

4.3 Cross-linking based alignment of TFB2 to TFB1 and TFIIB

With a set of promoter DNAs radiolabelled at several positions in the transcribed and non-transcribed strands, and with several TFB2 N-terminus variants, I could more precisely determine the positions of N-terminus amino acids relative to DNA in the transcription initiation complex. TFB2 R54 cross-linked transcribed strand of DNA at positions -6 and -4 (Fig. 18B) although the cross-linking with -4T was observed clearly only in the presence of TFE (Fig.18B lane 2). In comparison, TFB1 W44 also showed cross-linking with -6 and -4 positions of transcribed strand of DNA (figure 18B). Both TFB2 R54 and TFB1 W44 showed stronger cross-link with -6 position than with -4 position. Since the TFB2 R54 amino acid behaves similarly to TFB1 W44, and based on the prediction that TFB1 W44 is part of the B-reader helix, it is possible that TFB2 R54 is part of a structure that forms a B-reader helix too. However, another amino acid of TFB1, R52, is predicted to be a part of the B-reader loop, and also cross-linked at -6 and -4 positions of transcribed strand (Fig. 18C). Therefore, TFB2 R54 could be part of either the B-reader helix or the B-reader loop.

Does TFB2 contain the B-reader helix, the B-reader loop, or some other novel structure that interacts with DNA? The Clustal X alignment shows that *P. furiosus* TFB2
contains sequences that align fairly well with the conserved B-reader loop sequences, suggesting that the B-reader helix is missing (Figure 1). Since the B-reader helix is proposed to participate in start site selection (Kostrewa et al, 2009), one prediction is that lack of this region would cause alterations in transcription start site for TFB2 relative to TFB1. At most promoters, TFB2 and TFB1 appear to have similar start sites, although at tRNA promoters, start site selection appears to be altered for TFB2 (Micorescu et al, 2008). Next, the TFB B-reader helix appears to be structurally analogous to the bacterial sigma region 3.2, which is required for transcription at low NTP concentrations (Kostrewa et al, 2009 and Pupov et al, 2014). Recent experiments have shown that TFB2 is more sensitive to low concentrations of NTP than TFB1 (Robyn Eustis, personal communication), suggesting that it is missing the region of TFB analogous to sigma region 3.2. Taken together these data point to TFB2 lacking the B-reader helix but containing a structure that may resemble the B-reader loop, which is where I predict TFB2 R54 resides.

The other prediction that I made was that the TFB2 P62 will behave similarly to TFB1 R52, and will therefore cross-link the -6 and -4 positions of transcribed strand. However, TFB2 P62 did not cross-link with any transcribed strand probes (Fig. 18 C). Instead TFB2 P62 cross-linked with the -3 position of non-transcribed strand (Fig. 19B) where it was behaving similarly to TFB1 E74 (Fig. 19A), which is an amino acid in the B-linker strand. Thus I concluded that TFB2 P62 may be a part of the B-linker strand region.
Next, I performed cross-linking experiments with 3 amino acids TFB2 K57, TFB2 N59 and TFB2 I68, to pinpoint the locations of other amino acids relative to R54 and P62 and DNA in the open complex. TFB2 K57 was found to cross-link -4 position of the transcribed strand, but was also seen to cross-link to downstream +6 and +11 positions of transcribed strand, and to -3 and to -7 positions of the non-transcribed strand (Fig. 20A and 20B). The cross-linking intensity to -4T and -7NT was similar in absence of TFE (lanes 5 and 9 in 20A) but it was strongest at the -4T position in the presence of TFE (lane 8 in 20A). Comparing the cross-linking intensity of -6T and -4T, the strongest cross-link is seen with -4T (compare lanes 3 and 9 in figure 20A). Thus I propose that TFB2 K57 is a part of the B-reader loop. Since K57 cross-links to both the transcribed and non-transcribed DNA strand probes, it might be that this part of TFB2 is slightly mobile in the time scale of the experiment. The partial alignment generated by Clustal X (Fig. 1) shows that K57 is near a gap at the C-terminal side (relative to TFB1), so part of the B-reader loop might be missing from TFB2.

TFB2 N59 was also observed to cross-link both the transcribed and non-transcribed promoter strands, upstream of the transcription start site, although very weakly. For TFB2 N59 in the presence of TFE, cross-linking increased for the downstream transcribed strand position +6T, and was much stronger for +11T compared to the upstream transcribed DNA strand position -4T (Fig. 21A and 21B). Comparing cross-linking of TFB2 N59 with TFB2 K57 to -4T, K57 has strong cross-links with -4T both in the absence and presence of TFE (lanes 8 and 9 in 20A), while N59 shows almost no cross-linking with -4T in absence of TFE (lane 9 in 21A) and weak cross-link in
presence of TFE (lane 8 in 21A). This suggests that TFB2 N59 is not a part of the B-reader loop, in contrast to TFB2 K57. In addition cross-linking of N59 was seen with both non-transcribed strand probes -7NT and -3NT (lanes 4 and 6 in 21A). Therefore, N59 is likely not a part of B-reader loop or B-linker strand but instead a part of linker region joining these two regions. The strong TFE dependent cross-link of N59 to +11T was surprising, and may illustrate movement of the downstream DNA, which may not be fixed in position by the RNAP jaws prior to transcription initiation.

TFB2 I68 shows cross-linking only with non-transcribed DNA strand probes and not with transcribed strand probes, -4T, +6T and +11T (Fig.22). The cross-linking was strongest with -7 NT (compare lane 7 in 22A to lane 1 in 22B). Thus I concluded, based on the cross-linking results and also based on the conservation of amino acids in B-linker helix region, that TFB2 I68 is in the loop between the B-linker strand and the B-linker helix region.

To confirm that the radioactivity signals transferred from DNA to protein, was result of pBpa cross-linking, I conducted experiments with wild type TFB1 and wild type TFB2. No cross-linking signal was seen with wild type TFB1 or wild type TFB2 (Figure 18 and 19) which confirms that the signal we observed is authentic and is not due to any background cross-linking. Although I could see background radioactivity in all experiments, most of those appeared as bands less than 25 KD in size. These signals could be from the incomplete digestion of DNA, or from autolabelling of nuclease during the nuclease treatment (data not shown).
Observing the results obtained with TFB2 K57 and TFB2 N59, and considering the structural transition from B-reader loop region to the B-linker region (TFB2 is a part of linker joining B-reader loop and B-linker strand), I predicted that part of the B-reader loop region and B-reader strand region are missing in TFB2. The missing B-reader sequences might have caused TFB2 to have a promoter opening defect or might lead to formation of an unstable transcription bubble. In short the low transcription efficiency of TFB2 may be explained in terms of missing B-reader loop region, compared to TFB1 which is predicted to have an intact B-reader loop region. Based on the information I collected from my experiments, I made a new possible alignment of *P. furiosus* TFB2 (Figure 23).

**Figure 23:** New alignment of *P. furiosus* TFB2 N-terminus. The alignment in a fourth row is the new manual alignment. The amino acids which were mutated to unnatural amino acid pBpa are indicated by bold font. The domain boundaries of the amino acids are based on the *S. cerevisiae* TFIIB structure ([Sainsbury et al, 2012](#)). The N-terminus of *S. cerevisiae* TFIIB is shown and the amino acids aligning with the mutated amino acids of TFB1 and TFB2 of *P. furiosus* are indicated by bold letter font and pointed by arrow. These amino acid positions in TFIIB were used to represent the respective mutated positions of TFB2, while making TFB2 models.

Using the new alignment I constructed models of TFB2 with DNA strands using the coordinates of open complex model by *Kostrewa et al, 2009* and pointed out the amino acids R54, K57, N59, P62 and I68 and the positions -6, -4, +1, +6 and +11 in the
transcribed strand. The part of the non-transcribed strand in the transcription bubble was missing in the coordinates so, a possible connection was drawn and potential positions -3 and -7 were labelled (Figure 24, 25 and 27).

Figure 24: TFB2 R54 and TFB2 P62 with transcribed and non-transcribed DNA. The positions of TFB2 R54 and TFB2 P62 were predicted from sequence alignment (Fig. 23). TFB2 R54 = TFB1 R54 = TFIIB G73 and TFB2 P62 = TFB1 E74 = TFIIB S93. (A) TFB2 R54 lies in the B-reader loop region of TFB2 and interacts with -6T and -4T positions close to TSS. The images were drawn using VMD software 9.1 using the open complex model by Kostrewa et al, 2009 and lack a part of NT strand in transcription bubble. (B) TFB2 P62 located in B linker strand region far from TFB2 R54 and explains why no cross-linking was seen with -6T and -4T positions but instead with -3NT. Due to lack of complete structure of transcription bubble the -3NT and -7NT positions could not be shown in image. The path of non-transcribed strand in the transcription bubble region is suggested with the dotted line in the images, and -3NT and -7NT positions are labelled.

I modeled R54 as amino acid G73 in TFIIB motif (Fig. 24A). The figure shows the position of R54 and that it is close to -6 and -4 positions of template strand. The cross-linking result is consistent with the structural image and shows that R54 is likely to
be close to transcription start site. The cross-linking result, the manual alignment and the structural model support the idea that there is an intact B-reader loop region in TFB2.

I modeled P62 as amino acid S93 in the TFIIB motif, which is the part of B-linker strand region (see Fig. 24B). In this model, P62 is far from -6 and -4 positions but close to the inferred -3NT position. This completely complies with the experimental results I obtained.
Figure 25: TFB2 K57, TFB2 N59 and TFB2 I68 with transcribed and non-transcribed DNA. The images were drawn using VMD software 9.1 using the open complex model by Kostrewa et al, 2009 and lack a part of NT strand in transcription (A) Predicted TFB2 K57 position on TFB2 B reader loop. TFB2 K57= TFB1 R57=TFIIB P76. (B) TFB2 N59 on the end of B reader loop based on the experimental results. TFB2 N59= TFB1 G59= TFIIB R58. (C) TFB2 I68 in the B-linker strand region. TFB2 I68= TFB1 L81= TFIIB E100. Due to lack of complete structure of transcription bubble the -3NT and -7NT positions could not be shown in image. The path of the non-transcribed strand in the transcription bubble region is suggested with the dotted line in the images, and the -3NT and -7NT positions are labelled.

I next modeled K57 as amino acid P76 in TFIIB motif and N59 as amino acid R78 in TFIIB motif respectively (Fig. 25A and 25B). Figure 25 A shows TFB2 K57 in the B-reader loop region and Fig. 25B shows TFB2 N59 in the region between the B-reader strand and the B-linker strand. A portion of the B-reader loop might be missing, which could result in mobility of TFB2 K57 and TFB2 N59, explaining why they interact with both the transcribed and the non-transcribed strand. Lastly I modeled I68 as E100 amino acid in TFIIB motif (Fig. 25C). It shows that TFB2 I68 would lie in the B-linker strand/helix region, which is closest to the non-transcribed strand in the transcription bubble, modeled as a dotted line.
In the archaeal system for *in vitro* transcription, TBP, RNAP, TFB, DNA and nucleotides are sufficient for reconstitution of initiation. TFE is not needed for conducting transcription *in vitro* although it has been shown to increase transcription by TFB2 at some promoters (Ming Hsiao Wu personal communication). Because TFE has been shown to increase TFB2 transcription, and assist TFB2 in promoter opening, it seemed likely that it would influence the placement of the TFB2 N-terminus, and therefore the cross-linking of TFB2 to DNA. With nearly all TFB2 variants and, surprisingly with the TFB1 variants, TFE induced an increase in cross-linking signal with the exception of TFB1 F192 and TFB2 V175. Since no increase in cross-linking signal with -19T was seen in presence of TFE it seems likely that TFE does not interact with the C-terminus of TFB, but does interact with the TFB N-terminus, consistent with previous results (Grunberg *et al.*, 2007 and Naji *et al.*, 2007).

### 4.4 Structural orientation of TFB2 in transcription initiation complex

There is no structure of the archaeal transcription initiation complex with the TFB2 protein. Since the amino acid sequence of TFB2 is similar to TFB1 and eukaryotic TFIIB, although it is missing 27 amino acids in the N-terminus, we can consider the structure to be similar to structure of TFIIB with RNAP II and other transcription factors shown by Kostrewa *et al.*, 2009 and Sainsbury *et al.*, 2012. As shown in Fig.16 and Fig. 17, the C-terminus of TFIIB/TFB1/TFB2 is close to -19 position in template strand DNA. Merging the mutational analysis results with the structure of TFIIB gives a model for TFB2 (Figure 26A and 26B).
Figure 26: Structure of TFB2 showing the missing structures that are normally present in TFIIB/TFB1. The image was made using the TFIIB coordinates from the RNAPII-TFIIB co-crystal structure by Sainsbury et al, 2012 (PDB: 4BBS). The missing regions were determined based on the (A) Clustal X alignment (Fig. 1) and (B) our cross-linking experimental results and manual alignment (Fig. 23). Using the Clustal X alignment, residues 40-49, 59, 60 and 61 and 70-79 in \textit{Pfu} TFB1/59-68, 78, 79, 80 and 91-99 in \textit{S. cerevisiae} TFIIB) are missing (colored pink) which are a part of B-reader helix, a part of B-reader loop and B-linker strand. (b) From my cross-linking data I found that TFB2 has a part of B-reader loop and B-linker strand but missing B reader helix and B-reader strand. Residues 50-59 and 72-83 are missing in \textit{Pfu} TFB2 which are residues 40-49 and 62-73 in \textit{Pfu} TFB1 and residues 59-68 and 81-92 in \textit{S. cerevisiae}. The TFB2 predicted structure is modeled in green and missing residues are modeled pink in both images.

Compiling all the structural information from this cross-linking study, the possible missing regions were plotted (Figure 27), which allows modeling of the orientation of TFB2 N-terminus during open complex formation.
Figure 27: TFB2 N-terminus mutations shown along with transcribed and non-transcribed DNA. The image was drawn using VMD software 9.1 and the open complex model by Kostrewa et al, 2009 and lacks a part of NT strand. The missing regions of TFB2 based on the experimental results are color coded pink. Due to lack of complete structure of transcription bubble the -3NT and -7NT positions could not be shown in image. The path of the non-transcribed strand in the transcription bubble region is suggested with the dotted line in the images, and the -3NT and -7NT positions are labelled.

Since the complete DNA is not present in the RNAP-TFIIB structure by Sainsbury et al, 2012, the TFB2 orientation shown in the Figure 27 should be considered the best current hypothesis, given the data. However, a more complete structure and more biochemical experiments in the future will help to draw a more accurate orientation, and this opens a door for further research.

In summary, I can confirm that the C-terminus of Pyrococcus furiosus TFB2 is oriented similarly to Pyrococcus furiosus TFB1 and eukaryotic TFIIB C-terminus during transcription initiation, and that the N-terminus is oriented differently than TFB1 N-terminus domain at the gdh promoter (Figure 28).
My results suggest that the orientation of *Pyrococcus furiosus* TFB2 N-terminus is different in transcription initiation complexes compared to *Pyrococcus furiosus* TFB1 and eukaryotic TFIIB due to the missing B-reader helix and B-reader strand regions. The results from my experiments have helped to broadly determine TFB2 orientation, in a transcription initiation complex containing TFE. More extensive cross-linking experiments with additional pBpa mutants, and additional radiolabelled positions in the DNA are needed to find out the exact path of TFB2 N-terminus in the initiation complex.
REFERENCES


61. Luse DS. Rethinking the role of TFIIF in transcription initiation by RNA polymerase II. *Transcription*. 2012; 4: 156-159.


77. Reeve JN, Sandman K, Daniels CJ. Archaeal histones, nucleosomes and transcription initiation. *Cell.* 1997; 87: 999-1002.


