ANALYSIS OF PROTEIN-DNA INTERACTIONS WITHIN THE TRANSCRIPTION PREINITIATION COMPLEX OF *PYROCOCCUS FURIOSUS*

by

MATTHEW BLAIN RENFROW

Under the Direction of Robert A. Scott

ABSTRACT

Initiation of transcription requires the assembly of a multi-protein preinitiation complex (PIC) on promoter DNA. The protein-DNA interactions within the transcription PIC of the archaeal hyperthermophile, *Pyrococcus furiosus* (*Pf*) were probed by site-specific DNA-to-protein photocrosslinking. This involved the assembly of the PIC including *Pf* RNA polymerase (*Pf* RNAP) and general transcription factors *Pf* TATA-binding protein (*Pf* TBP) and *Pf* transcription factor B (*Pf* TBP) on 41 azido-phenacyl derivatized DNA promoter fragments containing the *Pf* glutamate dehydrogenase upstream promoter + gene. The region from -40 to +1 (transcription start site is +1) of both the template and nontemplate DNA promoter was probed. *Pf* RNAP-DNA promoter interactions span the entire probed region. *Pf* TFB-DNA promoter interactions occur upstream and downstream of the TATA box promoter element. *Pf* TBP-DNA interactions center around the TATA box promoter element. Results provide evidence of the homology between eukaryal RNA polymerase II gene transcription system and the archaeal gene transcription system. *Pf* TFB-DNA interactions are more extensive in the archaeal PIC than those seen in the analogous eukaryal system between transcription factor IIB (TFIIB) and promoter DNA. The *Pf* TFB-DNA interactions reported here provide support for a mechanism by which *Pf* TFB carries out its roles in PIC assembly of recruiting *Pf* RNAP to the complex and accurate start site selection.

INDEX WORDS: Archaea, *Pyrococcus furiosus*, basal transcription, preinitiation complex, transcription factor B, TFB, transcription factor IIB, TFIIB, TATA-binding protein, TBP, RNA polymerase, RNAP, photocrosslinking, azido-phenacyl, phenyl azide, protein-DNA crosslinking, molecular topography

ANALYSIS OF PROTEIN-DNA INTERACTIONS WITHIN THE TRANSCRIPTION PREINITIATION COMPLEX OF *PYROCOCCUS FURIOSUS*

by

MATTHEW BLAIN RENFROW

B.S., Western Kentucky University, 1996

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2002

ANALYSIS OF PROTEIN-DNA INTERACTIONS WITHIN THE TRANSCRIPTION PREINITIATION COMPLEX OF *PYROCOCCUS FURIOSUS*

by

MATTHEW BLAIN RENFROW

Major Professor: Robert A. Scott

Committee: James Omichinski
Anne O. Summers
Claiborne Glover III
Robert Woods

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2002

DEDICATION

This work is dedicated to my best friend and wife Tamara. To my parents for just expecting me to do my best and to do it right. To my brother Steven and my sisters Karen and Kendra. The bar always seemed to be set high, but we have enjoyed seeing each other jump over it in our own ways. And to the notion that intelligence is one third smarts, one third effort, and one third follow through.

ACKNOWLEDGEMENTS

Graduate school is 90% organization, 10% everything else. No wonder it took me so long. There are many people to thank. This has by no means been an individual effort.

Working for Dr. Scott has been great. So many times when I wanted to talk forrest, you made me talk trees (or vice versa). Not that this is a bad thing, I've come to learn. Thanks to the members of the Scott lab who laid the ground work for the transcription project (Qiandong, Michelle, Chris and H.-T.). Especially H.-T. for his encouragement and example. Thanks to Marly for always being an incredibly reliable source of knowledge. I cannot forget the Richard Ebright Lab at Rutgers University (the state university of New Jersey) for all their help in getting my research off the ground. Especially Nikolai, for his patience and thoroughness in teaching me photocrosslinking techniques. "This is not good. Not good." Everyone at some point in their life should drive to New Jersey during the winter and learn lab techniques from a Russian. Da!

Thanks to my old room mates (Josh, Murray, Dave, and Harris) who made things interesting always. Thanks to all my friends and fellow Christians at Oglethorpe Avenue. Especially Marc and Laurie, for their understanding, encouragement, and friendship.

Special thanks to my family for never letting me forget who I am and what is really important. I am a product of your love and guidance.

Then there is my wife Tamara, the most unexpected and wonderful surprise of Georgia. From the very beginning you have been there. This would not have been

finished without your help, hard work, constant support, encouragement, and love. I enjoy my life with you so much.

Above all else, because of the many blessings I have been given, I acknowledge, praise, and thank God. Six, twenty-four hour days, that's all it took Him.

TABLE OF CONTENTS

Page

CHAPTER 1

TRANSCRIPITON PREINITIATION COMPLEX AND ANALYSIS OF PROTEIN-DNA INTERACTIONS

1.1. Introduction

To synthesize a protein, a cell must first have a blueprint. Within the cell, DNA is the blueprint for proteins as well as other cellular machinery. Transcription is the process by which the blueprint for an individual gene is transcribed into a single molecule of RNA based on the DNA template. Transcription involves multi-protein-DNA complexes. These complexes carry out the three major events of transcription: initiation, elongation, and termination. Initiation involves the assembly of structural proteins and enzymes at the correct starting position (start site) for a given gene within a genome. This is followed by the polymerization of the first few ribonucleotides of the RNA molecule. After transcription is initiated, the assembled complex transitions into the elongation phase of transcription where the RNA molecule is synthesized (or elongated) based on the DNA template at an accelerated rate compared to the initiation phase. At the end of the DNA blueprint (for a given gene) the DNA template contains a termination site to signal the disassembly of the transcribing complex of proteins.

At the center of these multi-protein-DNA complexes is the DNA-dependent enzyme RNA polymerase (RNAP), which synthesizes an RNA molecule based on a DNA template. Although it is the key component, RNAP requires several additional

proteins (known as general transcription factors or GTFs) to carry out all events of transcription. Initiating gene transcription requires the assembly of the RNAP and GTFs at specific promoter sequences of DNA. This assembly of proteins on the DNA promoter is the preinitiation complex (PIC[1]).

For example, bacterial σ–factor binds to bacterial RNAP core enzyme and facilitates binding to promoter sequences. Bacteria have several different σ–factors which allow recognition of different classes of promoters (reviewed in Xu and Hoover, 2001). In addition to DNA promoter sequences and general transcription factors, regulatory proteins referred to as activators or repressors control transcription initiation. Different regulatory proteins can interact with RNAP, DNA, or GTFs to enhance, diminish, or even prevent the initiation of transcription.

Eukaryal transcription is a more complicated and diverse process. There are three different RNA polymerases (RNAP I, II, and III) in Eukarya. RNAP I and its associated GTFs transcribe genes encoding for ribosomal RNA (rRNA). RNAP II and its associated GTFs are responsible for transcription of protein-encoding messenger RNA (mRNA). RNAP III and its associated GTFs transcribe transfer RNA (tRNA). For each RNAP system the generic requirements to initiate gene specific transcription are the same: promoter DNA, RNAP, and GTFs (Figure 1.1). Once again, it is the regulatory factors that direct the required components to assemble the PIC at a specific location in the genome to transcribe a specific gene.

Archaea, the third domain of life (Woese et al., 1990), have a transcription system that falls between bacteria and Eukarya in complexity. While there is only one RNAP as

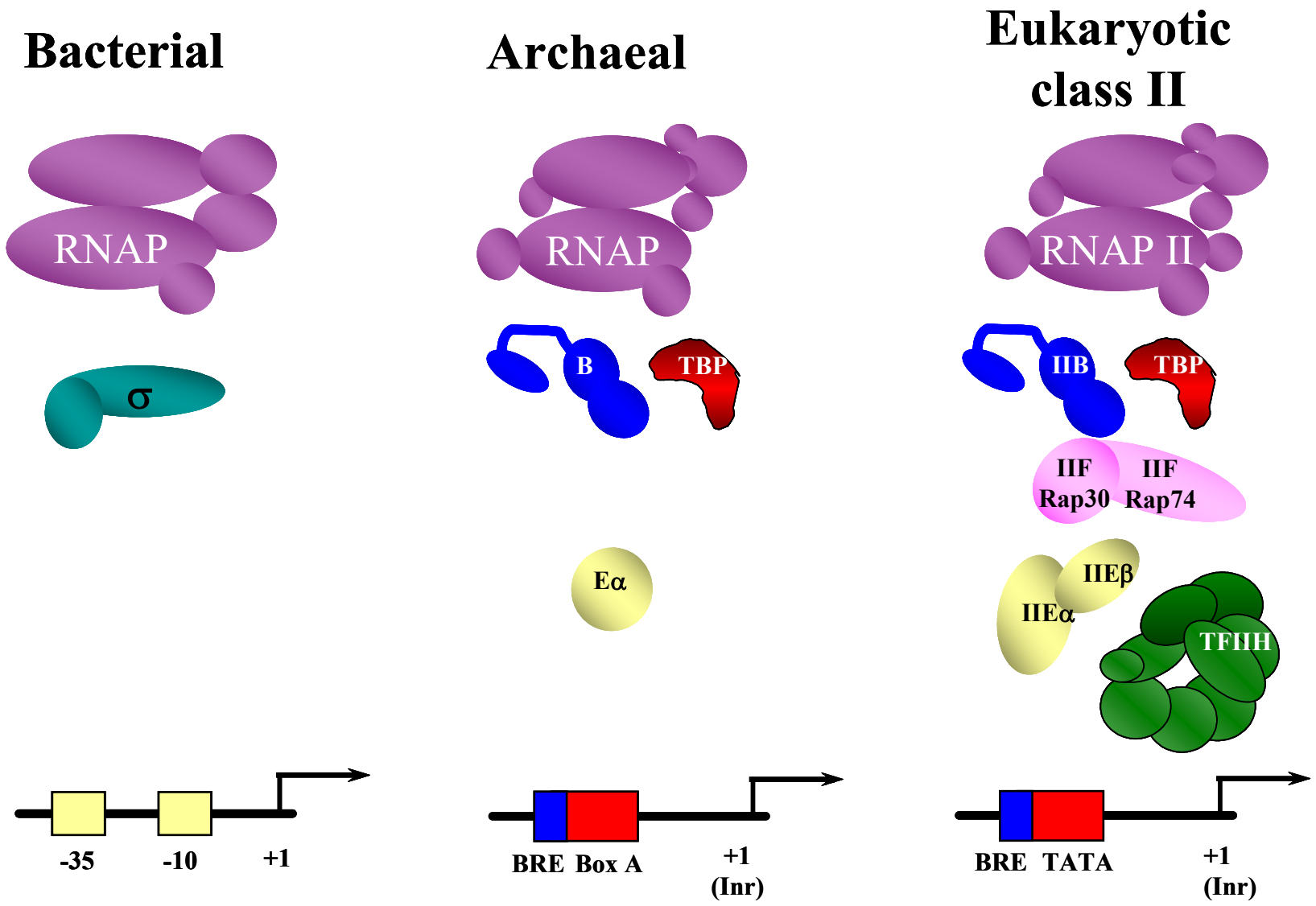
[1] Table 1.1 lists abbreviations used throughout

Table 1.1. Abbreviations used throughout

Most common abbreviations	
Pf or *P. furiosus*	*Pyrococcus furiosus*
TBP	TATA Binding protein
TFB	Transcription factor B
RNAP	RNA polymerase
aRNAP	Archaeal RNA polymerase
TFIIB	Transcription factor IIB from eukaryotic RNAP II system
TF(II)B	generic referral to transcription factor B from either system (eukarya or archaea)
TF(II)Bc	transcription factor B core or C-terminal domain
TFEα	Archaeal general transcription factor E (homologus to TFIIEα subunit)
PIC	Preinitiation complex
Gene promoter element terms	
TATA box	A/T-rich DNA gene promoter element for eukaryal genes typically 25-30 bp upstream of transcription start site
Box A	Archaeal equivalent to TATA box
BRE	transcription factor B recognition element
adMLP	adenovirus Major Late Promoter (eukaryal promoter)
Gdh	glutamate dehydrogenase promoter (archaeal promoter)
NT	Nontemplate strand of DNA promoter
T	Template strand of DNA promoter
*Pf*gdhP-M13mp18(19)	Cloned *Pf* promoter fragment
AP	Azido-phenacyl (photocrosslinking agent)
Eukaryal transcription terms	
Rpb1 - 12	Eukaryal RNA polymerase subunits
TFIIF	Eukaryal general transcription factor IIF (2 subunits)
RAP30	TFIIF subunit (RNA association protein)
RAP74	TFIIF subunit
TFIIE	Eukaryal general transcription factor IIE (2 subunits)
TFIIEα	TFIIE subunit
TFIIEβ	TFIIE subunit
TFIIH	Eukaryal general transcription factor IIH (9 subunits)
Pw	*Pyrococcus wosei*

Figure 1.1. Basal transcription machinery for the three domains of life.

Initiating gene specific transcription in all three domains of life requires promoter DNA, general transcription factors (GTFs), and a multi-subunit RNA polymerase (RNAP). In bacteria, -35 and -10 DNA promoter elements are bound by bacterial σ–factor which facilitates binding of bacterial RNAP (bRNAP) to initiate transcription. In archaea, the Box A promoter and transcription factor B recognition element (BRE) are recognized by archaeal GTF's TATA-binding protein (TBP) and transcription factore B (TFB). These two GTF's fascilitate binding of the archaeal RNAP (aRNAP) to DNA to initiate transcription. A third archaeal GTF TFEα has also been identified but its role in intitiating transcription is not clear. In the eukarya class II transcription system, the TATA box promoter and BRE are recognized by eukaryal GTF's TBP and transcription factor IIB (TFIIB). These two GTF's along with GTF TFIIF fascilitate binding of eukaryal RNAP II to promoter DNA. In addition, eukaryal GTF's TFIIE, and TFIIH are required for the accurate initiation of transcription.


Bacterial
RNAP
σ
-35
-10
+1
Archaeal
RNAP
B
TBP
Eα
BRE
Box A
+1
(Inr)
Eukaryotic
class II
RNAP II
IIB
TBP
IIF
Rap30
IIF
Rap74
IIEβ
IIEα
TFIIH
BRE
TATA
+1
(Inr)

in bacteria, this single RNAP and its associated GTFs are most similar to the eukaryal RNAP II transcription machinery (Soppa, 1999).

The transcription preinitiation complex is a landmark in the process of gene expression. Understanding the topography of this landmark is essential for understanding how the complex initiates transcription as well as how regulatory proteins interact with the complex, subcomplexes, or individual components to regulate gene expression. My graduate studies have utilized established research in the Scott group to better understand the arrangement of the preinitiation complex in the archaea, *Pyrococcus furiosus (Pf or P. furiosus)*. The current knowledge of the transcripition PIC in both eukarya and achaea will be briefly reviewed in the rest of this chapter with emphasis on the interactions of the transcriptional machinery with DNA. The development of an *in vitro* system for analyzing *P. furiosus* PIC will also be discussed. Protocols for construction and analysis of the *in vitro* system are presented in the following chapter. Site-specific protein-DNA photocrosslinking results, a proposed topographical map, and its implications for the mechanism of transcription initiation in archaea are discussed in the final two chapters.

1.2. Eukaryal transcription preinitiation complex

Eukaryal class II transcription initiation requires RNAP II and five general transcription factors (Figure 1.1, TBP, TFIIB, TFIIF, TFIIE, and TFIIH) assembling on a DNA promoter (Hampsey, 1998). This multi-protein-DNA assembly is called the preinitiation complex (PIC) and is required to achieve a basal level of accurate transcription. Core promoter elements within a gene promoter establish the site for assembly of the RNAP II PIC (Struhl, 1995). These include an AT-rich sequence

upstream of the transcription start site known as the TATA box element. While the distance from transcription start site has been located as far as 120 bp upstream (-120 bp) in yeast, the TATA element is most commonly located -25 to -30 bp from the start site. Other core promoter elements include a less defined initiator element surrounding the start site and a transcription factor IIB recognition element (BRE) immediately upstream of the TATA box (Lagrange et al., 1998).

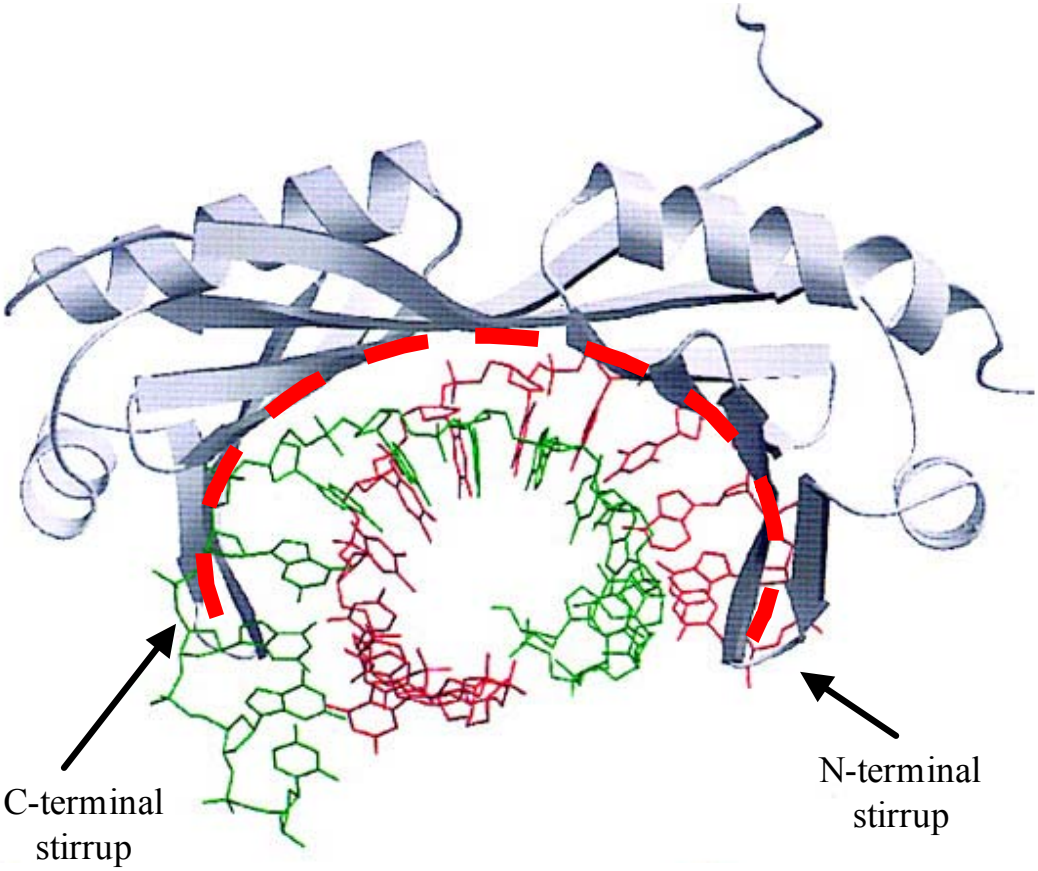
The TATA element is the binding site for the TATA binding protein (TBP). TFIIB acts as the functional bridge between TBP-bound promoter and the rest of the preinitiation complex. TFIIB interacts directly with TBP, the BRE, as well as RNAP II and TFIIF. To complete the eukaryotic PIC, TFIIE interacts with RNAP II and TFIIF and is required for the assembly of the final component of the complex, TFIIH.

1.2.1. Eukaryal TBP and TFIIB

TBP was originally identified as a subunit of TFIID, a large multisubunit complex also including TBP-associated factors, or TAFs (Pugh and Tjian, 1992). Yeast TBP was shown to be interchangeable with mammalian TBP in an *in vitro* transcription system (Buratowski et al., 1988). TBP was later identified to be an essential subunit of SL1 (Comai et al., 1994) and TFIIIB (Huet and Sentenac, 1992), GTFs for the RNA polymerase I and RNA polymerase III transcription systems, respectively. Its requirement for accurate basal transcripiton in all three systems led to it being dubbed the universal transcription factor.

Structural characterization of TBP revealed a unique DNA-binding protein fold (Figure 1.2). Sequence analysis showed two direct repeats. The crystal structure of

Figure 1.2. TATA-binding protein (TBP) bound to TATA element. Ribbon representation of *Arabidopsis thaliana* TATA-binding protein (TBP) bound to TATA box promoter DNA. The curved β-sheet is traced by a dashed line following the DNA-binding surface of TBP, which resembles the concave surface of a saddle (Figure from (Patikoglou et al., 1999).


C-terminal
stirrup
N-terminal
stirrup

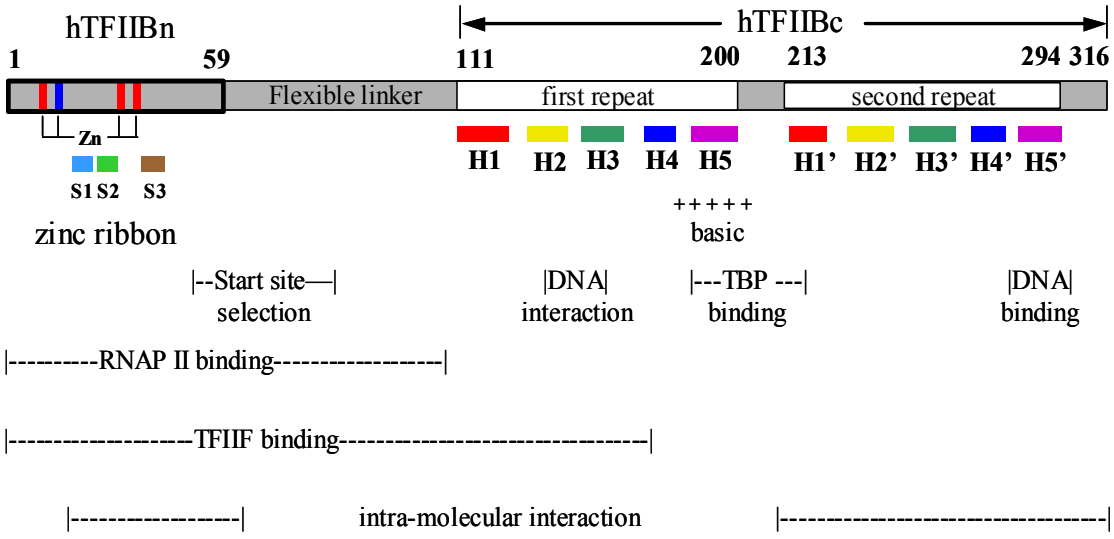
A. thaliana TBP revealed that the DNA binding surface is a curved β-sheet which resembles the concave surface of a saddle (Nikolov and Burley, 1997; Nikolov et al., 1992). This analogy is extended by the presence of a loop (or stirrup) at the ends of the long axis ofTBP. The top side of the saddle is composed of two alpha-helices, one in each repeat. This "protein saddle" binds to the DNA in the minor groove bending the DNA towards the major groove (away from the TBP surface). TBP's induced-fit mechanism of binding the minor groove is unique for a sequence-specific binding protein. Most sequence specific binding proteins interact with the major groove where the face of B-DNA is chemically variable depending on the base sequence. The minor groove is chemically monotonous. This allows for a degree of variability in the TATA element. Burley's group (Patikoglou et al., 1999) showed that TBP does not easily distinguish T:A from A:T base pairs in binding and can in fact function on many different A/T-rich sequences.

TFIIB is a structural and functional bridge between TBP and the rest of the PIC. Early sequence analysis suggested several structural motifs within the protein. Located at the N-terminus is a zinc binding motif. In addition, two imperfect repeats take up almost two thirds of the molecule (Figure 1.3.A, Malik et al., 1991). Between the two structural domains is a flexible linker region of undefined structure.

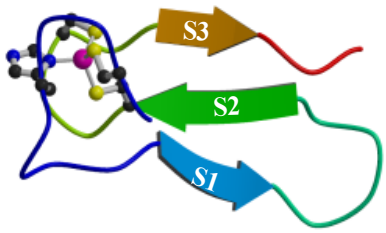
Gel mobility shift assays determined that the N-terminal region of TFIIB was required for interaction with RNAP II and TFIIF (Barberis et al., 1993). Therefore, without the N-terminal domain, PIC assembly cannot be completed. The two C-terminal repeats (referred to as C-terminal core or TFIIBc) are required for TFIIB-TBP interaction and the formation of a TFIIB-TBP-DNA ternary complex (Barberis et al., 1993;

Figure 1.3. Functional and structural domains of human TFIIB. A. The N- and C-terminal domains of human TFIIB contain independent structural motifs. At the N terminus is a zinc ribbon motif. The C terminus contains a set of two indirect repeats that fold into cyclin A-like motifs. The two domains are connected by a structurally undefined flexible linker. Functionally characterized regions of TFIIB are indicated by dashes enclosed by vertical bars. **B.** Ribbon representation of the human TFIIBn, a zinc ribbon motif (Chen et al., 2000b). **C.** Ribbon representation of human TFIIBc, two cyclin A-like folds (Bagby et al., 1995).

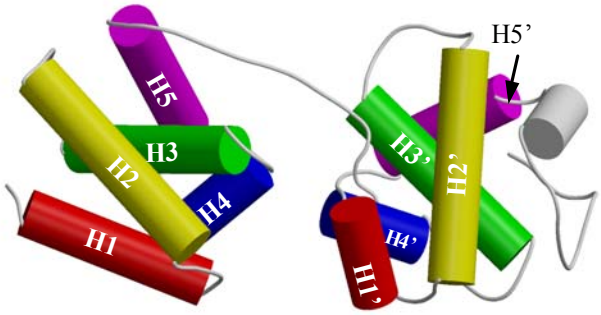
A. human TFIIB



B. hTFIIBn



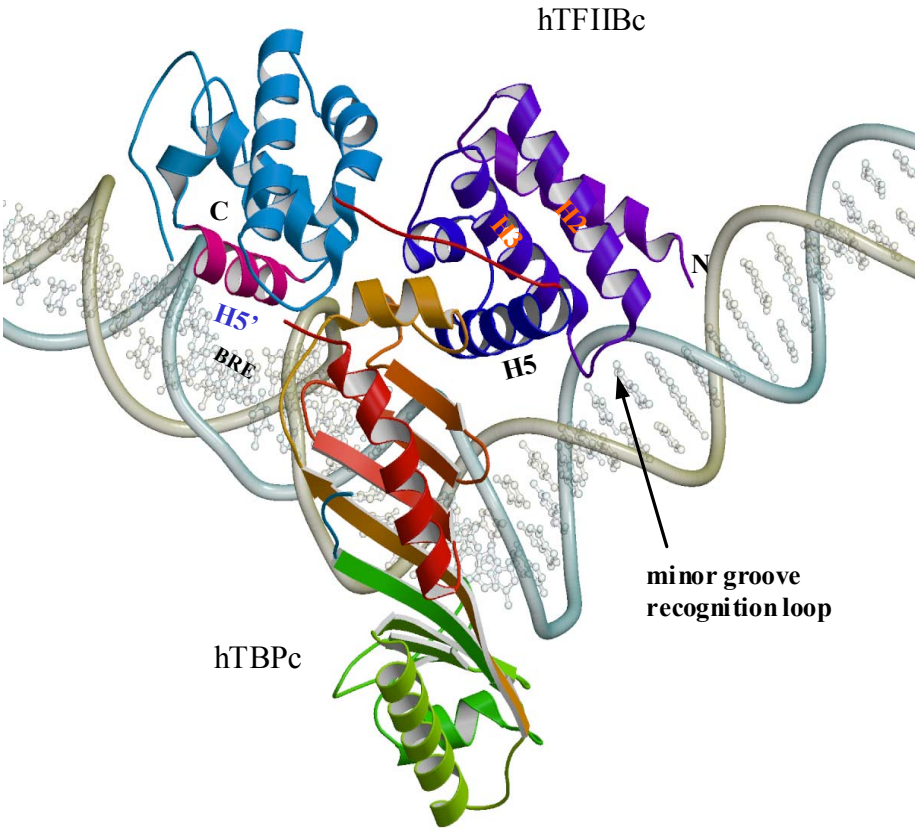
C. hTFIIBc



Buratowski and Zhou, 1993). TFIIB is also known to interact with transcription regulators including acidic activators (Manet et al., 1993; Roberts et al., 1993), hormone receptors (Masuyama et al., 1997) and HIV type 1 Tat transactivator (Veschambre et al., 1997). Roberts and Green (1994) demonstrated that the N- and C-terminal domains of TFIIB engage in an intramolecular interaction that is disrupted by some activators, such as VP16. Several experiments have suggested that TFIIB undergoes conformational changes that are part of PIC assembly (Roberts and Green, 1994) and (in some cases) activator-induced (Malik et al., 1993).

In 1995, Nikolov et al. published a crystal structure of a TFIIBc-TBP-TATA ternary complex (Figure 1.4). The imperfect repeats in the C-terminal core of TFIIB each fold into a cyclin A-like domain (Figure 1.3.C). The first repeat is composed of five α–helices while the second repeat has an additional sixth α-helix . Within the ternary complex, TFIIBc binds beneath and to one face of the TBP-TATA complex. The cleft region between TFIIBc's two α-helical domains clamps the TBP C-terminal stirrup. The crystal structure also revealed TFIIBc-DNA interactions both upstream and downstream of the TATA box. Later analysis of the TFIIB-DNA interactions revealed that the upstream interaction was a sequence specific helix-turn-helix interaction. Base substitution analysis of this upstream promoter element defined the transcription factor B recognition element (BRE, Lagrange et al., 1998). Unlike the TATA box, the BRE is not well defined. Lagrange et al., did report an 8 bp consensus sequence of 5'-G/C-G/C-G/A-C-G-C-C-3' which promotes the entry of TFIIB into the preinitiation complex. However, others (Evans et al., 2001) have reported that the BRE is actually a negative element that *does* promote TFIIB binding but *represses* basal transcription and that transcription

Figure 1.4. Crystal structure of human TFIIBc-TBPc-TATA-element complex. Ribbon representation of human TFIIBc-TBPc-TATA-element complex (Tsai and Sigler, 2000). The N- and C-terminal repeats of human TBPc are in green and orange, respectively. The first repeat of TFIIBc is in purple with H2 and H3 labeled to highlight the downstream minor groove recognition loop (discussed in section 3.2). The second repeat of TFIIC is in blue with the exception of H5' (in magenta) which recognizes the transcription factor B recognition element (BRE) in the major groove upstream of the TATA element (Lagrange et al., 1998).


hTFIIBc
C
H5'
BRE
H3
H2
N
H5
minor groove
recognition loop
hTBPc

activator (VP16) disruption of the TFIIB-BRE interaction is a means of transcriptional activation .

In 2000, Sigler's group published a second ternary crystal structure (Tsai and Sigler, 2000). This structure included both human TFIIBc and TBP (unlike the original structure that contained *A.thaliana* TBP) as well as a longer piece of promoter DNA centered around the TATA box. While the overall structure remained the same, a more extensive interaction between TFIIBc and DNA was seen than in the original structure. This interaction will be discussed further in a later section (Section 3.2).

The first structure of the N-terminal region of TFIIB came from an archaeal homolog (*Pyrococcus furiosus*, Zhu et al., 1996). The NMR solution structure revealed that the N-terminal region folds into a "zinc ribbon" domain similar to the Zn binding domain found in eukaryal transcription elongation factor TFIIS. The zinc is bound by four cysteines in a tetrahedral arrangement. Opposite the zinc coordination site is a short three-stranded anti-parallel β-sheet. Recently, an NMR solution structure has been published for the human TFIIB N-terminal domain (Figure 1.3.B, (Chen et al., 2000b). In hTFIIB the zinc ion is bound by three cysteines and one histidine. This zinc ribbon domain has been found to be a common structural element in transcription machinery. While various components of basal transcriptional machinery containing zinc ribbons have been characterized, the mechanism of the domain's function is still unknown.

Between the two structurally characterized domains of TFIIB there is a region of high homology that links the two. Although not structurally defined, its role in TFIIB's function is significant. Mutations in this region have indicated that it is vital for

transcription start site selection and TFIIB's intramolecular interaction (Hawkes et al., 2000). Mutational anlaysis of TFIIB will be discussed in a later section (Section 3.5).

1.2.2. TFIIF, TFIIE, and TFIIH

In the context of the studies reported here (on an archaeal system), the roles of the remaining eukaryal general transcription factors (TFIIF, TFIIE, and TFIIH) are of interest because of their absence in the archaeal transcription system. To date, only a homolog to the N-terminal 20 kDa of the α-subunit of eukaryal TFIIE has been identified in archaea (Bell et al., 2001a; Hanzelka et al., 2001). So the question remains, how are the functions attributed to these eukaryal GTFs carried out in the archaeal system? While the roles that the remaining GTFs play in eukaryal PIC formation and initiation of transcription have been well characterized, the exact mode of action for all three has not.

1.2.2.1. TFIIF

The two subunits of TFIIF were initially identified as RNAP II-associating proteins 30 (RAP30) and 74 (RAP74, Flores et al., 1991). Within the eukaryotic PIC, TFIIF is closely associated with RNAP II, and inhibits non-specific transcripition initiation. TFIIF's close association with RNAP II makes it a part of RNAP II recruitment to the PIC. TFIIF-RNAP binding to TFIIB-TBP-DNA has been reported as the rate limiting step of PIC assembly (Ferguson et al., 2001). In this RNAP II recruitment role, TFIIF also binds directly with TFIIB (Fang and Burton, 1996), and the TAF250 subunit of TFIID providing stability to the PIC (Ruppert and Tjian, 1995). Photocrosslinking experiments on the eukaryal PIC and sub-complexes have reported

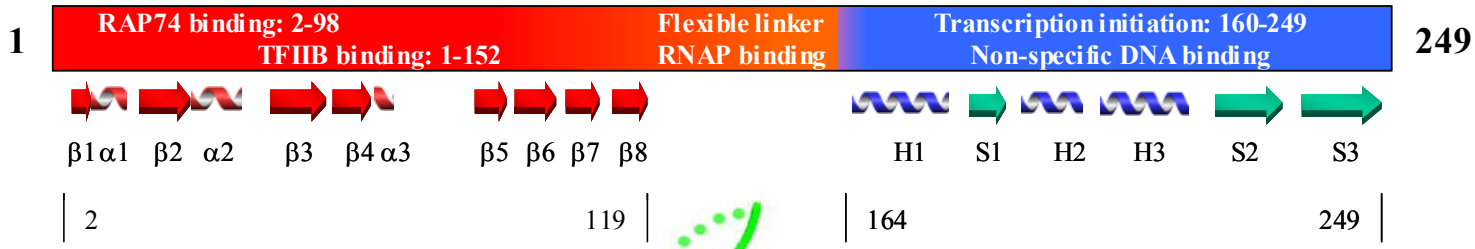
TFIIF-DNA interaction from the TATA box downstream to the transcripiton start site (-36 to +1) by both RAP30 and RAP74 (Kim et al., 1997). In contrast, other groups have reported photocrosslinking experiments in which RAP30 and RAP74 crosslink from -60 to +40 on promoter DNA (Douziech et al., 2000). Additionally, TFIIF is the link between RNAP and TFIIE-TFIIH to complete the eukaryal PIC. Once the PIC is formed, TFIIF and TFIIE together separate the promoter DNA strands near the transcription start site by promoting supercoiling of DNA around RNAP (Hampsey, 1998). TFIIF also remains complexed with RNAP II beyond initiation as the transcription complex transitions to the elongation complex where it has been reported to enhance polymerization rate and processivity (Gaiser et al., 2000). Figures 1.5.A and 1.6.A are linear schematics of both TFIIF subunits with functional segments as determined by mutational analysis indicated. Figures 1.5.B and C and 1.6.B and C include NMR solution and X-ray crystal structures of the globular domains of both subunits.

1.2.2.2. TFIIE

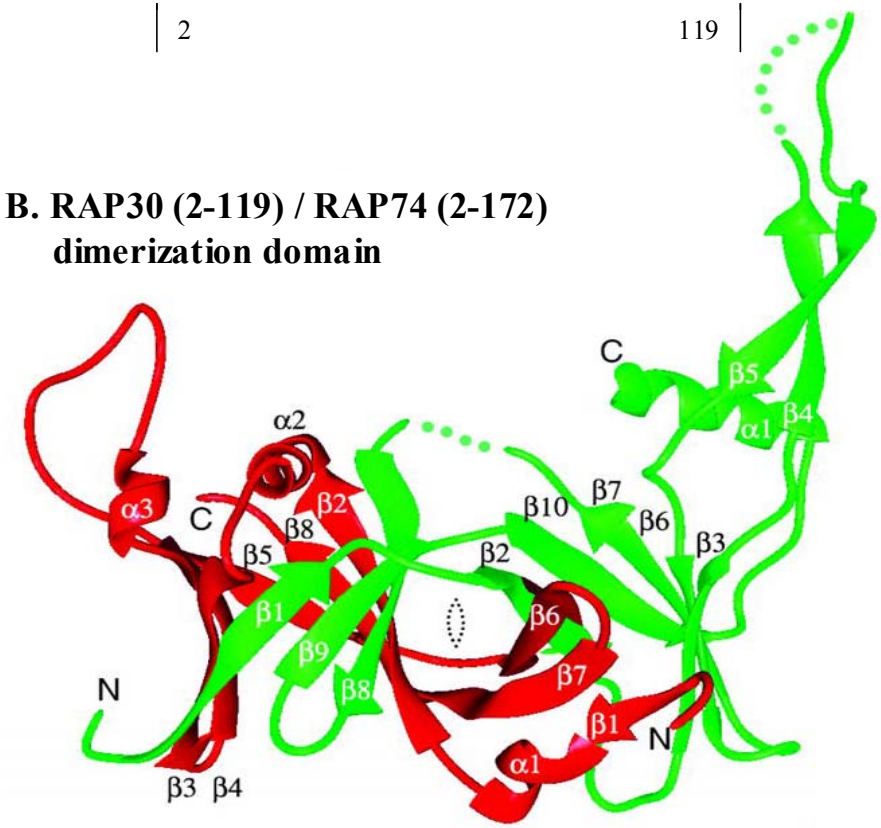
TFIIE was initially characterized as the basal transcription factor that must enter the PIC before TFIIH (Buratowski et al., 1989; Flores et al., 1989; Hampsey, 1998). In addition to recruiting TFIIH to the PIC, TFIIE may be involved (along with TFIIF) in the melting of promoter DNA at the transcription start site in an ATP-independent manner (Ren et al., 1999). TFIIE is known to interact with RNAP II, TFIIF, and DNA as part of its role in the final stages of PIC assembly and transcription initiation (Kuldell and Buratowski, 1997; Ohkuma et al., 1995). TFIIE also has been reported to have a role in the early stages of PIC formation through interaction with TBP (Yokomori et al., 1998).

Figure 1.5. TFIIF RAP30 functional and structural domains. A. Schematic representation of TFIIF RAP30 functional domains (Gaiser et al., 2000). **B.** Ribbon representation of TFIIF RAP30 / RAP74 dimerization domain. **C**. Ribbon representation of RAP30 C-terminal domain, a winged helix domain (Groft et al., 1998).

A. TFIIF RAP30



B. RAP30 (2-119) / RAP74 (2-172) dimerization domain



C. RAP30-CTD (164-249)

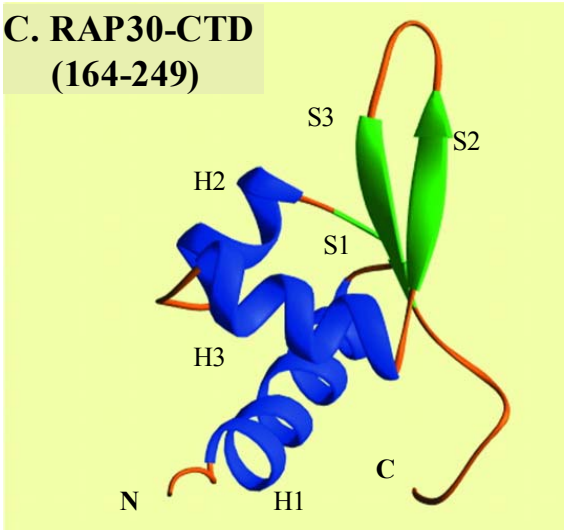
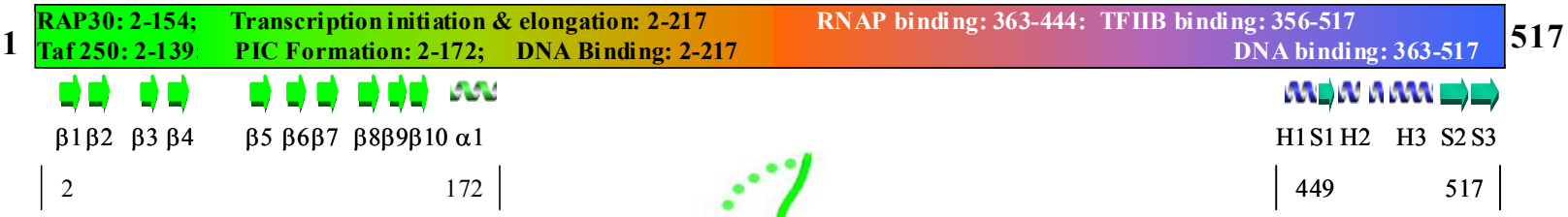
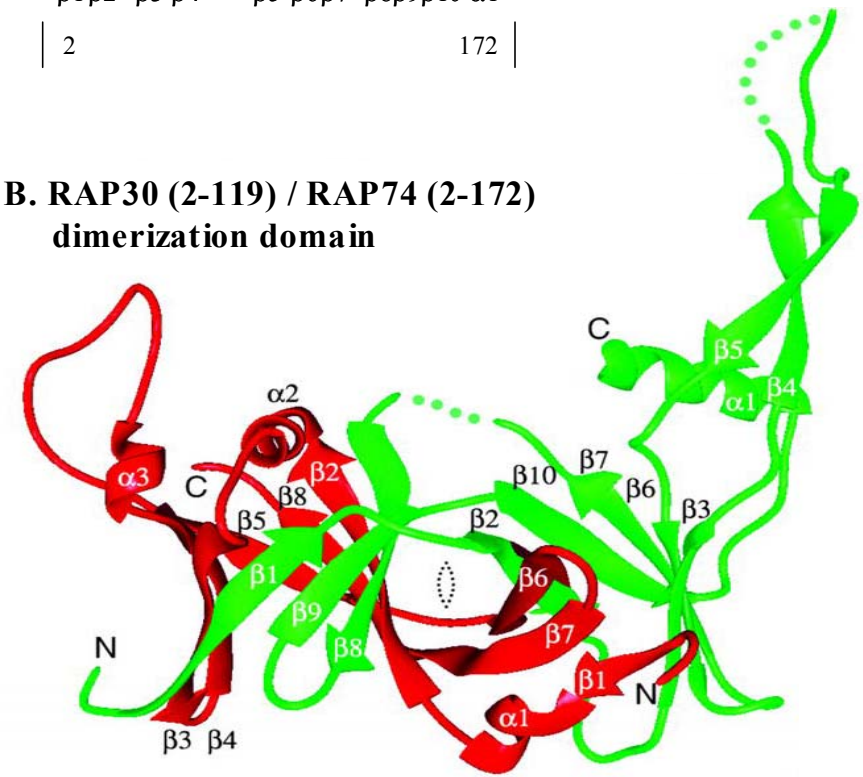


Figure 1.6. TFIIF RAP74 functional and structural domains. A. Schematic representation of TFIIF RAP74 functional domains. **B.** Ribbon representation of TFIIF RAP30 / RAP74 dimerization domain (Gaiser et al., 2000). **C**. Ribbon representation of RAP74 C-terminal domain, a winged helix domain (Kamada et al., 2001)the early stages of PIC formation through interaction with TBP (Yokomori et al., 1998).

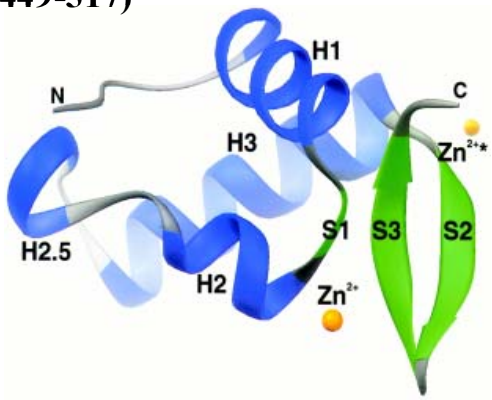
A. TFIIF RAP74



B. RAP30 (2-119) / RAP74 (2-172) dimerization domain



C. RAP74-CTD (449-517)



Similar to TFIIF, it is composed of two subunits called TFIIEα and TFIIEβ. Leuther et al. (1996) showed that TFIIE interacts with the active center of RNAP II.

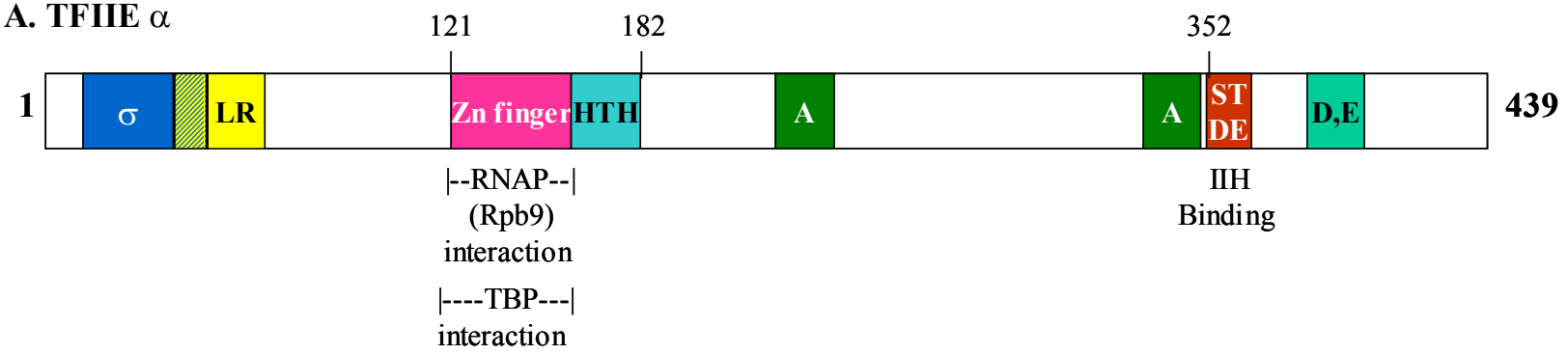
Photocrosslining results (Kim et al., 1997; Robert et al., 1996) showed that TFIIEβ interacts with promoter DNA from bp -14 to +4 of the adenovirus major late promoter (AdMLP). TFIIEβ has also been shown to bind RNAP II, TFIIB, and TFIIF subunit RAP30 (Okuda et al., 2000). Whether TFIIEα interacts with DNA is uncertain but it has been shown to bind TBP and TFIIH strongly (Okuda et al., 2000). This TFIIEα-TBP interaction has been reported to stimulate early stages of PIC formation thus implicating TFIIE at both initial and final stages of PIC formation (Yokomori et al., 1998). Figure 1.7.A and B show the delineated functions of TFIIEα and TFIIEβ as determined by deletion mutational analysis. Figure 1.7.C is the NMR solution structure of the TFIIEβ central core which is comprised of a winged helix motif similar to those seen in the TFIIF RAP74 and RAP30 C-terminal domains.

1.2.2.3. TFIIH

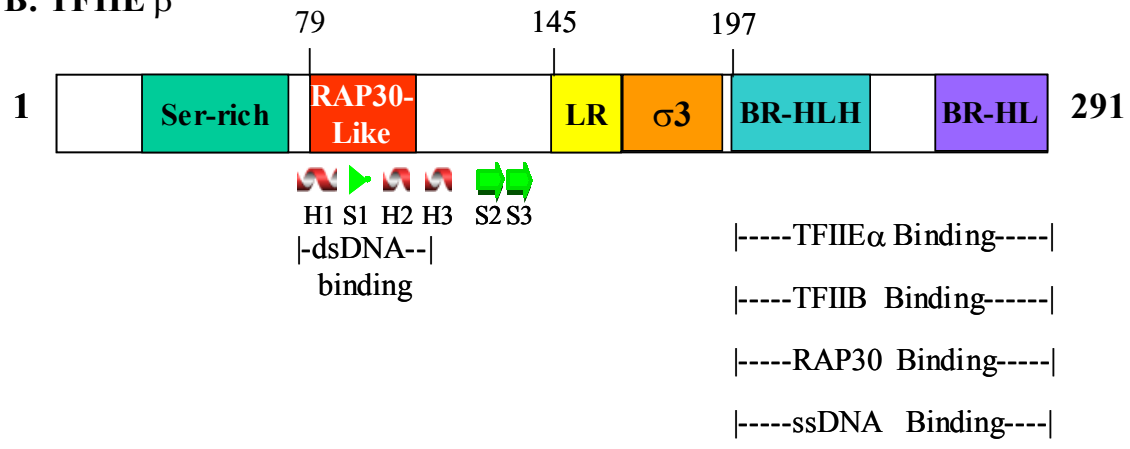
TFIIH is the only GTF with known enzymatic activities including DNA-dependent ATPase (Conaway and Conaway, 1989), ATP-dependent DNA helicase (Schaeffer et al., 1993), and RNAP II C-terminal domain kinase (Feaver et al., 1991). TFIIH is also involved with nucleotide excision repair (NER) and cell cycle progression (Drapkin et al., 1994). As would follow, TFIIH is the most complex GTF consisting of nine subunits of total mass ~ 500 kDa. TFIIH is commonly divided into two sub-complexes (Egly, 2001). TFIIH core is composed of the ATP-dependent XPB helicase and four structural proteins p62, p52, p44, and p34. XPB is the primary helicase of

Figure 1.7. TFIIE functional and structural domains. A. Schematic representation of TFIIEα. **B.** Schematic representation of TFIIEβ . **C**. Ribbon representation of TFIIEβ central domain, a winged helix domain (Okuda et al., 2000).

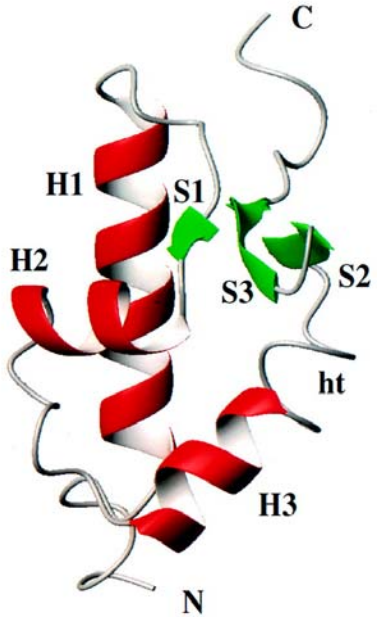
A. TFIIE α



B. TFIIE β



C. TFIIE β (66-146)

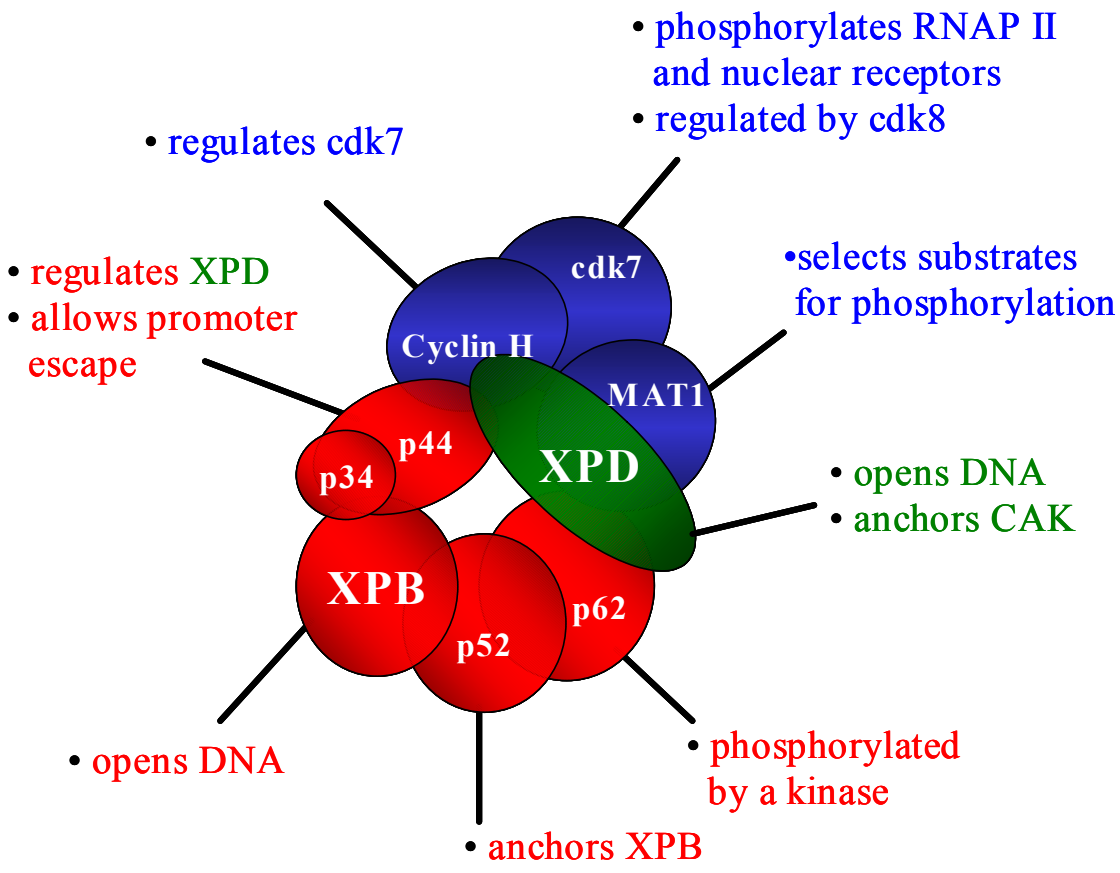


transcription initiation (Tirode et al., 1999). Once the preinitiation complex is formed, XPB opens the promoter around the start site. XPB is anchored to the TFIIH complex by structural protein p52 (Egly, 2001). The second helicase of TFIIH, XPD is connected to the TFIIH core complex by p44, which also is involved in promoter escape through a ring finger structure (Fribourg et al., 2000). XPD helicase activity is dispensable for transcription but it is a structural bridge to the other TFIIH subcomplex, cdk-activating kinase (CAK, Coin et al., 1998). CAK is composed of the kinase cdk7, cyclin H which regulates cdk7 and MAT1 which selects substrates for phosphorylation. CAK is responsible for phosphorylation of the C-terminal domain of RNAP II's largest subunit (as well as nuclear receptors, Chen et al., 2000a; Rochette-Egly et al., 1997), which acts as a switch in the transcription process from initiation to elongation. CAK also acts independently in cell cycle regulation. While the connection between all three of TFIIH's enzymatic activities is not clear, each is needed for basal transcription (Egly, 2001). Within the preinitiation complex, XPB has the primary role of melting the DNA duplex. Figure 1.8.A is a schematic of TFIIH and the known functions of its subunits.

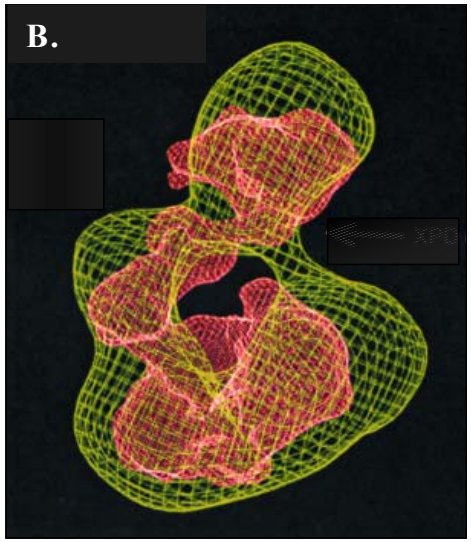
Current TFIIH structural information is currently only in parts. There are solution structures of subdomains of some of the subunits (MAT1 N-terminal RING finger domain (Gervais et al., 2001); cyclin H (Andersen et al., 1996); and p44 (Andersen et al., 1997; Fribourg et al., 2000). In 2000, Kornberg's group (Chang and Kornberg, 2000) reported a ~18 Å resolution 2D electron crystal structure of yeast TFIIH subcomplex composed of Ssl1, Rad3, TFB1, Tfb2, and TFB3 (corresponding to human p44, XPD, p62, p55, and MAT1 subunits, respectively). At the same time, Egly's group (Schultz et al., 2000) reported an ~38 Å resolution electron micrograph of single particles of human

Figure 1.8 TFIIH functional domains. A. Schematic representation of TFIIH (from (Egly, 2001). **B.** Overlay of TFIIH core ~18 Å resolution 2D electron crystal structure of Ssl1, Rad3, TFB1, Tfb2, and TFB3 (corresponding to human p44, XPD, p62, p55, and MAT1 subunits, respectively) with an ~38 Å resolution electron micrograph of single particles of human holo-TFIIH (Chang and Kornberg, 2000; Schultz et al., 2000).

A.


• phosphorylates RNAP II and nuclear receptors
• regulated by cdk8
• regulates cdk7
•selects substrates for phosphorylation
• regulates XPD
• allows promoter escape
cdk7
Cyclin H
MAT1
p44
p34
XPD
• opens DNA
• anchors CAK
XPB
p52
p62
• opens DNA
• phosphorylated by a kinase
• anchors XPB

B.



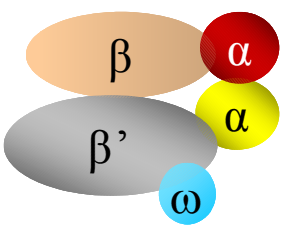
holo-TFIIH. Figure 1.8.B is an overlay of the two low resolution structures. The most notable feature of the structure is the general ring-like structure of the TFIIH subunits with a hole whose size could accommodate a double-stranded DNA molecule (Schultz et al., 2000).

1.3. RNA polymerases at the center of transcription

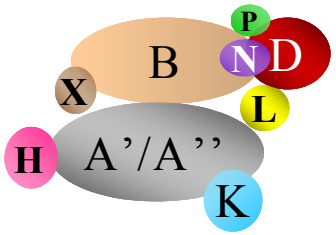
Regardless of the type of cell, transcription of genetic information is carried out by multisubunit RNA polymerases. Recently determined X-ray crystallographic structures of a bacterial RNAP (Zhang et al., 1999), bacterial RNAP bound to rifampicin (Campbell et al., 2001), RNAP II from yeast (Cramer et al., 2000), and a yeast RNAP II elongation complex (Gnatt et al., 2001) provide a detailed framework for understanding transcripiton in all cells. Comparison of these structures reveals what was suspected from sequence comparison of bacterial RNAP subunits and eukaryotic RNAP II subunits. There is a high degree of structural homology in all RNA polymerases. Reported RNAP-DNA interactions are best discussed with reference to these determined crystal structures. Figure 1.9 is from a concise review by Cramer (2002) on multisubunit RNA polymerases in the context of the recent crystal structures. Figure 1.9.A is a schematic representation and color code of RNAP subunits in bacterial, archaeal, and eukaryotic (yeast RNAP II) RNA polymerases. Figure 1.9.B and C are the three-dimensional structures of *T. aquaticus* RNAP and *S. cerivisiae* RNAP II. Table 1.3 lists the RNAP subunits from the three domains of life. Comparison of the structures reveals five central or core subunits make up the base RNAP architecture. The two largest subunits (B and B' for bRNAP and Rpb1 & 2 for yRNAP II make up the bulk of the enzyme and meet to make a central cleft

Figure 1.9 Architecture and structures of RNA polymerases (Cramer, 2002). **A.** Schematic of RNA polymerases from the three kingdoms of life, bacteria, archaea, and eukarya. **B**. and **C.** Three-dimensional structures of *T. aquaticus* RNAP (Zhang et al., 1999) and *S. cerivisiae* RNAP II (Cramer et al., 2000), respectively.

A.

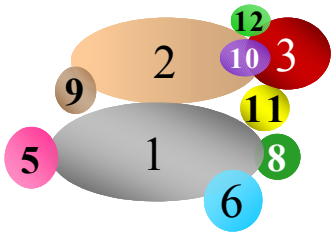


Bacteria



+ F & E
+ 1 other

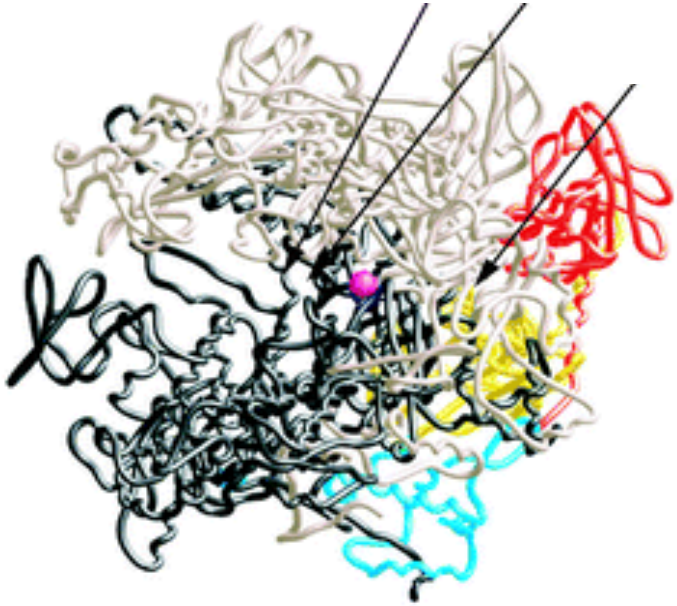
Archaea



+ 4 & 7

Eukaryotes

B



C.

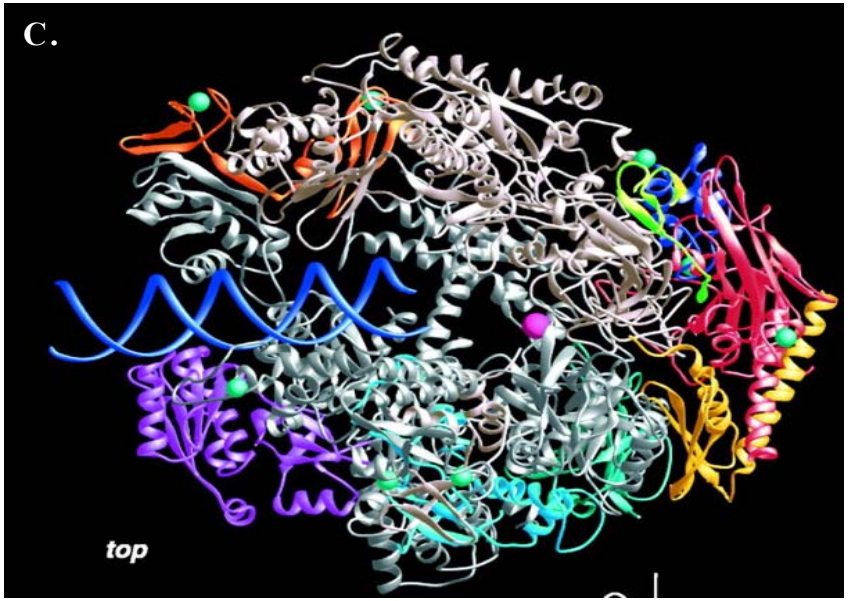


Table 1.3 RNA polymerase subunits[a]

Bacteria	Archaea	Eukaryotes RNAP II	
β'	A' + A"	Rpb1	core
β	B	Rpb2	core
α	D	Rpb3	core
α	L	Rpb11	core
ω	K	Rpb6	core
-	H	Rpb5	
-	-	Rpb8	
-	N	Rpb10	
-	P	Rpb12	
-	X (M)	Rpb9	
-	F	Rpb4[b]	
-	E	Rpb7[b]	
-	+one other	-	

[a](Cramer, 2002)

[b]Not included in yeast RNAP structure.

that is positively charged. Two smaller core subunits (α homodimer in bRNAP and Rbp3-Rbp11 heterodimer in yRNAP) anchor the two larger subunits at one end of the enzyme. This anchoring is further supported by the fifth core subunit (ω in bRNAP and Rbp6 in yRNAP). Protein-DNA crosslinking analysis of PIC's from all three domains of life (Bartlett et al., 2000; Kim et al., 1997; Naryshkin et al., 2000), have shown that RNAP-DNA crosslinks occur on both sides of the DNA face primarily to the two largest subunits in the preinitiation complex. This would place the DNA in the positively charged cleft between the two core subunits. The active center is marked by the presence of a Mg^{2+} ion bound by three conserved aspartate residues (Cramer, 2002). This active center is at the base or floor of the central cleft near the center of the enzyme. From the active site, there are channels which would allow entry of template DNA, dNTPs, and the exit of template DNA and newly synthesized RNA.

The benefit of crystal structures of multiple RNAP isomers revealed that the Rbp1 side of the central cleft is mobile. This mobility is induced by the binding of DNA in the central cleft. When bound, there is a 30° rotation of Rbp1 side that acts as a clamp to hold DNA within the cleft (yRNAP elongation complex), explaining the ability of RNAP to transcribe long pieces of DNA. In the absence of DNA, the clamp is open. As discussed in above sections, to bind and open promoter DNA at a given start site, RNAP's require GTF's. Based on the clamp position found in different RNAP complexes, it has been suggested that a partial role of GTF's is to bind to the surface of RNAP and "hold" the clamp in an open position as DNA is assembled and opened onto this large enzyme complex (Cramer, 2002). Secondly, the GTF's bind the surface of RNAP and DNA to align the transcription start site with the RNAP active center.

With such a large complex, determining the individual roles of RNAP subunits and GTF subunits and subdomains has been a large task. In the course of understanding transcription initiation, archaea have provided a unique opportunity to "simplify" the investigation. While RNAP subunit composition of archaeal RNAP (aRNAP) is highly homologus to eukaryotic RNAP II (Figure 1.9.A, Table 1.3), there are fewer GTF's to deal with (Figure 1.1).

1.4. Archaeal preinitiation complex

The archaeal preinitiation complex resembles the eukaryotic PIC but is less complicated (Figure 1.1). Homologs of 11 of 12 yeast RNAP II subunits have been identified in Archaea (Bell et al., 2001b). The first notable difference in the two PICs is the identification of fewer GTFs in the archaeal PIC. Homologs to eukaryal TBP and TFIIB have been found and characterized in Archaea (TBP and TFB, respectively, Hausner and Thomm, 1993; Ouzounis and Sander, 1992). A homolog to eukaryal TFIIEα has also been identified (Hanzelka et al., 2001).

1.4.1. Archaeal TBP and TFB

An archaeal homolog to TFIIB was first identified by Ouzounis and Sander (Ouzounis and Sander, 1992). Before this identification it had been determined that cell-free transcription was mediated by transcription factors in Archaea (Frey et al., 1990; Hudepohl et al., 1990). Two transcription factors from *Methanococcus thermolithotrophicus*, identified as aTFA and aTFB, were found to be required for transcription initiation (Hausner and Thomm, 1993). These two factors were identified as

homologs to eukaryal TFIIB and TBP, respectively. Homologs to these two general transcription factors were found to be in several different Archaea, including *Pyrococcus*, *Thermococcus*, and *Sulfolobus* (Hethke et al., 1996). Archaea were originally thought to be strictly "extreme-ophiles" existing in environments of extreme temperatures, pH, or salt concentrations. However they are also abundant in more moderate habitats (Soppa, 1999). Still those species which do exist under extreme conditions provide opportunity for studying cellular processes under "abnormal" conditions. The archaeal transcription homologs used in our experiments come from *Pyrococcus furiosus (Pf or P. furiosus)* which thrives at temperatures between 70°C and 100°C

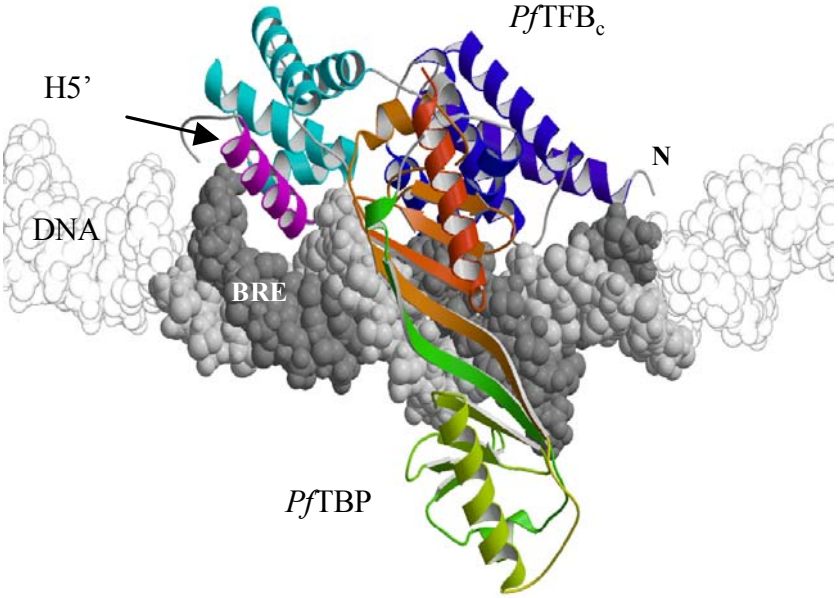
Archaeal transcription from *Pyrococcus furiosus* has been studied by both the Thomm (Institute for Allgemeine Mikrobiologie, Universitat Kiel, Kiel, Germany) and Scott (University of Georgia, Athens, Georgia) groups in the past few years. Thomm's group developed a cell-free transcription system including *Pf* RNAP as well as TBP and TFIIB homologues *Pf* TBP and *Pf* TFB, respectively, isolated from *Pf* cell extracts (Hethke et al., 1996). While *P. furiosus* grows optimally at 100°C, the *in vitro* assay system produced an RNA transcript between 60°C and 85°C. A second *in vitro* transcription system developed for the Archaea *Sulfolobus shibatae* farther confirmed that archaeal transcription required the presence of aRNAP, TBP and TFB for accurate transcription (Qureshi et al., 1997). Although less complex in its general transcription factor requirements, archaeal transcription is essentially homologous to eukaryal transcription. In 1997, a crystal structure of an archaeal ternary complex (TFB-TBP-DNA from *Pyrococcus wosei*) showed that the archaeal transcription factors were structural homologs to their eukaryotic counterparts (Kosa et al., 1997). Archaeal TBP

and TFB interact the same with *Pw* TFB clamping the C-terminal stirrup of *Pw* TBP. However, the orientation of the two proteins on the DNA promoter in the 1997 structure was reversed compared to the eukaryotic ternary complex. Although not included in the structure, the crystallized orientation suggested the N-terminal domain of TFB would be positioned *upstream* of the box A promoter (archaeal TATA-box equivalent), away from the start site. This was opposite the orientation of its eukaryotic homologue and seemed to contradict and complicate defining the N-terminal domain's role in transcription start site selection and RNAP recruitment. In 1999, Sigler's group reported a second crystal structure of the same ternary complex (Figure 1.10, Littlefield et al., 1999). This time the orientation of the two transcription factors was identical to that found in Eukarya. The difference was the promoter sequence used. The second structure included the transcription factor B recognition element (BRE) in the promoter sequence. This second structure allowed the BRE - TFBc interaction to define the orientation of the complex and therefore the direction of transcription.

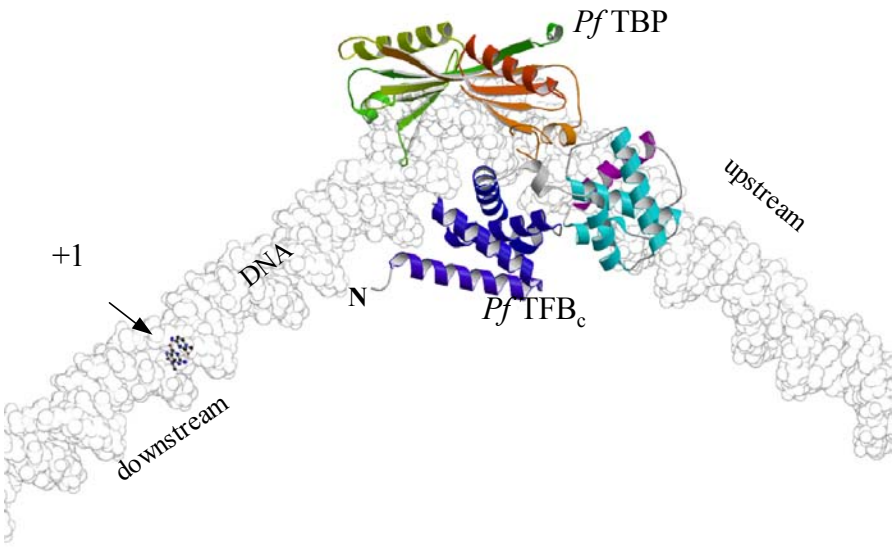
In 1996 the Scott group in collaboration with Mike Summers reported the NMR solution structure of the *Pf* TFB N-terminal domain (Figure 1.10.C, Zhu et al., 1996). As mentioned above the N-terminal domain forms a zinc ribbon similar to that found in eukaryal TFIIS. Wang et al., (1998) reported a third zinc ribbon structure found in aRNAP subunit 9 (RPB9) of the archaeon *Thermococcus celer*. These three structures found in different components of transcriptional machinery of both Archaea and Eukarya reveal the prominent role of this zinc ribbon motif in the architecture of transcriptional machinery.

Figure 1.10. *Pf* TFBc-TBP-DNA crystal structure and *Pf* TFB N-terminal zinc ribbon domain. A. Ribbon representation of *Pw*TFBc-TBP-DNA crystal structure as determined by Littlefield, et al. (Littlefield et al., 1999). The color scheme for TBP and TFBc are the same as for eukaryal TBPc and TFIIBc in figure 1.4. **B**. An extended view of the ternary complex with the transcription start site highlighted (+1, downstream). The original archaeal ternary complex was oriented such that the N-terminus of TFBc was oriented upstream. **C.** Ribbon representation of *Pf*TFB N-terminal zinc ribbon (Zhu et al., 1996).

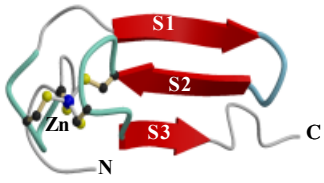
A.


PfTFB$_c$
H5'
N
DNA
BRE
PfTBP

B.


Pf TBP
upstream
+1
DNA
N
Pf TFB$_c$
downstream

C.


S1
S2
Zn
S3
C
N

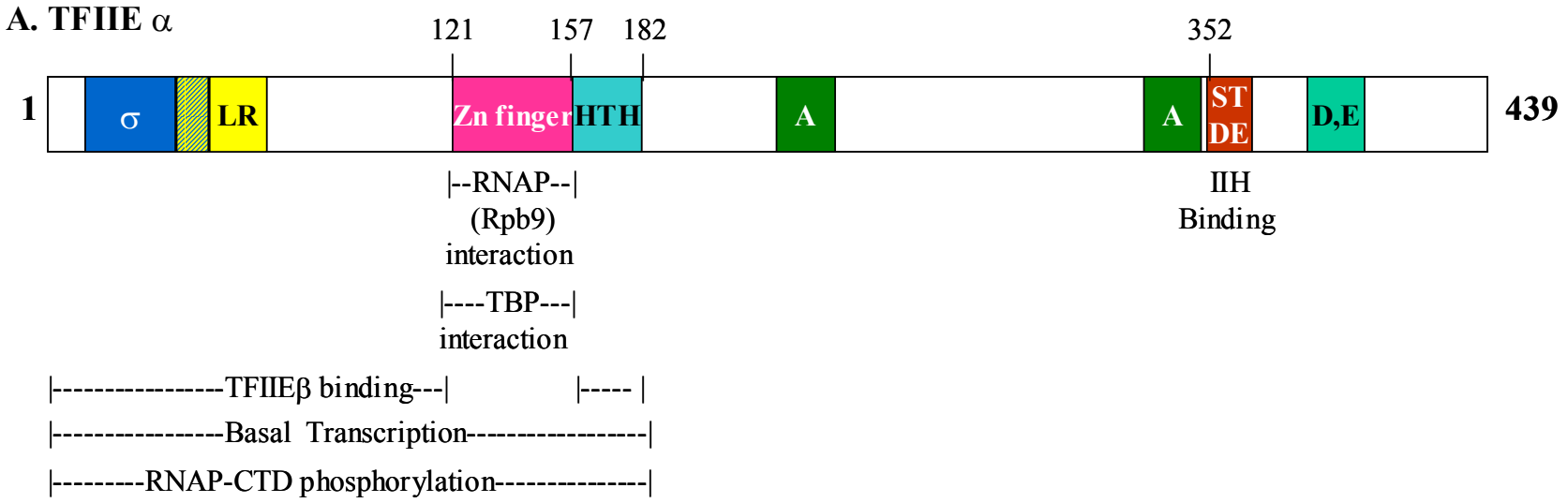
1.4.2. Archaeal TFEα

An archaeal homolog to the N-terminal 20 kDa of TFIIEα has recently been identified by two groups (Bell et al., 2001a; Hanzelka et al., 2001). This portion of TFIIEα has a leucine-rich helix-turn-helix motif and a putative zinc ribbon (Figure 1.11). The zinc ribbon has been proposed to bind RNAP II based on deletion mutagenesis (Ohkuma et al., 1995). Notably missing from the archaeal homolog is the C-terminal portion of TFIIEα which has been linked with TFIIH binding and therefore recruitment to the eukaryal PIC (Bell et al., 2001a; Ohkuma et al., 1995). Characterization of the archaeal homolog (TFEα) suggests that it may function in stimulating TBP binding to the TATA box promoter (Bell et al., 2001a). A second group reported similar findings but only at select gene promoters, thus suggesting that this TFIIEα homolog may not be a true "general" transcription factor (Hanzelka et al., 2001).

1.5. Site-specific DNA photocrosslinking

The Ebright group reported (Pendergrast et al., 1992) the use of a photoactivatible crosslinking agent (phenyl-azide, AP) to determine the orientation of a DNA binding motif in a protein-DNA complex. In this report, the photoactivatible crosslinking agent was covalently attached to single cysteine mutations of catabolite gene activator protein (CAP). In 1996, the Ebright and Reinberg groups adapted the technique to study protein-DNA interactions within a human TFIIA-TFIIB-TBP-DNA complex and sub-complexes (Lagrange et al., 1996). However, these studies covalently attached the photoactivatible crosslinking agent to a phosphorothioate incorporated at a single site in the DNA phosphate backbone. Initially, 36 site-specifically derivatized promoter DNA fragments

Figure 1.11. Eucharyal TFIIEα and an archaeal truncated homolog. A. Shematic representation of TFIIEα as in Figure 1.7. **B.** Schematic representation of archaeal TFEα. While homologus to eukaryal TFIIEα, archaeal TFEα has not been demonstrated to be necessary for basal transcription in archaea (Bell et al., 2001a; Hanzelka et al., 2001).



B. Archaeal TFE α

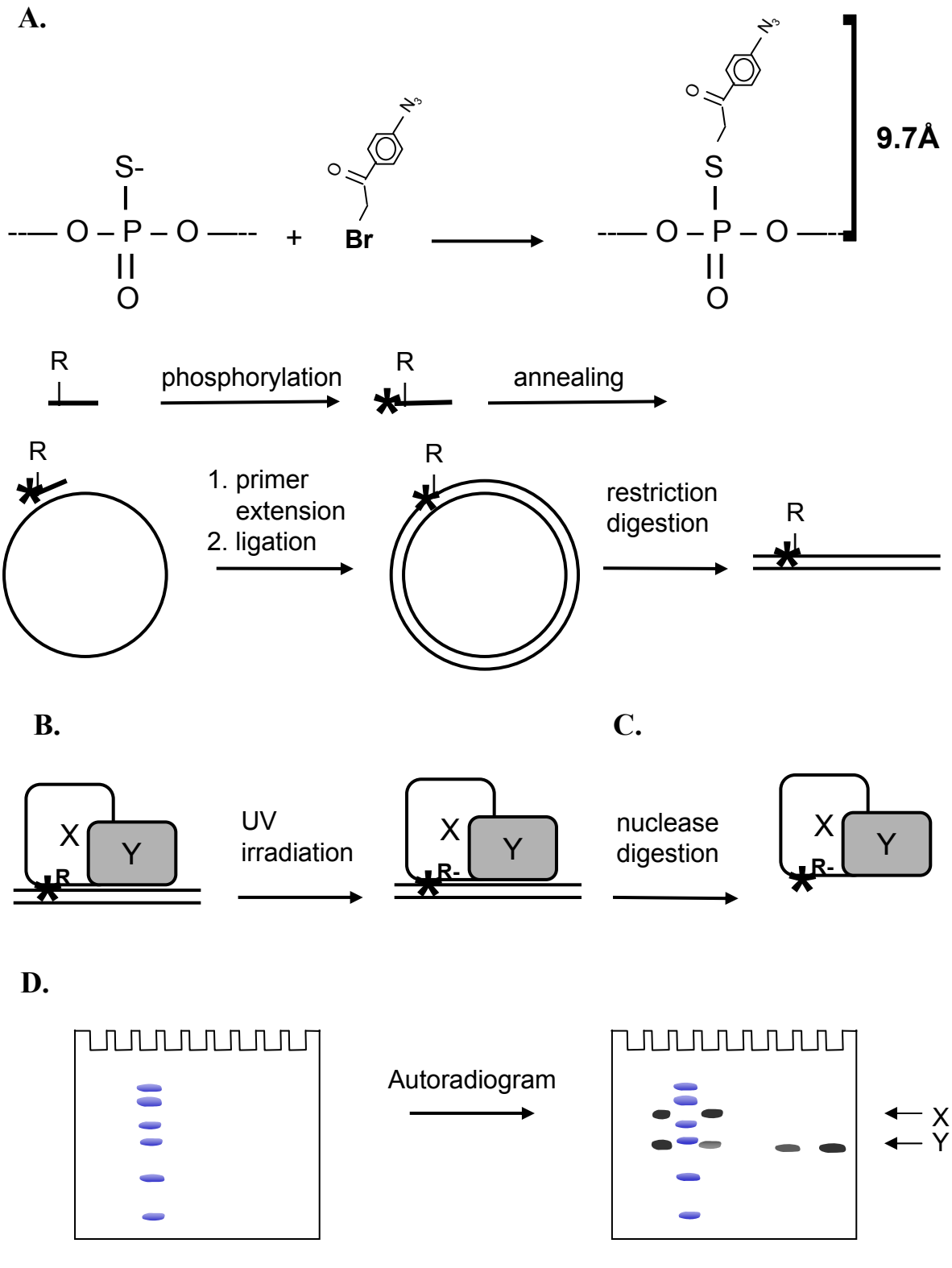
were used to analyze crosslinking in TBP-DNA, TBP-TFIIA-DNA, TBP-TFIIB-DNA, and TBP-TFIIA-TFIIB-DNA complexes (Lagrange et al., 1996). The survey of protein-DNA photocrosslinking in the complexes defined the positions of the proteins relative to the promoter DNA. While these initial results could in large part be compared to existing crystal structures, a later expansion of this crosslinking analysis was performed on an RNAP II-TFIIF-TFIIB-TBP-DNA complex and then RNAP II-TFIIF-TFIIE-TFIIH-TFIIB-TBP-DNA complex (Kim et al., 2000; Kim et al., 1997). Analysis of these latter large multi-protein complexes provided detailed topographical maps of protein-DNA interactions as well as protein positioning in the eukaryal PIC at a high molecular weight level in vitro. Ebright's lab has also extended this type of protein-DNA analysis to the bacterial RNAP system (Naryshkin et al., 2000).

The procedure for site-specific protein-DNA photocrosslinking consists of four parts (Figure 1.12, Naryshkin et al., 2001).

A) Chemical and enzymatic reactions are used to produce a DNA fragment containing a phenyl-azide photoactivatible crosslinking agent and an adjacent radiolabel incorporated at a single defined DNA phosphate (Figure 1.12.A), with a 9.7 Å linker between the photoreactive atom of the crosslinking agent and the phosphorus atom of the phosphate, and with an ~11 Å maximum "reach" between potential crosslinking targets and the phophorus atom of the phosphate.

B) The multiprotein-DNA complex of interest is assembled using the site-specifically derivatized DNA fragment, and the multiprotein-DNA complex is UV-irradiated initiating covalent crosslinking with proteins in direct physical proximity

Figure 1.12. Schematic of site-specific protein-DNA photocrosslinking. A. A single phosphorothioate is incorporated into a synthesized oligodeoxyribonucleotide at a specific site in the DNA backbone. Once synthesized, an Azido-phenacyl moiety (R) is covalently attached to the derivatized oligodeoxyribonucleotide. Enzymatic steps incorporate the derivatized oligodeoxyribonucleotide into a dsDNA fragment which is used to probe protein-DNA interactions. ^{32}P Phosphorylation (*) of the derivatized oligodeoxyribonucleotide is used for detection of photocrosslinked proteins. **B. and C.** The protein-DNA (proteins X and Y) complex of interest is assembled using the derivatized DNA fragment. Once formed, the complex is irradiated with UV-light, C. the DNA digested with various DNA nucleases. **D.** The crosslinked (or tagged) proteins (X and/or Y) are identified based on molecular weight in gel elution.


A.
S-
O – P – O
O
Br
9.7Å
R
phosphorylation
annealing
1. primer extension
2. ligation
restriction digestion
B.
X
Y
UV irradiation
C.
nuclease digestion
D.
Autoradiogram
X
Y

(within ~ll Å) to the crosslinking agent. Crosslinked proteins can be identified either by monoclonal antibodies or by mass spectrometry analysis (Figure 1.12.C.).

C) Extensive nuclease digestion is performed, eliminating uncrosslinked DNA and converting the crosslinked DNA to a crosslinked, radiolabelled 3-5 nucleotide "tag".

D) "Tagged" polypeptides are identified by denaturing polyacrylamide gel electrophoresis followed by autoradiography.

1.6. Analyzing protein-DNA interactions in an archaeal transcription PIC

The studies reported here were done in collaboration with the Ebright group at Rutgers University. A preinitiation complex from the hyperthermophilic archeaon *P. furiosus* was assembled on site-specifically derivatized DNA promoter fragments of *Pf* glutamate dehydrogenase gene promoter (positions -74 to +45 with respect to the transcription start site at +1). Recombinant *P. furiosus* transcription factors TBP (*Pf* TBP) and TFB (*Pf*TFB) were combined with partially purified *Pf* RNAP from *P. furiosus* cell extracts to form the complex. *In vitro* transcription assays for *P. furiosus* produce an RNA transcript between 60 °C and 95 °C (Lewis, 2000). While *P. furiosus* grows optimally at 100 °C, cell-free transcription systems have yet to be able to mimic *in vivo* conditions *in vitro*. Hethke et al. (1999) reported thermostability studies of *P. furiosus* components. *Pf* TBP showed a half-life of several hours at 100 °C. After 1 hour at 100 °C, *Pf* TBP did not lose any ability to support transcription activity. RNA polymerase lost 50% of its activity after 23 minutes at 100 °C. *Pf* TFB was >95% inactive after only 10 minutes at 100 °C. So while *Pf* TBP and aRNAP have some degree of intrinsic thermostability comparable to other proteins isolated from hyperthermophiles,

Pf TFB *in vitro* is not nearly as stable. Photocrosslinking analysis of the *Pf* PIC reported here was performed at 70°C. This falls well within the range of in vitro transcription results and is the lower end of known growth temperatures for *P. furiosus* (Segerer et al., 1993). Hethke et al., (1996) set up an *in vitro* transcription system using *P. furiosus* components on a plasmid template containing the *P. furiosus* glutamate dehydrogenase (*Pf* gdh) gene. This template was also used for DNase I footprinting analysis of *P. furiosus* transcription factors assembling at the box A promoter (Hausner et al., 1996). Just as the *Adenovirus* major late promoter has become the standard for eukaryal PIC studies, the studies reported here use the *Pf* gdh gene to maintain the context of archaeal PIC studies.

CHAPTER 2

MATERIALS AND METHODS

2.1. Cloning *Pf* gdh promoter into M13 phage

Plasmid pLUW479 was obtained from the Thomm group (Hethke et al., 1996). Plasmid pLUW479 contains a 0.6 kbp fragment of *P. furiosus* genomic DNA including the sequence of the glutamate dehydrogenase gene, its terminator sequence, and 200 bp prior to the transcription start site (including Box A promoter). PCR primers were designed to amplify positions -75 → +45 (+1 is the transcription start site) of the *Pf* gdh promoter (*Pf*gdhP, Table 2.1). EcoR I and Sph I restriction sites were included in the primers at the -75 and +45 ends (Table 2.1), respectively, for ligation into filamentous phage vectors, M13mp18 and M13mp19 (New England Biolabs, NEB) replicative form (RF) DNA. M13mp18 and M13mp19 differ only in the orientation of their respective cloning sites (Figure 2.1). This allowed for isolation of ssDNA of either the template (T) or nontemplate (NT) strand of the *Pf* gdh promoter as the template for photocrosslinking probe construction (Sambrook et al., 1989), discussed further below).

2.1.1. Cloning of *Pf* gdh promoter into M13mp19

Upon examination of the pLUW479 sequence, an EcoR I restriction site was discovered 381 bp beyond the inserted fragment of *P. furiosus* genomic DNA (Figure 2.2). Given that pLUW479 was to be used as a template for PCR, it was possible

Table 2.1

***Pyrococcus furiosus* glutamate dehyrogenase cloned promoter fragment (-75 → +45)**

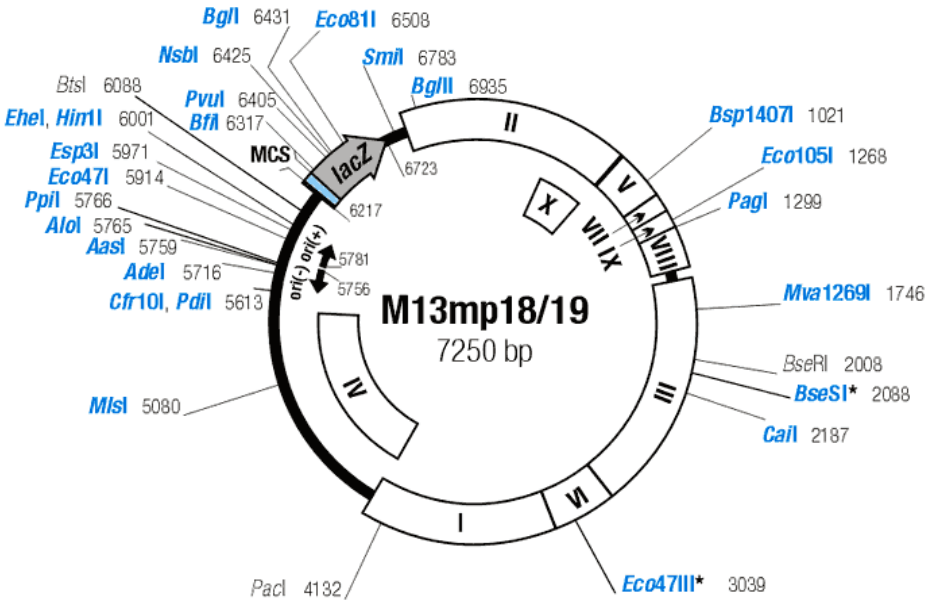
```
CTAATCAAATAAACAAAAGGATTTCCACTCTTGTTTACCGAAAGCTTTATATAGGCTATT
      -70           -60           -50           -40           -30           -20
GATTAGTTTATTTGTTTTCCTAAAGGTGAGAACAAATGGCTTTCGAAATATATCCGATAA

GCCCAAAAATGTATCGCCAATCACCTAATTTGGAGGGATGAACATGGTTGAGCAAGACCC
      -10           +1            +10           +20           +30           +40
CGGGTTTTTACATAGCGGTTAGTGGATTAAACCTCCCTACTTGTACCAACTCGTTCTGGG
```

Primer	Sequence	Purpose
PfgdhEcorI-M13	5'-TAGATTCTGAATTCCTAATCAAT-3'	Forward PCR primer (section 2.1)
gdhSphI-reverse	5'-CTCGTTCTGGGCGTACGTTAACAATAA-3'	Reverse PCR primer (section 2.1)
M13 Forward (-41)[a]	5'-GGTTTTCCCAGTCACGAC-3'	PCR plaque screening (section 2.1.2)
M13 Reverse (-48)[a]	5'-AGCGGATAACAATTTCACAC-3'	PCR plaque screening (section 2.1.2)
M13 Forward (-20)[a]	5'-GTAAAACGACGGCCAGTG-3'	PCR plaque screening and sequencing (section 2.1.2)
M13 Reverse (-27)[a]	5'-GGAAACAGCTATGACCATG-3'	PCR plaque screening and sequencing (section 2.1.2)
Probe synthesis upstream primer	5'-AGCGCCATTCGCCATTCAGGC-3'	Probe synthesis (section 2.3)
Beyond Ava II (Bava)	5'-CAACCCTATCTCGGGCTATTC-3'	PCR primer for testing cloned promoter fragment

[a]Integrated DNA Technologies (IDT) Ready-Made Primer

Figure 2.1. Filamentous phage vector M13mp18 & 19. A. Circular map of M13mp18/19 RF DNA. **B.** Multiple cloning site of M13mp18. **C.** Multiple cloning site of M13mp19.

A.



B. M13mp18

M13/pUC reverse sequencing primer (-26), 17-mer

CAG GAA ACA GCT **ATG** ACC ATG ATT

Met Thr Met Ile

6231 EcoRI | Ecl136II SacI | Acc65I KpnI | Cfr9I SmaI | BamHI | XbaI | HincII SalI XmiI | PstI SdaI | PaeI | HindIII 6287

ACG AAT TCG AGC TCG GTA CCC GGG GAT CCT CTA GAG TCG ACC TGC AGG CAT GCA AGC TTG

Thr Asn Ser Ser Ser Val Pro Gly Asp Pro Leu Glu Ser Thr Cys Arg His Ala Ser Leu

GCA CTG GCC GTC GTT TTA CAA

M13/pUC sequencing primer (-20), 17-mer

Ala Leu Ala Val Val Leu Gln → LacZ

C. M13mp19

M13/pUC reverse sequencing primer (-26), 17-mer

CAG GAA ACA GCT **ATG** ACC ATG ATT ACG

Met Thr Met Ile Thr

6234 HindIII | PaeI | PstI SdaI | HincII SalI XmiI | XbaI | BamHI | Cfr9I SmaI | Acc65I KpnI | Ecl136II SacI | EcoRI 6290

CCA AGC TTG CAT GCC TGC AGG TCG ACT CTA GAG GAT CCC CGG GTA CCG AGC TCG AAT TCA

Pro Ser Leu His Ala Cys Arg Ser Thr Leu Glu Asp Pro Arg Val Pro Ser Ser Asn Ser

CTG GCC GTC GTT TTA CAA

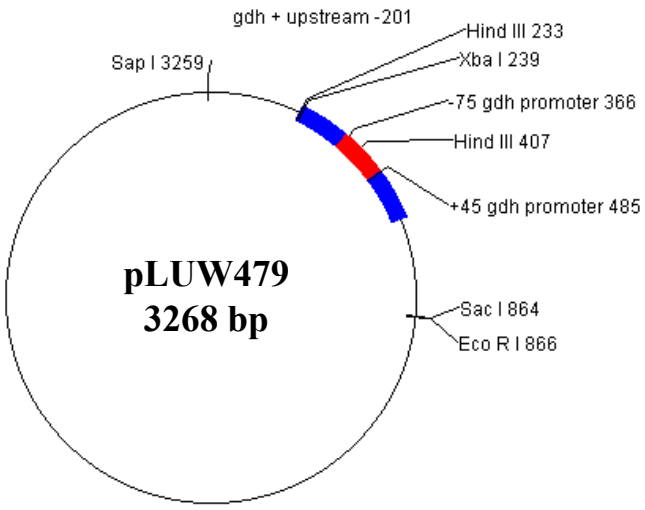
M13/pUC sequencing primer (-20), 17-mer

Leu Ala Val Val Leu Gln → LacZ

Figure 2.2. Cloning of *Pf* glutamate dehydrogenase promoter fragment.

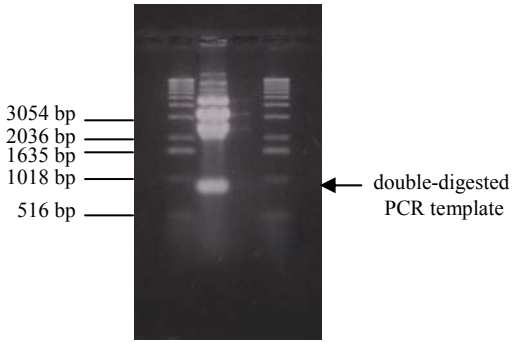
A. Circular map of pLUW479 obtained from Thomm group. The presence of an EcoR I site required that a fragment of the pLUW479 be restriction enzyme "cut out" of the plasmid for use in PCR amplifying the -75 to +45 segment for cloning into M13mp19 RF DNA. **B**. ~800 bp fragment restriction cut (Sap I and Sac I) from pLUW479. **C.** Screening of M13mp19 plaques by PCR using isolated ssDNA template from ssDNA preparations. Lanes on the ethidium bromide stained gel are designated by template and primers used in standard PCR amplifications. Lanes 1, 2, 3, and 4 show PCR products using M13mp18 ssDNA and M13mp19 RF DNA stock solutions as template for PCR. Lanes 5, 6, 7, and 8 show PCR products using isolated ssDNA from plaques 24 and 26. The larger PCR product from lanes 5-6 (compared to lanes 1-4) indicated insertion of the *Pf* gdh promoter fragment. Correct insertion was confirmed by DNA sequencing of large scale DNA preparations from plaques 24 and 26.

A.



B.

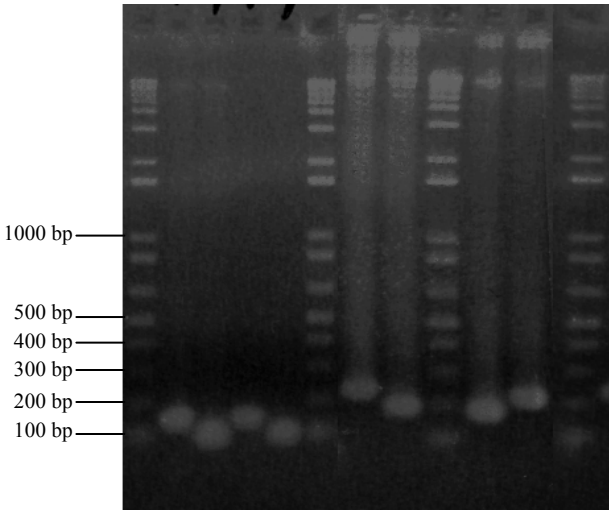
1. Sap I, Sac I digested pLUW 479



1

C.

1. M13mp18 ssDNA + -41 & -48 primers
2. M13mp18 ssDNA + -20 & -27 primers
3. M13mp19 dsDNA + -41 & -48 primers
4. M13mp19 dsDNA + -20 & -27 primers
5. Plaque 24 + -41 & -48 primers
6. Plaque 24 + -20 & -27 primers
7. Plaque 26 + -20 & -27 primers
8. Plaque 26 + -41 & -48 primers



that the EcoR I site could cause erroneous ligation in the subsequent restriction digest and ligation reaction steps. To prevent erroneous ligation, an ~800 bp fragment of pLUW479 was restriction cut with Sap I and Sac I (10 u / μg DNA, 3 h, 37°C, Promega) and then isolated by gel extraction (Qiagen) for use as a template in PCR amplification (Figure 2.2). Using this fragment as a template for PCR, the -75 → +45 region of the *P. furiosus* gdh promoter/gene was amplified. The PCR product was purified using the QIAquick PCR purification kit (Qiagen). The resulting PCR product was then restriction enzyme cut using EcoR I (10 u / μg DNA, 3 h, 37°C, heat inactivated 65°C, 5 min, NEB). The digestion reaction was desalted by ethanol precipitation and then restriction enzyme cut with Sph I (10 u / μg DNA, 3 h, 37°C, heat inactivated 65°C, 5 min, NEB). M13mp19 RF DNA (NEB) was digested in like fashion, and then both double-digested fragments were isolated by gel extraction (QIAquick gel extraction kit, Qiagen). The isolated fragments were quantified using the Hoefer DNAQuant Fluorimeter (gdh PCR: 10 ng/μL, 99 ng/pmol, 30 μL total; M13mp19 RF: 14 ng/μL, 4700 ng/pmol, 30 μL total). Ligation reactions (Ligation reaction:10 μL gdh PCR + 3 μL M13mp19 RF + 2.3 μL Ligase (3 u / μL, Roche) + 1.7 μL 10× reaction buffer) were performed using a 3:1 insert to vector molar ratio and incubated for 24 h at 15°C.

Three microliters of the ligation reaction mix was transformed into 40 μL of JM109 electrocompetent cells . For manipulation of M13 filamentous phage, the ligated M13 RF DNA had to be transformed into a strain of *E.coli* containing the F' episome. Protocols for electroporation were provided by the supplier. After electroporation, the cells were allowed to recover for 1 h at 37°C with shaking, centrifuged at 4000 × g for 5 min, decanted, and then resuspended (gently) in 200 μL of LB broth (per liter, 10g NaCl,

10g, bactotryptone, 5 g yeast extract, deionized water to final volume of 1 liter, autoclaved). Fifty μL of resuspended transformed cells was combined with 1000 μL of LB followed by nine serial dilutions (100 μL in 1000 μL total of LB). Each dilution was combined with 3 mL of SOC top agar (Sambrook et al., 1989) and 200 μL of non-infected JM109 cells, vortexed (gently) briefly, and then plated on 37 °C equilibrated LB agar plates. The top agar was allowed to solidify at room temperature and then the plates were incubated at 37°C overnight.

2.1.2. Screening of M13mp19 plaques for positive insertion of *Pf* gdh promoter

Unless otherwise stated, all ssDNA and RF DNA preparations of M13 phage followed protocols from Sambrook et al (1989). Forty M13 plaques from the ligation transformation were plugged using flame sterilized pasteur pipettes and placed in 1.5 mL microcentrifuge tubes containing 1 mL of LB and incubated at room temperature for 2 h. The tubes were then spun at 14,000 × g for 10 min and the supernatant transferred to new microcentrifuge tubes and placed at -20°C. This plaque suspension is stable at -20°C for up to 6 months.

To screen the isolated plaques, 100 μL of each plaque suspension was combined with 50 μL of non-infected JM109 DE3 cells plus 2 mL of LB and incubated at 37°C with shaking for 5 h. The culture was then spun at 3000 × g for 10 min and the supernatant transferred to new microcentrifuge tubes. M13 ssDNA was isolated from the LB supernatant following standard protocols (Sambrook et al., 1989). Each ssDNA preparation was re-suspended in 50 μL of TE buffer (stored at -20°C indefinitely).

Each ssDNA preparation was used as a template for PCR using Integrated DNA Technologies (IDT) Ready-made oligonucleotides M13 Forward (-41), M13 Reverse (-48), M13 Forward (-20), and M13 Reverse (-27) as primers (Table 2.1). Stock solutions of M13mp18 ssDNA (NEB) and M13mp19 RF DNA were used as control templates for the PCR screening. Polymerase chain reactions were performed by combining *Taq* PCR Master Mix solution (Qiagen) with the above mentioned templates (ssDNA preparations) and primers. Plaques 24 and 26 were identified as possible positives for the *Pf* gdh insert.

Large scale ssDNA preparations were obtained from plaques 24 and 26 using the original plaque suspension (stored at -20°C) to transfect a 4 mL culture of JM109 non-infected cells. This culture was used to inoculate a larger 500 mL culture. The single-stranded DNA preparation was obtained according to standard protocols (Sambrook et al., 1989). Correct insertion of the -75 → +45 region of the *Pf* gdh gene was confirmed by sequencing (MGIF, UGA).

2.1.3. Cloning of *Pf* gdh promoter into M13mp18

A different approach was taken for inserting the *Pf* gdh promoter fragment into the multi-cloning site of M13mp18 RF DNA (NEB), utilizing the first promoter fragment clone (M13mp19-*Pf*gdhP). Large-scale preparations of M13 RF DNA were performed on M13mp19-*Pf*gdhP and M13mp18 following standard protocols using *E. coli* XL1-blue (Stratagene) as the host strain. M13mp19-*Pf*gdhP was digested sequentially with EcoR I and Sph I and then loaded onto a 5% TBE gel (BioRad) and electrophoresed at 18 V/cm for 10 min. The gel was then stained with ethidium bromide to reveal the restriction

enzyme-digested product of ~120 bp. The stained band was excised with a scalpel, crushed using the tip of a sterile 1000 μL pipette tip, and eluted overnight in a microcentrifuge tube with 400 μL elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, pH 7.5, 1 mM EDTA). M13mp18 RF DNA (NEB) was digested with EcoR I and Sph I in similar fashion and desalted into TE using a CHROMA SPIN+TE-100 (ClonTech) column according to the supplier's protocol. The spin column eluate was then ethanol precipitated and resuspended in 20 μL of sterile deionized H_2O.

The eluted insert and digested M13 vector from the gel and CHROMA SPIN column, respectively, were then combined in a ligation reaction as described above. Ligation reactions were transformed into XL1-Blue electrocompetent cells (Stratagene) following supplier's suggested protocol for filamentous phage transformation. Plaques were screened (as above), and the correct M13mp18-*Pf*gdhP construction was confirmed by sequencing.

As needed, subsequent large scale ssDNA preparations of both promoter fragment clones were performed by either transfecting viable plaque suspensions of a given M13 clone or by transforming (by electroporation) a dilute (1:1000) solution of a ssDNA stock into XL1-blue cells.

2.2. Oligonucleotide synthesis and derivatization

2.2.1. Oligodeoxyribonucleotide synthesis

Forty-one oligodeoxyribonucleotides were synthesized for analysis of protein-DNA interactions in the *Pf* preinitiation complex (PIC) using an ABI392 DNA/RNA synthesizer (Applied Biosystems). An additional 3 oligodeoxyribonucleotides were

synthesized for use as controls in the analysis. Each oligodeoxyribonucleotide was 26 nucleotides in length with a phosphorothioate incorporated between the second and third nucleotide (from the 5' end). The synthesized oligodeoxyribonucleotides are listed in Tables 2.2 and 2.3. The following protocol was used for the synthesis of each oligodeoxyribonucleotide (R. H. Ebright Lab, Waksman Institute, Rutgers University, Piscataway, New Jersey).

Oligodeoxyribonucleotide synthesis protocol

1. Perform 24 standard cycles of solid-phase β-cyanoethylphosphoramidite oligodeoxyribonucleotide synthesis to prepare CPG-linked precursor containing residues 3-26 of desired oligodeoxyribonucleotide. Use the following settings: cycle, "1.0 μmol CE"; DMT, "on"; end procedure, "manual".
2. Replace iodine/water/pyridine/tetrahydrofuran solution (bottle 15) by tetraethylthiuram disulfide/acetonitrile solution. Perform one modified cycle of solid-phase β-cyanoethylphosphoramidite oligodeoxyribonucleotide synthesis to add base 2 and phosphorothioate linkage. Use the following settings: cycle, "1.0 μmol sulfur"; DMT, "on"; end procedure, "manual".
3. Replace tetraethylthiuram disulfide/acetonitrile solution (bottle 15) with iodine/water/pyridine/tetrahydrofuran solution. Place collecting vial on the DNA synthesizer. Perform one standard cycle of solid-phase β-cyanoethylphosphoramidite oligodeoxyribonucleotide synthesis to add residue 1. Use the following settings: cycle, "1.0 μmol CE"; DMT, "on"; end procedure, "CE". Remove collecting vial, screw cap tightly, and incubate 10 h at 55°C to remove "blocking" groups from the synthesized oligodeoxyribonucleotide side chains. Steps

Table 2.2. Non-Template Strand synthesized[a] derivatized oligodeoxyribonucleotides

Promoter strand	Probe position	Sequence (5' - 3')
NT[b]	-70[c]	AT(PSO$_3$[d])CAAATAAACAAAAGGATTTCCACT
NT	-40	TT(PSO$_3$)TACCGAAAGCTTTATATAGGCTAT
NT	-38	TA(PSO$_3$)CCGAAAGCTTTATATAGGCTATTG
NT	-36	CC(PSO$_3$)GAAAGCTTTATATAGGCTATTGCC
NT	-34	GA(PSO$_3$)AAGCTTTATATAGGCTATTGCCCA
NT	-32	AA(PSO$_3$)GCTTTATATAGGCTATTGCCCAAA
NT	-30	GC(PSO$_3$)TTTATATAGGCTATTGCCCAAAAA
NT	-28	TT(PSO$_3$)TATATAGGCTATTGCCCAAAAATG
NT	-26	TA(PSO$_3$)TATAGGCTATTGCCCAAAAATGTA
NT	-24	TA(PSO$_3$)TAGGCTATTGCCCAAAAATGTATC
NT	-22	TA(PSO$_3$)GGCTATTGCCCAAAAATGTATCGC
NT	-20	GG(PSO$_3$)CTATTGCCCAAAAATGTATCGCCA
NT	-18	CT(PSO$_3$)ATTGCCCAAAAATGTATCGCCAAT
NT	-16	AT(PSO$_3$)TGCCCAAAAATGTATCGCCAATCA
NT	-14	TG(PSO$_3$)CCCAAAAATGTATCGCCAATCACC
NT	-12	CC(PSO$_3$)CAAAAATGTATCGCCAATCACCTA
NT	-10	CA(PSO$_3$)AAAATGTATCGCCAATCACCTAAT
NT	-8	AA(PSO$_3$)AATGTATCGCCAATCACCTAATTT
NT	-6	AA(PSO$_3$)TGTATCGCCAATCACCTAATTTGG
NT	-4	TG(PSO$_3$)TATCGCCAATCACCTAATTTGGAG
NT	-2	TA(PSO$_3$)TCGCCAATCACCTAATTTGGAGGG
NT	+1	TC(PSO$_3$)GCCAATCACCTAATTTGGAGGGAT
NT	+45	

[a]All oligodeoxyribonucleotides synthesized using ABI392 DNA/RNA synthesizer (Applied Biosystems)
[b]Non-Template (NT) strand of promoter DNA
[c]Position of phosphorothioate with respect to *Pf* gdh promoter transcription start site (+1)
[d]Phosphorothioate (PSO_3) position within the synthesized oligodeoxyribonucleotide

Table 2.3. Template Strand synthesized derivatized oligodeoxyribonucleotides

Strand	Probe position	Sequence (5' - 3')
T	-66	TT(PSO_3[d])TATTTGATTAGGAATTCGTAATCA
T	-40	GT(PSO_3)AAACAAGAGTGGAAATCCTTTTGA
T	-38	CG(PSO_3)GTAAACAAGAGTGGAAATCCTTTT
T	-36	TT(PSO_3)CGGTAAACAAGAGTGGAAATCCTT
T	-34	CT(PSO_3)TTCGGTAAACAAGAGTGGAAATCC
T	-32	AG(PSO_3)CTTTCGGTAAACAAGAGTGGAAAT
T	-30	AA(PSO_3)AGCTTTCGGTAAACAAGAGTGGAA
T	-28	AT(PSO_3)AAAGCTTTCGGTAAACAAGAGTGG
T	-26	AT(PSO_3)ATAAAGCTTTCGGTAAACAAGAGT
T	-24	CT(PSO_3)ATATAAAGCTTTCGGTAAACAAGA
T	-22	GC(PSO_3)CTATATAAAGCTTTCGGTAAACAA
T	-20	TA(PSO_3)GCCTATATAAAGCTTTCGGTAAAC
T	-18	AA(PSO_3)TAGCCTATATAAAGCTTTCGGTAA
T	-16	GC(PSO_3)AATAGCCTATATAAAGCTTTCGGT
T	-14	GG(PSO_3)GCAATAGCCTATATAAAGCTTTCG
T	-12	TT(PSO_3)GGGCAATAGCCTATATAAAGCTTT
T	-10	TT(PSO_3)TTGGGCAATAGCCTATATAAAGCT
T	-8	AT(PSO_3)TTTTGGGCAATAGCCTATATAAAG
T	-6	AC(PSO_3)ATTTTTGGGCAATAGCCTATATAA
T	-4	AT(PSO_3)ACATTTTTGGGCAATAGCCTATAT
T	-2	CG(PSO_3)ATACATTTTTGGGCAATAGCCTAT

[a]All oligodeoxyribonucleotides synthesized using ABI392 DNA/RNA synthesizer (Applied Biosystems).
[b]Template (T) strand of promoter DNA.
[c]Position of phosphorothioate with respect to *Pf* gdh promoter transcription start site (+1).
[d]Phosphorothioate (PSO_3) position within the synthesized oligodeoxyribonucleotide.

1-3 yield an oligodeoxyribonucleotide 26 nt in length with a phosphorothioate linkage between nucleosides 2 and 3. Occasionally (rarely), a low yield is observed in the primer-extension reaction (discussed below, Section 2.3); in such a case, an oligodeoxyribonucleotide with 1-10 additional nucleotides at the 3' end, and preferably with dG or dC at the 3' end, should be prepared.

2.2.2. Synthesized oligonucleotide purification and deprotection

After the synthesized oligodeoxyribonucleotides were deblocked overnight, one tenth of the eluted oligodeoxyribonucleotide was placed into a microcentrifuge tube and purified using an Oligonucleotide Purification Cartridge (Perkin Elmer, details below) following the supplier's protocol. The remaining amount of synthesized oligodeoxyribonucleotide was dried in a SpeedVac and stored at -80°C (stable indefinitely).

Oligonucleotide Purification Cartridge Protocol

1. Take one tenth aliquot of synthesized oligonucleotide and dilute to 1 mL with water.
2. Label OPC columns well with the corresponding oligonucleotide.
3. Pass 5 mL of 100% acetonitrile over the column. This step washes impurities from the column. This and other specified steps should be done with a vacuum manifold. Attach a 10 mL syringe to top of the OPC column. A small adapter is attached to the bottom of the column for attachment to the vacuum manifold.
4. Pass 5 mL of 2 M TEAA over the column (vacuum manifold).
5. Pass diluted oligonucleotide over the OPC column at a rate of 1 drop/second. This is done manually by detaching the OPC column and 10 mL syringe from the vacuum

manifold. The flow rate affects the binding efficiency of the oligonucleotide to the column. Use the following technique to push the diluted solution through into a sterile 1.7 mL microcentrifuge tube.

a. Place the diluted oligonucleotide in the attached 10 mL syringe.
b. Insert the syringe plunger just inside the syringe barrel and wait (this initial pressure will initiate the flow of the diluted oligonucleotide through the OPC column).
c. Hold the barrel of the syringe with your hand (when the barrel is wrapped in hand, the flow rate of the solution can be slightly regulated by unwrapping and wrapping the barrel within the hand).
d. Only after most of the solution has entered the OPC column, push the plunger slightly just past the plunger "catch" of the barrel.

6. Pass the diluted oligonucleotide through the OPC column a second time (manually as before).
7. Wash with 5 ml of 1.5 M NH_4OH (1/10 dilution of concentrated NH_4OH) followed by 10 mL of millipore water (use vacuum manifold).
8. Fill syringe barrel with 5 mL of 3% TFA in water (this step removes tritylated side groups for the synthesized oligonucleotide). Pass through 1 mL and then let stand for 1 min. Pass through 2 mL and let stand for 2 min. Pass the remaining amount of 3% TFA. The column should turn a noticeable red color as the TFA passes through (vacuum manifold).
9. Pass 10 mL of millipore water over the column (vacuum manifold).

10. Elution of oligonucleotide (done manually with syringe). Elute and collect the purified detritylated oligonucleotide by slowly passing 1 mL of 35% acetonitrile (follow same procedure for elution as described for Step 5). Thrty-five percent acetonitrile is the recommended concentration for elution of oligonucleotides containing phosphorothioates: for oligonucleotides without phosphorothioates, elution with 20% acetonitrile is recommended.
11. Measure the OD_{260} of the eluted oligonucleotide (requires a 10:500 or 20:750 dilution).
12. Speedvac dry the eluted oligonucleotide. To shorten drying time, the 1 mL can be aliquoted. Spin 30 min with Speedvac lid ajar and with no vacuum (allowing evaporation of ammonia). Close Speedvac lid, apply vacuum, and dry.
13. Store at -20 ˚C (stable for at least 2 y).

2.2.3. Derivatization of oligodeoxyribonucleotide

After purification and deprotection using the OPC column, each phosphorothioate oligodeoxyribonucleotide was derivatized with azidophenacyl bromide (AP, Sigma) using the following protocol.

1. Dissolve 10 mg (42 μmol) azidophenacyl bromide in 1 mL chloroform. Transfer 100 μL aliquots (4.2 μmol) to 1.5 mL siliconized polypropylene microcentrifuge tubes, and dry in Speedvac. Wrap tubes with aluminum foil, and store desiccated at 4°C (stable indefinitely). This and all subsequent steps should be carried out under subdued lighting. Fluorescent light and daylight must be excluded. Low to moderate

levels of incandescent light (e.g., from single task lamp with 60 W tungsten bulb) are acceptable.

2. Dissolve 50 nmol aliquot of phosphorothioate oligodeoxyribonucleotide (Section 2.2.2, Step 12) in 50 µL water, and re-suspend 42 µmol aliquot of azidophenacyl bromide (Section 2.2.3, Step 1) in 220 µL methanol.
3. Mix 50 µL (50 nmol) phosphorothioate oligodeoxyribonucleotide solution, 5 µL 1 M potassium phosphate (pH 7.0), and 55 µL (1 µmol) azidophenacyl bromide solution in 1.5 mL siliconized polypropylene microcentrifuge tube. Incubate 3 h at 37°C.
4. Precipitate derivatized oligodeoxyribonucleotide by adding 11 µL 3 M sodium acetate (pH 5.2), and 275 µL ice-cold 100% ethanol. Invert tube several times, and place at -80°C for 30 min. Centrifuge 5 min at 13,000 × g at 4°C. Remove supernatant, and wash pellet with ice-cold 70% ethanol. Air dry 15 min at RT. Store at -20°C (stable for at least 1 y).

2.2.4. Purification of derivatized oligonucleotide

After each oligodeoxyribonucleotide was derivatized, it was purified from the underivatized oligodeoxyribonucleotide using HPLC under the following conditions.

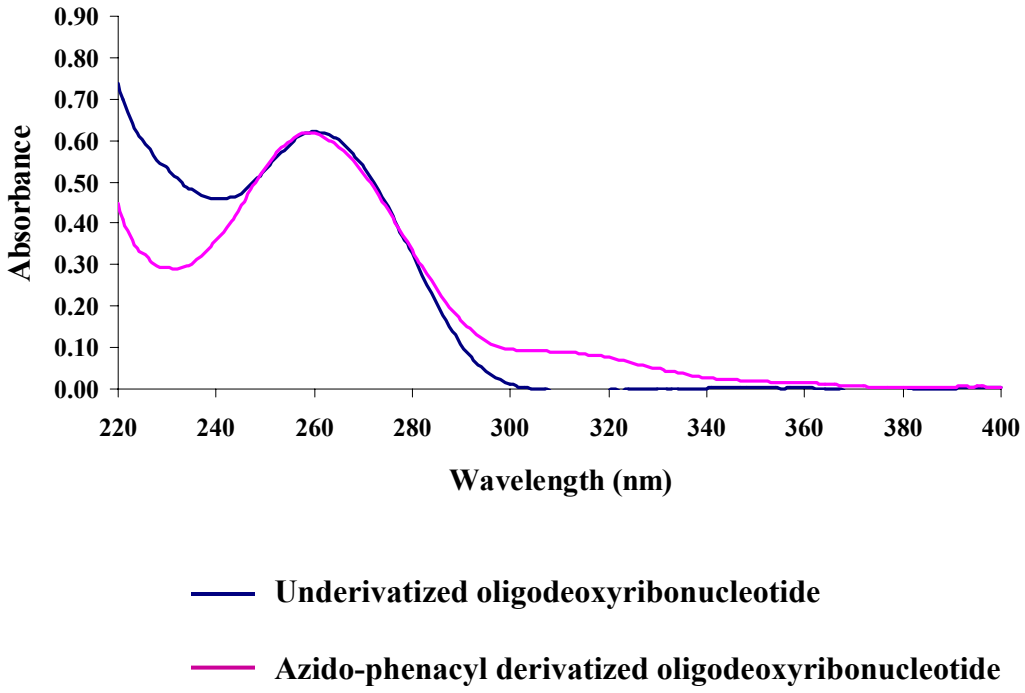
1. Re-suspend derivatized oligodeoxyribonucleotide (from Section 2.2.3, Step 4) in 100 µL 50 mM triethylammonium acetate (pH 7.0).
4. Analyze 5 µL aliquot by C18 reversed-phase HPLC to confirm efficiency of derivatization reaction. Use LiChrospher 100 RP-18 C18 reversed-phase HPLC column (5 µm), with solvent A = 50 mM triethylammonium acetate (pH 7.0), 5% acetonitrile; solvent B = 100% acetonitrile; and flow rate = 1 ml/min. Equilibrate

column with 10 column volumes solvent A before loading sample. After loading sample, wash column with 6 column volumes solvent A, and elute with 15 min gradient of 0-35% solvent B in solvent A. Derivatized and underivatized oligodeoxyribonucleotides elute at ~25% solvent B and ~16% solvent B, respectively. Elution points vary depending on the composition of the oligodeoxyribonucleotide. This makes the 5 μL aliquot analysis critical until familiarity with the elution profile is achieved. If possible, a UV-visible absorbance spectrum (220 - 350 nm) should be recorded of the oligodeoxyribonucleotides as they elute from the column. The underivatized oligodeoxyribonucleotides will have a primary absorbance peak at 260 nm. The derivatized oligodeoxyribonucleotides will have the same absorbance peak at 260 nm plus a distinct smaller peak (shoulder) at ~300 nm (Figure 2.3).

5. If derivatization efficiency is ≥50%, purify remainder of sample using procedure of step 2, collecting derivatized oligodeoxyribonucleotide peak fractions.

   Pool peak fractions, and dry in Speedvac. Store desiccated at -20°C in the dark (stable for at least 1 y). The derivatization procedure yields two diastereomers in an approximately 1-to-1 ratio: one in which azidophenacyl is incorporated at the sulfur atom corresponding to the phosphate O1P, and one in which azidophenacyl is incorporated at the sulfur atom corresponding to the phosphate O2P (Mayer and Barany, 1995). Depending on oligodeoxyribonucleotide sequence and HPLC conditions, the two diastereomers may elute as a single peak, or as two peaks (e.g., at 24% and 25% solvent B). In most cases, no effort was made to resolve the two diastereomers, and experiments are performed using the unresolved diastereomeric mixture. This permits simultaneous probing of protein-DNA interactions in the DNA

Figure 2.3. UV-visible absorbance spectrum (220 – 400 nm) of derivatized and underivatized oligodeoxyribonucleotide. The underivatized oligodeoxyribonucleotide (dark blue, spectrum of oligdeoxyribonucleotide after synthesis, Section 2.2.1) has a primary absorbance at 260 nm. The azido-phenacyl derivatized oligodeoxyribonucleotide (magenta line, spectrum of oligodeoxyribonucleotide after azido-phenacyl derivatization and HPLC isolation, Sections 2.2.3 and 2.2.4) has a primary absorbance at 260 nm as well as a smaller peak (shoulder) at ~300 nm.


0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Absorbance
220
240
260
280
300
320
340
360
380
400
Wavelength (nm)
Underivatized oligodeoxyribonucleotide
Azido-phenacyl derivatized oligodeoxyribonucleotide

minor groove (probed by the O1P-derivatized diastereomer) and the DNA major groove (probed by the O2P-derivatized diastereomer) (Lagrange et al. 1996; Naryshkin et al. 2000).

After the derivatized oligodeoxyribonucleotides were isolated and dried, they were dissolved in distilled deionized water and OD_{260} was measured (20:750 dilution in deionized water) to determine concentration (concentration calculated based on the equation Conc. (pmol/μL) = Abs_{260nm} * [100/(1.5 * n_A + 0.71 * n_C + 1.2 * n_G + 0.84 * n_T)] where n_A, n_C, n_G, and n_T are the number of adenines, cytosines, guanines, and thymidines, respectively). Then the oligodeoxyribonucleotides were divided into four 200 pmol aliquots and three larger aliquots (ranging from 1000 to 4000 pmol), dried in Speedvac, and stored at -80°C in the dark.

2.3. Probe synthesis

Using the ssDNA preparations of M13mp18-*Pf*gdhP, M13mp19-*Pf*gdhP (from Section 2.1), and the derivatized oligodeoxyribonucleotides (from Section 2.2.4), primer extension reactions were carried out to construct a 596 bp photocrosslinking probe. Single-stranded M13mp19-*Pf*gdhP contained the template (T) strand of the *Pf* gdh promoter fragment and was therefore used as the template for constructing probes to analyze protein-DNA interactions with the non-template (NT) strand of the *Pf* gdh promoter. Likewise M13mp18-*Pf*gdhP contained the NT strand of the *Pf* gdh promoter fragment and was used as the template for T strand probe construction. Ten to twenty individual probes (each containing a different derivatized oligodeoxyribonucleotide) were constructed at a time for photocrosslinking analysis. The following is the standard

protocol for DNA probe construction for a single probe. In practice, "master mixes" were used at each individual step (where applicable) to minimize repetitive pipeting. The concentrations of derivatized oligodeoxyribonucleotide and single-stranded template are critical for steps 1-6 of the protocol. Following standard protocols for large scale preparation of single-stranded DNA (discussed above in Section 2.1.2), the final re-suspended solution of ssDNA was not as concentrated as needed (based on OD_{260}) for the probe construction protocol. Both ssDNA templates were precipitated with ethanol and 3 M sodium acetate (pH 4.5), re-suspended in a smaller volume of water, and then OD_{260} was measured to determine the new concentration (target concentration 0.25-0.5 μM ssDNA template).

Derivatized promoter DNA fragment probe protocol (see Figures 2.4, 2.5, and Table 2.4)

1. Resuspend 200 pmol aliquot of derivatized oligodeoxyribonucleotide in 40 μL deionized water. Place 3 μL (15 pmol) in a 1.5 mL siliconized polypropylene microcentrifuge tube. The remaining amount of resuspended derivatized oligodeoxyribonucleotide can be stored at -20°C for up to one year and used repeatedly.
2. Add 2 μL 10× phosphorylation buffer (500 mM Tris-HCl, pH 7.6, 100 mM $MgCl_2$, 15 mM β-mercaptoethanol), 5 μL [γ^{32}P]ATP (50 μCi, 6000 Ci, 222 TBq/mmol, Perkin Elmer), and 1 μL (10 units) T4 polynucleotide kinase (NEB). Incubate 30 min at 37°C. This and all subsequent steps should be carried out under subdued lighting. Daylight must be excluded. Low to moderate levels of incandescent light (e.g., from single task lamp with 60 W tungsten bulb) are acceptable.
3. Add 10 μL water and mix thoroughly.

Figure 2.4. Derivatized DNA promoter fragment synthesis (Section 2.3, Steps 1-12). Diagram representation of derivatized DNA promoter fragment synthesis as detailed in Section 2.3. The protocol begins with the aliquoted, purified, and derivatized oligodeoxyribonucleotide from Section 2.2.4. The oligodeoxyribonucleotide is the substrate for a kinase reaction to covalently attach a radioactive phosphate group (^{32}P, Step 2) to the 5' end. The labeled oligodeoxyribonucleotide is then desalted (Step 4) and combined with M13mp18-PfgdhP (or mp19) ssDNA template and a second oligodeoxyribonucleotide for primer annealing (Step 6 and 7) for the subsequent primer extension reaction (Step 8). Once the primer extension reaction is complete, the dsDNA is digested with restriction enzymes AvaII and FspI (Step 10). The resulting DNA fragment with AvaII digested 3' recessed end is filled in by incubation with T4 DNA polymerase I (NEB), Large (Klenow) fragment and dNTPs (Step 12). When immobilized DNA probe photocrosslinking was done (Section 2.5.2), this Klenow fragment step included Biotin-14-CTP (Gibco BRL) to allow binding of the DNA probe to steptavidin-coated magnetic beads. Bold numbers correspond to steps detailed in Section 2.3.

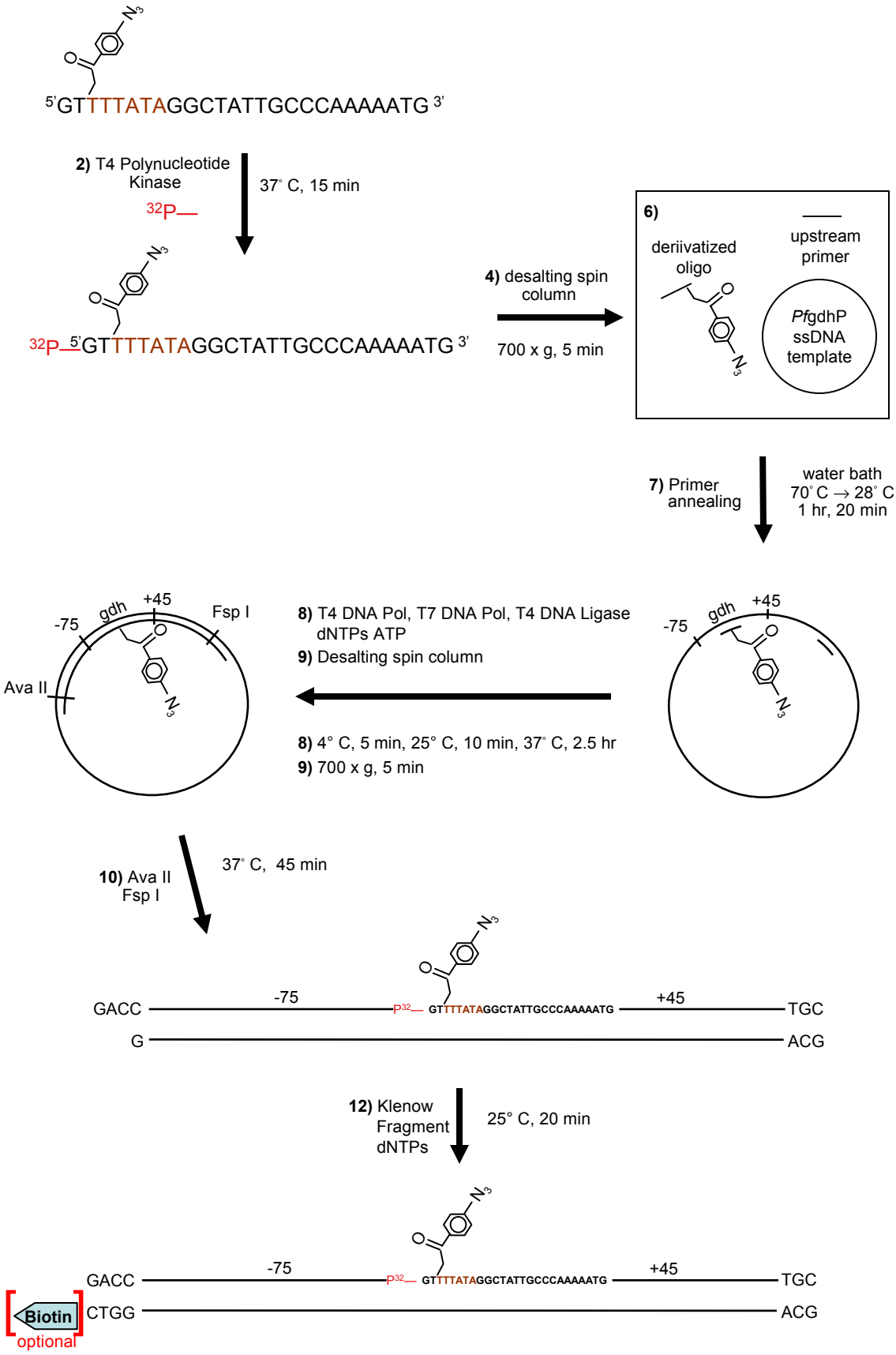

5'GTTTTATAGGCTATTGCCCAAAAATG 3'
2) T4 Polynucleotide Kinase
32P—
37° C, 15 min
32P—5'GTTTTATAGGCTATTGCCCAAAAATG 3'
4) desalting spin column
700 x g, 5 min
6)
deriivatized oligo
upstream primer
PfgdhP ssDNA template
7) Primer annealing
water bath
70° C → 28° C
1 hr, 20 min
-75
gdh
+45
Fsp I
Ava II
8) T4 DNA Pol, T7 DNA Pol, T4 DNA Ligase
dNTPs ATP
9) Desalting spin column
8) 4° C, 5 min, 25° C, 10 min, 37° C, 2.5 hr
9) 700 x g, 5 min
10) Ava II
Fsp I
37° C, 45 min
GACC
-75
P32— GTTTTATAGGCTATTGCCCAAAAATG
+45
TGC
G
ACG
12) Klenow Fragment dNTPs
25° C, 20 min
GACC
CTGG
Biotin
optional

Figure 2.5. Derivatized DNA promoter probe synthesis (Section 2.3, Steps 13-21). Diagram representation (continued) of derivatized DNA promoter probe synthesis as detailed in Section 2.3. After the restriction digested recessed ends are filled in (Step 12), the DNA fragment is ethanol precipitated and resuspended in sterile H_2O (Step 13). Calf Intestinal alkaline phosphatase is then added to remove 5' phosphate groups from the DNA fragment (Step 14). This step removes ^{32}P from the derivatized oligonucleotide primer (Steps 6-9) that was not ligated in the primer extension reaction (Step 8). The DNA fragment is then loaded onto a 5% TBE gel and electrophoresed (Step 15). After exposing the gel to X-ray film, the gel is superimposed on the developed film (Step 16) and the corresponding band of the DNA fragment is cut out (Step 16). The gel piece is placed in a microcentrifuge tube, crushed using a pipet tip, and then eluted overnight with an elution buffer (Step 17). The eluted DNA fragment is separated from the gel pieces using a Spin-X filter microcentrifuge tube (Costar), ethanol precipitated, and finally resuspended in TE buffer at a final concentration of 10,000 cpm / μL. Bold numbers correspond to steps detailed in Section 2.3.

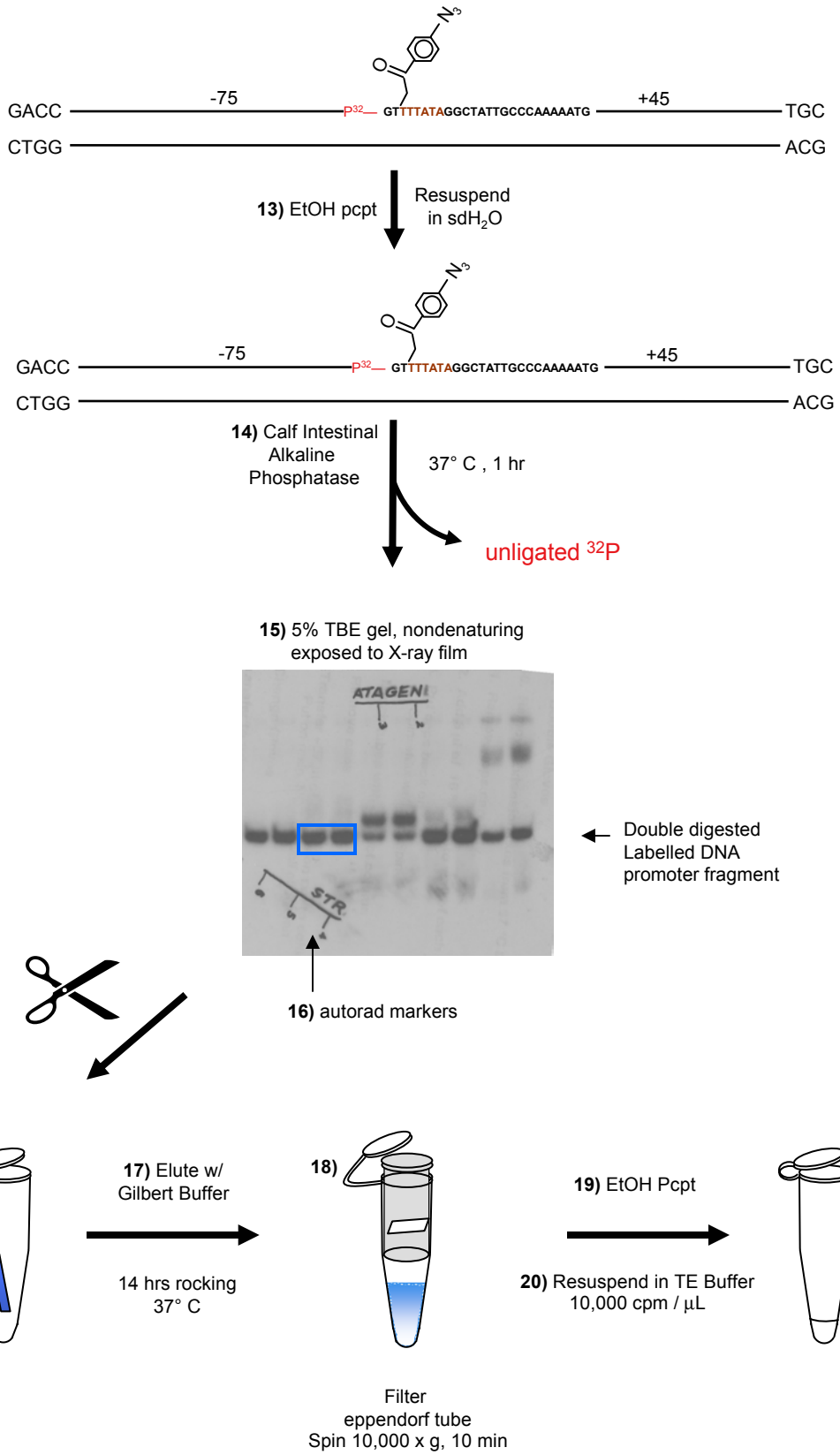

GACC
-75
P32— GTTTTATAGGCTATTGCCCAAAAATG
+45
TGC
CTGG
ACG
13) EtOH pcpt
Resuspend in sdH2O
14) Calf Intestinal Alkaline Phosphatase
37° C , 1 hr
unligated 32P
15) 5% TBE gel, nondenaturing exposed to X-ray film
ATAGEN
STR
Double digested Labelled DNA promoter fragment
16)
16) autorad markers
17) Elute w/ Gilbert Buffer
14 hrs rocking 37° C
18)
Filter eppendorf tube Spin 10,000 x g, 10 min
19) EtOH Pcpt
20) Resuspend in TE Buffer 10,000 cpm / µL

Table 2.4. Buffers for DNA probe synthesis (Section 2.3)

10× phosphorylation buffer	500 mM Tris-HCl, pH 7.6, 100 mM $MgCl_2$, 15 mM β-mercaptoethanol
10× annealing buffer	400 mM Tris-HCl, pH 7.9, 500 mM NaCl, 100 mM $MgCl_2$
20× restriction endonuclease digestion buffer	1 M potassium acetate, 400 mM Tris-acetate, 200 mM magnesium acetate, 10 mM βME
Elution buffer	0.5 M ammonium acetate, 10 mM magnesium acetate, pH 7.5, 1 mM EDTA
Nondenaturing loading buffer	0.3% bromophenol blue, 0.3% xylene cyanol, 30% glycerol, in water
0.5× TBE	45 mM Tris-borate, pH 8.3, 1 mM EDTA
Low-EDTA TE buffer	10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA
TE buffer	10 mM Tris-HCl, pH 7.6, 1 mM EDTA

4. Desalt radiophosphorylated derivatized oligodeoxyribonucleotide into TE using CHROMA SPIN+TE-10 spin column (ClonTech) according to the supplier's protocol.
5. Immediately proceed to next step or, if necessary, store radiophosphorylated derivatized oligodeoxyribonucleotide solution at -20°C in the dark (stable for up to 24 h).
6. In 1.5 mL siliconized polypropylene microcentrifuge tube, mix 30 µL radiophosphorylated derivatized oligodeoxyribonucleotide, 2 µL 10 µM upstream primer (20 pmol, Table 2.1), 5 µl 0.4 µM single-stranded template (T or NT strand template), and 4.5 µL 10× annealing buffer (400 mM Tris-HCl, pH 7.9, 500 mM NaCl, 100 mM $MgCl_2$).
7. Heat 5 min at 70°C. Transfer tubes to 600 mL beaker containing 400 mL water at 70°C, and place beaker at room temperature to permit slow cooling (70°C to 25°C in ≈ 60 min). The rate of cooling should be slowed by covering the beaker with aluminum foil initially for 10 min, followed by venting the foil on one side for 15 min, and then completely removing the foil.
8. Add 2 µL DNA polymerization mix (20 mM each dATP, dGTP, dCTP, and TTP), 1 µL 100 mM ATP, 0.5 µL (10 units) T7 DNA polymerase, 1 µL (3 units) T4 DNA polymerase, 1 µL (5 units) T4 DNA ligase, and 1 µL (1 µg/ml) single-stranded binding protein (single-stranded binding protein optional). Incubate 5 min on ice, followed by 10 min at room temperature, followed by 2-3 h at 37 °C. Terminate reaction by adding 3 µL 10% SDS.

9. Desalt into TE using CHROMA SPIN+TE-100 spin column according to supplier's protocol (ClonTech).
10. Restriction digestion. Combine the CHROMA SPIN+TE-100 eluate with 2.5 μL of 20× digestion buffer, 10 units of Ava II (NEB), and 10 units of Fsp I (NEB). Incubate 1 h at 37°C.
11. Heat inactivate the restriction endonucleases by incubating 5 min at 65°C.
12. Allow the solution from step 11 to cool to room temperature and then add 2.5 μL 1 mM biotin-labelled dCTP (Gibco BRL), 0.25 μL 10 mM each other dNTP (Sigma, not dCTP), and 0.3 μL (3 units) DNA polymerase I Klenow fragment (Promega). Incubate 20 min at 25°C. When constructing derivatized DNA fragment probes for use in solution photocrosslinking (Section 2.5.4 below), biotin-labelled dCTP is not needed. Simply add 0.25 μL 10 mM of all dNTP's (including dCTP).
13. Precipitate the DNA by adding 11 μL of 3 M sodium acetate and 150 μL of ice-cold 100% ethanol. Invert tube several times, and place at -20°C for 30 min. Centrifuge 10 min at 13,000 × g at 4°C. Remove supernatant, wash pellet with ice-cold 95% ethanol. Air dry 15 min at room temperature.
14. Resuspend pellet in 25 μL of Low-EDTA TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) and add 3 μL of 10x CIP reaction buffer and 2 μL of calf-intestinal alkaline phosphatase (Promega). Incubate 30 min to 1 h at 37°C.
15. Add, in order, 3 μL 0.5 M EDTA (pH 8.0) and 10 μL nondenaturing loading buffer. Apply to *nondenaturing* 5% polyacrylamide (29:1 acrylamide:bisacrylamide), 0.5× TBE slab gel (Bio-Rad, 10 x 7 x 0.15 cm). Electrophorese at 25 V/cm until Bromophenol Blue has completely run off gel.

16. Remove one glass plate, and cover gel with plastic wrap. Attach two autorad markers (Stratagene) to gel. Expose to X-ray film for 1-2 min at room temperature, and process film. Cut out portion of film corresponding to derivatized DNA fragment. Using white paper, superimpose cut-out film on gel, using autorad markers as alignment reference points. Using disposable scalpel, excise portion of gel corresponding to derivatized DNA fragment.
17. Place excised gel slice in 1.5 mL siliconized polypropylene microcentrifuge tube, and crush with 1 mL pipette tip. Add 400 µL elution buffer, vortex briefly, and rock 12 h at 37°C.
18. Transfer supernatant to Spin-X centrifuge filter tubes (Costar, Fisher), and centrifuge 1 min at 13,000 × g at room temperature in fixed-angle microcentrifuge.
19. Transfer filtrate to 1.5 mL siliconized polypropylene microcentrifuge tube. Precipitate derivatized DNA fragment by addition of 800 µL ice-cold 100% ethanol. Invert tube several times, and place at -20°C for 30 min. Centrifuge 5 min at 13,000 × g at 4°C in fixed-angle microcentrifuge. Remove and discard supernatant, wash pellet with 200 µL ice-cold 95% ethanol, and air dry 20 min at room temperature.
20. Resuspend the pellet so that the final concentration of radioactive-labelled probe is ~10,000 cpm/µL.
21. Store derivatized DNA fragment at -20°C in the dark (stable for ~20 days).

2.4. *Pf* RNAP and *Pf* Transcription Factor purification

2.4.1. Expression and purification of *Pf* TBP and *Pf* TBP N-terminal HisTAG

Recombinant *Pf* TBP was expressed and purified as described by H.-T. Chen (2000). *Pf* TBP N-terminal HisTAG was expressed and purified following the same methods as established for *Pf* TBP purification.

Stock solutions of the *Pf* transcription factors were diluted in transcription complex binding buffer [transcription buffer, 40 mM HEPES pH 6.5, 250 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 10 mM β-mercaptoethanol, 1× Protease Inhibitor (10x protease inhibitor, 10 μg/mL aprotinin; 60 μg/mL, chymostatin; 4 μg/mL, leupeptin; 4 μg/mL pepstatin; 2mM AEBSF), 8 % glycerol]. *Pf* TBP and *Pf* TBP N-terminal HisTAG were diluted to a final concentration of 2 nM or 4 nM (based on OD_{260}) and divided into 200 μL aliquots in siliconized microcentrifuge tubes, frozen, and stored at -80 °C. For photocrosslinking experiments, aliquots were thawed on ice just before use.

2.4.2. Expression and purification of *Pf*TFB and *Pf*TFB C-terminal HisTAG

Recombinant *Pf*TFB was expressed and purified as described by H.-T. Chen (2000). DNA coding for *Pf*TFB was cloned into bacterial expression vector pET-21b (Novagen) to produce a clone containing six histidine codons at the 3' end of the gene (Lewis, 2000). *Pf*TFB/pET-21b was transformed into *E. coli* BL-21 DE3 cells, plated on rich media Luria Broth (LB) agar plates containing ampicillin, and incubated at 37°C overnight. Single colonies were streaked on new LB agar plates and incubated at 37°C overnight. A 10 mL culture was inoculated and incubated at 37°C with shaking

overnight. Three mL of the overnight culture was used to inoculate two 50 mL cultures in rich media LB. The 50 mL cultures were incubated at 37°C with vigorous shaking for 2.5 h and then used to inoculate 2 L of rich media LB. Cells were grown to an OD_{600} of about 0.9 and subsequently induced with 0.1 mM (final concentration) isopropyl β-D-thiogalactopyranoside (IPTG). The culture was grown an additional 4 h and harvested by centrifugation at 3000 × g. Cells were frozen and stored at -80°C until they were used.

Cell paste (~6 g) from a 2 L growth was thawed by suspension in 30 mL lysis buffer (50 mM potassium phosphate (KP_i), pH 7.8, 10 mM 2-mercaptoethanol, 1 mM phenylmethanesulfonyl fluoride (PMSF), 1 µL/mL leupeptin, and 1 µg/mL DNase I). The cells were lysed by sonication and the supernatant was clarified by centrifugation at 30,000 × g for 1 h at 4°C. The supernatant was loaded onto a 2.5 × 20-cm phosphocellulose column (Whatman) and washed with 5 column volumes of 50 mM potassium phosphate buffer (KP_i), pH 7.8. *Pf*TFB protein was eluted with 500 mM potassium chloride (KCl) in 50 mM KP_i pH 7.8. The eluate was approximately 60% *Pf* TFB by SDS-PAGE.

The eluate was subsequently equilibrated in 500 mM KCl in 50 mM potassium phosphate, pH 7.8 and concentrated from 150 mL to 20 mL using an Amicon ultrafiltration unit fitted with a YM3 membrane at 4 °C. The resulting protein solution in 50 mM KP_i buffer containing 500 mM KCl was loaded onto a 1 × 5-cm HisBind Ni-NTA resin column (Novagen) and washed with 5 column volumes of 50 mM KP_i buffer containing 500 mM KCl. The *Pf*TFB protein was eluted at 300 mM imidazole in 50 mM KP_i buffer containing 500 mM KCl. The eluate was ca. >95% pure by SDS-PAGE. Eluted fractions containing 300 mM and 600 mM imidazole 50 mM KP_i buffer

containing 500 mM KCl were pooled and buffer-exchanged into 50 mM KP_i buffer containing 500 mM KCl and then concentrated to ~9 mL. The resulting *Pf* TFB solution was determined to be 4 μM (OD_{280}) and divided into 1 mL aliquots that were frozen and stored at -80 °C until they were used.

For photocrosslinking experiments, *Pf* TFB and *Pf* TFB C-terminal HisTAG were diluted to a final concentration of 20 nM and divided into 25 μL aliquots in siliconized microcentrifuge tubes, flash frozen in liquid nitrogen, and then stored at -80 °C. Aliquots were thawed on ice just before use.

2.4.3. *Pf* RNAP purification

P. furiosus RNA polymerase (*Pf* RNAP) was partially purified from *Pf* whole cell extracts by L. M. Lewis in the Scott lab [(Lewis, 2000) concentration 600 ng/μL as determined by Bradford assay, Perkin Elmer]. Its functionality at high temperatures was confirmed using in vitro nonspecific and specific assays, which included recombinant *Pf*TFB and *Pf* TBP as well (Lewis, 2000). This partially purified *Pf* RNAP was used in the photocrosslinking experiments discussed below.

Pf RNAP was diluted in transcription binding buffer containing 30% glycerol to a final concentration of 150 ng/μL and divided into 25 μL aliquots, frozen and stored at -80°C as above. For photocrosslinking experiments, aliquots were thawed on ice just before use.

2.5. Photocrosslinking protocol

Protein-DNA interactions within the *P. furiosus* transcription preinitiation complex were analyzed by assembling recombinant transcription factors *Pf* TBP and *Pf* TFB and partially purified native *Pf* RNA polymerase on the above-described derivatized DNA fragment probes (Section 2.3). Protocols for two methods of photocrosslinking analysis (immobilized DNA probe, Section 2.5.2 and solution photocrosslinking, Section 2.5.3) are presented below.

2.5.1. Experimental Setup

Assembly and photocrosslinking of the *P. furiosus* transcription preinitiation complex was done at 70°C. Each individual experiment was conducted in a single siliconized or high-density polymer microcentrifuge tube. This required a combination of two dry baths with heat blocks and a water bath to maintain a constant temperature of 70°C from steps 8 - 20 for immobilized DNA probe ("on-bead") photocrosslinking and from steps 4 - 12 for solution photocrosslinking. The first heat block was maintained at 70°C for the assembly of the transcription preinitiation complex. The second heat block contained 13 x 100-mm borosilicate glass culture tubes filled with water and maintained at a temperature of approximately 85°C so that the water inside the tubes was 70°C. The second heat block and test tubes were used to maintain the high temperature while the tubes were being UV-irradiated. The water bath was maintained at 70°C and held the magnetic particle concentrator (Dynal) when the immobilized DNA probe photocrosslinking was done. In the case of solution photocrosslinking the water bath was not needed.

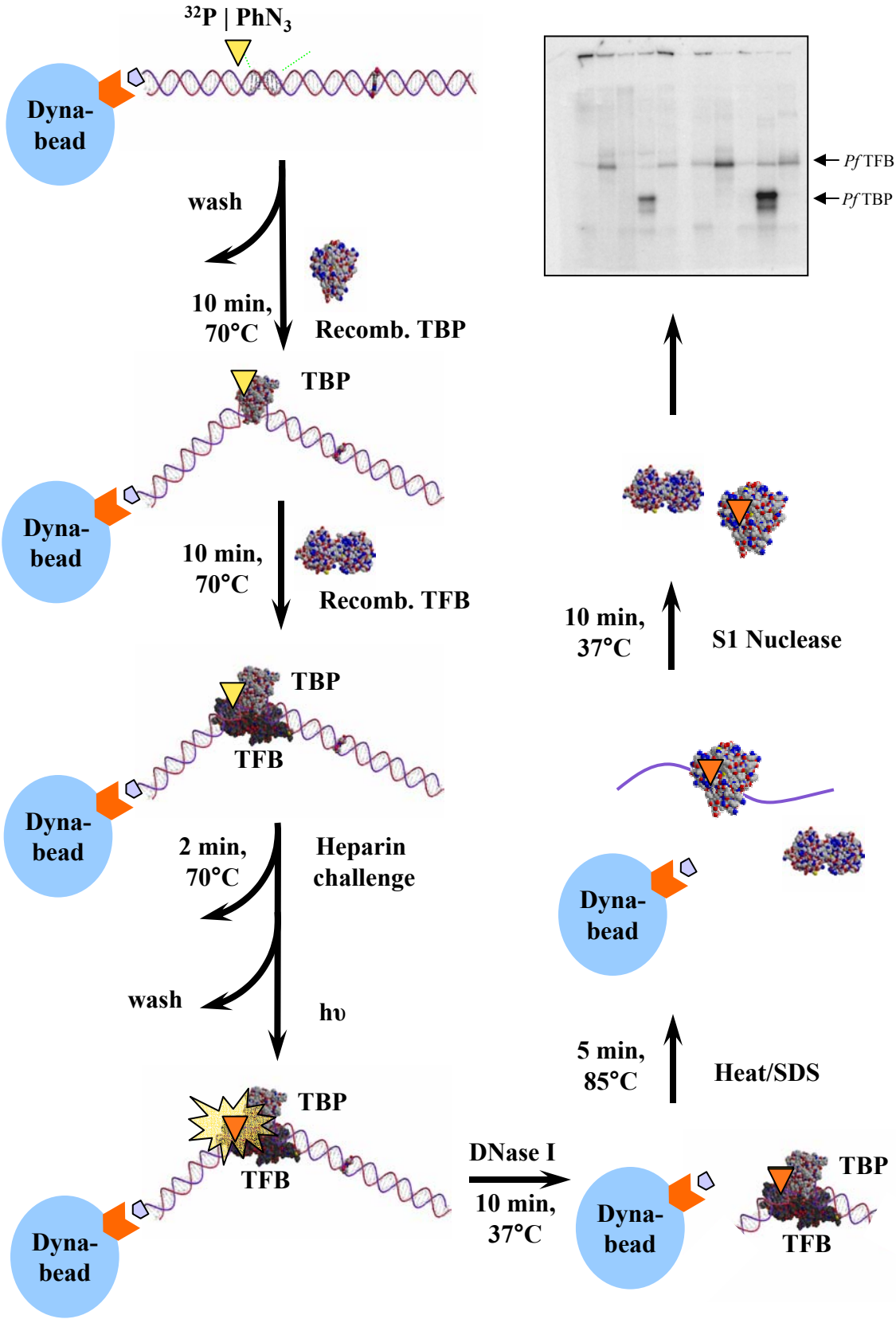
2.5.2. Immobilized DNA probe photocrosslinking

Immobilized DNA probe ("on-bead") photocrosslinking (Figure 2.6) utilized streptavidin- coated magnetic beads (Dynabeads, Dynal) to immobilize biotinylated, derivatized promoter fragment DNA (Figure 2.4, Step 12).

(This protocol is apportioned for 5 photocrosslinking experiments done simultaneously.)

1. Resuspend paramagnetic beads (Dynabeads), by gently shaking storage vial. Transfer 10 μL aliquot to 1.5 mL siliconized polypropylene microcentrifuge tube. Spin briefly in mini-centrifuge. Place tube into magnetic particle concentrator, wait 30 s, and remove supernatant.
2. Resuspend beads in 100 μL 1× transcription buffer (40 mM HEPES, pH 6.5, 250 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 10 mM β-mercaptoethanol, 1x protease inhibitor cocktail, 8% glycerol, 50 μg/mL BSA), and mix gently using micropipette. Incubate 2 min at room temperature. Spin briefly in mini-centrifuge. Place tube in magnetic particle concentrator, wait 30 s, and remove supernatant. Repeat once.
3. Resuspend bead pellet in 75 μL 1× transcription buffer. Aliquot 15 μL of bead suspension into each of five siliconized microcentrifuge tubes.
4. Add 2 μL of biotin-labelled, site-specifically derivatized DNA fragment in low-EDTA TE (Section 2.3). Incubate 15 min at 25°C, mixing occasionally by gently tapping tube. Place tube into magnetic particle concentrator and wait 30 s. Transfer each supernatant to a new 1.5 mL siliconized polypropylene microcentrifuge tube. Estimate binding efficiency of biotin-labelled, site-specifically derivatized DNA fragment by comparing radioactivity of bead pellet and supernatant. This and all subsequent steps should be carried out under subdued lighting. Direct daylight

Figure 2.6. Immobilized DNA probe ("on-bead") photocrosslinking (Section 2.5.2). Diagram of "on-bead" photocrosslinking experiments of the *Pf* PIC. The bead-immobilized DNA probe allows the PIC to be "washed" at different stages while it is assembled on the promoter DNA. Once the promoter is attached to the beads, recombinant *Pf* TBP is added to the solution and incubated for 10 min at 70°C. Subsequently, recombinant *Pf* TFB (or recombinant *Pf* TFB plus partially purified *Pf* RNAP) is added to the complex and incubated for 10 min at 70°C. Once the complex is formed, buffer containing heparin is added (2 min, 70°C) to compete with DNA for non-specific binding of proteins. After the heparin challenge buffer is removed, the "on-bead" complex is irradiated with UV light (365 nm) to induce photocrosslinking. The crosslinked complex is then digested with DNase I followed by a heat/SDS treatment to further denature the complex. The solution is digested a second time with S1 nuclease and then loaded onto an Tris-glycine SDS-PAGE (Bio-Rad) and electrophoresed. The gel is dried, exposed to a phosphor screen (Molecular Dynamics), and then visualized with a phosphorimager. Multiple photocrosslinking experiments can be visualized on a single gel.


^{32}P | PhN$_3$
Dyna-bead
wash
10 min, 70°C
Recomb. TBP
TBP
Dyna-bead
10 min, 70°C
Recomb. TFB
TBP
TFB
Dyna-bead
2 min, 70°C
Heparin challenge
wash
hυ
TBP
TFB
Dyna-bead
DNase I
10 min, 37°C
Dyna-bead
TBP
TFB
5 min, 85°C
Heat/SDS
Dyna-bead
10 min, 37°C
S1 Nuclease
Pf TFB
Pf TBP

must be avoided. Low to moderate levels of incandescent light (e.g., from single task lamp with 60 W tungsten bulb) are acceptable.

5. Resuspend bead pellets in 20 μL of 1× transcription buffer. Spin briefly in mini-centrifuge. Place tube into magnetic particle concentrator, wait 30 s, and remove supernatant.
6. Remove a single aliquot (2 nM, 200 μL) of *Pf* TBP and place on ice to thaw.
7. Incubate *Pf* TBP at 70°C for 5 min. Steps 6 and 7 can be done simultaneously with Steps 4 and 5.
8. Resuspend each bead pellet from step 5 with 20 μL of the pre-incubated *Pf* TBP and immediately spin/mix the resuspended beads using minicentrifuge. Incubate for 10 min at 70°C. Mix occasionally by gently tapping tube (it is recommended that the tubes be lifted out of the heat block, tapped lightly against the heat block, and then quickly returned, to maintain the 70°C incubating temperature).
9. When the *Pf* TBP-DNA-bead solution has incubated ~4 min, remove the aliquot(s) of *Pf* TFB or *Pf* TFB and *Pf* RNAP from -80°C and place on ice to thaw.
10. Dilute *Pf*TFB and *Pf* RNAP to a final concentration of 2 nM and 4 ng/μL, respectively, in a total volume of 150 μL. (When only *Pf* TFB-TBP-DNA complex is being analyzed, omit *Pf* RNAP.)
11. When *Pf* TBP has incubated for ~9 min, place the diluted solution of *Pf* TFB + *Pf* RNAP at 70°C for 30 to 40 s (this should provide just enough time to open the 5 microcentrifuge tubes from step 8).
12. Quickly add 20 μL of *Pf* TFB + *Pf* RNAP solution to each of the 5 tubes from step 8 and quickly close the caps (this is done without removing the siliconized

microcentrifuge tubes from the heat block). Gently tap the tubes as suggested in step 8 and incubate the bead solution at 70°C for 10 min. Mix occasionally during incubation as before in step 8.

13. While the *Pf* RNAP + TFB + TBP + DNA-beads solution is incubating, place ~300 μL of 1× transcription buffer in a single micocentrifuge tube. In a second tube place 200 μL of trancription buffer containing 50 μg/mL heparin. Incubate both tubes at 70°C until needed (these two tubes may be prepared ahead of time but do not need to be incubated at 70°C until this step).
14. Place tubes containing the bead solution into magnetic particle concentrator (which is in a water bath at 70°C), wait 10 s, and remove supernatant.
15. Resuspend beads in 25μL of 1 × transcription buffer containing 50 μg/mL heparin.
16. Return the tubes to the heat block and mix gently as before. Incubate at 70°C for 2 min. Place tubes into magnetic particle concentrator, wait 10 s, and remove supernatant.
17. Resuspend beads in 25 μL of 70°C equilibrated 1× transcription buffer. Place tubes into magnetic particle concentrator, wait 10 s, and remove supernatant.
18. Resuspend beads in 25 μL of 70°C equilibrated 1× transcription buffer. Return the tubes to the heat block and mix gently as before.
19. Place microcentrifuge tubes inside 13 × 100-mm borosilicate glass culture tube filled with water equilibrated to 70°C (in second heat block). Place the metal heat block containing the culture tubes and siliconized microcentrifuge tubes in Fisher Scientific UV box.

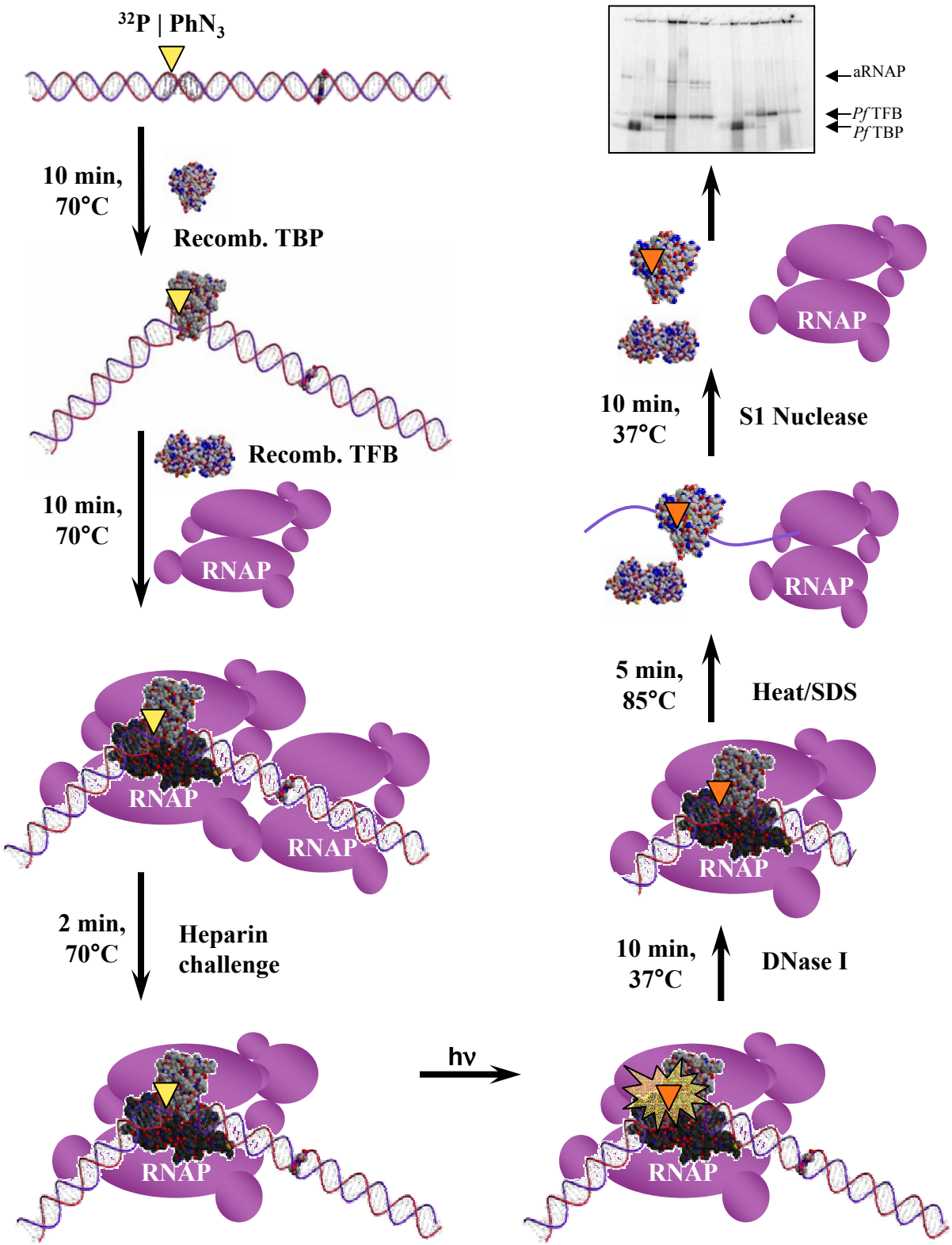
20. UV-irradiate 20 s (11 mJ/mm^2 at 350 nm). Immediately proceed to nuclease digestion (Section 2.6.1).

2.5.3. Solution photocrosslinking (Figure 2.7)

(This protocol is apportioned for 10-12 photocrosslinking experiments done simultaneously)

1. Place 2 μL of each derivatized DNA fragment probe from Section 2.3 into a siliconized microcentrifuge tube.
2. Remove a single aliquot (2 nM, 200 μL) of *Pf* TBP and place on ice to thaw.
3. Incubate *Pf* TBP at 70°C for 5 min.
4. Add 18 μL of *Pf* TBP to each tube containing the derivatized DNA fragment probe.
5. Incubate for 10 min at 70°C
6. When the *Pf* TBP-DNA solution has incubated ~ 4 min, remove the aliquot(s) of *Pf*TFB or *Pf*TFB and *Pf* RNAP from -80°C and place on ice to thaw.
7. Dilute *Pf*TFB and *Pf* RNAP to a final concentration of 2 nM and 4 ng/μL, respectively, in a total volume of 150 μL. (When only *Pf*TFB-*Pf* TBP-DNA complex is being analyzed omit *Pf* RNAP.)
8. When *Pf* TBP has incubated for ~9 min, place the diluted solution of *Pf*TFB + *Pf*RNAP at 70°C for 30 to 40 s (this should provide just enough time to open the 5 microcentrifuge tubes from Step 5).
9. Quickly add 20 μL of *Pf*TFB + *Pf* RNAP solution to each of the 5 tubes from step 8 and quickly close the caps (this is done without taking the siliconized

Figure 2.7 Solution photocrosslinking (Section 2.5.3). Diagram of solution crosslinking of *Pf* PIC. The derivatized DNA promoter fragment is combined in solution with recombinant *Pf* TBP and incubated for 10 min at 70°C. Subsequently partially purified *Pf* RNAP and recombinant *Pf* TFB (or just recombinant *Pf* TFB) are added to the solution and incubated for 10 min at 70°C. A heparin solution is added to compete for protein binding non-specifically (2 min, 70°C) . The solution is then UV irradiated (365 nm) to induce photocrosslinking. The derivatized DNA promoter fragment is then digested by the addition of DNase I followed by heat/SDS treatment and a second nucleic acid digestion with S1 nuclease. The solution is then loaded onto Tris-glycine SDS-PAGE (Bio-Rad) and electrophoresed. The gel is then dried, exposed to a phosphor screen, and then visualized with a phorphorimager. Multiple photocrosslinking experiments can be visualized on a single gel.


^{32}P | PhN$_3$
10 min, 70°C
Recomb. TBP
Recomb. TFB
10 min, 70°C
RNAP
2 min, 70°C
Heparin challenge
RNAP
RNAP
hν
RNAP
RNAP
10 min, 37°C
DNase I
RNAP
5 min, 85°C
Heat/SDS
RNAP
10 min, 37°C
S1 Nuclease
RNAP
aRNAP
Pf TFB
Pf TBP

microcentrifuge tubes from the heat block). Gently tap the tubes as suggested in step 8 and incubate the bead solution at 70°C for 10 minutes.

10. Add 3 μL of 70°C equilibrated heparin solution (583 μg/mL stock solution). Incubate at 70°C for 2 minutes.
11. Place polystyrene tube inside 13 × 100 mm borosilicate glass culture tube filled with water preincubated at desired temperature (in second heat block). Place the metal heat block containing the culture tubes and siliconized microcentrifuge tubes in Fisher Scientific UV box.
12. UV-irradiate 20 s (11 mJ/mm^2 at 350 nm). Immediately proceed to nuclease digestion (Section 2.6.1).

2.6. Photocrosslinking analysis

2.6.1. Nuclease Digestion

1. For immobilized DNA fragment photocrosslinking: To 25 μL of bead suspension (Section 2.5.2, Step 20), add 1.25 μL 100 mM $CaCl_2$ and 1 μL (10 units) DNase I.

   For solution crosslinking: To 40 μl of solution (Section 2.5.3, Step 12) add 2 μL 100 mM $CaCl_2$, 2 μL (10 units) DNase I.

   Incubate 15 min at 37°C. Terminate reaction by adding 2 μL 20% SDS and heating 5 min at 70°C.

2. Remove 15 μL of bead suspension or protein-DNA solutions and place in a new siliconized polypropylene microcentrifuge tube. After allowing sample to cool to room temperature, add 1.5 μL 25 mM $ZnCl_2$, 1.5 μL 1 M acetic acid, and 1 μL (30

units) S1 nuclease. Incubate 15 min at 37°C. Terminate reaction by adding 8 µL 4× SDS loading buffer and heating 3 min at 70°C.

2.6.2. Identification of crosslinked protein(s)

1. Apply entire sample (Section 2.6.1, Step 2), to polyacrylamide (37.5:1 acrylamide:bisacrylamide, Criterion gel, Bio-Rad) slab gel. As marker, load into adjacent lane 5 µL prestained protein molecular weight markers (Bio-Rad). Electrophorese in SDS running buffer at 18 V/cm for 100 min.
2. Dry gel. [Gels for the experiments reported here were dried using the Hoefer SE 1200 Easy Breeze air gel drying system between two precut cellophane sheets (Pharmacia Biotech). After electrophoresing, gels were immediately placed between the two cellophane sheets within the gel drying frames, taking care to remove air bubbles between the sheets. Gels were dried for 2 h in the air dryer with regular turning to prevent cracking (approximately every 10 minutes). Alternatively gels were dried between the cellophane sheets within the gel drying frame on the benchtop overnight). Gels were never equilibrated in supplier's suggested buffers since the gels were not to be kept (only dried and exposed in the phosphor screen).]
3. PhosphorImage. [Gels for the experiments reported here were exposed in a 35 × 43 cm phosphor screen (Molecular Dynamics) for 14-24 hours. Only completely dry gels were exposed. Exposed phosphor screens were then visualized.]

CHAPTER 3

Pyrococcus furiosus PHOTOCROSSLINKING RESULTS

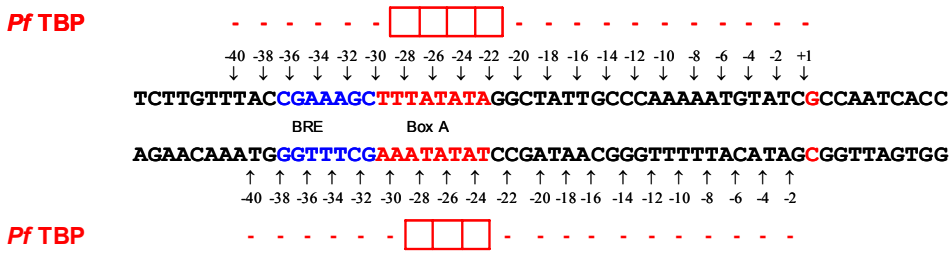
3.1. Overview of photocrosslinking results of *Pyrococcus furiosus* preinitiation complex and sub-complexes

Photocrosslinking analysis of protein-DNA interactions within the transcription preinitiation complex (PIC) from *Pyrococcus furiosus* (*Pf*) was performed utilizing 41 uniquely derivatized promoter fragments (Chapter 2, Table 2.1) which all contained the -75 → +45 segment of the *Pf* glutamate dehydrogenase gene (gdh). The position of azido-phenacyl (AP) incorporation into the phosphate backbone was defined as being either a nontemplate (NT) or template (T) strand probe and then numbered as the phosphate 5' to a given base. For example, the archaeal TATA box promoter begins at -30 with respect to the transcription start site. The derivatized promoter fragment with the AP incorporated at the phosphate between position -31 and -30 on the NT strand was designated NT-30. T-30 corresponds to the probe with AP incorporated between bases -30 and -29 on the template strand. Results of protein-DNA photocrosslinking in archaeal *Pf* TBP-DNA, *Pf* TFB-TBP-DNA, and *Pf* RNAP-TFB-TBP-DNA complexes are summarized in Figure 3.1. Detailed gel results are provided in Appendix A.1, A.2, and A.3 and are referenced in the discussion below. Crosslinked polypeptides were identified based on molecular weight in gel elution and in some cases substitution of protein derivatives (*Pf* TFBHisTAG) in the complex. Three DNA promoter positions

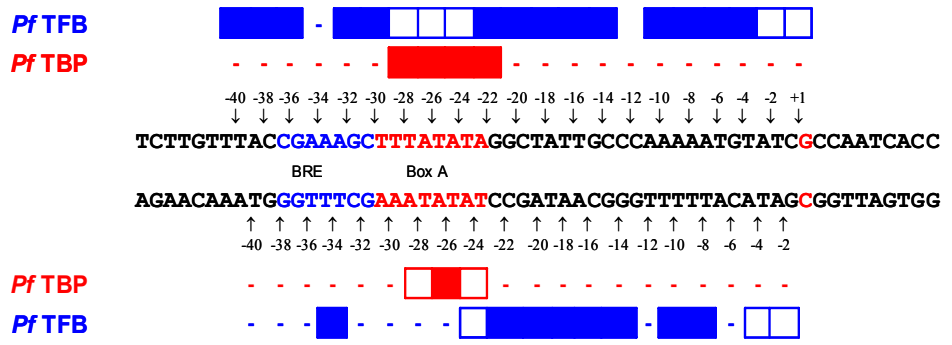
Figure 3.1. Summary of archaeal preinitiation complex protein-DNA photocrosslinking (results for NT strand above sequence; results for T strand beneath sequence). **A**. Summary of results from *Pf* TBP alone on promoter DNA. **B.** Summary of results from *Pf* TFB-TBP bound to promoter DNA. **C.** Summary of results from *Pf* RNAP-TFB-TBP bound to promoter DNA. Sites exhibiting reproducible crosslinking are indicated by solid boxes; sites exhibiting less reproducible crosslinking are indicated by unshaded boxes (scoring based on data from 2-5 independent standard photocrosslinking experiments). The BRE, Box A element, and transcription start site are indicated by blue and red letters, respectively within the promoter sequence.

Figure 3.1

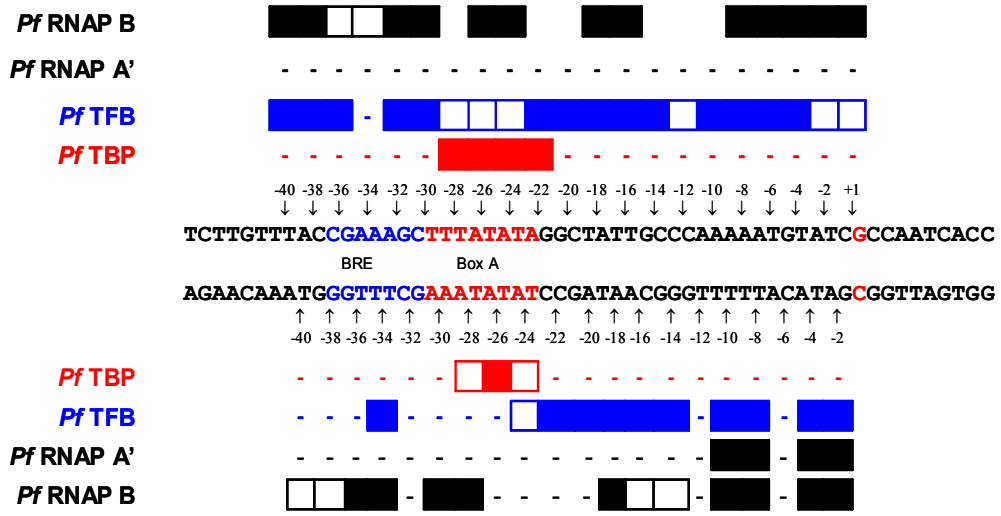
A.



B.



C.



were chosen outside the range of expected protein-DNA interactions within the archaeal PIC to serve as control positions for photocrosslinking. These 3 positions (NT-70 and NT-66 upstream of the promoter and T+45 downstream of the promoter, numbering based on the transcription start site, +1) were chosen based on existing knowledge of protein-DNA interactions in both the archaeal and eukaryal PIC's (Hausner and Thomm, 2001; Hausner et al., 1996; Kim et al., 1997). Protein-DNA photocrosslinking did not occur at these three control positions in all complexes examined by photocrosslinking reported here. (Appendices A.1, A.2, and A.3). Within the probed promoter region, a total of four different polypeptides were identified to be photocrosslinked by the derivatized promoter fragments. These were *Pf* TBP, *Pf* TFB, *Pf* RNAP subunit A' and *Pf* RNAP subunit B (Appendix A.1, A.2, and A.3). *Pf* TBP-DNA crosslinking occurred at the archaeal Box A promoter element. *Pf* TFB-DNA crosslinking occurred at positions both upstream and downstream of the TATA box. *Pf* RNAP subunit B-DNA crosslinking occurred at various positions across the entire probed region (-40 → +1) of the *Pf* gdh promoter fragment. Crosslinking to *Pf* RNAP subunit A' occurred at positions near the transcription start site. While *Pf* TBP-DNA photocrosslinks remained the same in all three complexes, the addition of *Pf* RNAP to the *Pf* TFB-TBP-DNA complex did alter the *Pf* TFB-DNA pattern of crosslinks. In the rest of this chapter, I will discuss the crosslinked positions for each identified protein.

3.2. TBP crosslinks

Pf TBP-DNA interactions were analyzed in three separate complexes (*Pf* TBP-DNA, *Pf* TFB-TBP-DNA, and *Pf* RNAP-TFB-TBP-DNA). *Pf* TBP-DNA crosslinking

occurred at 7 of the DNA fragment probes tested in each of the complexes. With *Pf* TBP alone, crosslinking occurred at positions NT-28, -26, -24, and -22 and positions T-30, -28, and, -26 (TATA box -30→-23, Appendix A.1, A, B, C, and D). Given the physical limitations of the AP covalently attached to the DNA phosphate backbone, *Pf* TBP residues are within ca. 11 Å of these phosphates in each complex in solution. These results are in agreement with other studies of *Pf* TBP-DNA interactions. DNase I footprint assays of the *Pf* TBP-DNA complex protected 15 nucleotides centered on the TATA box promoter (-34 → -20) of the same *Pf* gdh promoter region (Hausner and Thomm, 2001; Hausner et al., 1996). Similar results have been reported in DNase I footprinting experiments of TBP-DNA complexes in *Sulfolobus* and *Methanococcus* (Hausner and Thomm, 2001; Qureshi and Jackson, 1998). Also, *Pf* TBP-DNA crosslinks mimic those seen in eukaryal photocrosslinking analysis of eukaryal transcription components by the Ebright and Reinberg labs where TBPc crosslinked to *Adenovirus* major late promoter (AdMLP) DNA exclusively at or within 1 bp of the TATA box (Lagrange et al., 1996).

The addition of *Pf* TFB to produce the *Pf* TFB-TBP-DNA ternary complex did not change the positions where *Pf* TBP-DNA crosslinking occurred. There was a significant change in the *Pf* TBP binding (Appendix A.1 and A.2). Crosslinking of the *Pf* TBP-DNA complex with a *Pf* TBP concentration of 1 nM produced only low levels of crosslinked *Pf* TBP (Figure A.1, B). Increasing the *Pf* TBP concentration to 2 nM modestly increased the amount of tagged *Pf* TBP (Figure A.1, A). When *Pf* TFB (1 nM) was included, 1 nM *Pf* TBP was sufficient to produce a much stronger signal of "tagged" *Pf* TBP (Appendix A.2, B and C). This suggests *Pf* TFB aids in *Pf* TBP-DNA binding.

These results are consistent with gel mobility shift assays conducted in the Scott lab in which *Pf* TBP-DNA alone does not result in shift. When *Pf* TFB is added, a supershift of probe DNA is seen that is dependent on both *Pf* TFB and *Pf* TBP concentration (Chen, 2000).

The only crystal structures of archaeal TBP-DNA is of *Pyrococcus woesei* (*Pw*) TFBc-TBP-DNA. In both structures, archaeal TBP binds in the minor groove of the 8 bp TATA element exclusively (Littlefield et al., 1999; Patikoglou et al., 1999). The agreement of photocrosslinking results for *Pf* TBP-DNA interactions with existing structural information suggests an accurate PIC was formed *in vitro*.

3.3. *Pf* TFB crosslinks

Pf TFB-DNA crosslinking occurred both upstream and downstream of the Box A promoter. Tagged *Pf* TFB was identified by either its molecular weight or by substitution of *Pf* TFBHisTAG in the photocrosslinking experiment. When *Pf* TFBHisTAG was included in the experiment, adjacent lanes of the same crosslinking experiment containing *Pf* TFB or *Pf* TFBHisTAG were loaded onto SDS-PAGE (Section 2.6.2). The higher molecular weight of *Pf* TFBHisTAG "shifts" the position of crosslinked polypeptide (if it is *Pf* TFB that is crosslinked). In the *Pf* TFB-TBP-DNA complex, crosslinking occurred at positions NT-40, -38, 36, -32, -30, -28, -26, -24, -22, -20, -18, -16, -14, -10, -6, and -4 (Appendix A.2, B) and at positions T-34, -22, -20, -18, -16, -14, -10, and -8 and to a lesser extent T-4 and T-2 (Appendix A.2, C). Addition of *Pf* RNAP resulted in no significant change in the NT TFB-DNA crosslinks (with the exception of position NT-12 where crosslinking of *Pf* TFB did occur to some extent (Figure 3.1 and

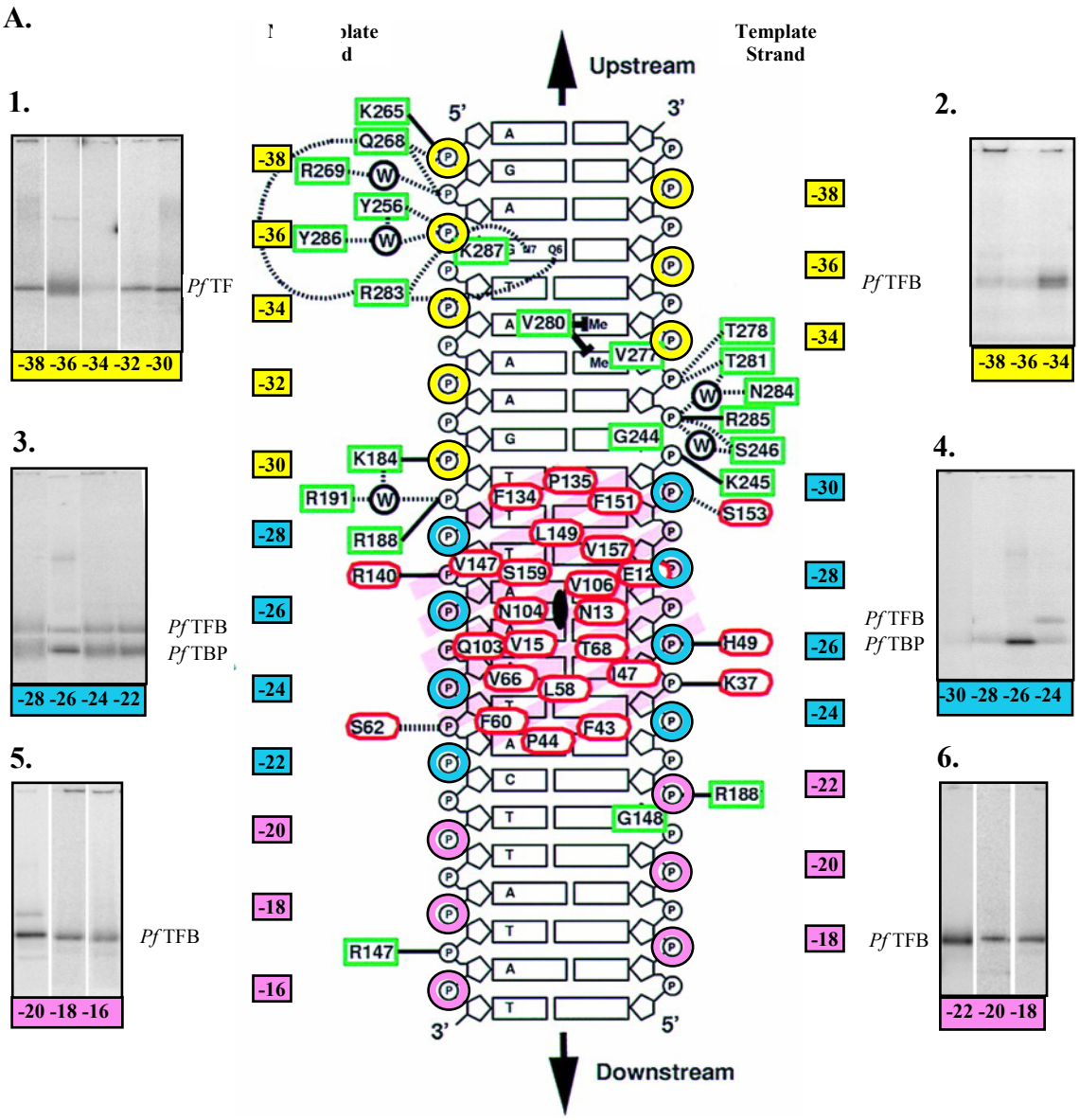
Appendix A.3, A and B)). On the T strand, crosslinking was noticeably enhanced at positions T-4 and -2 (compare Appendix A.2, C with A.3, C). Once again, given the physical limitations of the AP covalently attached to the DNA phosphate backbone, *Pf* TFB residues are within approximately 11 Å of these phosphates in each complex in solution.

3.3.1. Comparison of *Pf* TFB-DNA photocrosslinking to *Pyrococcus woesei* TFBc-TBP-DNA crystal structure.

Comparison of *Pf* TFB-DNA crosslinks to crytallographic data is limited for two reasons. First, the *Pyrococcus wosei* (*Pw*) TFBc-TBP-DNA crystal structures only contain the C-terminal 200 residues of *Pw* TFB that interact with the upstream BRE and *Pw* TBP. Secondly, the DNA in the crystal structure was 24 bp in length (-39 → -16). Still, the structures are helpful in analyzing archaeal TFB-DNA interactions within this region. Figure 3.2 is a schematic of TBP-DNA and TFB-DNA contacts in the *Pw* TFBc-TBP-DNA crystal structure (Littlefield et al., 1999) along with representative gels of photocrosslinking analysis for the portion of the DNA promoter included in the crystal structure. *Pw* TFBc-DNA interactions (green squares) upstream of the Box A promoter are dominated by the interaction of TFBc helix BH5' with the major groove of DNA in the BRE (Littlefield et al., 1999). The pattern of *Pf* TFB-DNA photocrosslinking upstream of the TATA box correlates well with the determined structure (Figure 3.2.A gels 1 and 2 with yellow highlighted phosphates). *Pw* TBP-DNA interactions (red ovals) dominate the Box A region (-30 → -23). Likewise, *Pf* TBP-DNA photocrosslinking occurs exclusively at phosphates within or immediately adjacent to the Box A element

Figure 3.2. Correlation between site-specific photocrosslinking and crystal structure of archaeal *Pw* TFBc-TBP-DNA complex. A. Schematic of *Pw* TFBc-DNA (green squares) and *Pw* TBP-DNA interactions (red ovals, Littlefield et al., 1999) compared with *Pf* TFB-TBP-DNA photocrosslinking analysis (4-20% SDS-PAGE, Biorad) of the regions upstream of the Box A element (gels 1 and 2 with yellow highlighted phosphates), within the Box A element (gels 3 and 4 with blue highlighted phosphates), and downstream of the Box A element (gels 5 and 6 with magenta highlighted phosphates). Photocrosslinking results correspond to summary in Figure 3.1 (B). **B.** Archaeal promoter DNA in *Pw* TFB-TBP-DNA ternary crystal structure. **C.** Corresponding portion of archaeal promoter fragment used in *Pf* TFB-TBP-DNA site-specific photocrosslinking analysis.

Figure 3.2



B. 5'- **AGAGTAAGTTTAAATACTTATAT** -3' (NT strand)

C. 5'- **CCGAAAGCTTTATATAGGCTATT** -3' (NT strand)

(Figure 3.2.A gels 3 and 4 with blue highlighted phosphates). In contrast, all NT photocrosslinking positions (-28, -26, and -24) and one T position (-24) within the Box A element crosslink *Pf* TFB as well. This discrepancy could be attributed to the fact that photocrosslinking experiments were done *in vitro* at 70°C, which allows for more molecular motion and flexibility than the crystal structure lattice.

In the ternary crystal structure, there are only a few *Pw* TFBc-DNA interactions downstream of the Box A promoter (-22 → -16). In contrast, there are several NT positions (-20, -18, and -16) and T positions (-22, -20, and -18) downstream of the Box A element that crosslink *Pf* TFB (Figure 3.2.A gels 5 and 6 with magenta highlighted phosphates). *Pf* TFB-DNA crosslinks at positions NT-18 and -16 and at positions T-22 and -20 can be justified by the increased flexibility of the proteins at high temperatures *in vitro*. The additional crosslinking at positions NT-20 and -22 as well as T-18 and -16 would indicate a higher level of *Pf* TFB-DNA interaction than the Littlefield, et al., crystal structure shows. Additionally, several positions in preliminary photocrosslinking experiments (NT-20, -18 and -16 as well as positions T-22 and -20) produced an intense tagged-*Pf* TFB double band pattern (on SDS-PAGE) that could not be digested to a single band with a cocktail of DNase I, micrococcal, and Omnicleave nucleases (Appendices A.2.A and A.3.A). Substitution of *Pf* TFBHisTAG confirmed that both tagged protein-bands were *Pf* TFB. Only digestion with DNase I, subsequent heat treatment with 1% SDS, and a final digestion with S1 nuclease could resolve the tagged-*Pf* TFB double-band pattern into a single band. Again, these results would suggest closer contact of *Pf* TFB with DNA downstream of the Box A than the *Pw* ternary crystal structure suggests.

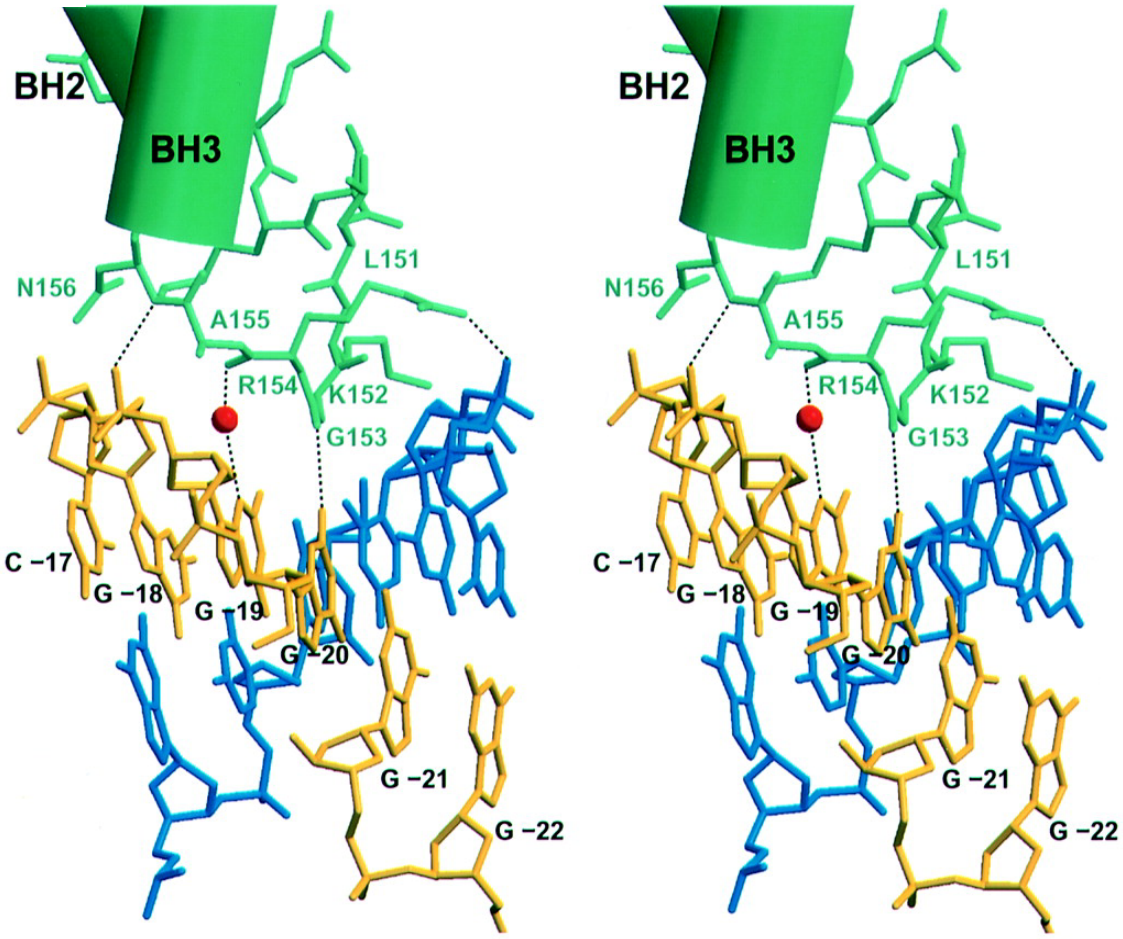
In 2000, the Sigler group reported a "revised" eukaryotic crystal structure of the TFIIBc-TBPc-DNA ternary complex. Unlike the original structure, this complex included human TFIIBc and human TBPc as well as a longer piece of DNA containing 3 additional bp downstream and 7 additional bp upstream of the TATA box (Tsai and Sigler, 2000). While the general structure remained the same, the longer DNA produced a crystal lattice where the DNA packed end-to-end in perfect helical register. This extension of the DNA helix revealed that in addition to contacts with the phosphate backbone TFIIBc makes base specific contacts with the minor groove immediately downstream of the TATA box through a loop that links helices BH2 and BH3 (Figure 3.3). This minor groove interaction was initially proposed by Ebright and Reinberg based on RNAP II-TFIIF-TFIIB-TBP-DNA photocrosslinking results (Kim et al., 1997). A simple comparison of the eukaryotic ternary complex protein-DNA contact schematic with that of the archaeal ternary complex schematic (Figure 3.4) suggests a similar *Pf* TFB-minor groove interaction would exist given the correct geometry of the DNA. Also, Tsai and Sigler's examination of the minor groove recognition loop showed a high homology among TF(II)B's from several species including archaea (Figure 3.4.B).

3.3.2. Comparison of *Pf* TFB-DNA photocrosslinking with other structural work

Other structural characterizations of archaeal and eukaryal preinitiation complexes are useful for comparing *Pf* TFB-DNA photocrosslinking. As mentioned above (Section 3.2), DNase I footprint assays of the *Pf* TBP-DNA complex protected 15 nucleotides centered on the TATA box promoter (-34 → -20) of the *Pf* gdh promoter (Hausner and Thomm, 2001; Hausner et al., 1996). The addition *Pf* TFB increased the

Figure 3.3. (Originally from Tsai and Sigler, 2000) A. Stereo view depicting the molecular interactions between the minor groove recognition-loop of hTFIIB (residues 152-156 of hTFIIBc) and adenovirus Major Late Promoter (adMLP) immediately downstream of the TATA box. Nontemplate strand residues are labeled (yellow). **B.** Alignment of the minor groove recognition-loop sequences of eukaryal and archaeal TF(II)Bs. Green highlighting indicates conserved, and yellow highly similar residues. An asterisk indicates a residue in hTFIIBc that contacts the minor groove in the human TFIIBc-TBPc-DNA crystal structure (Tsai and Sigler, 2000).

A.


BH2
BH3
N156
L151
A155
R154
K152
G153
C −17
G −18
G −19
G −20
G −21
G −22
BH2
BH3
N156
L151
A155
R154
K152
G153
C −17
G −18
G −19
G −20
G −21
G −22

B.

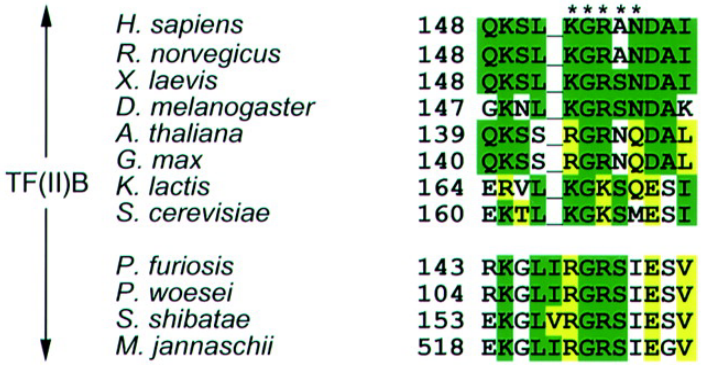
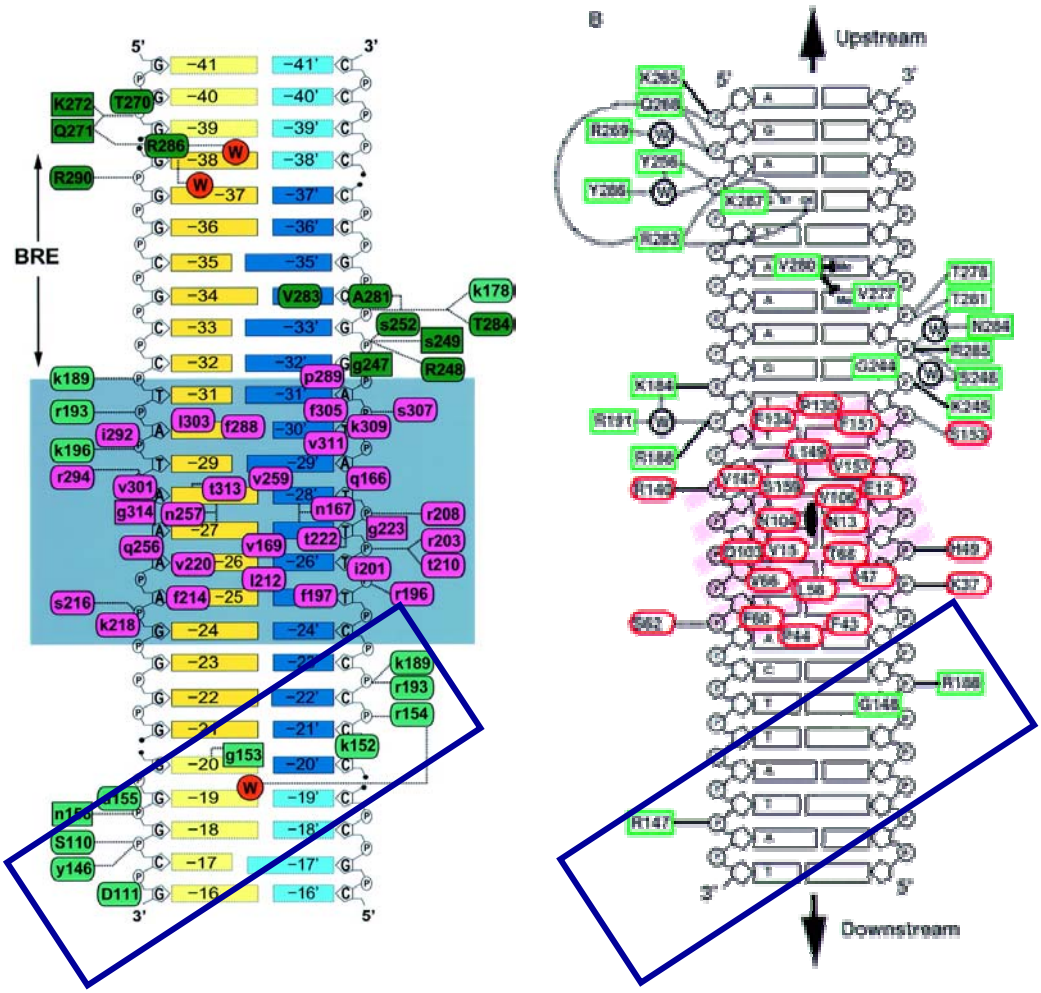
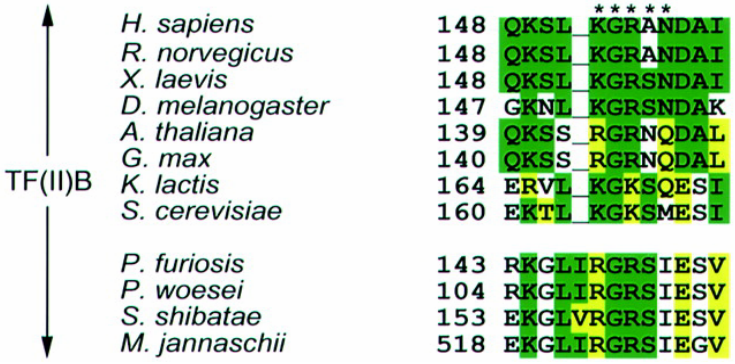

TF(II)B
H. sapiens 148 QKSL_KGRANDAI
R. norvegicus 148 QKSL_KGRANDAI
X. laevis 148 QKSL_KGRSNDAI
D. melanogaster 147 GKNL_KGRSNDAK
A. thaliana 139 QKSS_RGRNQDAL
G. max 140 QKSS_RGRNQDAL
K. lactis 164 ERVL_KGKSQESI
S. cerevisiae 160 EKTL_KGKSMESI
P. furiosis 143 RKGLIRGRSIESV
P. woesei 104 RKGLIRGRSIESV
S. shibatae 153 EKGLVRGRSIESV
M. jannaschii 518 EKGLIRGRSIEGV

Figure 3.4. Comparison of Protein-DNA interactions within human and *Pyrococcus woesei* TF(II)Bc-TBP(c)-DNA crystal structures. A. Schematics of protein-DNA contacts in hTFIIBc-TBPc-DNA crystal structure (left, Tsai and Sigler, 2000) and *Pyrococcus woesei* (*Pw*) TFBc-TBP-DNA (right, Littlefield, et al., 1999). Human TFIIBc-DNA and *Pw* TFBc-DNA contacts are represented by green ovals and green rectangles, respectively. Human TBPc-DNA and *Pw* TBP-DNA contacts are represented by magenta ovals and red ovals, respectively. The nontemplate strand is on the left side of each sequence. The TATA box is shaded in either blue (human) or pink stripes (*Pw*). The blue rectangle highlights the minor groove recognition-loop contacts (Figure 3.3) in the human schematic as well as where a same interaction might occur in *Pw*. *Pf* photocrosslinking results (reported here, Figures 3.1, Appendix A.2, B, and A.3, B) suggest the same TFB-DNA minor groove interaction exists in archaea although not seen in *Pw* crystal structure. **B.** Alignment of minor groove recognition loop as in Figure 3.4.B.

A.


5'
3'
BRE
W
Upstream
Downstream

B.

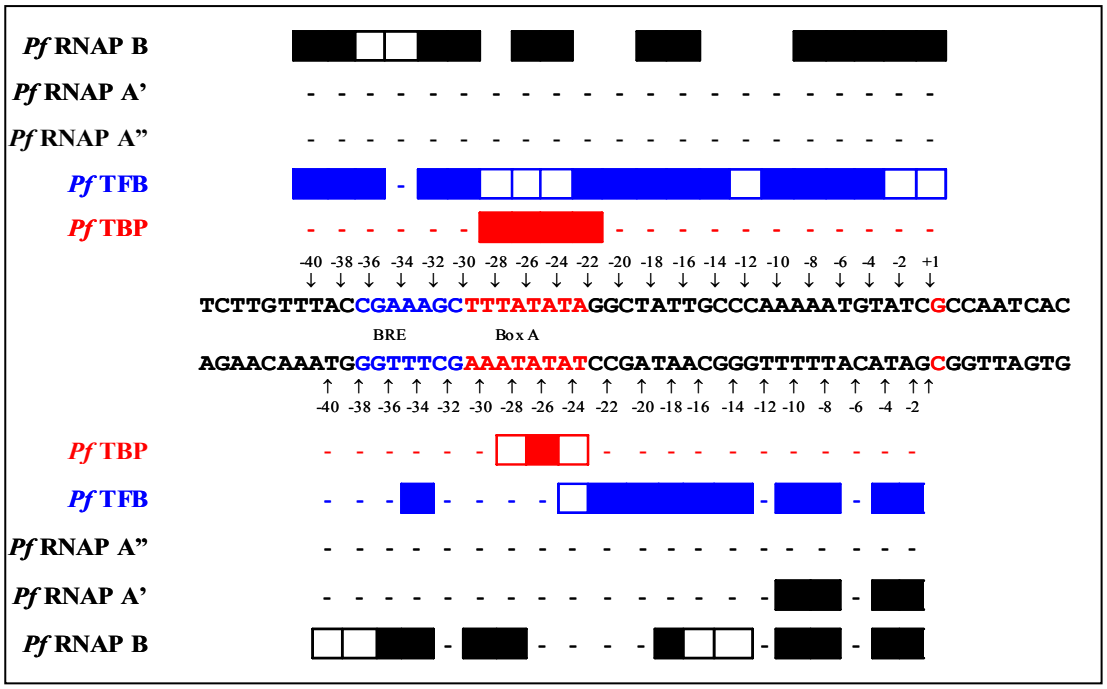

TF(II)B
H. sapiens 148 QKSL_KGRANDAI
R. norvegicus 148 QKSL_KGRANDAI
X. laevis 148 QKSL_KGRSNDAI
D. melanogaster 147 GKNL_KGRSNDAK
A. thaliana 139 QKSS_RGRNQDAL
G. max 140 QKSS_RGRNQDAL
K. lactis 164 ERVL_KGKSQESI
S. cerevisiae 160 EKTL_KGKSMESI
P. furiosis 143 RKGLIRGRSIESV
P. woesei 104 RKGLIRGRSIESV
S. shibatae 153 EKGLVRGRSIESV
M. jannaschii 518 EKGLIRGRSIEGV

protection from DNase I digestion from -42 → -19 around the TATA box (Hethke et al., 1996). In similar experiments with the *Methanococcus* TFB-TBP-DNA complex, positions -40 → -14 were protected from DNase I digestion (Hausner and Thomm, 2001). While the *Pf* TFB-TBP-DNA photocrosslinking data suggest interaction of *Pf* TFB from -40 → -14 of the promoter DNA, there are other positions crosslinked farther downstream of the TATA box (-10 → -2, Figure 3.1, B and C) that are not reported for DNase I footprinting assays for the archaeal ternary complex.

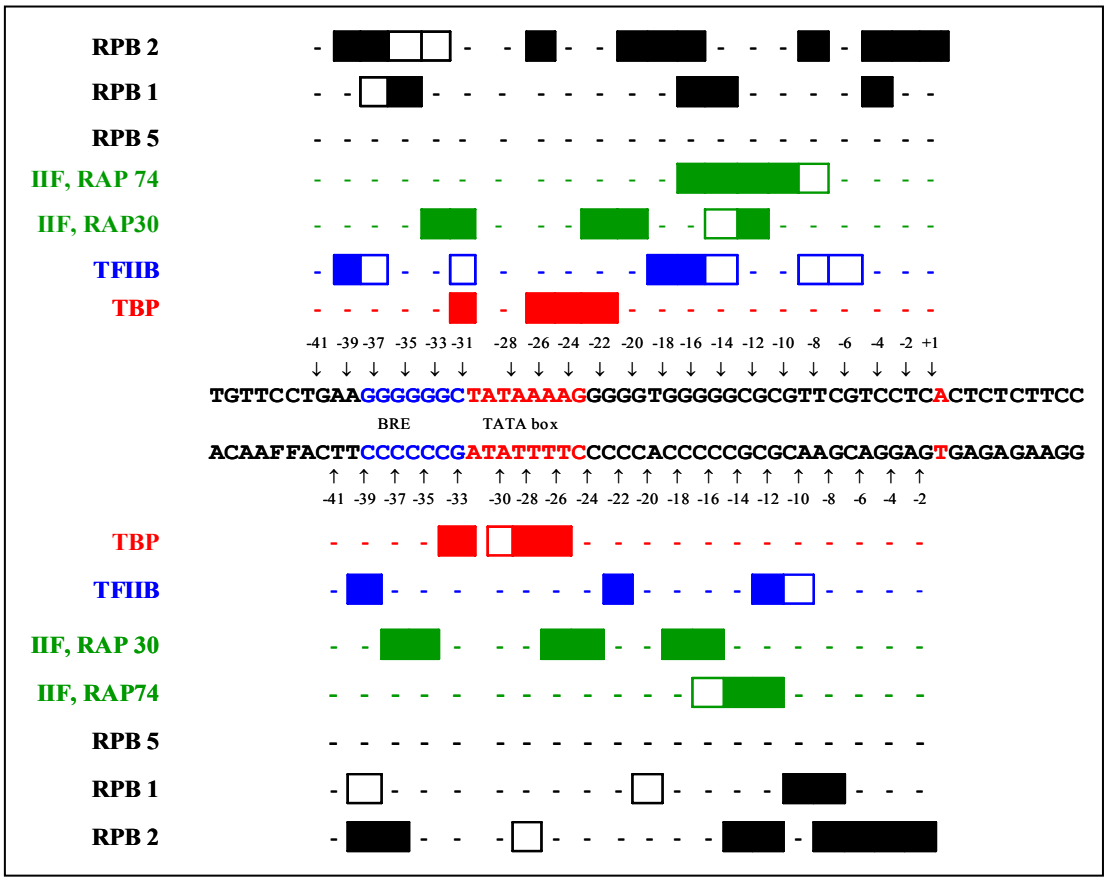
Site-specific photocrosslinking allows for a more extensive comparison of nontemplate (NT) and template (T) strand positions crosslinked in the respective systems. *Pf* TFB-DNA crosslinking occurs at additional positions compared to eukaryotic TFIIB-DNA-promoter crosslinking data. In eukaryotic photocrosslinking experiments, TFIIB-DNA-promoter crosslinks were found in distinct areas upstream and downstream of the TATA box with a greater percentage of crosslinking on the NT strand than the T strand. In the eukaryotic TFIIB-TBP-DNA complex, the TFIIB-DNA crosslinks encompassed a span of the DNA promoter 6 bp upstream to 12 bp downstream of the TATA box (-37 → -12) (Lagrange et al., 1996). In the RNAP II-TFIIF-TFIIB-TBP-DNA complex, the TFIIB-DNA crosslinks extended farther in both directions (upstream and downstream, -39 → -6, Figure 3.5.B, blue bars). Likewise in my experiments, *Pf* TFB-DNA crosslinks were found upstream and downstream of the TATA box. There is also a greater percentage of crosslinking on the NT strand than the T strand. The addition of *Pf* RNAP (forming the Pf RNAP-TFB-TBP-DNA complex) to the complex also extends the span of *Pf* TFB-DNA crosslinks compared to the *Pf* TFB-TBP-DNA complex (compare Appendix A.2, C with A.3, C).

Figure 3.5. Comparison of *Pyrococcus furiosus* and eukaryal protein-DNA photocrosslinking results. A. *Pyrococcus furiosus* (*Pf*) PIC photocrosslinking results from the studies reported here (nontemplate positions -40 → +1 and template positions -40 → -2 as indicated by arrows). *Pf* RNAP subunit-DNA crosslinking indicated by black bars. *Pf* TFB-DNA crosslinking indicated by blue bars. *Pf* TBP-DNA crosslinking indicated by red bars. **B.** Eukaryal PIC photocrosslinking results (for corresponding positions) as reported by Kim, et al., (2000). RNAP II subunit-DNA crosslinking indicated by black bars. TFIIF subunit (RAP 30 and RAP 74)-DNA crosslinking indicated by green bars. TFIIB-DNA crosslinking indicated by blue bars. TBP-DNA crosslinking indicated by redbars. In all cases, solid bars indicate positions where strong crosslinking occurred, unshaded bars indicate positions where weak crosslinking occurred, and dashes indicate no crosslinking detected. With the absence of a TFIIF homologue in the archaeal system, *Pf* TFB takes on a greater role in stabilizing the PIC by interacting with DNA more than seen with its eukaryal counterpart (discussed in section 3.4). TF(II)B-DNA interactions at positions downstream of -12 in both systems are discussed extensively in Section 3.5.

A.


Pf RNAP B
Pf RNAP A'
Pf RNAP A"
Pf TFB
Pf TBP
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 +1
TCTTGTTTACCGAAAGCTTTATATAGGCTATTGCCCAAAAATGTATCGCCAATCAC
BRE
Box A
AGAACAAATGGGTTTCGAAATATATCCGATAACGGGTTTTTACATAGCGGTTAGTG
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2
Pf TBP
Pf TFB
Pf RNAP A"
Pf RNAP A'
Pf RNAP B

B.


RPB 2
RPB 1
RPB 5
IIF, RAP 74
IIF, RAP30
TFIIB
TBP
-41 -39 -37 -35 -33 -31 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 +1
TGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCACTCTCTTCC
BRE
TATA box
ACAAFFACTTCCCCCCGATATTTTCCCCCACCCCCGCGCAAGCAGGAGTGAGAGAAGG
-41 -39 -37 -35 -33 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2
TBP
TFIIB
IIF, RAP 30
IIF, RAP74
RPB 5
RPB 1
RPB 2

The major difference between the two sets of photocrosslinking data (archaeal PIC vs. eukaryal PIC) is the distance of the crosslinks downstream of the Box A/TATA box promoter. *Pf* TFB-DNA crosslinks in the *Pf* TFB-TBP-DNA complex extend from 10 bp upstream to 17 bp downstream of the Box A (-40 → -6). In the *Pf* RNAP-TFB-DNA complex, the *Pf* TFB-DNA crosslinking extended farther to 21 bp downstream of the Box A promoter (-40 → -2, Figure 3.5.A, blue bars). Reported eukaryotic PIC photocrosslinking results showed eukaryal TFIIB-DNA photocrosslinking no farther than position NT-6 in a RNAP II-TFIIF-TFIIB-TBP-DNA complex (Figure 3.5.B, blue bars (Kim et al., 1997)). There is a 1 bp difference in the positioning of the TBP-binding sites (eukaryotic TATA box starts at –31 where archaeal Box A starts at –30). Still, this does not account for archaeal TFB-DNA crosslinks found 4 bp farther downstream than those reported for TFIIB-DNA crosslinking (nearly at the +1 transcription start site).

3.4. *Pf* RNAP crosslinks

Pf RNAP-DNA promoter crosslinking was found at several positions on two subunits (Subunit A' and B). *Pf* RNAP subunit B-DNA crosslinking in the *Pf* RNAP-TFB-TBP-DNA complex occurred at positions NT-40, -38, -36, -34, -32, -30, -26, -24, -20, -18, -16, -6, -4, -2, and +1 (Appendix A.3, B) and at positions T-36, -34, -30, -28, -26, -22, -18, -10, -8, -4 and -2 (Appendix A.3, C). *Pf* RNAP subunit A'-DNA crosslinking occurred at positions T-8, -4, -2 and to a lesser extent at position T-10.

Existing information on archaeal RNAP-DNA promoter interactions comes primarily from two techniques: DNase I footprinting and site-specific photocrosslinking. Hausner and Thomm reported DNase I footprinting analysis of an archaeal PIC from

Methanococcus where RNAP, TFB, and TBP bound to promoter DNA produced a footprint which extended from -40 to +17 (Hausner and Thomm, 2001). Only the -13 to +17 portion of the footprint could be assigned to the addition of archaeal RNAP to the complex given that inclusion of archaeal TBP and TFB already produced a footprint from -40 to -14.

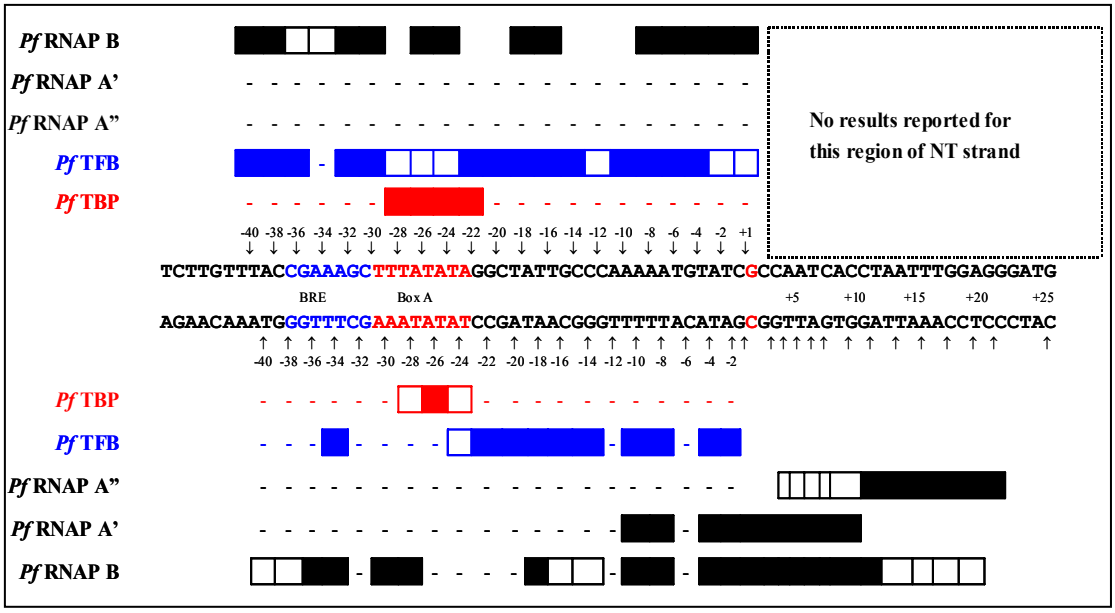
Bartlett, et al. (2000) reported site-specific photocrosslinking of the *Pf* PIC to the -5 to +24 region of the *Pf* gdh promoter (the same promoter used in the studies reported here, DNA fragment probes were of the template strand only). *Pf* RNAP subunit B was reported to produce strong crosslinking at positions T-5, -1, +2, +3, +4, +5, +6 and +8, and weaker crosslinking at positions T+10, +12, +14, +16, and +18. *Pf* RNAP subunit A' produced strong crosslinking at positions T-5, -1, +2, +3, and +4, and weaker crosslinking at positions T+5, +6, +8, +10, +12 and +14. *Pf* RNAP subunit A" produced crosslinking from positions T+2 to +20 with the strongest crosslinking occurring at positions T+8, +10, +12 and +14. The results reported by Bartlett, et al overlap a portion of the DNA promoter positions probed with photocrosslinking in my experiments. The results in this overlapped region (around the transcription start site -5 → -1) are in agreement with one another. The agreement between the two sets of data provides some interesting points of reference concerning *Pf* RNAP-DNA contacts. First, in my experiments, crosslinking to subunit A' was only seen at positions T-8, -6, -4, and -2. Likewise, the farthest upstream Bartlett, et al. reported crosslinking to subunit A' was position T-5. Secondly, in my experiments, crosslinking to subunit A" (-40 to -2) was not seen. Bartlett, et al reported crosslinking to subunit A" starting at T+2 and then farther downstream, beyond the location of my farthest downstream DNA fragment probe (T-2).

One point of contrast is the extent of subunit B crosslinking. Bartlett, et al., reported subunit B-DNA crosslinking as far upstream as position NT-28 but none at position NT-40 whereas my studies showed subunit B crosslinking at position NT-40. Combined, these results suggest RNAP-DNA interactions extend from at least -40 to +20 (60 bp) on the *Pf* gdh promoter.

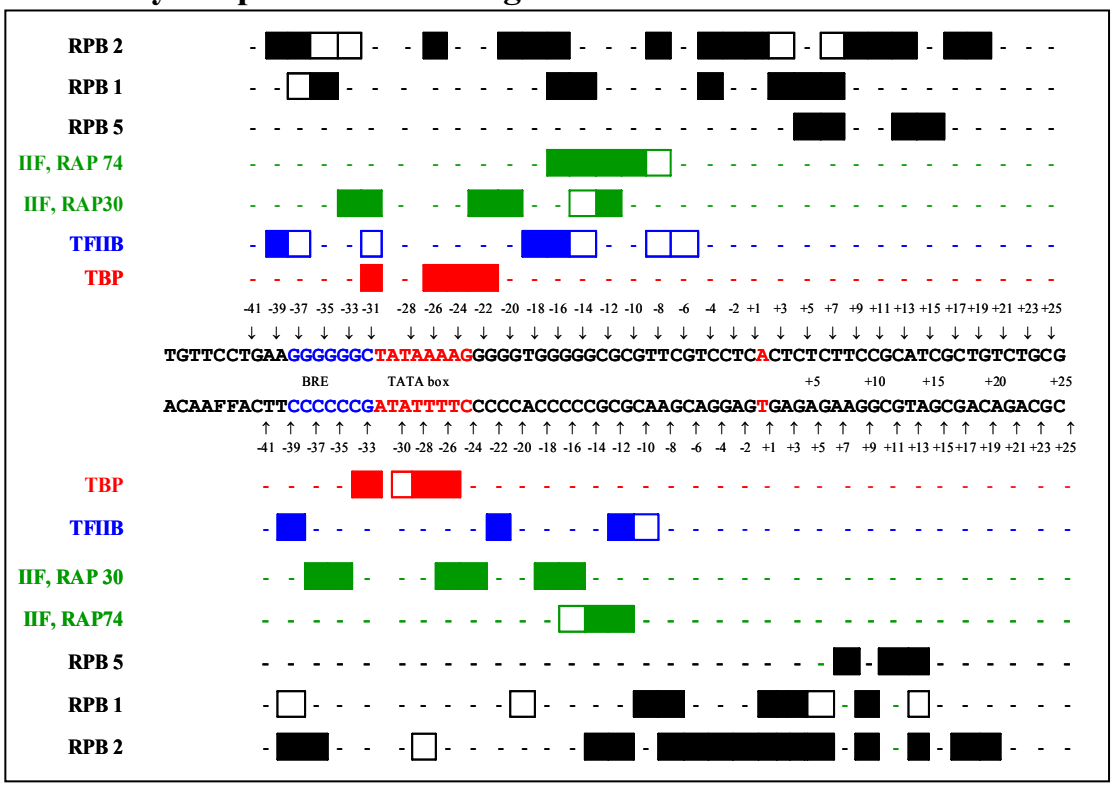
Figure 3.6 aligns the combined *Pf* photocrosslinking results with those reported by the Ebright and Reinberg groups for the eukaryotic PIC (Kim et al., 1997). Eukaryotic PIC photocrosslinking was reported from –55 → +25 of the AdMLP eukaryotic promoter however, Figure 3.6 only extends to position –40 upstream. Crosslinking to RNAP II subunit RPB1 (homologous to *Pf* RNAP subunits A' and A") was reported from -40 to +10 with the most reproducible crosslinks falling from -10 to +10 around the transcription start site. This coincides well with the combined *Pf* photocrosslinking results (Figure 3.6.A). RNAP II subunit RPB2 (homologous to *Pf* RNAP B) crosslinked DNA more extensively than RPB1 across the -40 to +20 region of the promoter with more extensive crosslinking reported for the NT strand than the T strand. Also, several of the crosslinked positions are within the -10 to +10 region of the promoter. Again this is in agreement with the combined *Pf* crosslinking results. My studies showed *Pf* RNAP subunit B crosslinking at more positions on the NT strand than T strand. Also the strongest crosslinking occurred within the -10 to +1 region of the promoter and Bartlett, et al., reported the strongest crosslinking of RNAP subunit B from positions -5 to +10. The parallels between these three sets of photocrossliking experiments provide further evidence of the similarity between the eukaryotic and archaeal transcription systems.

Figure 3.6. Comparison of *Pyrococcus furiosus* and eukaryal protein-DNA photocrosslinking results. A. Combined *Pyrococcus furiosus* (*Pf*) PIC photocrosslinking results from the studies reported here (nontemplate positions -40 → +1 and template positions -40 → -2 as indicated by arrows) and results reported by Bartlett, et al., (2000, template positions +2 → +25 as indicated by arrows). The color scheme is the same as in Figure 3.6.A. **B.** Eukaryal PIC photocrosslinking results as reported by Kim, et al., (2000). RNAP II subunit-DNA crosslinking indicated by black bars. The color scheme is the same as in Figure 3.5.B. In all cases, solid bars indicate positions where strong crosslinking occurred, unshaded bars indicate positions where weak crosslinking occurred, and dashes indicate no crosslinking detected.

A. Combined *Pyrococcus furiosus* photocrosslinking results



B. Eukaryotic photocrosslinking results



In contrast to eukaryotic crosslinking results, *Pf* RNAP subunit B crosslinked at nearly every derivatized position on the NT strand in my experiments and to some extent at all positions in the results reported by Bartlett, et al. A degree of variability in RNAP subunit-DNA contacts is reasonable considering it is not bound specifically to DNA and must proceed beyond initiation into the elongation phase of transcription. Also, the archaeal system lacks GTFs that are in the eukaryal system (TFIIF, TFIIH, and larger subunit of TFIIE) all of which bind promoter DNA. Most notably TFIIF binds within the -40 to -6 region of the adenovirus major late promoter (AdMLP, Figure 3.6.B green bars). In the absence of additional transcription factors in the archaeal system, binding and orienting DNA in the PIC to initiate transcription is determined by the RNAP, and existing GTFs (TBP and TFB).

3.5. Protein-DNA contacts not predicted by eukaryal crosslinking or structural work.

Comparison of *Pf* TFB-DNA crosslinking results (in the *Pf* TFB-TBP-DNA complex) with the *Pw* TFBc-TBP-DNA ternary complex crystal structure (Littlefield et al., 1999) would ascribe the positions where crosslinked *Pf* TFB can be identified as most likely being interactions of the C-terminal cyclin A folds of *Pf* TFB with DNA. Still there remain other *Pf* TFB-DNA crosslinked positions farther downstream from the TATA box that cannot logically be ascribed to the C-terminal domain's interaction with DNA. DNase I footprint analysis of archaeal transcription machinery on archaeal promoters have reported protection as far as 9 bp downstream from the TATA box upon addition of TFB to TBP-DNA in solution (Hausner and Thomm, 2001). Still, this does

not account for the extreme downstream positions of derivatized archaeal promoter DNA that crosslinked *Pf* TFB in a TFB-TBP-DNA complex in my work. Addition of RNAP does extend the DNase I footprint to position +16 on an archaeal promoter (Hausner and Thomm, 2001), but the addition of a large 12-subunit enzyme with high affinity for DNA would preclude identifying specific protein-DNA interactions. Finally, comparison of *Pf* TFB-DNA photocrosslinking to TFIIB-DNA crosslinking does provide some support for TFB-DNA interaction as far downstream as position NT-6 and T-10 when in an RNAP II-TFIIF-TFIIB-TBP-DNA complex (Figure 3.6.B blue bars). These contacts, although less reproducible than other TFIIB-DNA crosslinks, provide two interesting comparisons. First, both eukaryotic and archaeal crosslinking results show segments of TFIIB/TFB crosslinking (positions NT-18, -16, -14 in eukarya and NT-20, -18, -16 in archaea) that would appear to correspond to the minor groove binding immediately downstream of the TATA box reported by Tsai and Sigler (2000, Figure 3.4). Secondly, after a small gap where crosslinking does not occur, both sets of results have additional positions that crosslink TFIIB or TFB closer to the transcription start site (NT-8 and -6 on the AdMLP and positions NT-10, -8 and -6 and positions T-10, -8, -4 and -2 on an archaeal gdh promoter). Given the localized TFIIBc-DNA interactions in the TFIIBc-TBPc-DNA crystal structure (Tsai and Sigler, 2000), there are two possibilities to explain why TFIIB would crosslink to DNA promoter positions farther downstream: (1) DNA downstream of position -16 is somehow bent back towards the TATA box in the RNAP II-TFIIF-TFIIB-TBP complex to allow TFIIBc to interact with promoter DNA as far down as -6 on the NT strand; or (2) Some other part of TFIIB is in the vicinity of DNA near the transcription start site, namely the N-terminal domain residues of TFIIB.

The same two possibilities would seem to exist for *Pf* TFB-DNA interactions given the structural homology between the two systems and the fact that *Pf* TFB-DNA crosslinks occur even farther downstream on the template strand within 2 bp of the transcription start site.

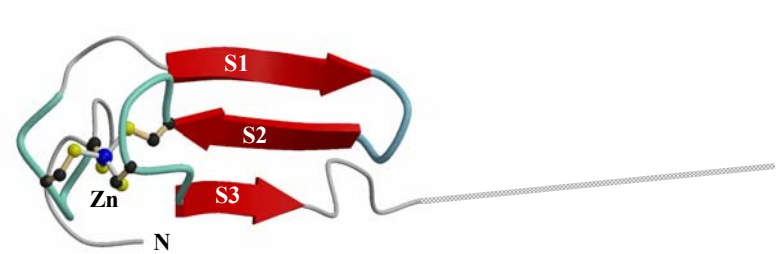
Several pieces of evidence diminish the possibility that promoter DNA downstream of position -16 is bent back towards the TATA box to allow farther interaction of the TFIIB C-terminal domain. First, there is DNA bending that is known to exist as a result of TBP binding the minor groove of DNA at the TATA box (Figure 1.4 & 1.10). TFIIB aids in this bending of the DNA with its upstream interaction with the BRE and downstream interaction with the minor groove. Logically the downstream arm of the DNA angle is towards the active site of RNAP so that the +1 transcription start site is in position for the polymerization of the first ribonucleotides. Secondly, photocrosslinking results for eukaryotic PIC established that RNAP II interacts with ~240 Å of promoter DNA from upstream of the TATA box to well beyond the transcription start site. This expanse of DNA interaction is more than twice the distance of the longest dimension of RNAP II (Kim et al., 1997). As a result, Ebright and Reinberg proposed that DNA wrapping around RNAP would be the only way to produce such extensive RNAP II-DNA interactions. This DNA bending and wrapping model around RNAP II has been extended to include prokaryotic RNA polymerases (Naryshkin et al., 2000) and by structural homology archaeal RNA polymerases (this model has been reviewed by Coulombe and Burton, (Coulombe and Burton, 1999). The initial model based on photocrosslinking data also utilized work by the Kornberg group that determined the location of TFIIB in a TFIIB-RNAPII complex (Kim et al., 1997; Leuther et al., 1996).

Locating TFIIB relative to RNAP II also located TBP bound to the TATA box. This low-resolution crystal structure determined the distance between the TFIIB density and the RNAP II active site to be ~110 Å. This would correspond to 32 bp of B-form DNA. Given that the distance between the TATA box and start site is usually 30 bp (in archaea as well), Leuther et al., proposed that the DNA follows a straight path on the surface of the polymerase to the active center. This last constraint of the Ebright and Reinberg model does the most to diminish the possibility that DNA downstream of position -16 on a eukaryotic promoter bends back towards the TATA box to account for TFIIB interaction at positions downstream of -10.

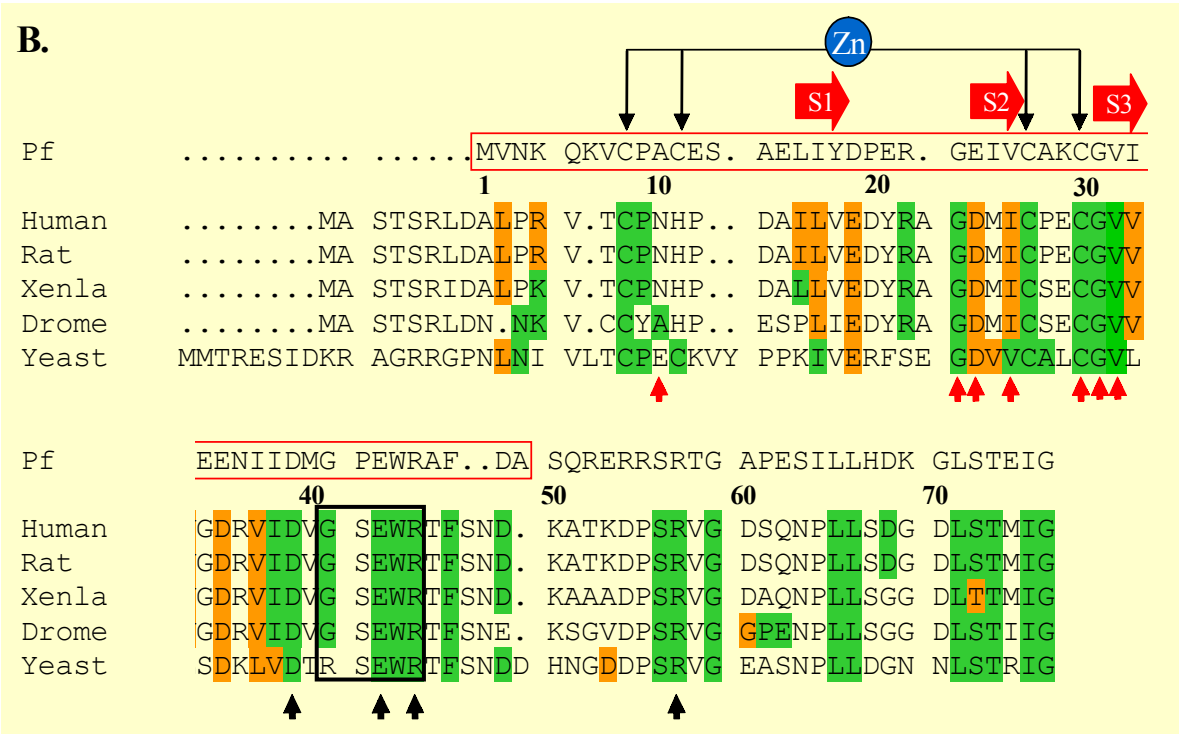
The second possibility is that a portion of TFIIB or TFB other than the C-terminal cyclin A folds interacts with downstream DNA (beyond -10) and this is supported by several pieces of evidence. First, mutational analysis of the N-terminal domain TFB/TFIIB (archaeal, human, and yeast) suggests three roles for this domain which are mapped to two specific areas of this protein (Figure 3.7). Role 1 is selection of the transcription start site. Yeast TFIIB mutant E62K and a corresponding archaeal mutant in *Sulfolobus acidocaldarius* TFB (E46K) both affect transcription *in vitro* by shifting the transcription start site (Bangur et al., 1997; Bell and Jackson, 2000; Pinto et al., 1994). Role 2 for the N-terminal domain is an intramolecular interaction with the C- terminal domain. Mutations which shift the transcription start site also affect this intramolecular interaction. Mutation of hTFIIB residues E51 (same as E62 in yeast and E46 in *Sulfolobus*), D47, and R66 also affect N- and C-terminal interactions. Replacing the charged residues in these positions with neutral or oppositely charged amino acids dramatically affects the affinity of the intramolecular interaction (Hawkes et al., 2000).

Figure 3.7. *Pyrococcus furiosus* TFB N-terminal domain. A. Ribbon representation of Pf TFB Zinc ribbon domain (Zhu et al., 1996). **B.** Sequence alignment of the N-terminal domain of TFB and TFIIB from archaea and eukaryal sources (numbered with repect to the *Pf* TFB sequence). Identical residues are highlighted in green. Similar residues are highlighted in orange. Zinc coordinating residues are indicated by black arrows above the *Pf* sequence. The zinc ribbon β-strands are indicated above the *Pf* sequence (thick red arrows). TFIIB and TFB mutational analysis 1) localizes TFB/TFIIB's role in RNAP/RNAP II recruitment (to the PIC) to the TFIIB zinc ribbon domain (red triangles below sequences) and 2) localizes TFIIB's role in start site selection to the adjacent high homology segment of the flexible linker (black triangles below sequences).

A.


S1
S2
S3
Zn
N

B.


Zn
S1
S2
S3
Pf ....MVNK QKVCPACES. AELIYDPER. GEIVCAKCGVI
1 10 20 30
Human .....MA STSRLDALPR V.TCPNHP.. DAILVEDYRA GDMICPECGVV
Rat .....MA STSRLDALPR V.TCPNHP.. DAILVEDYRA GDMICPECGVV
Xenla .....MA STSRIDALPK V.TCPNHP.. DALLVEDYRA GDMICSECGVV
Drome .....MA STSRLDN.NK V.CCYAHP.. ESPLIEDYRA GDMICSECGVV
Yeast MMTRESIDKR AGRRGPNLNI VLTCPECKVY PPKIVERFSE GDVVCALCGVL
Pf EENIIDMG PEWRAF..DA SQRERRSRTG APESILLHDK GLSTEIG
40 50 60 70
Human GDRVIDVG SEWRTFSND. KATKDPSRVG DSQNPLLSDG DLSTMIG
Rat GDRVIDVG SEWRTFSND. KATKDPSRVG DSQNPLLSDG DLSTMIG
Xenla GDRVIDVG SEWRTFSND. KAAADPSRVG DAQNPLLSGG DLTTMIG
Drome GDRVIDVG SEWRTFSNE. KSGVDPSRVG GPENPLLSGG DLSTIIG
Yeast SDKLVDTR SEWRTFSNDD HNGDDPSRVG EASNPLLDGN NLSTRIG

A fourth mutant at position R53 in hTFIIB and R64 in yTFIIB enhances the affinity of the intramolecular interaction as well as the formation of TFIIB-TBP-DNA ternary complex (Bangur et al., 1999). All of these mutations lie within the linker region between the N- and C- terminal structural domains of TFIIB (Figure 1.3.A & 3.7.B., black triangles). More specifically, at the N-terminal end of this linker (Figure 3.7.B). Mutations in this high homology segment of the flexible linker do not affect the N-terminal region's third role. Role 3 is the recruitment of RNAP II or archaeal RNAP (and TFIIF in eukarya) to the PIC. None of the mutations that affected role one and role two hinder TFIIB or TFB from its role in recruiting RNA polymerase to the preinitiation complex. Mutations in yeast TFIIB zinc ribbon domain and its RNAP III system homologue (BRF), which disrupt the β–sheet-facing of the zinc ribbon or the metal-binding motif, affect growth phenotypes dramatically (Hahn and Roberts, 2000). It was proposed that these mutants disrupted interactions of the zinc ribbon with other cellular targets. Several studies using deletion mutants and point mutations within the zinc ribbon domain show that transcriptional activity and/or RNAP II-TFIIF binding was most severely affected by disruptions of the zinc ribbon's β-sheet (Figure 3.7.B. red triangles, Hahn and Roberts, 2000).

The roles of the N-terminal domain of TFB/TFIIB in recruitment of RNAP/RNAP II to the PIC and accurate start site selection have been localized to regions adjacent to one another in the amino acid sequence. The mutational analysis shows that the N-terminal zinc ribbon and the adjacent high homology segment of the flexible linker are both required for accurate basal levels of transcription. However, mutations in the high homology segment of the flexible linker affect accurate transcription start site selection,

but do not affect PIC formation. This indicates that the zinc ribbon must play a crucial role in RNAP (and TFIIF in eukarya) interaction/recruitment to the preinitiation complex. Between the localized TFIIBc-TBP-TATA box complex (relative to RNAP II) and the RNAP II active site is a proposed stretch of promoter DNA (-16 to +1) where TFIIB-DNA and *Pf* TFB-DNA crosslinks have occurred at specific positions in both eukaryal (Kim et al., 1997) and archaeal (reported here) systems. I propose that the N-terminal domain of TFIIB or TFB make both protein-protein and protein-DNA interactions within the preinitiation complex of their respective transcription systems. The protein-protein interactions are between the zinc ribbon domain of TFIIB and a RNAP II subunit (most likely RPB2/subunit B or RPB1/subunit A'). The protein-DNA interactions are between the high homology segment of the flexible linker and positions -10 to -2 of the DNA promoter. Based on our photocrosslinking analysis, it would appear that the protein-DNA interaction of archaeal TFB N-terminal domain with promoter DNA is more extensive than that of Eukaryal TFIIB with its promoter DNA. In contrast to the eukaryal system, archaeal TFB makes protein-DNA contacts within the -10 to -2 region even in a TFB-TBP-DNA complex. However, this interaction is probably weak considering the inability to see the archaeal DNase I footprint extended to this region (Hausner and Thomm, 2001). Upon addition of RNAP to the complex, the *Pf* TFB N-terminal domain-DNA interaction is attenuated and extended in the archaeal system (to positions T-4 and -2).

Based on previous studies, it is clear that the zinc ribbon domain of TFB/TFIIB plays an important role in RNAP/RNAP II recruitment to the PIC. In addition, the high homology domain of the TFB/TFIIB's flexible linker is crucial for start site selection.

Photocrosslinking data for TFB/TFIIB-DNA interactions demonstrates a TFB/TFIIB-DNA interaction downstream of -10 in a *Pf* TFB-TBP-DNA complex that is attenuated by the addition of RNAP and is reported for TFIIB-DNA photocrosslinking in an RNAP II-TFIIF-TFIIB-TBP-DNA complex. Based on these results, I propose that recruitment of RNAP/RNAP II to the transcription PIC is initiated by a protein-protein interaction between the TFB/TFIIB zinc ribbon domain and RNAP/RNAP II. This interaction induces/produces an additional loop in TFB/TFIIB immediately C-terminal to the zinc ribbon that interacts with the minor groove of DNA to dictate start site selection. This type of interaction is similar to the TFIIB C-terminal domain minor groove recognition loop (Tsai and Sigler, 2000). This induced N-terminal domain loop interacts with DNA through a geometrical alignment of the DNA, rather than recognition of a specific sequence.

The higher degree of *Pf* TFB-DNA interaction near the transcription start site (-10 to -2) could be attributed to the simplified nature of the archaeal transcription system. In eukarya, TFIIF also plays a vital role in RNAP II recruitment to the PIC and stabilization of the PIC by binding to RNAP II, TFIIB, and promoter DNA. There is no TFIIF homolog in the archaeal system, thus leaving the responsibility of archaeal RNAP recruitment, stabilization of the PIC (under extreme conditions), and start site selection to *Pf* TFB.

CHAPTER 4

PREINITIATION COMPLEX: THE BIG PICTURE IN ARCHAEA AND EUKARYA

4.1 The preinitiation complex in archaea and eukarya

The preinitiation complex is a functional protein scaffold built near the transcription start site of a gene to provide RNAP/RNAP II a beginning point and an opening to transcribe the DNA blueprint. In archaea and eukarya, DNA promoter elements (TATA box, BRE) provide a reference from within the blueprint as to the position of the start site. TBP recognizes the minor groove of the TATA box to become the first anchor of the scaffold that begins to position the DNA for transcription. Eukaryal TFIIE (specifically TFIIEα) and possibly archaeal TFEα assist TBP in its initial binding to the TATA box. TFB/TFIIB binds the TBP anchor as well as DNA to 1) further position DNA for transcription through TFBc/TFIIBc interaction upstream and downstream of TATA box, 2) provide a link between the TBP anchored to the DNA (TFBc/TFIIBc) and the RNAP/RNAP II (TFB/TFIIB N-terminal zinc ribbon), and 3) orient the RNAP/RNAP II within the scaffold for correct initiation of transcription via a TFB/TFIIB-DNA interaction of the high homology segment of the flexible linker.

In eukarya, TFIIF also serves as an additional link by interacting with RNAP II, TFIIB, and DNA. Additionally, TFIIF links DNA-TBP-TFIIB-RNAP to TFIIE which (along with TFIIF) further positions the DNA for transcribing by assisting DNA wrapping around RNAP II and begins the untwisting of DNA for transcription initiation.

This allows for the polymerization of the first nucleotides. TFIIE is also where the final piece of the scaffold assembles, TFIIH. TFIIH has several functions within the eukarya PIC scaffold. Its helicase activity at the front end of the PIC opens or unwinds the DNA to clear the way for the transcribing complex. At the same time, its kinase activity phosphorylates RNAP II subunit Rpb1 C-terminal domain to trigger the transition between an immobilized scaffold (PIC) to a mobile transcribing scaffold (elongation complex).

The absence of TFIIF, TFIIH, and the majority of TFIIE homologs in archaea leaves their respective functions either to the known GTFs (TBP and TFB) or to other yet unidentified components. By way of elimination, TBP's role does not seem to stretch beyond anchoring and generic localization of the PIC. TFIIF's eukaryal roles of linking RNAP to TFIIB-TBP-DNA and stabilizing the PIC could be ascribed to archaeal TFB, especially considering the higher degree of TFB-DNA interaction reported here between the TATA box and the transcription start site (Kim et al., 1997).

This would leave further wrapping of DNA around RNAP/RNAP II (ascribed to TFIIF and TFIIE in eukarya), untwisting DNA to form an open complex (ascribed to TFIIF and TFIIE), and helicase unwinding of DNA (TFIIH) unassigned to *Pf* PIC components. To begin with the last, there are several putative helicase genes based on sequence homology in the *P. furiosus* genome. Secondly, in the context of the *Pf* cell, DNA is not linear. Archaeal histones have been identified that pack and unpack DNA (as in eukarya, (Pereira and Reeve, 1998). A gene expression system in which transcriptional machinery and DNA packaging machinery work in closer proximity might eliminate the need for additional GTFs (TFIIE and TFIIF) that would assist in wrapping

or untwisting DNA. There is also the fact that the inherent higher temperature at which *P. furiosus* exists (70 - 100°C) may result in melting of DNA and eliminate the need for mechanical untwisting of DNA by additional transcription factors. In such a gene expression system, one might also find the appropriate helicase working with the DNA packaging machinery rather than the transcriptional machinery.

4.2 *Pf*TFB interaction with DNA and future work

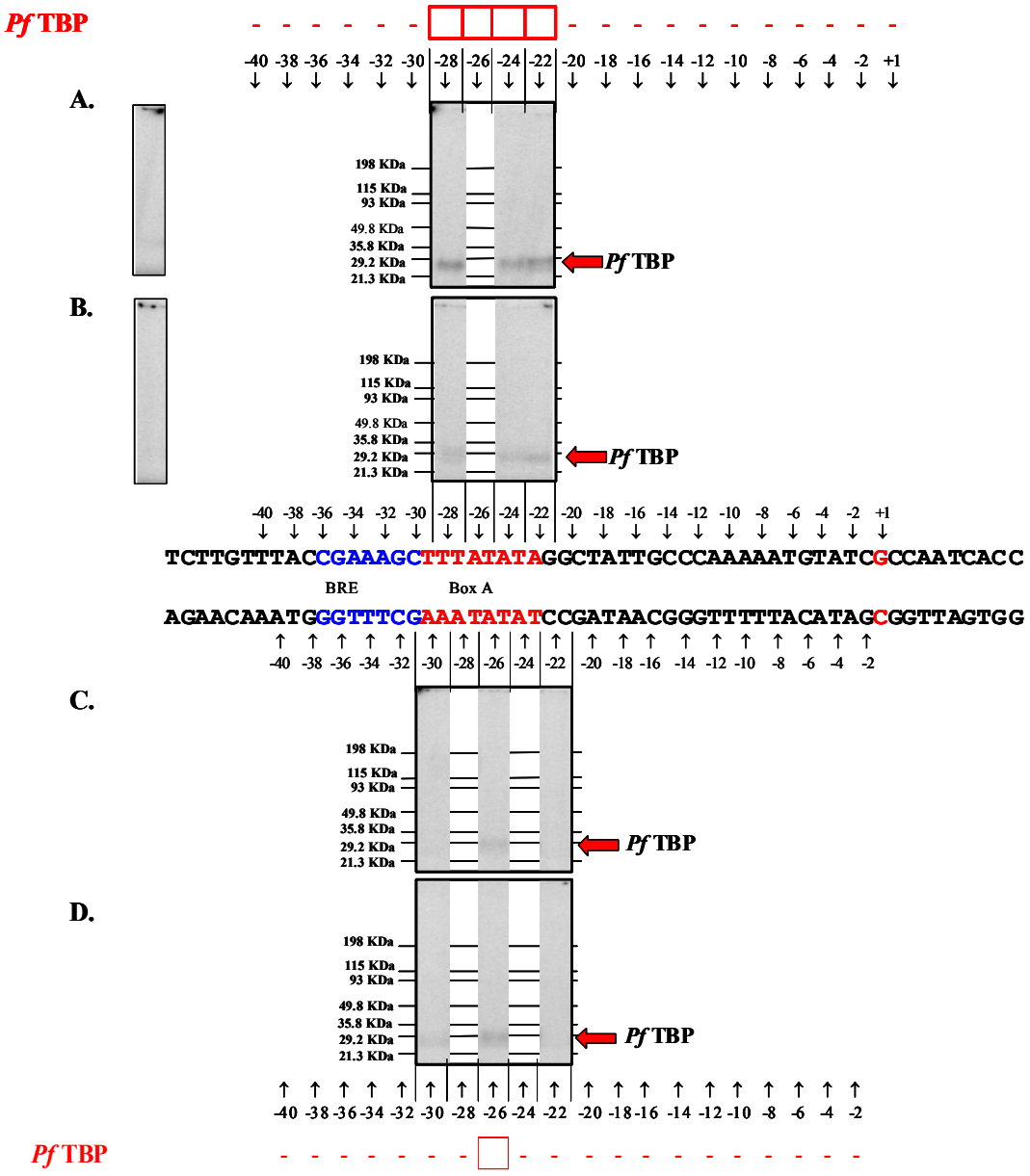
Pyrococcus furiosus Pf TFB-DNA interactions reported here point toward a mechanism by which the N-terminal domain of *Pf* TFB recruits RNAP to the PIC and positions the RNAP for accurate start-site selection. However, while *Pf* TFB-DNA crosslinks identify at what position on the promoter *Pf* TFB is interacting with DNA, no information in these experiments identifies what portion or domain of *Pf* TFB is making that contact. Ebright's group utilized this site-specific photocrosslinking technique along with site-specific protein chemical cleavage methods to attribute protein-DNA crosslinks to identifiable segments of cleaved bacterial RNA polymerase in their analysis of the bacterial transcription initiation complex (Naryshkin et al., 2000). Attempts to use chemical cleavage on *Pf* TFB to identify which domain is interacting with DNA positions near the transcription start site were unsuccessful. Another possibility would be to conduct the same photocrosslinking experiments with unlabeled derivatized DNA promoter fragment and then to analyze trypsinized protein fragments by mass spectrometry. Theoretically, crosslinked "tagged" fragments would have an additional mass consistent with the molecular weight of 2-4 nucleotides of the promoter DNA. The limitation however would be in the amount of sample needed to detect tagged

polypeptides. Derivatized DNA promoter fragment synthesis typically yielded (at the levels used in the experiments reported here) 10-15 femtomols derivatized DNA. Secondly, the efficiency of photocrosslinking by UV irradiation may still limit further the amount of tagged polypeptide in a given experiment. Still with mass spectrometry reaching the attomol to femtomol / μL range of detection, a small scale-up of the experiments reported here maybe all that is needed to map what portion of *Pf* TFB is crosslinked. If possible, this would greatly increase the value of protein-DNA photocrosslinking as a whole.

In the context of the archaeal transcription project in the Scott group, one set of photocrosslinking data for the *Pf* PIC on the glutamate dehydrogenase promoter may be enough. However, as new promoters and/or transcriptional regulators are found, a new set of photocrosslinking probes for a different promoter might provide valuable information. For a given span of 100 bp of DNA, even derivatized DNA promoter fragments for every fourth phosphate (rather than every second) could quicken the time to produce a topographical map of a new complex. This may be especially relevant if the archaeal TFEα is found to be a selective transcription factor that enables PIC formation at only a certain class of archaeal gene promoters.

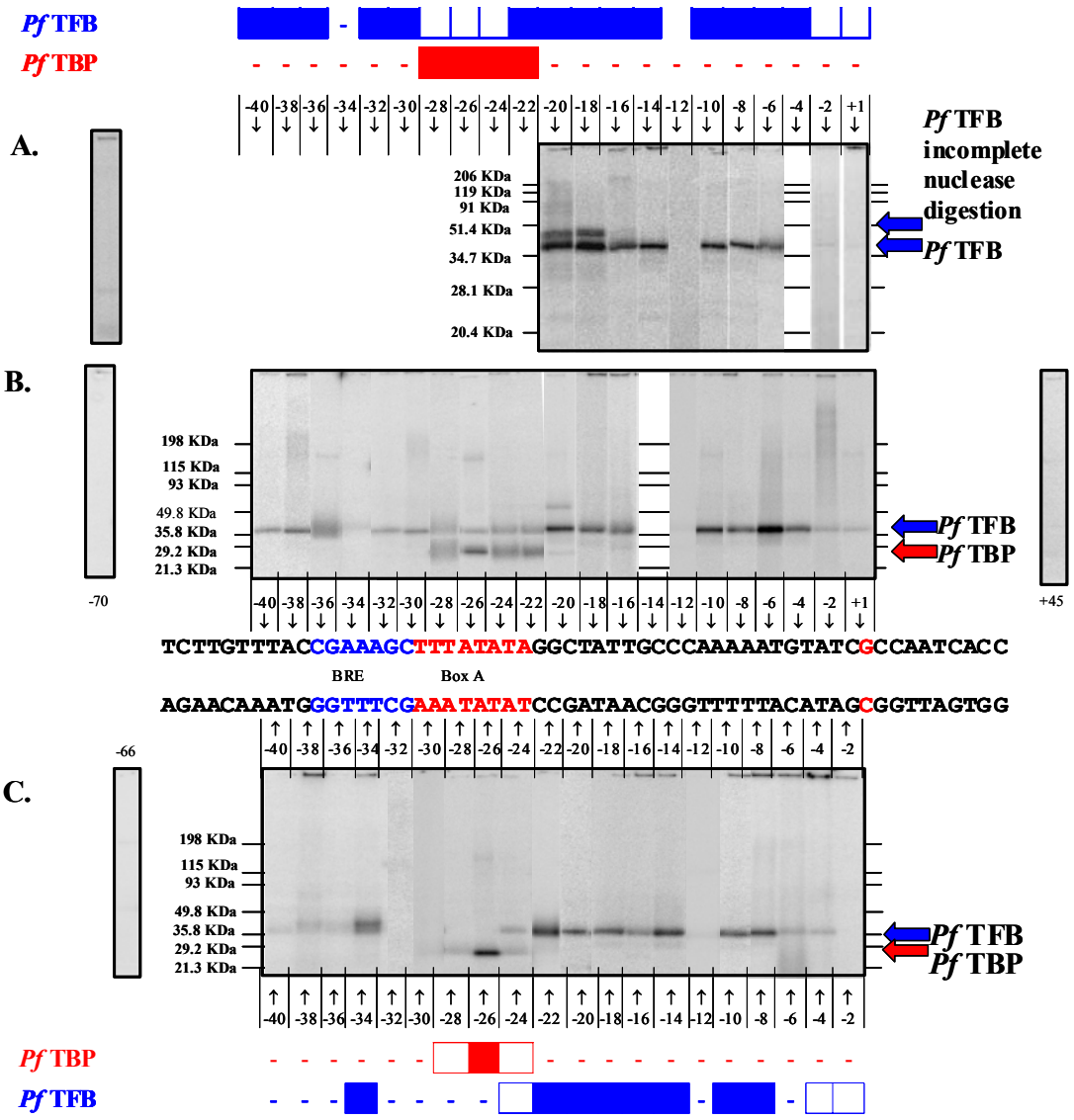
APPENDIX A.1

Figure A.1. Photocrosslinking analysis of *Pf* TBP-DNA interactions. **A.** *Pf* TBP-DNA photocrosslinking at nontemplate (NT) postions NT-70, -28, -24, and -22 with 2 nM TBP final concentration. **B.** *Pf* TBP-DNA photocrosslinking at postions NT-70, -28, -24, and -22 with 1 nM TBP final concentration. **C.** *Pf* TBP-DNA photocrosslinking at template (T) postions T-30, -26, and -22 with 1 nM TBP final concentration. **D.** *Pf* TBP-DNA photocrosslinking at postions T-30, -26, and -22 with 2 nM TBP final concentration. Each gel lane represents a single experiment conducted using a single site-specific derivatized promoter DNA fragment probe (probe synthesis protocol, section 2.3) in photocrosslinking (protocol, section 2.5.3). All gel lanes (4-20% TRIS-glycine SDS-PAGE, Criterion gel, Bio-Rad) are aligned with respect to the derivatized promoter DNA fragment used in each experiment and with summarized results as seen in Figure 3.1, A.


Pf TBP
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 +1
A.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf TBP
B.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf TBP
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 +1
TCTTGTTTACCGAAAGCTTTATATAGGCTATTGCCCAAAAATGTATCGCCAATCACC
BRE
Box A
AGAACAAATGGGTTTCGAAATATATCCGATAACGGGTTTTTACATAGCGGTTAGTGG
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2
C.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf TBP
D.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf TBP
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2
Pf TBP

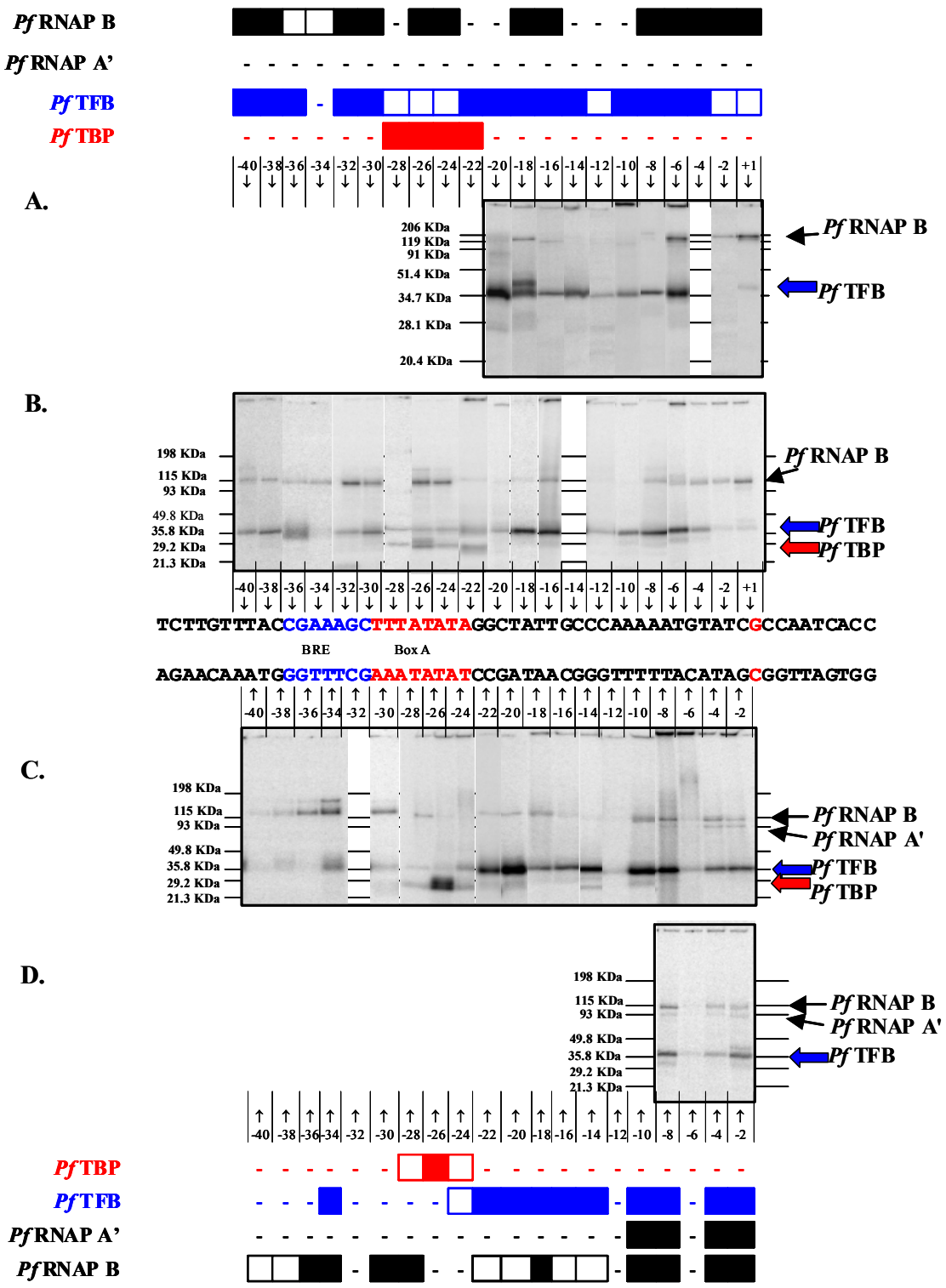
APPENDIX A.2

Figure A.2. Photocrosslinking analysis of *Pf* TFB-TBP-DNA interactions. **A.** *Pf* TFB-TBP-DNA photocrosslinking at nontemplate (NT) postions NT-70 and NT-20 through NT+1. **B.** *Pf* TFB TBP-DNA photocrosslinking at postions NT-70, NT-40 through NT+1, and NT+45. **C.** *Pf* TFB-TBP-DNA photocrosslinking at template (T) postions T-66 and T-40 through T-1. Each gel lane represents a single experiment conducted using a single site-specific derivatized promoter DNA fragment probe (probe synthesis protocol, section 2.3) in photocrosslinking experiments (protocol, section 2.5.3). All gel lanes (15% or 4-20% TRIS-glycine SDS-PAGE, Criterion gel, Bio-Rad) are aligned with respect to the derivatized promoter DNA fragment used in each experiment and with summarized results as seen in Figure 3.1, B.


Pf TFB
Pf TBP
A.
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 +1
206 KDa
119 KDa
91 KDa
51.4 KDa
34.7 KDa
28.1 KDa
20.4 KDa
Pf TFB
incomplete
nuclease
digestion
Pf TFB
B.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf TFB
Pf TBP
-70
+45
TCTTGTTTACCGAAAGCTTTATATAGGCTATTGCCCAAAAATGTATCGCCAATCACC
BRE
Box A
AGAACAAATGGGTTTCGAAATATATCCGATAACGGGTTTTTACATAGCGGTTAGTGG
-66
C.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf TFB
Pf TBP
Pf TBP
Pf TFB

APPENDIX A.3

Figure A.3. Photocrosslinking analysis of *Pf* RNAP-TFB-TBP-DNA interactions. **A.** *Pf* RNAP-TFB**-**TBP-DNA photocrosslinking at nontemplate (NT) postions NT-20 through NT+1. **B.** *Pf* RNAP-TFB-TBP-DNA photocrosslinking at postions NT-70, NT-40 through NT+1, and NT+45. **C.** *Pf* TFB**-**TBP-DNA photocrosslinking at template (T) postions T-66 and T-40 through T-1. **D.** *Pf* TFBHisTAG**-**TBP-DNA photocrosslinking at postions T-8 through T-2. Each gel lane represents a single experiment conducted using a single site-specific derivatized promoter DNA fragment probe (probe synthesis protocol, section 2.3) in photocrosslinking experiments (protocol, section 2.5.3). All gel lanes (15% or 4-20% TRIS-glycine SDS-PAGE, Criterion gel, Bio-Rad) are aligned with respect to the derivatized promoter DNA fragment used in each experiment and with summarized results as seen in Figure 3.1, C.


Pf RNAP B
Pf RNAP A'
Pf TFB
Pf TBP
A.
-40 -38 -36 -34 -32 -30 -28 -26 -24 -22 -20 -18 -16 -14 -12 -10 -8 -6 -4 -2 +1
206 KDa
119 KDa
91 KDa
51.4 KDa
34.7 KDa
28.1 KDa
20.4 KDa
Pf RNAP B
Pf TFB
B.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf RNAP B
Pf TFB
Pf TBP
TCTTGTTTACCGAAAGCTTTATATAGGCTATTGCCCAAAAATGTATCGCCAATCACC
BRE
Box A
AGAACAAATGGGTTTCGAAATATATCCGATAACGGGTTTTTACATAGCGGTTAGTGG
C.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf RNAP B
Pf RNAP A'
Pf TFB
Pf TBP
D.
198 KDa
115 KDa
93 KDa
49.8 KDa
35.8 KDa
29.2 KDa
21.3 KDa
Pf RNAP B
Pf RNAP A'
Pf TFB
Pf TBP
Pf TFB
Pf RNAP A'
Pf RNAP B

REFERENCES

BioRad Laboratories, 2000 Alfred Nobel Drive, Hercules CA 94547.

Clontech Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230.

Dynal Biotech ASA, Oslo, Norway.

Fisher Scientific Worldwide, One Liberty Lane, Hampton, NH 03842.

Integrated DNA Technologies (IDT), 1710 Commercial Park, Coralville, IA 52241, U.S.A.

Invitrogen Corporation, Life Technologies Gibco BRL, 1600 Faraday Ave., P.O. Box 6482, Carsbad, CA 92008.

New England Biolabs, Inc., 32 Tozer Road, Beverly MA 01915.

Novagen Inc., 597 Science Drive, Madison, WI 53711.

PerkinElmer Life Sciences, 9238 Gaither Road, Gaithersburg, Maryland 20877.

Promega, 2800 Woods Hollow Road, Madison, WI 53711-5399.

Qiagen, 28159 Avenue Stanford, Valencia, CA 91355.

Roche Molecular Biochemicals, 9115 Hague Rd., P.O. Box 50414, Indianapolis, IN 46250-0414.

Sigma, 3050 Spruce Street, St. Louis, MO 63103.

Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92037.

Andersen, G., Busso, D., Poterszman, A., Hwang, J. R., Wurtz, J. M., Ripp, R., Thierry, J. C., Egly, J. M., and Moras, D. (1997). The structure of cyclin H: common mode of kinase activation and specific features. Embo J *16*, 958-967.

Andersen, G., Poterszman, A., Egly, J. M., Moras, D., and Thierry, J. C. (1996). The crystal structure of human cyclin H. FEBS Lett *397*, 65-69.

Bagby, S., Kim, S., Maldonado, E., Tong, K. I., Reinberg, D., and Ikura, M. (1995). Solution structure of the C-terminal core domain of human TFIIB: similarity to cyclin A and interaction with TATA-binding protein. Cell *82*, 857-867.

Bangur, C. S., Faitar, S. L., Folster, J. P., and Ponticelli, A. S. (1999). An interaction between the N-terminal region and the core domain of yeast TFIIB promotes the formation of TATA-binding protein-TFIIB-DNA complexes. J Biol Chem *274*, 23203-23209.

Bangur, C. S., Pardee, T. S., and Ponticelli, A. S. (1997). Mutational analysis of the D1/E1 core helices and the conserved N- terminal region of yeast transcription factor IIB (TFIIB): identification of an N-terminal mutant that stabilizes TATA-binding protein-TFIIB-DNA complexes. Mol Cell Biol *17*, 6784-6793.

Barberis, A., Muller, C. W., Harrison, S. C., and Ptashne, M. (1993). Delineation of two functional regions of transcription factor TFIIB. Proc Natl Acad Sci U S A *90*, 5628-5632.

Bartlett, M. S., Thomm, M., and Geiduschek, E. P. (2000). The orientation of DNA in an archaeal transcription initiation complex. Nat Struct Biol *7*, 782-785.

Bell, S. D., Brinkman, A. B., van der Oost, J., and Jackson, S. P. (2001a). The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. EMBO Rep *2*, 133-138.

Bell, S. D., and Jackson, S. P. (2000). The role of transcription factor B in transcription initiation and promoter clearance in the archaeon Sulfolobus acidocaldarius. J Biol Chem *275*, 12934-12940.

Bell, S. D., Magill, C. P., and Jackson, S. P. (2001b). Basal and regulated transcription in Archaea. Biochem Soc Trans *29*, 392-395.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.

Buratowski, S., Hahn, S., Sharp, P. A., and Guarente, L. (1988). Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature *334*, 37-42.

Buratowski, S., and Zhou, H. (1993). Functional domains of transcription factor TFIIB. Proc Natl Acad Sci U S A *90*, 5633-5637.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell *104*, 901-912.

Chang, W. H., and Kornberg, R. D. (2000). Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. Cell *102*, 609-613.

Chen, D., Riedl, T., Washbrook, E., Pace, P. E., Coombes, R. C., Egly, J. M., and Ali, S. (2000a). Activation of estrogen receptor alpha by S118 phosphorylation involves

a ligand-dependent interaction with TFIIH and participation of CDK7. Mol Cell *6*, 127-137.

Chen, H.-T. (2000) Structural study of Eukaryal and Archeal transcription factors TF(II)B, Ph.D. Dissertation, University of Georgia, Athens.

Chen, H.-T., Legault, P., Glushka, J., Omichinski, J. G., and Scott, R. A. (2000b). Structure of a (Cys3His) Zinc Ribbon, a Ubiquitous Motif in Archaeal and Eucaryal Transcription. Prot Sci, submitted for publication.

Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M., and Egly, J. M. (1998). Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. Nat Genet *20*, 184-188.

Comai, L., Zomerdijk, J. C., Beckmann, H., Zhou, S., Admon, A., and Tjian, R. (1994). Reconstitution of transcription factor SL1: exclusive binding of TBP by SL1 or TFIID subunits. Science *266*, 1966-1972.

Conaway, R. C., and Conaway, J. W. (1989). An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. Proc Natl Acad Sci U S A *86*, 7356-7360.

Coulombe, B., and Burton, Z. F. (1999). DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. Microbiol Mol Biol Rev *63*, 457-478.

Cramer, P. (2002). Multisubunit RNA polymerases. Curr Opin Struct Biol *12*, 89-97.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000).

Architecture of RNA polymerase II and implications for the transcription mechanism [see comments]. Science *288*, 640-649.

Douziech, M., Coin, F., Chipoulet, J. M., Arai, Y., Ohkuma, Y., Egly, J. M., and Coulombe, B. (2000). Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. Mol Cell Biol *20*, 8168-8177.

Drapkin, R., Reardon, J. T., Ansari, A., Huang, J. C., Zawel, L., Ahn, K., Sancar, A., and Reinberg, D. (1994). Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. Nature *368*, 769-772.

Egly, J. M. (2001). The 14th Datta Lecture. TFIIH: from transcription to clinic. FEBS Lett *498*, 124-128.

Evans, R., Fairley, J. A., and Roberts, S. G. (2001). Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes Dev *15*, 2945-2949.

Fang, S. M., and Burton, Z. F. (1996). RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. J Biol Chem *271*, 11703-11709.

Feaver, W. J., Gileadi, O., Li, Y., and Kornberg, R. D. (1991). CTD kinase associated with yeast RNA polymerase II initiation factor b. Cell *67*, 1223-1230.

Ferguson, H. A., Kugel, J. F., and Goodrich, J. A. (2001). Kinetic and mechanistic analysis of the RNA polymerase II transcrption reaction at the human interleukin-2 promoter. J Mol Biol *314*, 993-1006.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z. F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A *88*, 9999-10003.

Flores, O., Maldonado, E., and Reinberg, D. (1989). Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. J Biol Chem *264*, 8913-8921.

Frey, G., Thomm, M., Brudigam, B., Gohl, H. P., and Hausner, W. (1990). An archaebacterial cell-free transcription system. The expression of tRNA genes from Methanococcus vannielii is mediated by a transcription factor. Nucleic Acids Res *18*, 1361-1367.

Fribourg, S., Kellenberger, E., Rogniaux, H., Poterszman, A., Van Dorsselaer, A., Thierry, J. C., Egly, J. M., Moras, D., and Kieffer, B. (2000). Structural characterization of the cysteine-rich domain of TFIIH p44 subunit. J Biol Chem *275*, 31963-31971.

Gaiser, F., Tan, S., and Richmond, T. J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. J Mol Biol *302*, 1119-1127.

Gervais, V., Busso, D., Wasielewski, E., Poterszman, A., Egly, J. M., Thierry, J. C., and Kieffer, B. (2001). Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family. J Biol Chem *276*, 7457-7464.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876-1882.

Groft, C. M., Uljon, S. N., Wang, R., and Werner, M. H. (1998). Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. Proc Natl Acad Sci U S A *95*, 9117-9122.

Hahn, S., and Roberts, S. (2000). The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. Genes Dev *14*, 719-730.

Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev *62*, 465-503.

Hanzelka, B. L., Darcy, T. J., and Reeve, J. N. (2001). TFE, an archaeal transcription factor in Methanobacterium thermoautotrophicum related to eucaryal transcription factor TFIIEalpha. J Bacteriol *183*, 1813-1818.

Hausner, W., and Thomm, M. (1993). Purification and characterization of a general transcription factor, aTFB, from the archaeon Methanococcus thermolithotrophicus. J Biol Chem *268*, 24047-24052.

Hausner, W., and Thomm, M. (2001). Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. J Bacteriol *183*, 3025-3031.

Hausner, W., Wettach, J., Hethke, C., and Thomm, M. (1996). Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. J Biol Chem *271*, 30144-30148.

Hawkes, N. A., Evans, R., and Roberts, S. G. (2000). The conformation of the transcription factor TFIIB modulates the response to transcriptional activators in vivo. Curr Biol *10*, 273-276.

Hethke, C., Bergerat, A., Hausner, W., Forterre, P., and Thomm, M. (1999). Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of Pyrococcus transcription. Genetics *152*, 1325-1333.

Hethke, C., Geerling, A. C., Hausner, W., de Vos, W. M., and Thomm, M. (1996). A cell-free transcription system for the hyperthermophilic archaeon Pyrococcus furiosus. Nucleic Acids Res *24*, 2369-2376.

Hudepohl, U., Reiter, W. D., and Zillig, W. (1990). In vitro transcription of two rRNA genes of the archaebacterium Sulfolobus sp. B12 indicates a factor requirement for specific initiation. Proc Natl Acad Sci U S A *87*, 5851-5855.

Huet, J., and Sentenac, A. (1992). The TATA-binding protein participates in TFIIIB assembly on tRNA genes. Nucleic Acids Res *20*, 6451-6454.

Kamada, K., De Angelis, J., Roeder, R. G., and Burley, S. K. (2001). Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. Proc Natl Acad Sci U S A *98*, 3115-3120.

Kim, T. K., Ebright, R. H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science *288*, 1418-1422.

Kim, T. K., Lagrange, T., Wang, Y. H., Griffith, J. D., Reinberg, D., and Ebright, R. H. (1997). Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. Proc Natl Acad Sci U S A *94*, 12268-12273.

Kosa, P. F., Ghosh, G., DeDecker, B. S., and Sigler, P. B. (1997). The 2.1-A crystal structure of an archaeal preinitiation complex: TATA- box-binding protein/transcription factor (II)B core/TATA-box. Proc Natl Acad Sci U S A *94*, 6042-6047.

Kuldell, N. H., and Buratowski, S. (1997). Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. Mol Cell Biol *17*, 5288-5298.

Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev *12*, 34-44.

Lagrange, T., Kim, T. K., Orphanides, G., Ebright, Y. W., Ebright, R. H., and Reinberg, D. (1996). High-resolution mapping of nucleoprotein complexes by site-specific protein-DNA photocrosslinking: organization of the human TBP-TFIIA- TFIIB-DNA quaternary complex. Proc Natl Acad Sci U S A *93*, 10620-10625.

Leuther, K. K., Bushnell, D. A., and Kornberg, R. D. (1996). Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. Cell *85*, 773-779.

Lewis, L. M. (2000) Structural and Functional Characterization of the Transcription Preinitiation Complex from Pyrococcus furiosus, Ph.D. Dissertation, University of Georgia, Athens.

Littlefield, O., Korkhin, Y., and Sigler, P. B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci U S A *96*, 13668-13673.

Malik, S., Hisatake, K., Sumimoto, H., Horikoshi, M., and Roeder, R. G. (1991). Sequence of general transcription factor TFIIB and relationships to other initiation factors. Proc Natl Acad Sci U S A *88*, 9553-9557.

Malik, S., Lee, D. K., and Roeder, R. G. (1993). Potential RNA polymerase II-induced interactions of transcription factor TFIIB. Mol Cell Biol *13*, 6253-6259.

Manet, E., Allera, C., Gruffat, H., Mikaelian, I., Rigolet, A., and Sergeant, A. (1993). The acidic activation domain of the Epstein-Barr virus transcription factor R interacts in vitro with both TBP and TFIIB and is cell- specifically potentiated by a proline-rich region. Gene Expr *3*, 49-59.

Masuyama, H., Jefcoat, S. C., Jr., and MacDonald, P. N. (1997). The N-terminal domain of transcription factor IIB is required for direct interaction with the vitamin D receptor and participates in vitamin D-mediated transcription. Mol Endocrinol *11*, 218-228.

Mayer, A. N., and Barany, F. (1995). Photoaffinity cross-linking of TaqI restriction endonuclease using an aryl azide linked to the phosphate backbone. Gene *153*, 1-8.

Naryshkin, N., Kim, Y., Dong, Q., and Ebright, R. H. (2001). Site-specific protein-DNA photocrosslinking. Analysis of bacterial transcription initiation complexes. Methods Mol Biol *148*, 337-361.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R. H. (2000). Structural organization of the RNA polymerase-promoter open complex. Cell *101*, 601-611.

Nikolov, D. B., and Burley, S. K. (1997). RNA polymerase II transcription initiation: a structural view. Proc Natl Acad Sci U S A *94*, 15-22.

Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G., and Burley, S. K. (1992). Crystal structure of TFIID TATA-box binding protein [see comments]. Nature *360*, 40-46.

Ohkuma, Y., Hashimoto, S., Wang, C. K., Horikoshi, M., and Roeder, R. G. (1995). Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. Mol Cell Biol *15*, 4856-4866.

Okuda, M., Watanabe, Y., Okamura, H., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2000). Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. Embo J *19*, 1346-1356.

Ouzounis, C., and Sander, C. (1992). TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes [letter]. Cell *71*, 189-190.

Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T., and Burley, S. K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev *13*, 3217-3230.

Pendergrast, P. S., Chen, Y., Ebright, Y. W., and Ebright, R. H. (1992). Determination of the orientation of a DNA binding motif in a protein-DNA complex by photocrosslinking. Proc Natl Acad Sci U S A *89*, 10287-10291.

Pereira, S. L., and Reeve, J. N. (1998). Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. Extremophiles *2*, 141-148.

Pinto, I., Wu, W. H., Na, J. G., and Hampsey, M. (1994). Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. J Biol Chem *269*, 30569-30573.

Pugh, B. F., and Tjian, R. (1992). Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. J Biol Chem *267*, 679-682.

Qureshi, S. A., Bell, S. D., and Jackson, S. P. (1997). Factor requirements for transcription in the Archaeon Sulfolobus shibatae. Embo J *16*, 2927-2936.

Qureshi, S. A., and Jackson, S. P. (1998). Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength. Mol Cell *1*, 389-400.

Ren, D., Lei, L., and Burton, Z. F. (1999). A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly. Mol Cell Biol *19*, 7377-7387.

Robert, F., Forget, D., Li, J., Greenblatt, J., and Coulombe, B. (1996). Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. J Biol Chem *271*, 8517-8520.

Roberts, S. G., and Green, M. R. (1994). Activator-induced conformational change in general transcription factor TFIIB. Nature *371*, 717-720.

Roberts, S. G., Ha, I., Maldonado, E., Reinberg, D., and Green, M. R. (1993). Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. Nature *363*, 741-744.

Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J. M., and Chambon, P. (1997). Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7. Cell *90*, 97-107.

Ruppert, S., and Tjian, R. (1995). Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. Genes Dev *9*, 2747-2755.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd. edn (Plainview, Cold Spring Harbor Lab. Press).

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P., and Egly, J. M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor [see comments]. Science *260*, 58-63.

Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J. M. (2000). Molecular structure of human TFIIH. Cell *102*, 599-607.

Segerer, A. H., Burggraf, S., Fiala, G., Huber, G., Huber, R., Pley, U., and Stetter, K. O. (1993). Life in hot springs and hydrothermal vents. Orig Life Evol Biosph *23*, 77-90.

Soppa, J. (1999). Transcription initiation in Archaea: facts, factors and future aspects. Mol Microbiol *31*, 1295-1305.

Struhl, K. (1995). Yeast transcriptional regulatory mechanisms. Annu Rev Genet *29*, 651-674.

Tirode, F., Busso, D., Coin, F., and Egly, J. M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol Cell *3*, 87-95.

Tsai, F. T., and Sigler, P. B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. Embo J *19*, 25-36.

Veschambre, P., Roisin, A., and Jalinot, P. (1997). Biochemical and functional interaction of the human immunodeficiency virus type 1 Tat transactivator with the general transcription factor TFIIB. J Gen Virol *78*, 2235-2245.

Wang, B., Jones, D. N., Kaine, B. P., and Weiss, M. A. (1998). High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. Structure *6*, 555-569.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A *87*, 4576-4579.

Xu, H., and Hoover, T. R. (2001). Transcriptional regulation at a distance in bacteria. Curr Opin Microbiol *4*, 138-144.

Yokomori, K., Verrijzer, C. P., and Tjian, R. (1998). An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. Proc Natl Acad Sci U S A *95*, 6722-6727.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution [see comments]. Cell *98*, 811-824.

Zhu, W., Zeng, Q., Colangelo, C. M., Lewis, M., Summers, M. F., and Scott, R. A. (1996). The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc ribbon [letter]. Nat Struct Biol *3*, 122-124.