

Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

1-1-2010

The Function of an Alternate TFB from *Pyrococcus furiosus* and the Orientation of the TFB B-reader within Archaeal Transcription Initiation Complexes

Michael Micorescu
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Let us know how access to this document benefits you.

Recommended Citation

Micorescu, Michael, "The Function of an Alternate TFB from *Pyrococcus furiosus* and the Orientation of the TFB B-reader within Archaeal Transcription Initiation Complexes" (2010). *Dissertations and Theses*. Paper 278.

<https://doi.org/10.15760/etd.278>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

The Function of an Alternate TFB from *Pyrococcus furiosus* and the Orientation of the
TFB B-reader within Archaeal Transcription Initiation Complexes

by

Michael Micoreescu

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

Doctor of Philosophy
in
Biology

Dissertation Committee:
Michael S. Bartlett, Chair
Kenneth Stedman
Justin Courcelle
Niles Lehman
Dirk Iwata-Reuyl

Portland State University
©2010

ABSTRACT

The genome of the hyperthermophile archaeon *Pyrococcus furiosus* encodes two transcription factor B (TFB) paralogs, one of which (TFB1) was previously characterized in transcription initiation. The second TFB (TFB2) is unusual in that it lacks recognizable homology to the archaeal TFB/eukaryotic TFIIB B-reader (also called the B-finger) motif. TFB2 functions, though poorly, in promoter-dependent transcription initiation. Domain swaps between TFB1 and TFB2 showed that the low activity of TFB2 is determined mainly by its N terminus. The low activity of TFB2 in promoter opening and transcription can be partially relieved by transcription factor E (TFE). The results indicate that the TFB N-terminal region, containing conserved Zn ribbon and B-finger motifs, is important in promoter opening and that TFE can compensate for defects in the N terminus through enhancement of promoter opening.

Archaeal RNA polymerase requires two transcription factors for initiation: TBP, which binds to TATA boxes, and TFB, which binds TBP and DNA, recruits RNAP and helps initiate transcription. Archaeal TFBs usually contain a conserved B-reader sequence homologous to the eukaryotic B-reader motif in their N-terminal domains. This region is involved in the assembly of the transcription complex, promoter melting and in transcription start site determination but its position and orientation relative to promoter DNA during initiation is not clear. In this study the positioning of the TFB B-reader relative to DNA was determined by cross-linking using TFB variants substituted with photoactivatable unnatural amino acids. The

acids. The results demonstrate that the B-reader is in close proximity to the transcription start site on the template but not the non-template strand in transcription initiation complexes. Furthermore, the position of the B-reader varies between closed and open promoter complexes, and between open promoter and early initiation complexes. Thus the archaeal B-reader sequence is poised to interact with promoter DNA in a dynamic fashion, and is likely playing a role in positioning the template-strand in an open pre-initiation complex.

Table of Contents

Abstract	i
List of Figures	v
CHAPTER I-Archaeal transcription factors and transcription in the	
Archaea.....	1
The archaeal transcriptional machinery.....	1
Structure and function of the TFIIB family of transcription factors	2
The role of the TFB N-terminus in transcription initiation	3
The position of the TFB N-terminus during transcription initiation.....	4
Overview of archaeal transcription initiation	6
References	8
CHAPTER II-The function of TFB2 – an alternate <i>Pyrococcus furiosus</i>	
transcription factor	11
Introduction	11
Results	15
Transcription by TFB2: activity and promoter selectivity	15
Domain swapping between TFB2 and TFB1.....	21
Discussion.....	25
Evolutionary distribution of the B-finger motif.....	25
Functional impact of the B-finger motif.	26
Interplay between TFB and TFE.....	27
Role of multiple TFBs.	28
Experimental Procedures	30
References	34
CHAPTER III-An investigation into the position and orientation of the archaeal	
B-reader region using protein-DNA cross-linking	36
Introduction	37
Results	41
Purified pBpa TFB1 mutants are active in transcription.....	41
pBpa protein-DNA cross-linking confirms the proximity of F192 to the template strand	43
p-Bpa cross-linking shows B-reader surfaces are proximal to specific promoter positions.....	45
The B-reader is not proximal to promoter DNA in the absence of RNAP	49
The TFB1 B-reader is close to the template strand.....	51
The position of the B-reader changes during the formation of the open complex....	54
The position of the B-reader changes upon transcription initiation	56
pBpa cross-linking as an assay for detecting the state of the RNAP clamp?	59
Discussion.....	60
Role of TFB in transcription: post-assembly events leading to promoter escape	60
A novel method for mapping protein-DNA interacting surfaces.....	61
The B-reader: a role for transcribed strand placement	62
The position of TFB1 during the initiation of transcription.....	64
Experimental Procedures	71

Mutagenesis and protein purification.....	71
Standard transcription assays.....	72
Protein-DNA Cross-linking.....	73
References	74
Chapter IV-Conclusions and future directions	76
Conclusions	76
Unanswered questions and future experiments	78
Does the archaeal B-reader play a role in the formation of the initial phosphodiester bond ?	78
Does the archaeal B-reader play a role in abortive transcription?	79
References	83
Terminal References	84
Appendix-An investigation into the role of TFE in transcription complex assembly.....	90
Introduction	90
Results	92
Discussion.....	98
Experimental Procedures	101
References	103

List of Figures

Figure 2.1 - Partial alignment of archaeal TFBs and eukaryotic TFIIBs.....	15
Figure 2.2 - Transcription of the <i>gdh</i> promoter using TFB1 or TFB2.....	17
Figure 2.3 - <i>P. furiosus</i> promoters for comparing TFB1 and TFB2 activities.	18
Figure 2.4 - TFB2 is active in transcription initiation with a variety of promoters.	20
Figure 2.5 - Comparison of transcription activities of hybrid TFBs.....	22
Figure 2.6 - Activation of TFB hybrids by TFE.	24
Figure 3.1 - An overview of archaeal transcription initiation.	38
Figure 3.2 - A. TFB1 R52 <i>p</i> -Bpa expression is <i>p</i> -Bpa dependent.	42
Figure 3.3 - Comparison of the transcription efficiency of substituted TFBs versus wild type TFB1.	
Figure 3.4 - F192, a residue in the TFB C-terminus, is located in close proximity to -19 on the template strand.....	44
Figure 3.5 - The archaeal B-reader is close to the template strand.	46
Figure 3.6 - <i>p</i> Bpa crosslinking is TBP dependent.....	47
Figure 3.7 - The B-reader loop is in close proximity to the template strand and the transcription start site in RFP _{init} but not RFP _O	48
Figure 3.8 - Cross-linking of R52 <i>p</i> -Bpa TFB to the transcribed (T) and non-transcribed (NT) strands in FP versus RFP _O at 65°C.	50
Figure 3.9 - Cross-linking of R52 <i>p</i> -Bpa and W44 <i>p</i> -Bpa TFBs to the non-transcribed strand in FP versus RFP _O at 65°C.	51
Figure 3.10 - Cross-linking W44 <i>p</i> -Bpa, R52 <i>p</i> -Bpa and T58 <i>p</i> -Bpa TFBs to the transcribed strand in RFP _O at 65°C.	53
Figure 3.11 - Cross-linking W44 <i>p</i> -Bpa, R52 <i>p</i> -Bpa and T58 <i>p</i> -Bpa TFBs to the transcribed strand in RFP _C at 50°C.	55
Figure 3.12 - Cross-linking of R52 <i>p</i> -Bpa, W44 <i>p</i> -Bpa and T58 <i>p</i> -Bpa TFBs to the transcribed or non-transcribed strand in the presence of saturating rATP, rCTP and rGTP.	57
Figure 3.13 - Cross-linking of R52 <i>p</i> -Bpa, W44 <i>p</i> -Bpa and T58 <i>p</i> -Bpa TFBs to the transcribed or non-transcribed strand in the presence of rATP, rCTP and rGTP at 50°C..	59

Figure 3.14 - The effect of S1 nuclease treatment on pBpa cross-linking signal in RFP _{init} at 65°C.	60
Figure 3.15 - The archaeal B-reader is close to the template strand at 65°C.	63
Figure 3.16 - The archaeal B-reader during transcription initiation. A.....	66
Figure 3.17 - The archaeal B-reader during transcription initiation in the presence of initiating nucleotides at 65°C	69

CHAPTER I

ARCHAEAL TRANSCRIPTION FACTORS AND TRANSCRIPTION IN THE ARCHAEA

Archaea constitute the third major branch of the universal tree of life (1). Archaea occupy niches that define the biological extremes of temperature, pH, pressure and salinity, but also thrive under less extreme conditions and are ubiquitous in soil and marine environments, as well as being part of the normal human flora. Although as a domain Archaea are morphologically prokaryotic, their information processing systems display structural and functional homology with both bacterial and eukaryotic components (2-5) and afford a unique opportunity to study how the mechanisms of gene expression and regulation have evolved.

The archaeal transcriptional machinery

In Archaea a single RNA polymerase (RNAP) that is very similar to eukaryotic RNAP II at the level of subunit identity and sequence homology catalyzes RNA synthesis (6, 7). Three extrinsic general transcription factors, sufficient for transcription initiation *in vitro* at many promoters, guide transcription initiation by archaeal RNAP (6, 8). The archaeal general transcription factors, TATA binding protein (TBP), transcription factor B (TFB), and transcription factor E (TFE) display high levels of structural and functional conservation with their eukaryotic counterparts TBP, TFIIB (and related BRF1) and the α subunit of TFIIE (9). The eukaryotic system uses additional general transcription factors absent in Archaea, that confer additional

levels of regulation (10). The functional and structural homology evident in the central components supports the evolutionary origin of eukaryotic-type transcription before the phylogenetic split of the archaeal and eukaryotic domains.

Structure and function of the TFIIB family of transcription factors

Based on sequence, structural and functional conservation, archaeal TFB is part of a larger TF(II)B protein family that includes eukaryotic transcription factors TFIIB and the RNAP III accessory factor Brf1. The structure of archaeal and eukaryotic TFBs can be divided into two domains. The C-terminal domain (CTD) is globular and composed of two imperfect direct repeats (11, 12) each of which adopt a cyclin fold (13). The CTD binds to a surface of the TBP-promoter complex by utilizing the bent DNA conformation induced by TBP (14). This allows TFB to make sequence specific contacts both upstream (BRE^U) and downstream (BRE^D) of the TATA box in both eukaryotic and archaeal promoters (15-18) and imparts directionality to pre-initiation complex (PIC) assembly by properly orienting TFB with respect to the transcription start site. In addition, the TFB CTD contains surfaces that bind TBP and thereby facilitate the formation of a stable ternary pre-initiation complex. The TFB NTD is connected to the CTD by a flexible linker and is predicted to bind DNA downstream of the TATA box upon the binding of the CTD to the TBP-DNA complex, a hypothesis supported by DNA site-specific photochemical cross-linking with *Pyrococcus furiosus* TFB in the presence of TBP (19). The N-terminal domain of all TFBs contains a zinc-coordinating ribbon motif (20, 21) that recruits RNAP to the promoter by interacting with the Rpb1/RpoA dock domain, close to the

RNA exit pore (22-25). Similarly, the N-terminal domain of RNAP III initiation factor Brf1 also cross-links to promoter regions encompassing the transcription start site and has been shown to play a post-recruitment role in promoter opening in addition to binding TBP and directing PIC assembly (26, 27). Structural information describing the linker that connects the two TFB domains has been difficult to obtain. The linker may be mobile, and might achieve a stable conformation in the presence of nucleic acids, absent from even the most recent crystal structures. However, even in the absence of the structure and position of the linker region, genetic and biochemical studies over the past decade have highlighted its importance in the transcription initiation mechanism.

The role of the TFB N-terminus in transcription initiation

Early genetic studies of TFIIB function used yeast strains with mutations in the leader region of the *cyc1* gene. The mutations caused translation from an alternate ATG start codon and lead to cold-sensitivity. Mutations in the *SUA7* gene were isolated as suppressors of this phenotype (28). Further characterization of the *SUA7* revertants revealed that the suppression mechanism involved transcription initiation from a secondary site, downstream of the alternate ATG codon. It was shown that *SUA7* encoded TFIIB (29) and that the altered transcription start site phenotype was elicited by either of two amino acid substitutions in the linker region between the zinc ribbon and the CTD, E62K and R78C. Since reverse of charge mutations at both positions eliminated the suppressor phenotype, the existence of a salt bridge between E62 and R78 was proposed. Similar charge-reversal experiments in the *Sulfolobus*

acidocaldarius TFB NTD (22) also showed linker residues are essential for transcription start site determination and promoter-specific transcription initiation in Archaea. Taken together these studies revealed that in addition to RNAP recruitment, the TF(II)B NTD and specifically the linker region is involved in additional molecular events such as promoter opening and the determination of the transcription start site. The linker region immediately C-terminal to the zinc-ribbon contains a highly conserved sequence block and was later termed the “B-finger” (now also called the B-reader) and contains many of the residues shown to be involved in determining the transcription start site in eukaryotic and archaeal TFBs (22, 29-32).

The position of the TFB N-terminus during transcription initiation

A series of photochemical cross-linking and structural studies beginning in 2003 have shed considerable light on the position of the TFB N-terminus in the context of the transcription initiation complex and have approximately defined the location of the conserved TFB linker region. The path of promoter DNA relative to the *Pyrococcus furiosus* initiation complex was first deduced by site-specific photochemical cross-linking (19, 33). As expected TFB cross-linked to the BRE, but also had extensive cross-links to DNA near the transcription start site. Subsequently, an Fe-BABE cleavage and hydroxy-radical cleavage study (34) showed that the TFIIB linker and CTD are located over the central cleft of RNAP. A portion of the linker region was glimpsed in 2004 when a yeast RNAP-TFIIB co-crystal was solved for the first time (23). This structure confirmed that the TFIIB zinc-ribbon domain contacts the RNAP “dock” and showed the TFB linker in a finger like conformation (“B-

finger”), which penetrated the core of RNAP via the RNA exit pore. This study and a subsequent Fe-BABE investigation (35) proposed that the B-finger, as suggested by the earlier genetic studies, may be the region of TFB N-terminus involved in molecular rearrangements following RNAP recruitment. Its location in the RNAP active site positions the B-finger in close proximity to promoter DNA regions near the transcription start site, which supports a role for the B-finger in promoter melting and/or transcription start site determination. It was also suggested the B-finger is poised to sterically clash with the nascent RNA strand past 8-9 nt, an event which may precipitate TFIIB release and promoter clearance. A more recent x-ray crystal structure confirmed the position of the B-finger region within the active cleft of RNAP and provided more specific information on its conformation (24). In this structure, many of the residues previously assigned to the B-finger are seen as an extension to the zinc-ribbon and form an α -helix and a mobile loop, referred to as the B-reader helix and B-reader loop, respectively, with proposed roles in template strand scanning, transcription start site determination and open complex formation. In addition, the structure revealed the presence of previously undetected electron density that represents a second α -helix C-terminal to the B-reader, termed the B-linker, proposed to also play a role in promoter melting given its likely proximity to the upstream edge of the transcription bubble (24).

Structure and function of Transcription factor E

TFE is a third archaeal transcription factor, and has homology to the amino-terminal domain of the α subunit of eukaryotic TFIIE (36). The structure of a

conserved winged-helix motif in the N-terminal half of archaeal TFE has been determined (37). TFE has been shown to stimulate transcription up to 3 fold at less than optimal TBP concentrations and in a promoter dependent fashion (36, 38, 39). Eukaryotic TFIIE α has been placed at the upstream edge of the transcription bubble by photochemical cross-linking (40) while the winged-helix of archaeal TFE has been shown to specifically cross-link to the non-template strand upstream of the transcription start site (41). TFIIE has been shown to enhance TBP-promoter binding (42) and facilitates the formation of the open complex by promoting transcription factor H (TFIIH) function, which has ATPase, CTD kinase and DNA helicase activities (43-46). Conserved core TFIIH subunits XPB and XPD have also been identified in archaeal species (45). Together these data suggest that like TFIIE, archaeal TFE might stimulate transcription by playing a role in promoter recognition, promoter melting or in the stability of the transcription bubble.

Overview of archaeal transcription initiation

Transcription in Archaea initiates at simple promoters, usually containing an A/T rich TATA box about 25 base pairs upstream of the transcription start site, with an adjacent, upstream TFB-recognition element (BRE) (17, 47, 48). During transcription initiation, pre-initiation complex formation begins when TBP binds the TATA box, followed by TFB, which binds the TBP-promoter complex and interacts with the BRE in a sequence specific manner (16, 49). The TBP-TFB-DNA complex recruits RNAP to the promoter via the TFB N-terminus which also plays a role in further initiation events (30-32, 50, 51). However, the specific role of the TFB N-

terminus in events following RNAP recruitment is unclear. TFE facilitates transcription in cases where TBP or TFB function is not optimal, at least in part by stabilizing the open complex, in which the DNA strands surrounding the transcription start site are separated (36, 38, 41, 52). However, the positioning and specific role of TFE in the archaeal transcription initiation complex are not yet clear.

The majority of archaeal genomes sequenced to date contain multiple ORFs encoding TFB. The genome of hyperthermophile *Pyrococcus furiosus* encodes two TFB paralogs (53, 54). The first of these (TFB1) has been well characterized. Previous work has shown the transcript of the second TFB (TFB2) is detectable *in vivo*, suggesting this gene is actively expressed, although the function of this alternate TFB had not been investigated. At the sequence level the two paralogs diverge in their N-termini. They share an evolutionarily conserved zinc-ribbon motif, found in both archaeal and eukaryotic TFBs. However, the B-reader, is absent from the N-terminus of TFB2.

I first characterized TFB2 by comparing its activity to TFB1 directed transcription using an *in vitro* system. In Chapter II I show that although less active, TFB2 is able to replace TFB1 in transcription in all promoters assayed, despite the absence of the B-reader. Therefore it appears the B-reader motif is not essential for TFB2 directed transcription (55). I have also examined the orientation and function of the B-reader in the PIC using a novel protein-DNA cross-linking approach and show in Chapter III that the B-reader cross-links to the template but not to the non-template strand. Using this approach I have examined the orientation of the B-reader helix and

loop under experimental conditions designed to isolate transcription initiation complexes at different steps in the transcription cycle. The data indicate that the position of the B-reader changes as the transcription initiation complexes transition from the closed to the open states and that the B-reader is repositioned in the presence of initiating nucleotides. The experiments presented in this thesis explore the hypothesis that the TFB linker region (including the B-finger/B-reader conserved sequence) is important for early steps in transcription initiation and may interact directly with DNA during the initiation process. The results confirm this hypothesis and provide details of proximity of the TFB N-terminus to DNA during the steps of initiation.

References

1. C. Woese, O. Kandler, M. Wheelis, *Proc Natl Acad Sci U S A* **87**, 4576 (Jun, 1990).
2. L. Aravind, E. V. Koonin, *Nucleic Acids Res* **27**, 4658 (Dec 1, 1999).
3. E. Perez-Rueda, J. Collado-Vides, *J Mol Evol* **53**, 172 (Sep, 2001).
4. Y. Korkhin *et al.*, *PLoS Biol* **7**, e102 (May 5, 2009).
5. A. Hirata, K. Murakami, *Curr Opin Struct Biol* **19**, 724 (Dec, 2009).
6. T. J. Darcy *et al.*, *J Bacteriol* **181**, 4424 (Jul, 1999).
7. D. Langer, J. Hain, P. Thuriaux, W. Zillig, *Proc Natl Acad Sci U S A* **92**, 5768 (Jun 20, 1995).
8. D. D. Leipe, L. Aravind, E. V. Koonin, *Nucleic Acids Res* **27**, 3389 (Sep 1, 1999).
9. N. A. Woychik, M. Hampsey, *Cell* **108**, 453 (Feb 22, 2002).
10. F. J. Asturias, *Curr Opin Struct Biol* **14**, 121 (Apr, 2004).
11. S. Buratowski, H. Zhou, *Proc Natl Acad Sci U S A* **90**, 5633 (Jun 15, 1993).
12. S. Malik, D. Lee, R. Roeder, *Mol Cell Biol* **13**, 6253 (Oct, 1993).
13. T. Gibson, J. Thompson, A. Blocker, T. Kouzarides, *Nucleic Acids Res* **22**, 946 (Mar 25, 1994).
14. D. B. Nikolov *et al.*, *Nature* **377**, 119 (Sep 14, 1995).

15. F. T. Tsai, P. B. Sigler, *Embo J* **19**, 25 (Jan 4, 2000).
16. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
17. S. A. Qureshi, S. P. Jackson, *Mol Cell* **1**, 389 (Feb, 1998).
18. W. Deng, S. G. Roberts, *Genes Dev* **19**, 2418 (Oct 15, 2005).
19. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
20. W. Zhu *et al.*, *Nat Struct Biol* **3**, 122 (Feb, 1996).
21. H. T. Chen, P. Legault, J. Glushka, J. G. Omichinski, R. A. Scott, *Protein Sci* **9**, 1743 (Sep, 2000).
22. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).
23. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
24. D. Kostrewa *et al.*, *Nature* **462**, 323 (Nov 19, 2009).
25. X. Liu, D. Bushnell, D. Wang, G. Calero, R. Kornberg, *Science* (Nov 12, 2009).
26. G. Kassavetis, G. Letts, E. Geiduschek, *EMBO J* **20**, 2823 (Jun 1, 2001).
27. G. Kassavetis, S. Han, S. Naji, E. Geiduschek, *J Biol Chem* **278**, 17912 (May 16, 2003).
28. I. Pinto, D. E. Ware, M. Hampsey, *Cell* **68**, 977 (Mar 6, 1992).
29. I. Pinto, W. H. Wu, J. G. Na, M. Hampsey, *J Biol Chem* **269**, 30569 (Dec 2, 1994).
30. C. S. Bangur, T. S. Pardee, A. S. Ponticelli, *Mol Cell Biol* **17**, 6784 (Dec, 1997).
31. N. A. Hawkes, R. Evans, S. G. Roberts, *Curr Biol* **10**, 273 (Mar 9, 2000).
32. N. A. Hawkes, S. G. Roberts, *J Biol Chem* **274**, 14337 (May 14, 1999).
33. M. B. Renfrow *et al.*, *J Biol Chem* **279**, 2825 (Jan 23, 2004).
34. H. T. Chen, S. Hahn, *Cell* **119**, 169 (Oct 15, 2004).
35. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
36. S. D. Bell, A. B. Brinkman, J. van der Oost, S. P. Jackson, *EMBO Rep* **2**, 133 (Feb, 2001).
37. A. Meinhart, J. Blobel, P. Cramer, *J Biol Chem* **278**, 48267 (Nov 28, 2003).
38. B. L. Hanzelka, T. J. Darcy, J. N. Reeve, *J Bacteriol* **183**, 1813 (Mar, 2001).
39. M. Ouhammouch, F. Werner, R. O. Weinzierl, E. P. Geiduschek, *J Biol Chem* **279**, 51719 (Dec 10, 2004).
40. D. Forget, M. F. Langelier, C. Therien, V. Trinh, B. Coulombe, *Mol Cell Biol.* **24**, 1122 (Feb, 2004).
41. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).
42. K. Yokomori, C. P. Verrijzer, R. Tjian, *Proc Natl Acad Sci U S A* **95**, 6722 (Jun 9, 1998).
43. T. I. Lee, R. A. Young, *Annu Rev Genet* **34**, 77 (2000).
44. Y. Ohkuma, S. Hashimoto, C. K. Wang, M. Horikoshi, R. G. Roeder, *Mol Cell Biol* **15**, 4856 (Sep, 1995).
45. Y. Ohkuma, R. G. Roeder, *Nature* **368**, 160 (Mar 10, 1994).

46. H. Serizawa, J. Conaway, R. Conaway, *J Biol Chem* **269**, 20750 (Aug 12, 1994).
47. M. Thomm, G. Wich, *Nucleic Acids Res* **16**, 151 (Jan 11, 1988).
48. M. S. Bartlett, *Curr Opin Microbiol* **8**, 677 (Dec, 2005).
49. S. D. Bell, P. L. Kosa, P. B. Sigler, S. P. Jackson, *Proc Natl Acad Sci U S A* **96**, 13662 (Nov 23, 1999).
50. C. S. Bangur, S. L. Faitar, J. P. Folster, A. S. Ponticelli, *J Biol Chem* **274**, 23203 (Aug 13, 1999).
51. T. S. Pardee, C. S. Bangur, A. S. Ponticelli, *J Biol Chem* **273**, 17859 (Jul 10, 1998).
52. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
53. M. Micorescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
54. K. R. Shockley *et al.*, *Appl Environ Microbiol* **69**, 2365 (Apr, 2003).
55. M. Micorescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).

CHAPTER II

THE FUNCTION OF TFB2 – AN ALTERNATE PYROCOCCLUS FURIOSUS TRANSCRIPTION FACTOR

Abstract

The genome of the hyperthermophile archaeon *Pyrococcus furiosus* encodes two transcription factor B (TFB) paralogs, one of which (TFB1) was previously characterized in transcription initiation. The second TFB (TFB2) is unusual in that it lacks recognizable homology to the archaeal TFB/eukaryotic TFIIB B-finger motif. TFB2 functions, though poorly, in promoter-dependent transcription initiation. Domain swaps between TFB1 and TFB2 showed that the low activity of TFB2 is determined mainly by its N terminus. The low activity of TFB2 in promoter opening and transcription can be partially relieved by transcription factor E (TFE). The results indicate that the TFB N-terminal region, containing conserved Zn ribbon and B-finger motifs, is important in promoter opening and that TFE can compensate for defects in the N terminus through enhancement of promoter opening.

Introduction

Transcription in archaea is catalyzed by a single RNA polymerase (RNAP) that is very similar to eukaryotic RNAP II at the level of subunit identity and sequence homology (1, 2). Initiation of transcription by archaeal RNAP is guided by at least three extrinsic factors, TATA binding protein (TBP), transcription factor B

(TFB), and transcription factor E (TFE), which display high levels of structural and functional conservation with their eukaryotic counterparts, TBP, TFIIB, and the TFIIE α subunit (3-8). Archaea apparently lack homologs of other RNAP II transcription initiation factors.

Transcription in archaea initiates at simple promoters, usually containing an AT-rich TATA box about 25 bp upstream of the transcription start site, with an adjacent TFB recognition element (BRE) (9-11). During transcription initiation, complex formation begins when TBP binds the TATA box, followed by TFB, which binds the TBP-promoter complex and interacts with the BRE in a sequence-specific manner (5, 12, 13). The TBP-TFB-DNA complex recruits RNAP to the promoter, and transcription initiates. TFE facilitates transcription in cases where the TBP or TFB function is not optimal, at least in part by stabilizing the open complex, in which the DNA strands surrounding the transcription start site are separated (3, 4, 14-16).

TFB in archaea and TFIIB in eukaryotes play a central role in recruiting RNAP and may also be involved in facilitating the structural rearrangements in the transcription complex that lead to initiation, but a detailed mechanism of action has not been determined for this transcription factor family. Like TFIIB, TFB contains a structurally complex, conserved N-terminal region that is connected by a linker to a globular C terminus. The C-terminal two-thirds of TFB contain a helix-turn-helix motif that mediates the sequence-specific recognition of the BRE, as well as surfaces that interact with TBP and make nonspecific DNA contacts downstream of the TATA box (13). The N terminus of TFB is close to the transcription start site, as

shown by photochemical cross-linking experiments (17, 18). Archaeal TFB and eukaryotic TFIIB N-terminal regions usually contain two conserved motifs, the zinc ribbon and the B-finger, which are important in RNAP recruitment and transcription start site selection (19, 20). The zinc ribbon interacts with the RNAP “dock” domain during RNAP recruitment (16, 21, 22), but the specific function of the B-finger in the transcription mechanism is not clear. Yeast RNAP II/TFIIB cocrystal and DNA-tethered Fe-BABE protein cleavage studies have indicated that the B-finger enters the RNAP main channel and is close to transcribed strand DNA immediately upstream of the transcription start site (21, 23). Therefore, this very highly conserved part of TFIIB and TFB may play a role in promoter opening or promoter escape by RNAP.

Two TFB paralogs, TFB1 and TFB2, are encoded by the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*, but the transcription activity of only TFB1 has been characterized (14, 24). TFB2 transcript levels rise following heat shock, suggesting that the TFB2 polypeptide is expressed and may be involved in the response to heat stress (24). The *tfb2* locus encodes a 283-amino-acid protein that is similar to TFB1 and other members of the TFIIB family (Figure 2. 1). The C terminus of TFB2 (amino acids 73 to 283) is 63% identical to the C terminus of TFB1 (amino acids 86 to 300) and is highly conserved in the helix-turn-helix motif that recognizes the BRE; two of the three amino acids that make base-specific contacts are identical. However, the TFB2 N terminus is not as well conserved; the putative zinc ribbon-containing portion of the N terminus (amino

acids 17 to 49) displays just 45% identity to the TFB1 zinc ribbon region (amino acids 7 to 39), and there is no recognizable B-finger motif.

Many archaeal species encode multiple TFBs. Most of the TFB sequences contain recognizable B-finger motifs. For example, *Thermococcus kodakaraensis* TFB1 and TFB2 each contain B-finger motifs and are nearly equivalent functionally in vitro, as well as at least partially redundant in vivo (25). However, *P.furiosus* TFB2 naturally lacks the B-finger motif and thus presents a unique opportunity to study the functional importance of a highly conserved transcriptional element with incompletely defined roles. In light of previous data concerning the B-finger motif, we predicted that TFB2 might differ from TFB1 in events following assembly of the transcription initiation complex. Here we investigated the role of TFB2 in the formation of transcription complexes using transcription assays and used domain swapping to define the regions of TFB1 and TFB2 important for transcription activity.

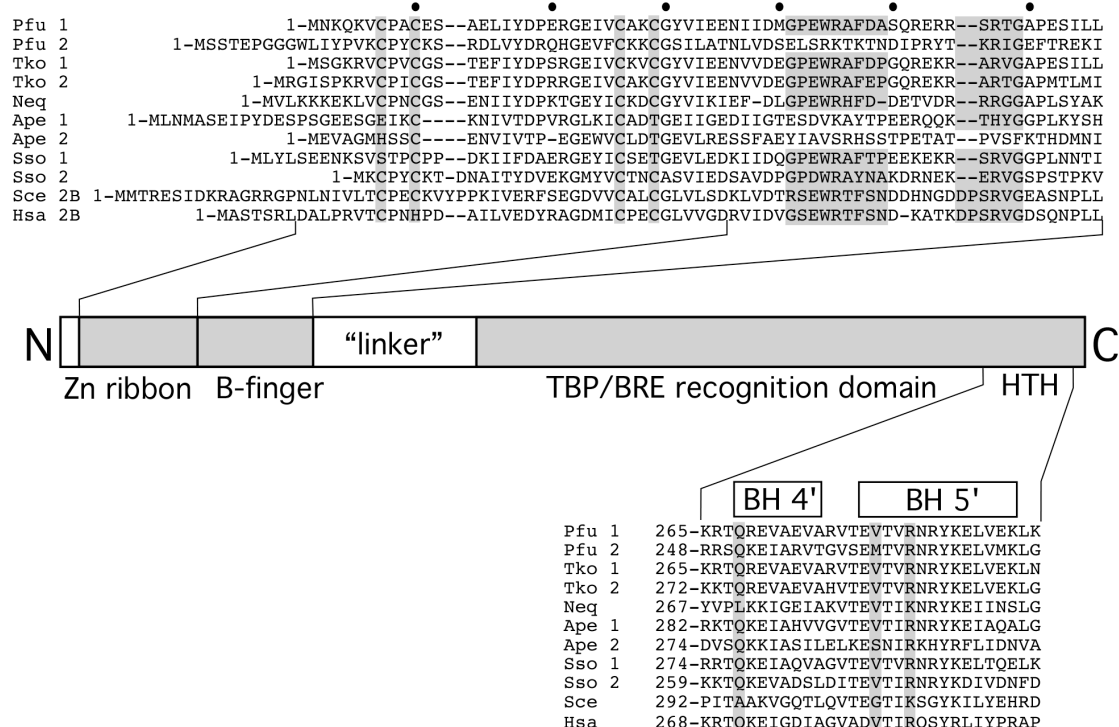


Figure 2.1 - Partial alignment of archaeal TFBs and eukaryotic TFIIBs. The four Cys/His residues defining the Zn ribbon motif are shaded, as are the conserved sequences defining the B-finger motif (7). *P. furiosus* TFB2 and both *A. pernix* TFBs lack homology to the B-finger sequence. A helix-turn-helix (HTH) motif makes sequence-specific contact with the BRE in the *P. furiosus* TBP-TFB1-DNA cocrystal; the amino acids responsible for these contacts (Q268, V280, R283, and the aligned amino acids) are indicated by shading. *Pfu*, *Pyrococcus furiosus*; *Tko*, *Thermococcus kodakarensis*; *Neq*, *Nanoarchaeum equitans*; *Ape*, *Aeropyrum pernix*; *Sso*, *Sulfolobus solfataricus*; *Sce*, *Saccharomyces cerevisiae*; *Hsa*, *Homo sapiens*; 1, TFB1; 2, TFB2; 2B, TFIIB. Reprinted from (14)

Results

Transcription by TFB2: activity and promoter selectivity.

We asked whether complexes formed by TFB2 were able to initiate transcription. Transcription reactions containing *gdh* promoter DNA, TBP and RNAP were performed under standard conditions in the absence or presence of saturating concentrations of TFB1 or TFB2 (Figure 2. 2). We observed that TFB2 directs

transcription from the same start site as TFB1, but with lower efficiency. Increasing TFB2 or TFB1 concentration did not increase transcription efficiency, confirming that both TFB concentrations were saturating (data not shown). One possible function for alternative TFBs is to selectively transcribe different promoters when present, analogous to the function of alternate sigma factors in bacteria (26). To determine whether TFB2 could function at other promoters, and to further compare its activity to TFB1, we used additional promoter regions whose genes are likely to be highly expressed or whose genes are expressed under heat shock conditions. Several criteria influenced our choice of novel promoters: 1) Predicted highly expressed genes (PHX) (27, 28); 2) *P.furiosus* microarray data (29, 30); 3) rRNA genes; 4) tRNA genes; 5) promoters containing recognizable TATA boxes; and 6) sequence conservation among three sequenced *Pyrococcus* species genomes. Thermosome genes (encoding chaperonins in the GroEL—Hsp60 family) are the most highly PHX genes in archaea (28).

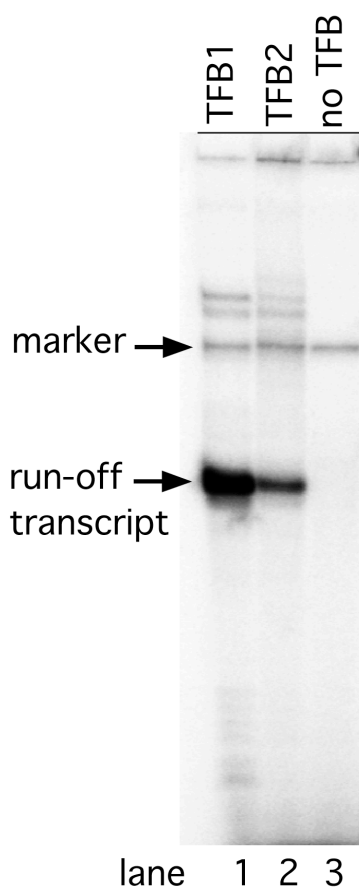


Figure 2.2 – Transcription of the *gdh* promoter using TFB1 or TFB2. In vitro transcription reaction mixtures were assembled using saturating levels of TFB1 or TFB2, as described in Materials and Methods. The position of the runoff transcript (37 nucleotides) is indicated by an arrow. A radiolabeled recovery marker (marker) was present in each lane. The weak, higher-molecular-weight bands in lanes 1 and 2 were likely the result of end-to-end template switching by RNAP. Reprinted from (14)

Figure 2.3a shows one example of aligned promoter regions for four euryarchaeal thermosome genes (3 *Pyrococcus* species and one closely related *Thermococcus* species). In addition to the highly conserved and readily identifiable TATA box and BRE promoter elements in thermosome gene upstream regions (labelled), the thermosome gene transcript was highly expressed in a microarray investigation of *Pyrococcus furiosus* ORFs (29). Taken together, these characteristics strongly suggested that this promoter would actively direct *in vitro* transcription, and thus serve as a second test case for comparing the transcriptional activity of the two TFBs.

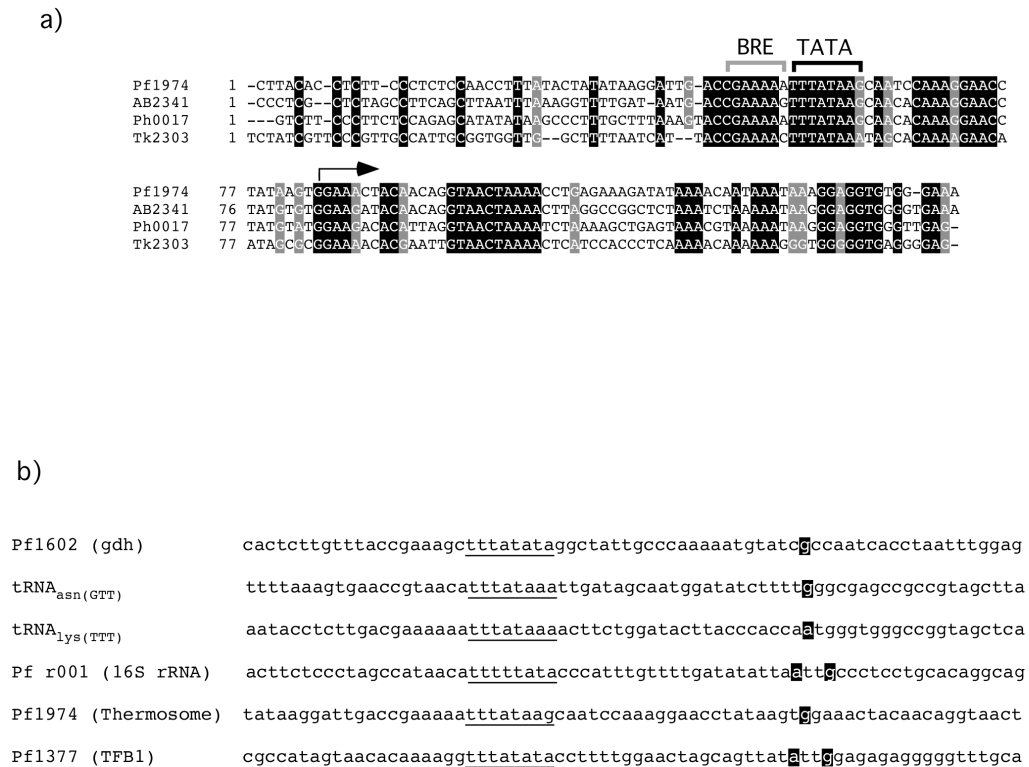


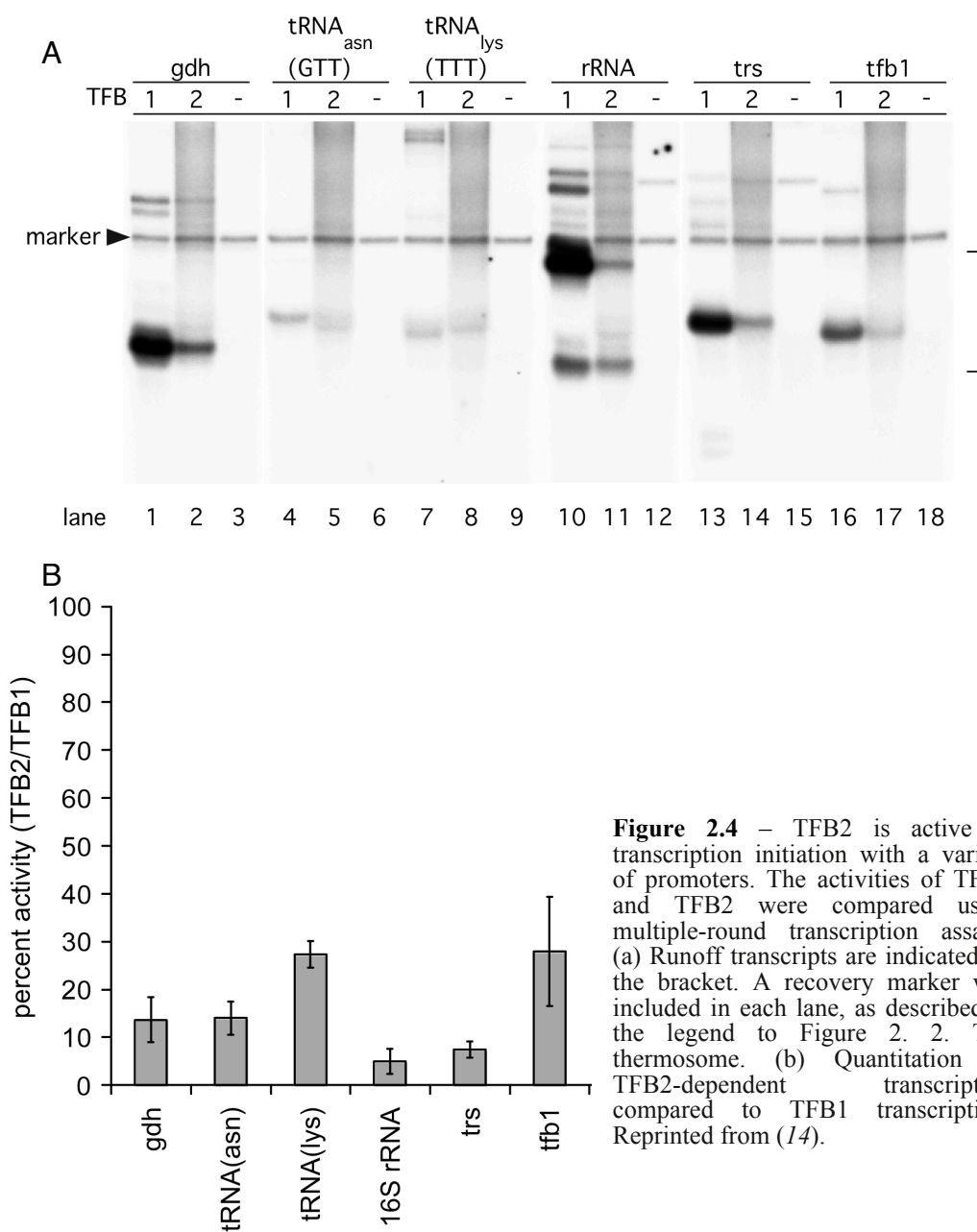
Figure 2.3 - *P. furiosus* promoters for comparing TFB1 and TFB2 activities. (a) Alignment of DNA regions immediately upstream of the initiating ATG of the thermosome gene from *P. furiosus* (Pf1974), *P. abyssi* (AB2341), *P. horikishii* (Ph0017), and *T. kodakarensis* (Tk2303). (b) Portions (70 bp) of the promoter sequences employed as transcription templates in this study. Predicted TATA boxes are underlined, and the potential transcription start sites are indicated by a black background. The predicted runoff transcript sizes based on initiation at the underlined start sites are as follows: Pf1602, 37 bp; tRNA_{Asn}(GTT), 45 bp; tRNA_{Lys}(TTT), 46 bp; Pf r001, 66 or 69 bp; Pf1974, 46 bp; Pf1377, 41 or 44 bp; Pf1882, 165 bp; Pf1883, 145 bp; and Pf1790, 135 bp. Reprinted from (14)

Portions of the intergenic, putative promoter regions that we selected for further study are shown in Figure 2.3b, aligned for comparison with the well-characterized glutamate dehydrogenase promoter. The two putative tRNA promoters contain sequences conserved among *Pyrococcales*, and direct the synthesis of tRNAs whose codons are highly abundant in *P. furiosus* (27, 28). Expression of the 16S rRNA gene is expected to be very high, in order to support the translational capacity required for the high growth rates achieved by *P. furiosus* (31). We identified the putative 16S

rRNA promoter by searching upstream of the 16S rRNA start (Genbank ncRNA annotation for Pf r001: bp 136,930 of the *P. furiosus* genome), and by comparing these regions from *P. furiosus*, *P. horikoshii*, and *P. abyssi*. The three genomic regions corresponding to *P. furiosus* sequences 136687 to 136930 had essentially identical sequences with high G/C content, and no clear TATA boxes. This region may represent rRNA leader sequences that are processed during rRNA maturation. Upstream of *P. furiosus* bp 136687, the three genomes are less similar, though there are conserved regions, with a likely TATA box for binding to TBP and nucleating transcription complex formation. Therefore, we predicted that *P. furiosus* sequences 136562-136687 contained the 16S rRNA promoter. The putative promoter for the TFB1 gene was chosen for analysis since it is highly conserved among *Pyrococcales*, and contains a canonical TATA box. The *gdhP* transcription start site is known (Figure 2. 4b, highlighted), and we identified potential starts site of the other promoters as purines approximately 30 bases downstream from the first T of the likely TATA box (Figure 2. 4b).

We compared the activities of TFB1 and TFB2 on six promoters in a standard transcription assay. TFB2 directs transcription from each template, but it is less efficient than TFB1 (Figure 2.4, panel a and data not shown). Quantitation of the primary transcript band intensity indicates that TFB2-dependent transcription of each promoter remains below 30% of that seen with TFB1 (Figure 2. 4b). The level of transcription achieved with both TFBS on the tRNA promoters was lower than from

the other 4 templates, and TFB2 directed transcripts have slightly shifted transcription start sites.



One hypothesis for TFB2 function is that it is a heat-shock specific transcription factor that may preferentially transcribe heat shock genes via selective promoter utilization. If this occurred, we predicted that TFB2 activity on heat-shock promoters would be higher than for non-heat-shock promoters, which would be observed as little or no reduction in transcription compared to TFB1. However, at each of the heat shock-inducible promoters tested (*trs* in Figure 2.4 and *hsp* in (14)), TFB2 was less efficient than TFB1 at directing transcription initiation (Figure 2.4b), suggesting that TFB2 does not selectively transcribe heat-shock genes when present in the cell.

Domain swapping between TFB2 and TFB1.

TFB2 is nearly as efficient as TFB1 in forming initiation complexes with TBP and RNAP as determined by photochemical cross-linking experiments (14), so we asked whether events following recruitment of RNAP were altered with TFB2, perhaps accounting for the low activity of TFB2 in transcription assays. The low activity of TFB2 relative to TFB1 could be determined by divergent amino acid sequence in the N-terminus, or by subtle but potentially important amino acid changes in the C-terminal region responsible for interactions with TBP and the BRE (Figure 2.1). To test this, we swapped the N-terminal regions of TFB1 and TFB2, and examined transcription of the *gdh* promoter using the hybrids. The TFB2 N-terminal region confers low activity on the TFB1 C-terminal region (hybrid 2::1), while the TFB1 N-terminal region confers high activity on the TFB2 C-terminal region (hybrid 1::2) (Figure 2. 5, lanes 3 and 4). To investigate whether the low activity conferred by the

TFB2 N-terminus is caused by the lack of a B-finger sequence motif, we swapped amino acid sequences encompassing the B-finger and linker regions between TFB1 and 2, and examined the transcriptional activity of these hybrids with the *gdh* promoter (Figure 2. 5, lanes 5 and 6). Both 2bf and 1bf had low activity, indicating that neither the TFB1 B-finger nor the Zn ribbon motif is alone sufficient to confer high activity to TFB2, and implying that the TFB1 B-finger requires its own Zn ribbon or nearby sequences for full function.

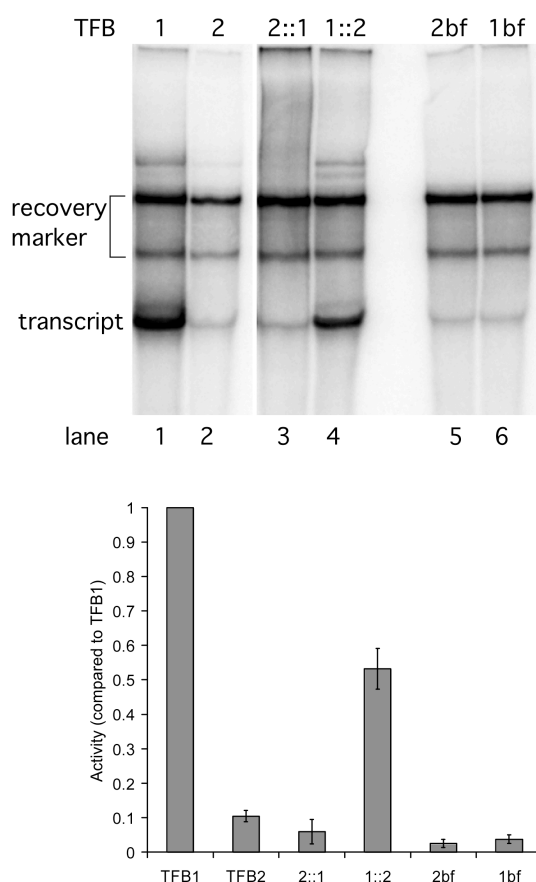


Figure 2.5 – Comparison of transcription activities of hybrid TFBS. Multiple-round transcriptions using the *gdh* promoter (positions –60 to 37) were performed with TFB1, TFB2, hybrid 2::1 (TFB2 amino acids 1 to 83 fused to TFB1 amino acids 101 to 300), hybrid 1::2 (TFB1 amino acids 1 to 100 fused to TFB2 amino acids 84 to 283), hybrid 2bf (TFB1 substituted with the TFB2 linker region), and hybrid 1bf (TFB2 substituted with the TFB1 B-finger and linker regions). The positions of the transcript and recovery marker are indicated. The levels of transcripts determined in several experiments are indicated in the bar graph (error bars, ± 1 standard deviation). Reprinted from (14).

We next asked whether TFE could compensate for the low transcriptional activity of TFB2 or any of the hybrids described in Figure 2. 5. Addition of TFE has little to no effect on transcription of the *gdh* promoter in the presence of TFB1 (Figure 2. 6, lanes 1 and 2), consistent with previous results (15). However, TFE activates transcription in the presence of TFB2 approximately 2-fold (Figure 2. 6, lanes 3 and 4), consistent with the compensatory role for TFE shown previously (Micorescu et al, Figs. 6 and 7). TFE also activates transcription in the presence of each of TFB hybrids, except for 1bf, the TFB2 variant engineered to contain the TFB1 B-finger motif (Figure 2. 6, lanes 5-12).

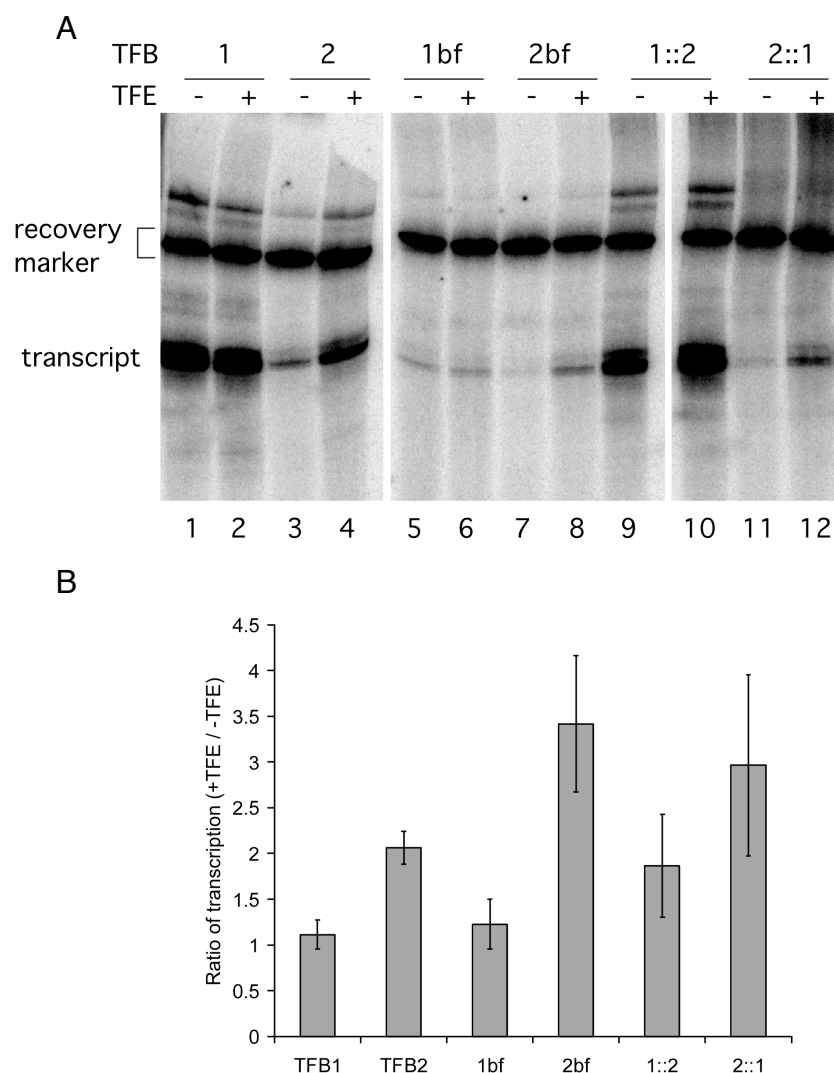


Figure 2.6 - Activation of TFB hybrids by TFE. Multiple-round transcriptions were performed as described in the legend to Figure 2. 5 in the absence or presence of TFE (240 nM). The positions of the transcript and recovery marker are indicated. The transcript levels from several experiments were quantitated, and the activation ratio for TFE with each TFB variant is indicated in the bar graph (error bars, ± 1 standard deviation). Reprinted from (14).

In addition, the magnitude of activation by TFE was marginally higher for 2::1 compared to 1::2 (compare lanes 11 and 12 to lanes 9 and 10, and see graph). Taken together, the data show that the presence of a B-finger motif in TFB reduces or outright prevents TFE-dependent activation of transcription at the *gdh* promoter.

Discussion

Evolutionary distribution of the B-finger motif.

The *P. furiosus* genome encodes two TFIIB family proteins, TFB1 and TFB2. Alignment of the *P. furiosus* TFBs with other archaeal TFBs indicated that TFB1 is most closely related to other TFBs, while the sequence of TFB2 has diverged (Figure 2. 1), particularly in the N-terminal one-third of the protein, suggesting that there is functional specialization of the two proteins. Interestingly, the TFBs from other *Pyrococcus* and *Thermococcus* species with more than one TFIIB family gene all contain the conserved B-finger sequence (Figure 2. 1). Thus, the TFB2 from *P. furiosus* is unusually divergent for this archaeal clade.

The B-finger is very highly conserved in TFIIB family members. TFB orthologs present in currently sequenced archaeal genomes almost always contain the B-finger, confirming its importance in transcription initiation. Notable exceptions, in addition to *P. furiosus* TFB2, include both TFBs encoded by the *Aeropyrum pernix* genome (Figure 2. 1) and several TFB orthologs encoded by sequences found in Sargasso Sea and other metagenome collections of sequences (M. Micorescu and M. Bartlett, unpublished observations). Thus, evolution occasionally leads to the loss of B-finger sequences in some archaeal TFBs, but the physiological significance surrounding the presence or absence of this region is unknown. It may be that the unrelated sequences in the divergent TFBs fold into B-finger-like structures or that transcription complex formation without a B-finger is advantageous under certain circumstances. An analog of *P. furiosus* TFB2 in eukaryotes is Brf, an RNAP III

transcription factor that is homologous to TFIIB in its N-terminal half, possessing an N-terminal Zn ribbon but lacking a conserved B-finger sequence (32-36).

Functional impact of the B-finger motif.

The experiments described here indicate that *P. furiosus* TFB2 is active in promoter-dependent transcription initiation. In addition, crosslinking experiments indicate that TFB2 forms transcription initiation complexes whose orientation is similar to that of complexes formed by TFB1, despite the lack of a recognizable B-finger sequence (14). The results reported here are the first characterization of an archaeal TFB that has evolved to lack a B-finger and are consistent with previous reports indicating that all or part of the archaeal TFB B-finger can be deleted without abolishing transcription activity (16, 37). While TFB2 is a functional transcription factor, it is much less efficient in directing runoff transcription than TFB1. Since TFB2 forms transcription initiation complexes nearly as efficiently as TFB1 but does not form open complexes efficiently (14), the lower activity is likely related to a role for the B-finger or other N-terminal TFB1 segments in events following assembly. The transcriptional defect of TFB2 is compensated for by a preopened transcription bubble, which strongly supports the idea that the TFB2 N terminus is less efficient in guiding promoter opening by RNAP. We predict that the difference in transcription initiation between TFB1 and TFB2 is conserved in vivo. Although an increased temperature in vivo could affect the nature and magnitude of the difference through thermal destabilization of the system's components (particularly DNA), increased temperature is likely to be modulated by intracellular

macromolecular crowding, along with osmo and thermoprotectants, which together would contribute to solution conditions in vivo that are very different from those used in this study.

The low activity of TFB2 can be compensated for by the TFB1 N terminus, which contains both a Zn ribbon and a B-finger motif. The TFB1 B-finger alone does not confer high activity to TFB2, implying that there are interactions, either direct or indirect, between the B-finger and Zn-ribbon motifs in the TFB1 N terminus. Such interactions could be important in positioning the B-finger so that it can fully stimulate promoter opening and thus transcription activity. The N terminus of TFB was previously shown to be important for recruitment of RNAP (16, 38). Our data indicate an additional role for the N terminus in post-RNAP recruitment steps and that TFB2 is deficient in this role. However, we cannot exclude the possibility that TFB2 also has altered RNAP recruitment properties, since the functions of the Zn ribbon (required for RNAP recruitment) and the B-finger may be linked, as suggested by the very low activity of both B-finger region swap TFBs (Figure 2. 6).

Interplay between TFB and TFE.

TFE can partially compensate for TFB2-dependent defects in transcription, which is consistent with previous data showing that TFE helps compensate for TFB defects caused by mutation or deletion of the B-finger and Zn ribbon motifs (16). Since TFE activation is most efficient with TFB variants that are missing the B-finger motif, this suggests that there is redundancy of function for TFE and

the TFB B-finger, at least in the context of the strong *gdh* promoter analyzed here. It also suggests a possible mechanism for TFE that allows TFB2 to maintain its function in vivo. The N-terminal portion of TFE is very close to the upstream edge of the transcription bubble, analogous to the position of the alpha subunit of TFIIE in eukaryotic transcription complexes (23, 39, 40). Thus, TFE may stabilize open complexes through interactions with the nontranscribed strand at the upstream end of the transcription bubble. Alternatively, TFE may contribute to strand opening allosterically, by stabilizing the closed-jaw conformation of RNAP that characterizes stable open and transcribing complexes.

Role of multiple TFBs.

The physiological importance of two TFBs in *P. furiosus* is not clear. Other archaeal species contain multiple homologs of TFB and TBP. For example, the *Halobacterium* NRC-1 genome contains six TBP and seven TFB open reading frames (41), and a recent analysis of transcription factor-promoter interactions in *Halobacterium* indicated that different combinations of TBP and TFB specify transcription from different promoter classes and thus regulate the expression of specific gene sets in different environments (26). The *T. kodakaraensis* genome also encodes two TFBs, either of which may be deleted without affecting cell growth under laboratory conditions. Both of these TFBs function in transcription initiation in vitro, but there is no apparent promoter selectivity (37). TFB2 transcript levels rise when there is a heat-shock, suggesting parallels with σ^E and σ^{32} , which are regulon-specific sigma factors involved in the bacterial heat shock response (24, 42).

However, TFB2 does not preferentially transcribe three known heat shock-induced promoters. Each of these promoters was transcribed using TFB2, but they were transcribed less efficiently than they were with TFB1 (Figure 2. 4), suggesting that promoter utilization by TFB2 is not selective for stress response genes under *in vitro* conditions. If TFB2 directs transcription of specific subsets of genes, as alternative bacterial sigma factors do, it must do so under conditions or with promoters other than those tested here. An alternative possibility is that TFB2 has unique properties that make it useful for changes in gene expression in the presence of fluctuating temperatures. In vitro, TFB2 is not more stable than TFB1 to high-temperature treatments (the half-life of both proteins at 95°C is about 5 min [data not shown]), so it seems unlikely that the expression of TFB2 at high temperatures is related to its thermostability. It could be that the utility of TFB2 is related to deficient promoter opening. For instance, at an abnormally high temperature, TFB2-dependent transcription may be enhanced by thermal effects on promoter melting. Under such conditions, TFB2 could be directed to a specific subset of stress response genes by an unknown mechanism or accessory factor. A return to a normal temperature would be accompanied by a rapid shutoff of TFB2-dependent transcription through destabilization of open complexes, thus preventing expenditure of energy on a response that is no longer required. Such a feedback system could provide an advantage in the fluctuating thermal environment of a marine hydrothermal vent.

Experimental Procedures

Gene cloning and protein purification. Recombinant *P. furiosus* TBP was prepared as described previously (43). *P. furiosus* TFB genes (Pf1377 for TFB1 and Pf0687 for TFB2) were amplified by PCR and cloned into the vector pET21b-H6-Nco (44), creating constructs that encoded proteins with a six-histidine tag at the N terminus. The overexpressed proteins were purified to near homogeneity (estimated by gel electrophoresis and Coomassie staining) by Ni⁺⁺ ion chromatography. Native RNAP used for the experiments whose results are shown in Figure 2. 2, 5b, 6, and 7 was purified from *P. furiosus* cells as described by Hethke et al. (43), while the native RNAP used for the experiments whose results are shown in Figure 2. 3, 5a, 8, and 9 was purified by the method described by Korkhin et al. (45).

Promoter DNA templates. Several criteria were used to identify promoters used in this study. For example, tRNA genes encoding tRNAs for abundant codons were predicted to have strong promoters to accommodate translation needs; the promoter for the single rRNA operon was predicted to be strong to accommodate the high growth rates attained by *P. furiosus*; and the TFB1 promoter was predicted to be strong because of its consensus TATA box. A more specific approach involved several genes that were predicted to be highly expressed (PHX genes) in *Pyrococcus abyssi* and *Pyrococcus horikoshii* (species closely related to *P. furiosus*) since their codon usage is similar to that of known highly expressed genes (27). The transcript abundance determined in microarray studies of *P. furiosus* gene expression provided independent support for high levels of expression of some *Pyrococcus* sp.

PHX genes (29, 30). Open reading frames with PHX genes that are preceded by intergenic (and presumably promoter-containing) sequences were chosen for further analysis. Putative promoter regions from *P. furiosus*, *P. horikoshii*, and *P. abyssi* were aligned using ClustalX. The alignments were examined for the presence of conserved TATA boxes and BREs. Predicted promoters were amplified from *P. furiosus* genomic DNA by PCR, as follows (sequence positions according to the annotation of Robb et al. (46)): for Pf1602, positions 1494929 to 1495025 (forward: 5' AAAGGATTTCCTACTCTTGTTTACCGAAAG, reverse: 5' CTCAACCATGTTTCATCCCTCCA); for tRNA^{Asn}(GTT), positions 1287414 to 1287519 (forward: 5' ATCACGAAGAGTTTTAAAGTGAAC reverse: 5' ATTAACAGTCCGGCGCTC); for tRNA^{Lys}(TTT), positions 508388 to 508493 (forward: 5' TGTTTATCACAATACCTCTTGACG reverse: 5' TTAAAAGCCCGGTGCTCTA); for Pf r001 (16S rRNA), positions 136562 to 136687 (forward: 5' CATACTAACTTCTCCCTAGCCAT reverse: 5' CACCCTATAGATAGCGAACCC); for Pf1974, positions 1823493 to 1823598 (forward: 5' CTTTATACTATATAAGGATTGACCGAA reverse: 5' ATTGTTTTATATCTTTCTCAGGTTTTAG); and for Pf1377, positions 1292896 to 1292998 (forward: 5' AGGTACTACCCGCCATAGTAACA reverse: 5' TTCACATCAACACCCCCAC). The promoters for Pf1882, Pf1883, and Pf1790 were described previously (47).

Standard transcription assays. Transcription reactions were performed essentially as described previously (17). The 12.5- μ l reaction mixtures contained 40 mM Na-

HEPES, pH 7.3, 250 mM NaCl, 2.5 mM MgCl₂, 0.1 mM EDTA, 5 mM beta-mercaptoethanol, 5% (v/v) glycerol, and 0.1 µg µl⁻¹ bovine serum albumin; 10 nM promoter DNA was combined with 60 nM TBP, 60 nM TFB1, or TFB2 and 20 nM RNAP. Increasing the TFB concentration to values greater than 60 nM resulted in no increase in transcription, indicating that 60 nM is saturating for TFB under these conditions. The reaction mixture was overlaid with mineral oil and incubated at 65°C for 40 min. Heparin was added (to a concentration of 50 µg ml⁻¹), followed 30 s later by ribonucleotide triphosphates (500 µM GTP, CTP and ATP, 10 µM [alpha-32P]UTP at ~40 Ci mmol⁻¹) to initiate transcription. Reactions were stopped after 20 min by addition of 80 µl of stop solution (20 mM EDTA containing a radiolabeled DNA recovery marker at a known concentration). Nucleic acids were purified by phenol-chloroform treatment, followed by ethanol precipitation. The transcripts were resolved by gel electrophoresis and analyzed by phosphorimaging, essentially as described previously (47, 48)

Hybrid TFBs. Megaprimer PCR was used to create hybrid TFBs. For the 2::1 and 1::2 hybrids, the N-terminal domain (NTD) sequence of the first TFB was amplified using a C-terminal primer containing a 24-bp tail sequence complementary to the second TFB. In the second round of PCR this product was used as the forward megaprimer for amplifying the C-terminal domain (CTD) sequence of the second TFB, resulting in fusion of the NTD sequence of the first TFB to the CTD of the second TFB. The hybrid genes were cloned into pET21b- H6-Nco (44), which inserted a six-histidine tag at the N terminus of the hybrid polypeptide. The constructs were

transformed into *Escherichia coli* Bl21, and the recombinant protein was overexpressed and purified using standard Ni²⁺ ion chromatography methods. The TFB variants with swapped B-finger and linker regions were created in a similar way, but NTD swap constructs previously created were used as templates for megaprimer PCR designed to swap the zinc ribbons. For the 2::1 hybrid, TFB2 amino acids 1 to 83 were fused to the TFB1CTD (amino acids 101 to 300), while for the 1::2 hybrid TFB1 amino acids 1 to 100 were fused to the TFB2 CTD (amino acids 84 to 283). The 1Bf hybrid was a fusion of TFB2 amino acids 1 to 49, TFB1 amino acids 41 to 100, and TFB2 amino acids 84 to 283. The 2Bf hybrid was a fusion of TFB1 amino acids 1 to 40, TFB2 amino acids 50 to 83, and TFB1 amino acids 101 to 300.

References

1. T. J. Darcy *et al.*, *J Bacteriol* **181**, 4424 (Jul, 1999).
2. D. Langer, J. Hain, P. Thuriaux, W. Zillig, *Proc Natl Acad Sci U S A* **92**, 5768 (Jun 20, 1995).
3. S. D. Bell, A. B. Brinkman, J. van der Oost, S. P. Jackson, *EMBO Rep* **2**, 133 (Feb, 2001).
4. B. L. Hanzelka, T. J. Darcy, J. N. Reeve, *J Bacteriol* **183**, 1813 (Mar, 2001).
5. W. Hausner, J. Wettach, C. Hethke, M. Thomm, *J Biol Chem* **271**, 30144 (Nov 22, 1996).
6. T. L. Marsh, C. I. Reich, R. B. Whitelock, G. J. Olsen, *Proc Natl Acad Sci U S A* **91**, 4180 (May 10, 1994).
7. S. A. Qureshi, P. Baumann, T. Rowlands, B. Khoo, S. P. Jackson, *Nucleic Acids Res* **23**, 1775 (May 25, 1995).
8. T. Rowlands, P. Baumann, S. P. Jackson, *Science* **264**, 1326 (May 27, 1994).
9. J. R. Palmer, C. J. Daniels, *J Bacteriol* **177**, 1844 (Apr, 1995).
10. S. A. Qureshi, S. P. Jackson, *Mol Cell* **1**, 389 (Feb, 1998).
11. M. Thomm, G. Wich, *Nucleic Acids Res* **16**, 151 (Jan 11, 1988).
12. S. D. Bell, P. L. Kosa, P. B. Sigler, S. P. Jackson, *Proc Natl Acad Sci U S A* **96**, 13662 (Nov 23, 1999).
13. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
14. M. Micoreescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
15. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).
16. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
17. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
18. M. B. Renfrow *et al.*, *J Biol Chem* **279**, 2825 (Jan 23, 2004).
19. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 31624 (Oct 13, 2000).
20. I. Pinto, W. H. Wu, J. G. Na, M. Hampsey, *J Biol Chem* **269**, 30569 (Dec 2, 1994).
21. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
22. H. T. Chen, S. Hahn, *Mol Cell* **12**, 437 (Aug, 2003).
23. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
24. K. R. Shockley *et al.*, *Appl Environ Microbiol* **69**, 2365 (Apr, 2003).
25. T. Santangelo, L. Cubonova, C. James, J. Reeve, *J Mol Biol* **367**, 344 (Mar 23, 2007).

26. M. T. Facciotti *et al.*, *Proc Natl Acad Sci U S A.* **104**, 4630 (Mar 13, 2007).
27. S. Karlin, J. Mrazek, *J Bacteriol* **182**, 5238 (Sep, 2000).
28. S. Karlin, J. Mrazek, J. Ma, L. Brocchieri, *Proc Natl Acad Sci U S A.* **102**, 7303 (May 17, 2005).
29. G. J. Schut, S. D. Brehm, S. Datta, M. W. Adams, *Journal of bacteriology.* **185**, 3935 (Jul, 2003).
30. G. J. Schut, J. Zhou, M. W. Adams, *J Bacteriol* **183**, 7027 (Dec, 2001).
31. J. DiRuggiero, L. A. Achenbach, S. H. Brown, R. M. Kelly, F. T. Robb, *FEMS microbiology letters.* **111**, 159 (Aug 1, 1993).
32. J. Rollins, I. Veras, S. Cabarcas, I. Willis, L. Schramm, *Int J Biol Sci* **3**, 292 (2007).
33. A. Saxena, B. Ma, L. Schramm, N. Hernandez, *Mol Cell Biol* **25**, 9406 (Nov, 2005).
34. D. Alexander *et al.*, *J Biol Chem* **279**, 32401 (Jul 30, 2004).
35. X. Zhao, L. Schramm, N. Hernandez, W. Herr, *Mol Cell* **11**, 151 (Jan, 2003).
36. Y. Huang, E. McGillicuddy, M. Weindel, S. Dong, R. Maraia, *Nucleic Acids Res* **31**, 2108 (Apr 15, 2003).
37. T. J. Santangelo, L. Cubonova, C. L. James, J. N. Reeve, *J Mol Biol* **367**, 344 (Mar 23, 2007).
38. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).
39. D. Forget, M. Langelier, C. Therien, V. Trinh, B. Coulombe, *Mol Cell Biol* **24**, 1122 (Feb, 2004).
40. S. Grunberg, M. S. Bartlett, S. Naji, M. Thomm, *J Biol Chem* **282**, 35482 (Dec 7, 2007).
41. N. S. Baliga *et al.*, *Mol Microbiol* **36**, 1184 (Jun, 2000).
42. P. E. Rouviere *et al.*, *Embo J* **14**, 1032 (Mar 1, 1995).
43. C. Hethke, A. C. Geerling, W. Hausner, W. M. de Vos, M. Thomm, *Nucleic Acids Res* **24**, 2369 (Jun 15, 1996).
44. G. A. Kassavetis, A. Kumar, E. Ramirez, E. P. Geiduschek, *Mol Cell Biol* **18**, 5587 (Sep, 1998).
45. Y. Korkhin, O. Littlefield, P. J. Nelson, S. D. Bell, P. B. Sigler, *Methods Enzymol* **334**, 227 (2001).
46. F. T. Robb *et al.*, *Methods Enzymol* **330**, 134 (2001).
47. G. Vierke, A. Engelmann, C. Hebbeln, M. Thomm, *J Biol Chem* **278**, 18 (Jan 3, 2003).
48. M. S. Bartlett, T. Gaal, W. Ross, R. L. Gourse, *J Bacteriol* **182**, 1969 (Apr, 2000).

CHAPTER III

AN INVESTIGATION INTO THE POSITION AND ORIENTATION OF THE ARCHAEAL B-READER REGION USING PROTEIN-DNA CROSS-LINKING

Abstract

Archaeal RNA polymerase requires two transcription factors for initiation: TBP, which binds to TATA boxes, and TFB, which binds TBP and DNA, recruits RNAP and helps initiate transcription. Archaeal TFBs usually contain a conserved B-reader sequence homologous to the eukaryotic B-reader (also called the B-finger) motif in their N-terminal domains. This region is involved in the assembly of the transcription complex, promoter melting and in transcription start site determination but its position and orientation relative to promoter DNA during initiation is not clear. In this study the positioning of the TFB B-reader relative to DNA was determined by cross-linking using TFB variants substituted with photoactivatable unnatural amino acids. The results demonstrate that the B-reader is in close proximity to the transcription start site on the template but not the non-template strand in transcription initiation complexes. Furthermore, the position of the B-reader varies between closed and open promoter complexes, and between open promoter and early initiation complexes. Thus the archaeal B-reader sequence is poised to interact with promoter DNA in a dynamic fashion, and is likely playing a role in positioning the template-strand in an open pre-initiation complex.

Introduction

The archaeal transcriptional apparatus is structurally and functionally homologous to its eukaryotic counterpart (1-7) despite the fact that these two domains of life are thought to have separated phylogenetically more than 3 billion years ago (10). It is thus most likely that both basal transcriptional systems have evolved from the same ancestral proteins. Archaeal TFB shares a high degree of sequence and structural homology to eukaryotic TFIIB, which functions in RNAP II transcription initiation. Like TFIIB and the N-terminal half of Brf, archaeal TFB contains two separate structural domains with distinct roles in the initiation process (11, 12). The C-terminal domain (CTD) consists of two cyclin repeats with a highly conserved helix-turn-helix motif involved in sequence specific recognition of a second promoter element, the BRE (13). Additional CTD surfaces are involved in non-specific interactions with promoter DNA as well as TBP. Thus the C-terminal domain recognizes the promoter of a gene and positions the TFB N-terminal domain (NTD) in close proximity to the transcription start site as shown by photochemical cross-linking (Figure 3.1, Step 1) (14, 15). Connected to the CTD by a flexible linker, the NTD plays a key role in recruiting RNAP to the promoter, making it essential in the formation of the RFP_C (Figure 3.1, Step 2) (16). Early genetic studies in yeast identified a conserved TFIID NTD sequence involved in the recruitment process (16, 17). NMR studies of the conserved *Pyrococcus* TFB NTD, show that this region folds into a Zn-ribbon motif (18). In both yeast and archaea this structure directly contacts

the dock domain of RNAP and mutations in this region abolish RNAP recruitment (19, 20).

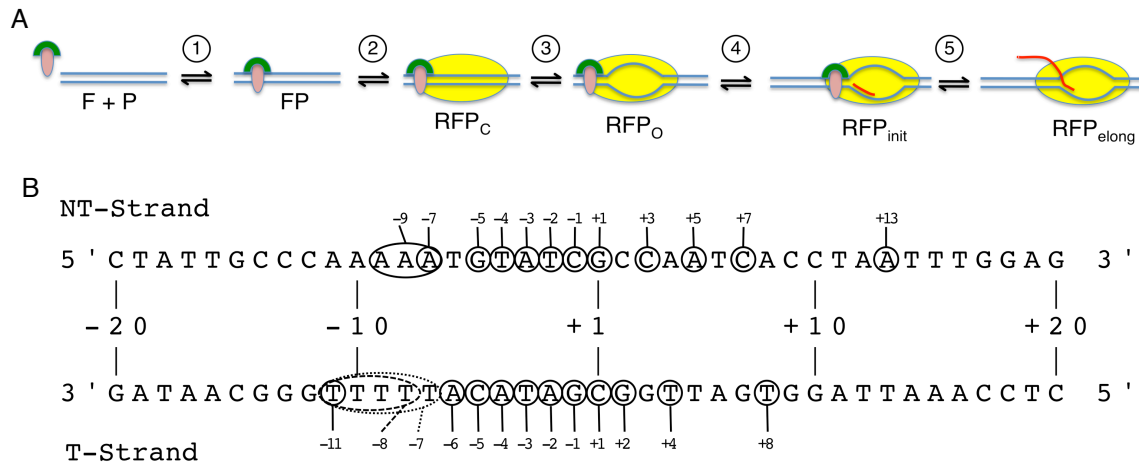


Figure 3.1. An overview of archaeal transcription initiation. **A.** The sequence of events in archaeal transcription initiation. The transitions are numbered according to the position in the sequence. Archaeal transcription begins with general transcription factors TBP and TFB, (designated “F” in Figure 3.1) recognizing the promoter of a gene, here denoted by “P”. Upon factor binding (Step 1), the DNA-TBP-TFB complex (FP,) recruits RNAP (yellow oval), a process mediated by the TFB N-terminus (Step 2) which leads to the formation of the closed transcription initiation complex (RFP_C). RNAP conformational changes, aided by interactions with the TFB B-reader, lead to the opening of the promoter DNA (Step 3), the loading of the template strand into the active site and the formation of the open transcription initiation complex (RFP_O). Productive RNA synthesis commences after a number of short abortive RNAs are released (Step 4) as the RNA strand sterically clashes with the TFB NTD in the mature transcription initiation complex (RFP_{init}). A stable RNAP elongation complex (RFP_{elong}) results after promoter clearance (Step 5). **B.** A section from the promoter region of the glutamate dehydrogenase gene from *Pyrococcus furiosus*. The radiolabeled probes used in this study are shown above the sequence (NT-Strand) or below (T-Strand) and for each probe the incorporated radiolabel is circled. Some probes (-9NT, -8T and -7T) contain multiple radiolabels as indicated by the ovals.

A large body of genetic and biochemical evidence suggests the TFB NTD is also important in steps following RNAP recruitment (Figure 3.1, Steps 3,4 and 5) and that the conserved sequences within the NTD C-terminal to the Zn ribbon may play a crucial role in transcription initiation by facilitating promoter melting or clearance. However, we currently lack a molecular explanation for this very key role in transcription initiation. Three yeast RNAP II/TFIIB co-crystals (8, 21, 22) and a

protein-tethered Fe-BABE cleavage study (23) show a structure called the eukaryotic B-finger (also called the B-reader) is located deep within the active cleft of the RNA polymerase. The 4.3 Å resolution crystal structure by Cramer and colleagues reveals that the TFIIB N-terminus enters the RNAP via the presumed RNA exit pore, and that the B-reader is located in close proximity to the active site (21). In the Cramer structure, the B-reader residues, previously assigned to the B-finger, form an α -helix termed the B-reader helix (residues 38 to 49 in *P.furiosus* TFB1) and a flexible B-reader loop (residues 50 to 63 *P.furiosus* TFB1). This suggests the possibility that the B-reader helix may play roles in DNA start site scanning while the B-reader loop may contribute to the stability of RFP_O (Figure 3.1, Step 3) and interfere with the growing RNA strand (Figure 3.1, Step 4). Although the Cramer structure did not contain nucleic acids, their model in conjunction with Fe-BABE cleavage data suggests that the B-reader is located in close proximity to the template strand and may play a role in transcription start site determination or promoter opening. Furthermore, the B-reader is predicted to sterically clash with the RNA transcript, which could explain its proposed role in abortive transcription and promoter clearance.

We have previously investigated the role of the *P.furiosus* TFB1 N-terminal domain in archaeal transcription by functionally characterizing TFB2, a paralog of the well-characterized *Pyrococcus furiosus* TFB1 that naturally lacks the conserved B-reader sequence. Our results showed that despite the absence of the B-reader, TFB2 is able to replace TFB1 in transcription across the promoters assayed, although its activity is consistently lower than TFB1. Therefore it appears the conserved B-reader

sequence is not essential for TFB function in transcription (5). However, potassium permanganate footprinting experiments revealed that transcription complexes formed with TFB2 are deficient in promoter opening and that this deficiency can be overcome by assembly of transcription complexes on partially melted DNA templates (5). Therefore, the low activity of TFB2 can be attributed to an impaired transition from RFP_C to RFP_O or from a failure to maintain RFP_O (Figure 3.1, Step 3).

Domain swapping experiments between *P.furiosus* TFB1 and TFB2 show that the TFB1 N-terminal domain conferred much higher activity on the TFB2 C-terminal domain, indicating that the N-terminus of TFB2 is responsible for this activity. Although our previous analysis of TFB1 and TFB2 helped to roughly define the TFB regions important for transcriptional activity, this work did not specifically reveal a structural and functional role for the B-reader during the steps defining the archaeal transcription initiation cycle, since the position and mobility of the B-reader during the initiation cycle are unknown. In order to answer this important question, we replaced three key TFB1 residues from the putative B-reader helix (W44 p -Bpa and T58 p -Bpa) and B-reader loop (R52 p -Bpa) with a cross-linkable amino acid analogue, p -benzoyl L-phenylalanine (p -Bpa). This approach relies on a genetic system composed of modified tRNAs and aminoacyl synthetases designed to insert unnatural amino acids at specific positions (24). In combination with site-specific radiolabeled promoter DNA, this approach detects specific promoter DNA positions that may interact with the TFB B-reader. Our protein-DNA cross-linking approach complements structural studies performed in the absence of DNA with functional data, confirming the

proximity of the B-reader to promoter DNA and illustrating the dynamic orientation of the B-reader relative to the promoter as transcription initiates.

Results

Purified pBpa TFB1 mutants are active in transcription

To define the position of the TFB1 B-reader relative to DNA, we incorporated the unnatural amino acid *p*-Bpa into specific sites in TFB1 using the heterologous tRNA/aaRS system developed by Schultz and colleagues (25). Briefly, this method involves site-directed mutagenesis to replace a specific residue's codon with an amber codon. The engineered tRNA specifically recognizes the amber codon and inserts *p*-Bpa at that position. We first wanted to verify that the amber codon allowed incorporation of *p*-Bpa when the modified TFB was expressed in its presence. Figure 3.2A shows that the expression of the amber-mutant TFB occurs only if the media is supplemented with *p*-Bpa. Since incorporation of *p*-Bpa by the engineered tRNA is not 100% efficient, a 61 aa termination product is expected for R52*p*-Bpa (including the 6 additional residues of the 6x-His tag). The expression level of R52*p*-Bpa is lower than the expression seen for wildtype TFB1, most likely due to low efficiency recognition of the amber codon by the modified tRNA, leading to prematurely terminated TFB polypeptide (Figure 3.2B). The same approach was used to make TFB1 substituted with *p*-Bpa at positions 44, 58 and 192.

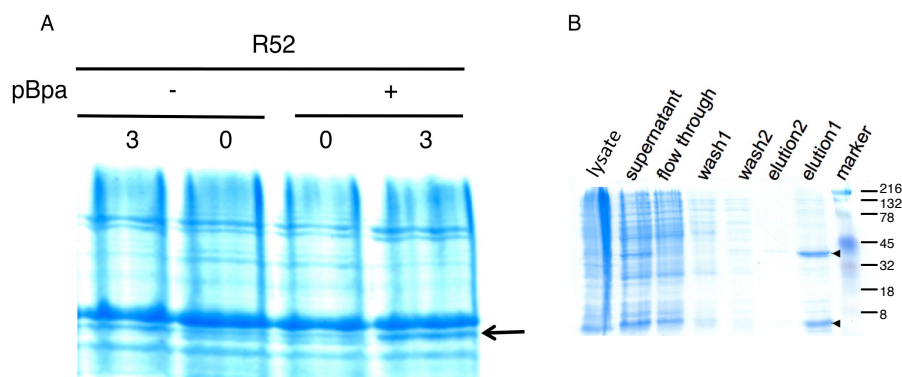


Figure 3.2. A. TFB1 R52*p*-Bpa expression is *p*-Bpa dependent. The gene cloning and mutagenesis protocols used to engineer the pBpa mutants are described in the methods section. For the lanes marked + expression was carried out in the presence of 1mM *p*-benzoyl-L-phenylalanine. Time samples were removed before induction at time 0 (lanes labeled 0) and after 3 hours of induction (lanes labeled 3). B. An example purification of TFB1 with pBpa inserted at R52. SDS-PAGE followed by Coomassie staining was used to follow the overexpression and purification of 6x-His tagged TFB using Ni^{++} chromatography. Recombinant, N-terminal 6xHis-tagged TFB1 with an amber codon engineered at R52 was overexpressed with *p*-Bpa inserted at the R52 position by an engineered tRNA/tRNA synthetase pair. The overexpressed full length TFB1 appears in the soluble fraction of the cell lysate (supernatant), and elutes at high purity from the Ni^{++} column (indicated by top arrow). The bottom arrow points to short polypeptides that failed to be extended past the amber at position 52.

We wanted to determine if TFBs containing *p*-Bpa are active in directing transcription initiation by RNAP. We compared the activity of W44*p*-Bpa and R52*p*-Bpa to unmodified TFB1 using a standard *in vitro* transcription assay (described in the methods section) using the *gdh* promoter.

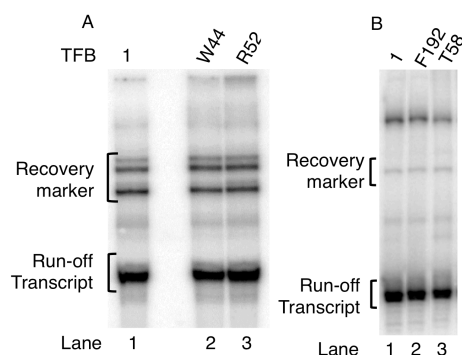


Figure 3.3. Comparison of the transcription efficiency of substituted TFBs versus wild type TFB1. Transcription reactions contained *gdh* promoter DNA (10 nM), TBP (20 nM), TFBs (60 nM), and RNAP (10 nM). Proteins and DNA were incubated at 65°C for 30 minutes to allow transcription complexes to form, NTPs were added (with $\alpha\text{-}^{32}\text{P}$ -UTP), and after 20 minutes reactions were stopped, processed, and analyzed by 12% PAGE. The run-off transcript (37 nucleotides) is indicated. The recovery marker is ^{32}P 3'-end-labeled DNA. **A.** W44*p*-Bpa and R52*p*-Bpa were compared to WT TFB1 **B.** F192*p*-Bpa and T58*p*-Bpa were compared to WT TFB1.

W44p-Bpa, R52p-Bpa, F192p-Bpa and T58p-Bpa cause RNAP to produce run-off transcript of correct size, at levels comparable to unmodified TFB1 (Figure 3.3).

pBpa protein-DNA cross-linking confirms the proximity of F192 to the template strand

In order to determine the feasibility of *p*-Bpa cross-linking, we chose to investigate the position of TFB1 F192*p*-Bpa relative to promoter DNA since the F192 position is part of the TFB1 C-terminal domain involved in promoter/TBP recognition, and a promoter DNA-TBP-TFB1 co-crystal structure shows this residue in close proximity to the -19 position of the transcribed strand ((9) and Figure 3.4A). A *gdh* promoter template was radio-labeled with ³²P specifically at this position on the template strand, complexes were assembled as described in the methods section, and the reactions were exposed to UV (365 nm) at 65 °C for 60 min. The reactions were then treated with nuclease to digest free DNA, as well as DNA covalently attached to protein by the initiation complexes. We reasoned that if the *p*-Bpa residue at the F192 position is located in close proximity to the -19 position on the template strand (Figure 3.4A), a short, radioactively labeled, DNA stub will remain cross-linked to the F192*p*-Bpa polypeptide after the nuclease treatment.

Figure 3.4B compares the cross-linking signal generated by reactions containing unmodified TFB1, F192*p*-Bpa or R52*p*-Bpa assembled on promoter labeled at -19T (“T” denotes the template strand) in the presence or absence of RNAP. In the absence

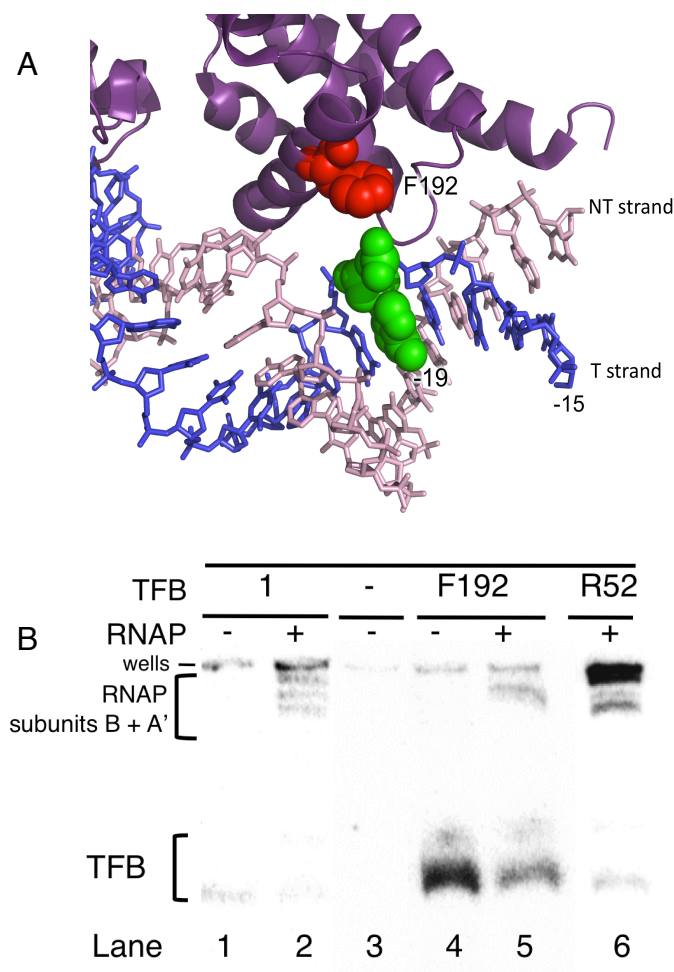


Figure 3.4. F192, a residue in the TFB C-terminus, is located in close proximity to -19 on the template strand. **A.** Model constructed using coordinates from the *Pyrococcus woesei* co-crystal structure of archaeal TBP and the C-terminal core of TFB (TFBc) in a complex with a TATA-box-containing promoter (9). **B.** Cross-linking to position -19, transcribed strand. Transcription complexes were assembled as described in the legend to Figure 3.3, except the *gdh* promoter DNA (labeled at -19T) was present at 1 nM, and no NTPs were added to the reactions. Reactions were exposed to 365 nm UV for 60 minutes, after which they were treated with DNase I and micrococcal nuclease, and then separated by 8-24% gradient SDS-PAGE.

of a *p*-Bpa substitution, a low level of background (cross-linker independent) radiolabelling of TFB was observed (lanes 1 and 2) presumably because aromatic residues from RNAP subunits and TFB1 can potentially cross-link to the probe (26, 27). Insertion of *p*-Bpa at F192 of TFB1 resulted in a large increase in radiolabeled TFB1, consistent with a transfer of label from the -19T position (lanes 3 and 4). In contrast, TFB1 R52*p*-Bpa was not radiolabeled (lane 6). Only background level cross-linking was observed with R52*p*-Bpa under these conditions, as expected for residues within the B-reader, which is expected to be far away from position -19. This

confirms the close proximity of the F192 position to the -19T position of the *gdh* promoter, and indicates that *p*-Bpa cross-linking is position specific in that the transfer of radioactive label to the TFB polypeptide is dependent on proximity between *p*-Bpa and the specific DNA position of the radiolabel.

***p*-Bpa cross-linking shows B-reader surfaces are proximal to specific promoter positions**

We next asked whether the TFB1 B-reader is close to DNA near the transcription start site, since the B-reader is predicted to be involved in determining the transcription start site and promoter melting (22). To determine this we used TFB1 substituted with *p*-Bpa at positions W44 and R52 (and, later, T58), located in the B-reader helix and the B-reader loop, respectively (Figure 3.5A). Reactions containing TFB1, W44*p*-Bpa and R52*p*-Bpa were assembled on *gdh* promoter DNA radioactively labeled at the -6 position on the template strand. While both W44*p*-Bpa and R52*p*-Bpa generated signal upon UV exposure, wild type TFB1 did not exhibit cross-linking (Figure 3.5, lanes 1-3).

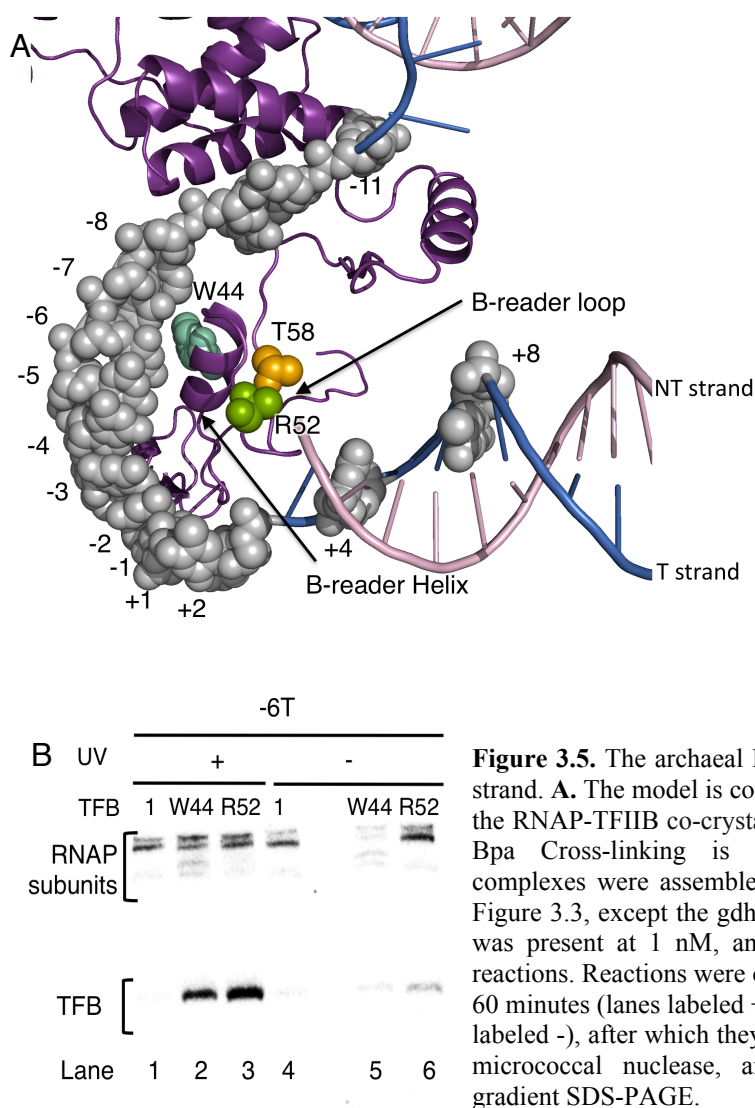


Figure 3.5. The archaeal B-reader is close to the template strand. **A.** The model is constructed using coordinates from the RNAP-TFIIB co-crystal structure reported in (8). **B.** *p*-Bpa Cross-linking is UV dependent. Transcription complexes were assembled as described in the legend to Figure 3.3, except the *gdh* promoter DNA (labeled at -6T) was present at 1 nM, and no NTPs were added to the reactions. Reactions were either exposed to 365 nm UV for 60 minutes (lanes labeled +) or incubated in the dark (lanes labeled -), after which they were treated with DNase I and micrococcal nuclease, and then separated by 8-24% gradient SDS-PAGE.

We also asked whether *p*-Bpa cross-linking occurs in the absence of UV exposure. A low level of background signal was observed for R52*p*-Bpa while no signal was detected for W44*p*-Bpa or TFB1 (lanes 4-6). Taken together these results show that efficient cross-linking is both *p*-Bpa and UV dependent.

In order to ensure the observed cross-linking signal was generated by correctly assembled initiation complexes and not by non-specific TFB-DNA interactions, we compared cross-linking reactions assembled in the presence or absence of TBP. Figure

3.6 shows that although W44p-Bpa cross-links strongly to the -6T position in the presence of TBP, no cross-linking signal occurs in its absence (compare lanes 1 and 3). This confirms our cross-linking protocol generates signal only in the context of complexes assembled in a TBP dependent manner.

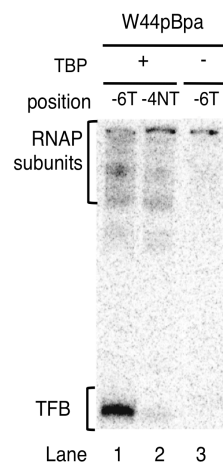


Figure 3.6. pBpa crosslinking is TBP dependent. Cross-linking reactions contained W44p-Bpa and radioactive probe labeled on -6 of the transcribed strand (T) or -4 of the non-transcribed (NT) strand. Transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3.

These results and those from Figure 3.5 suggest that both the B-reader helix and the B-reader loop are in close proximity to the -6T position of the *gdh* promoter in open complexes (RFP_O).

We next addressed whether the B-reader is also close to the non-transcribed strand of DNA in RFP_O, since the T and NT strands are expected to be separated and thus distant from each other in an open complex. RFP_O complexes assembled on *gdh* promoter containing a radiolabel at -4NT were compared to complexes assembled on *gdh* promoter labeled at -6T. Figure 3.6 indicates that the B-reader helix (the W44 position) may not be in close proximity to the non-template strand, as probe labeled at the -4 position on the non-template strand does not generate signal under these conditions. Taken together these results suggest the B-reader may interact with the

template but not the non-template strand immediately upstream of the transcription start site.

To round out our initial exploratory studies of engineered cross-linkable TFB, we addressed the question of TFB dynamics during the initiation cycle. We hypothesized *p*-Bpa cross-linking may detect changes in the orientation of the B-reader relative to promoter DNA as the complex transitions from a closed (RFP_C) to an open state (RFP_O) or as the template register changes during synthesis of a short RNA transcript (RFP_{init}). We investigated this possibility by comparing *p*-Bpa cross-linking signal in RFP_O versus RFP_{init}, using a +1T probe in the absence or presence of rATP, rCTP and rGTP ribonucleotides. When RNAP initiates transcription at the *gdh* promoter in the absence of UTP, transcription cannot proceed past the 5th base. Figure 3.7 shows that while no TFB1 cross-linking signal was detected in the absence of rNTPs, the +1T position generates substantial signal in the presence of the three nucleotides (compare lanes 1 and 2).

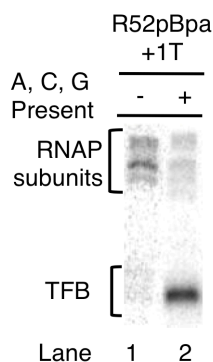


Figure 3.7. The B-reader loop is in close proximity to the template strand and the transcription start site in RFP_{init} but not RFP_O. Cross-linking reactions contained R52pBpa and radioactive probe labeled on +1T of the transcribed strand (T). Transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3.

The NTP-dependence of cross-linking between TFB1 R52*p*-Bpa and +1 of the DNA T-strand could be explained by movement of DNA relative to the RNAP active

site as RNA is being synthesized, or the complex may also undergo other conformational changes as the first nucleotides are added, which may bring the B-reader and the +1T position closer together. It is clear, however, that this method allows us to investigate the position of the B-reader in a variety of transcription initiation states, and that progress through the transcriptional cycle causes observable changes in the cross-linking signal patterns.

The B-reader is not proximal to promoter DNA in the absence of RNAP

In order to determine the position of the B-reader within transcription initiation complexes more precisely, a more extensive cross-linking investigation was undertaken with a series of radiolabeled probes spanning the region expected to undergo strand separation (-12 to +10). Radiolabels were placed at specific positions both on the T-strand (positions -11, -8, -7, -6, -5, -4, -3, -2, -1, +1, +2, +4, and +8) and the NT-strand (positions -9, -7, -5, -4, -3, -2, -1, +1, +3, +5, +7 and +13) (Figure 3.1). We first wanted to examine if the archaeal B-reader region is in close proximity to promoter DNA in the absence of RNAP, since RNAP independent cross-linking could potentially contribute to and thus confound the signal generated by complete initiation complexes. In addition, the position of the TFB NTD relative to DNA prior to RNAP recruitment could be important in the recruitment and subsequent isomerization steps. These experiments contained the same reaction components described in the methods section with the exception of RNAP. We assembled complexes with either R52p-Bpa or W44p-Bpa on promoter DNA labeled on either the template or the non-template strand. No significant cross-linking signal was

observed, with the exception of -9NT, indicating the B-reader helix and loop are not in close proximity to promoter DNA near the transcription start site in the absence of RNAP (Figures 8 and 9). Interestingly, while the -9NT position followed this general pattern and did not display significant signal in the presence of RNAP, it produced strong cross-links in its absence (compare lane 1 of 9A and 9B with lane 1 of 9C and 9B). This may indicate that before RNAP recruitment the B-reader is poised at the upstream edge of what will become the transcription bubble, perhaps as a way to correctly orient the promoter during RNAP recruitment.

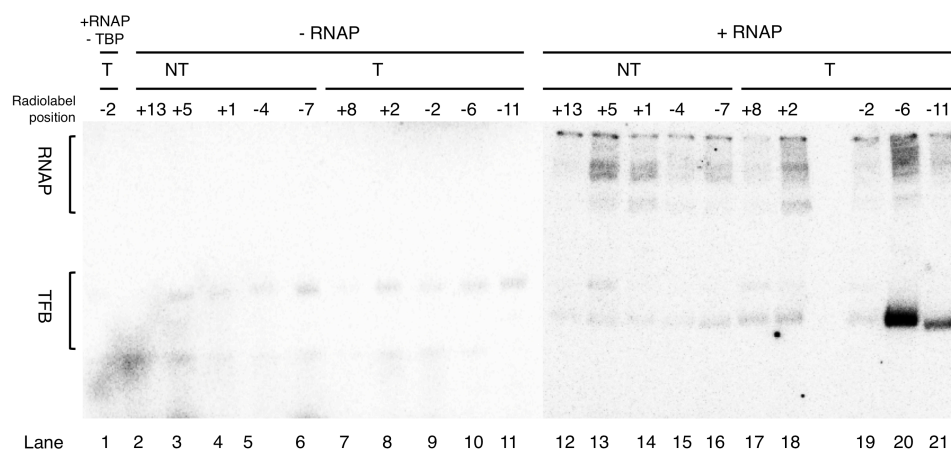


Figure 3.8. Cross-linking of R52p-Bpa TFB to the transcribed (T) and non-transcribed (NT) strands in FP versus RFP_O at 65°C. With the exception of omitting RNAP transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3. Lane 1 served as a control for cross-linking specificity and contained RNAP but lacked TBP.

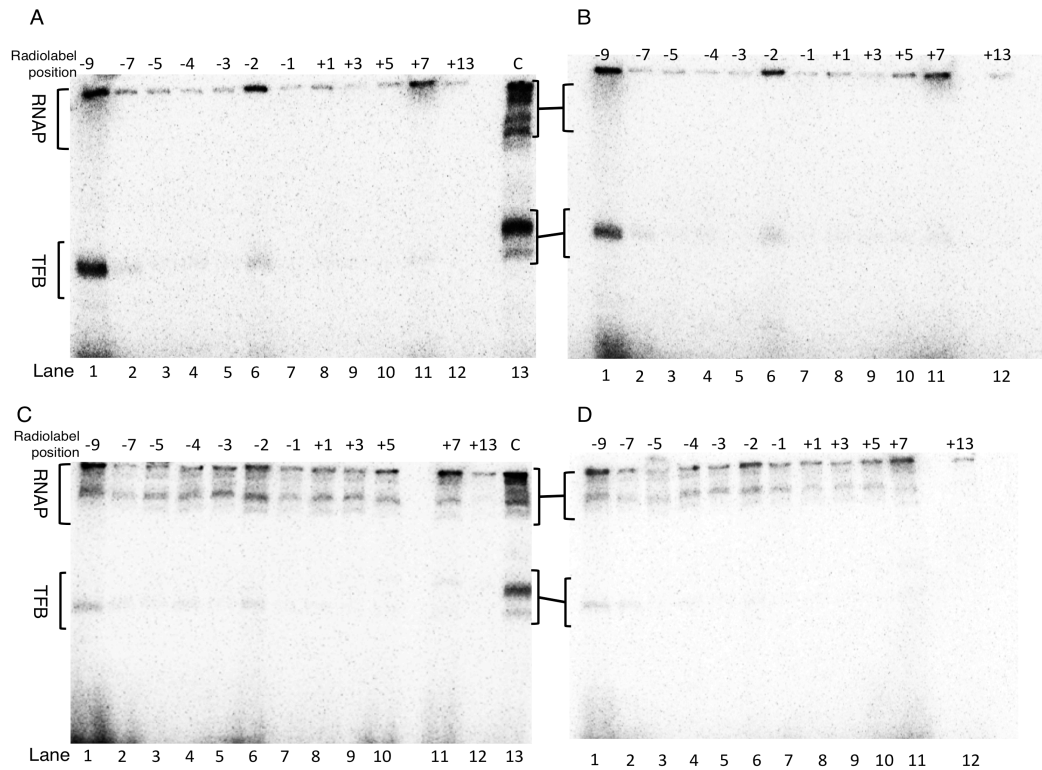


Figure 3.9. Cross-linking of R52p-Bpa and W44p-Bpa TFBs to the non-transcribed strand in FP versus RFP_O at 65°C. With the exception of omitting RNAP, transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3. Lanes labeled “C” contained a control reaction assembled on -2T probe, incubated at 65 °C in the presence of rATP, rCTP and rGTP. A. R52p-Bpa, -RNAP B. W44p-Bpa, -RNAP C. R52p-Bpa, + RNAP D. W44pBpa, + RNAP.

The TFB1 B-reader is close to the template strand

We next wanted to define the extent of cross-linking by the TFB1 B-reader to DNA sequences encompassing the predicted transcription bubble in RFP_O complexes. To do this, promoter DNA containing radiolabel (³²P) encompassing positions -11 to +8 on the template strand and -9 to +13 on the non-template strand, a region of promoter DNA known to span the transcription bubble ((28, 29) and Fig 1B). This set of radiolabelled promoter variants allows mapping of the general position of the B-reader within the archaeal transcription initiation complex. Functional information

from these experiments would thus help to validate or refute the existing models. Thus we attempted to more clearly define if the B-reader is located in close proximity to the template, non-template or to both promoter strands. We also aimed to determine where the B-reader helix and loop are located by comparing the W44*p*-Bpa (helix), R52*p*-Bpa and T58*p*-Bpa (loop) cross-linking signals generated under the same set of conditions. Figure 3.10 shows the results from cross-linking experiments performed using RFP_O at 65 °C. Complexes (RFP_O) containing R52*p*-Bpa (Figure 3.10B, compare with Figure 3. 9A) or W44*p*-Bpa (Figure 3.10A, compare with Figure 3. 9B) were assembled on promoter DNA radioactively labeled on the template strand while reactions containing R52*p*-Bpa and T58*p*-Bpa were assembled on both T and NT strand probes (Figure 3.8, lanes 12-21 and Figure 3.10, panel C). Following RFP_O complex formation and cross-linking the non-template strand positions produced little observable signal while the template strand positions produced relatively strong cross-linking signal, with the strongest cross-links centered around -7 for both R52*p*-Bpa and W44*p*-Bpa. The strongest cross-linking by T58*p*-Bpa was slightly downstream, with the highest signal intensity generated with -3T probe. Since T58 is located on the C-terminal side of the B-reader loop relative to R52, it may be that this residue is more proximal to downstream template positions under these conditions. Taken together these results indicate the B-reader is located in close proximity to the template but not to the non-template strand. In addition, the B-reader loop (which contains R52 and T58) appears to be oriented slightly downstream relative to the B-reader helix, since more downstream positions such as -3 and +4 display higher signal intensities when

compared to W44-*p*Bpa (compare lanes 7 in 10A and 10B to lane 8 in 10C and lanes 12 in 10A and 10B to lane 10 in 10C).

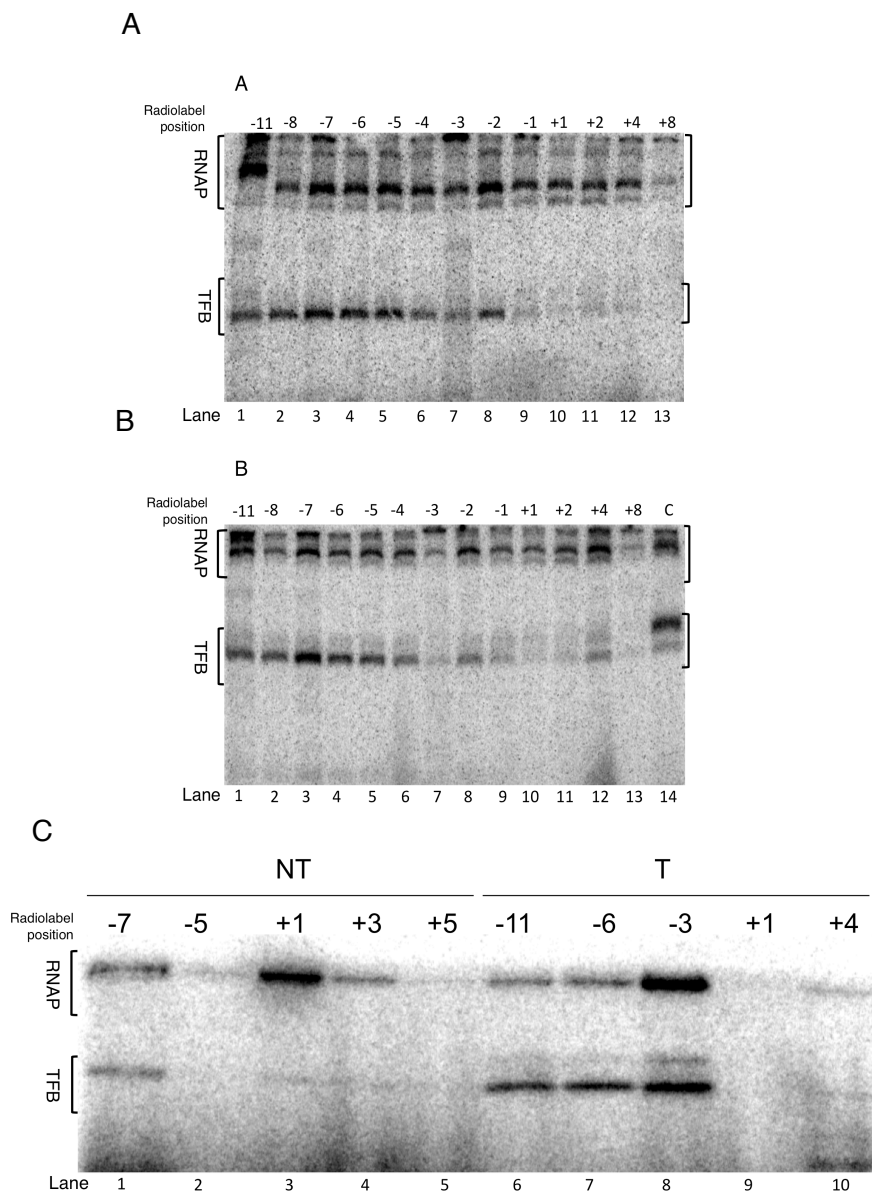


Figure 3.10. Cross-linking W44-*p*Bpa, R52-*p*Bpa and T58-*p*Bpa TFBs to the transcribed strand in RFP_O at 65°C. Transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3. Lanes labeled “C” contained a control reaction assembled on -2T probe, incubated at 65 °C in the presence of rATP, rCTP and rGTP. A. W44-*p*Bpa, template strand probes. B. R52-*p*Bpa, template strand probes. C. T58-*p*Bpa, template (T) or non-template (NT) strand probes.

The position of the B-reader changes during the formation of the open complex

During initiation the RNA polymerase structure must convert double-stranded DNA to the single-stranded transcription bubble (Figure 3.1, Step 3) to allow the active site access to T-strand DNA. We reasoned that the TFB1 N-terminus may change its orientation relative to DNA during this transition, as it performs its role in promoter opening. RFP_C complexes can be made by incubating factors, RNAP and promoter DNA at 50 °C, where promoter opening is not detectable. We predicted *p*-Bpa-dependent cross-linking will detect changes in the position of the B-reader with respect to the promoter in transcription initiation complexes trapped in the closed state (RFP_C, formed at 50 °C) versus the open state (RFP_O, formed at 65 °C) (Figure 3.1).

TFB1 B-reader contacts to DNA at 50 °C were measured with W44, R52 and T58 *p*-Bpa mutants. The general cross-linking pattern at 50 °C was similar to cross-linking at 65°C in that all three mutants generated signal with promoter DNA radiolabelled on the template but not non-template positions (Figure 3.11) (data not shown). However, we observed significant differences in cross-linking intensity between the two temperatures (Figure 3.11, compare to Figure 3.10). At 65 °C the W44p-Bpa signal was significantly more intense at upstream positions, with -7T cross-linking increasing 5X while -6T increased approximately 2X (Compare 10A to 11A, lanes 3 and 4). In addition, the weak signal generated by downstream positions at 50°C disappeared at 65°C. This may indicate that when the transcription complex transitions to a different conformation at 65°C, the B-reader helix is positioned slightly

differently, closer to upstream positions between -8T and -4T. We observed similar differences for the B-reader loop. Like W44*p*-Bpa, cross-linking with both R52*p*-Bpa and T58*p*-Bba at 65°C resulted in stronger signal with upstream probes such as -7T and -6T, whereas cross-linking to downstream positions (-4T to +4T) decreased significantly or disappeared. At 50°C W44*p*-Bpa cross-linking signal was highest for

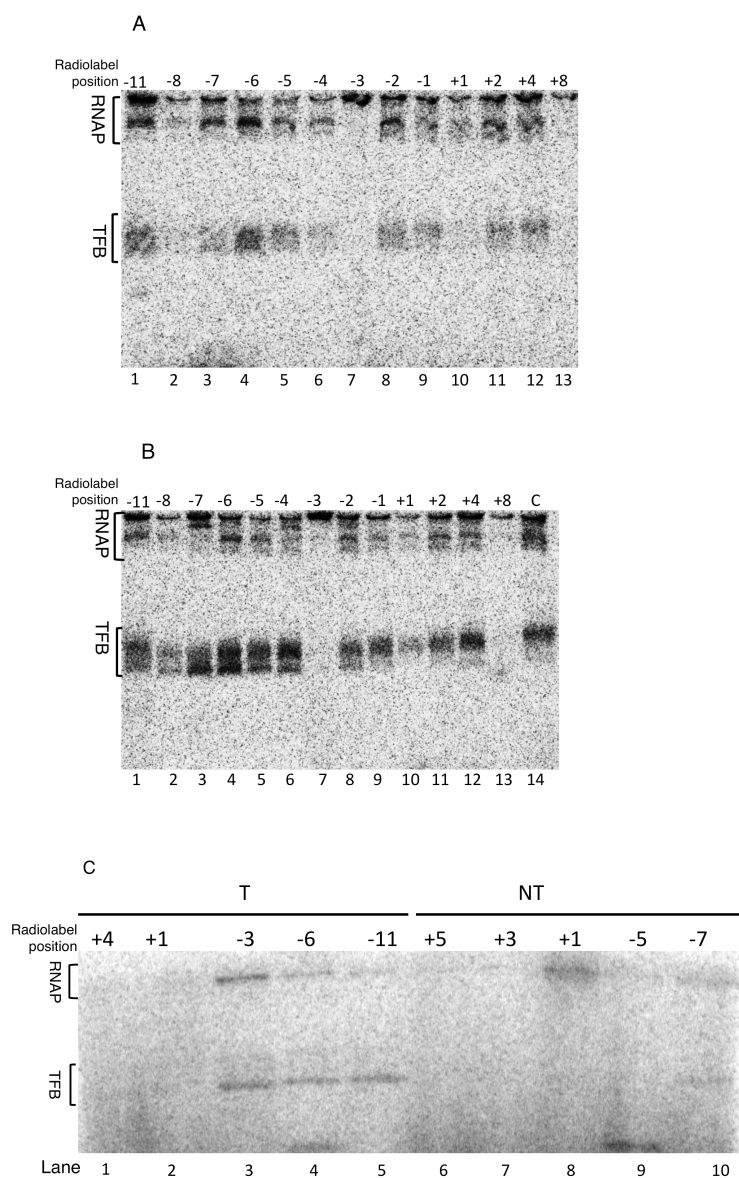


Figure 3.11. Cross-linking W44*p*-Bpa, R52*p*-Bpa and T58*p*-Bpa TFs to the transcribed strand in RFP_C at 50°C. Transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3 except the reactions were incubated at 50 °C. The reaction in lane “C” is described in the legend to Figure 3.10. A. W44*p*-Bpa B. R52*p*-Bpa C. T58*p*-Bpa cross-linking to the transcribed (T) strand probes or non-transcribed (NT) strand probes.

upstream probes centered around -6, but we also observed cross-linking to more downstream positions (-1T, +2T and +4T) (Figure 3.11, panel A). R52p-Bpa cross-linking to the template strand followed a similar pattern but generally the signal was about 2X more intense. The cross-linking signal peaked upstream of the transcription start site, but whereas W44p-Bpa cross-linking centered symmetrically around -6T, R52p-Bpa generated a much stronger signal with -5T and -4T probes, indicating that the B-reader loop may be positioned slightly further downstream than the helix. In addition, compared with W44p-Bpa, R52p-Bpa also generated significantly stronger signal with more downstream positions (-1T and +2T) and especially the +4T, where the signal approached the level exhibited by upstream positions. This suggests that unlike the B-reader helix, the loop may be in close proximity to downstream positions as well.

The position of the B-reader changes upon transcription initiation

When RNAP II begins RNA synthesis, the main channel becomes occupied with an RNA-DNA duplex, and after synthesis of an 8-9 nt transcript, the clamp domain of Rpb1/Rpb5 changes position, making a stable elongation conformation for RNAP (30-32). We reasoned that if the TFB1 B-reader is close to T-strand DNA prior to initiation, that synthesis of the RNA-DNA duplex may cause a shift in the position of the B-reader. To examine this we investigated the effect of rNTP addition by comparing cross-linking of TFB1 W44p-Bpa, R52p-Bpa and T58p-Bpa in transcription complexes allowed to synthesize 5 nt transcripts (RFP_{init}, Figure 3.1). At

65 °C R52*p*-Bpa generated strong cross-linking signal from positions downstream of the transcription start site on the template strand while maintaining similar levels of signal at upstream positions compared to reactions lacking the rNTPs (Figure 3.12B, compare to Figure 3.10B). T58*p*-Bpa displayed a similar pattern but while R52*p*-Bpa displayed similar signal intensity from upstream positions in the presence of NTPs, T58*p*-Bpa showed a significant increase from the -6T position. This may indicate the backside of the B-reader loop is positioned slightly differently under these conditions. Similar to R52*p*-Bpa, W44*p*-Bpa maintained cross-links to upstream positions, however, while new signal developed from most downstream positions, +1T and +4T displayed significantly higher signal intensity compared to R52*p*-Bpa (Figure 3.12A, compare to 12B and Figure 3.10).

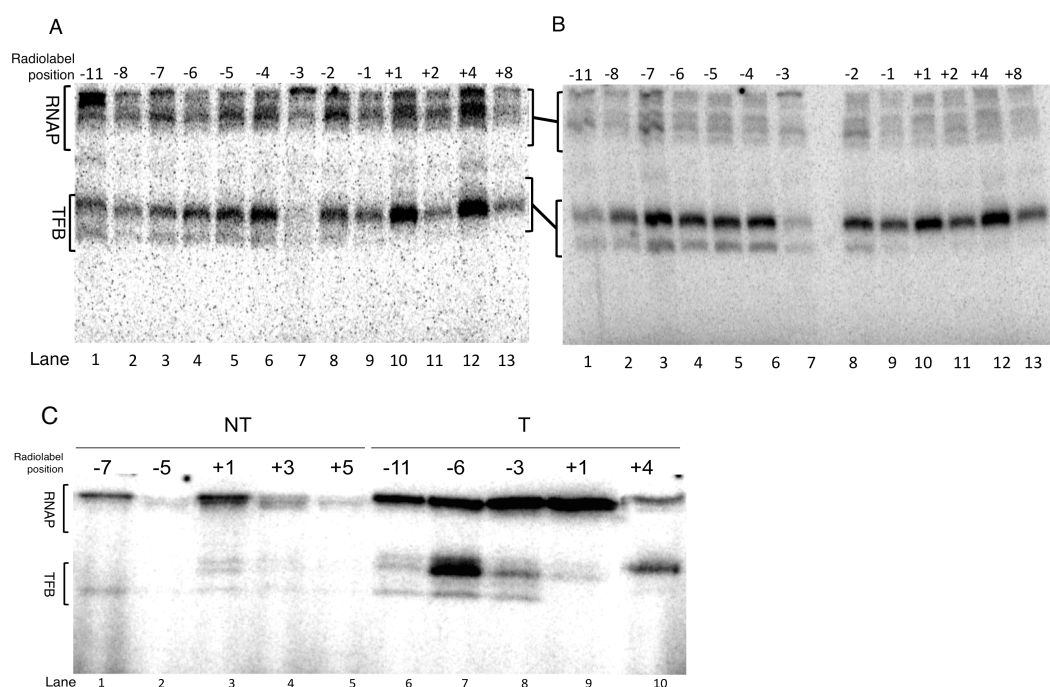
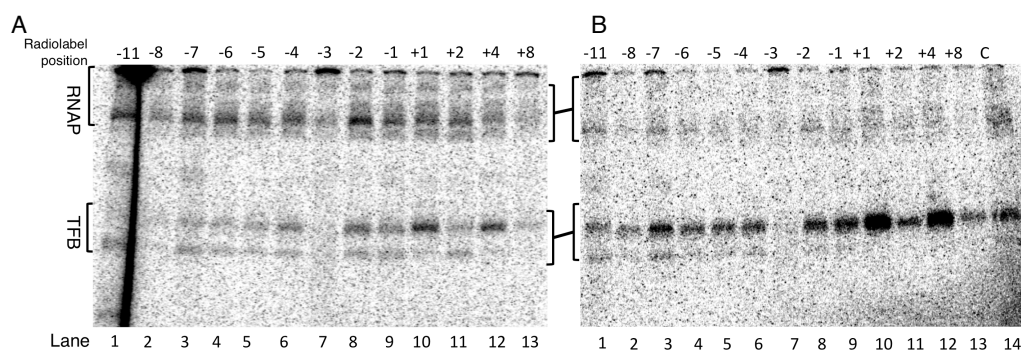


Figure 3.12. Cross-linking of R52*p*-Bpa, W44*p*-Bpa and T58*p*-Bpa TFBs to the transcribed or non-transcribed strand in the presence of saturating rATP, rCTP and rGTP. Transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3 except rATP, rCTP and rGTP were added to 500 μ M final concentration. A. W44*p*-Bpa 65 °C, transcribed

strand B. R52*p*-Bpa, 65 °C, transcribed strand C. T58*p*-Bpa cross-links to the transcribed strand (T) or to the non-transcribed (NT) strand at 65 °

At 50°C, transcription complexes are thought to be mainly in the RFP_C state, since KMnO₄ reactivity of T residues is very low compared to 65°C (Grunberg & Thomm, personal communication). However, transient promoter-opening events would likely be missed by KMnO₄ experiments, but could be trapped if NTPs are present. To test this, we compared cross-linking of TFB1 W44*p*-Bpa, R52*p*-Bpa and T58*p*-Bpa in transcription complexes allowed to synthesize 5 nt transcripts at 50°C. Under these conditions, R52*p*-Bpa generated the strongest signal from +1T and +4T, surpassing the signal generated by these positions at 65°C (Figure 3.13B, compare to Figures 12B) while T58*p*-Bpa displayed a moderate increase in signal intensity, especially from -6T and +4T (compare 13C to 11C). Interestingly, the addition of rNTPs had no significant effect on W44*p*-Bpa cross-linking signal at 50 °C, indicating that the B-reader helix may not be involved in T-strand interactions at low temperatures and that the B-reader loop may grip the T-strand prior to promoter opening.



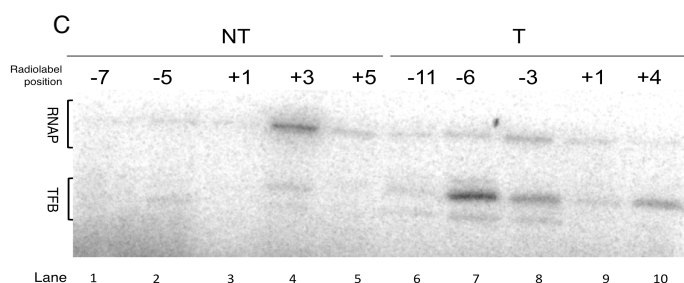


Figure 3.13. Cross-linking of R52*p*-Bpa, W44*p*-Bpa and T58*p*-Bpa TFBs to the transcribed or non-transcribed strand in the presence of rATP, rCTP and rGTP at 50°C.. Transcription complexes were formed, cross-linked, and analyzed as described in the methods and the legend to Figure 3.3 except rATP, rCTP and rGTP were added to 500 μ M final concentration. A. W44*p*-Bpa, transcribed strand B. R52*p*-Bpa, transcribed strand C. T58*p*-Bpa, transcribed strand (T) or non-transcribed (NT).

The presence of rNTPs did not have a significant effect on cross-linking to the NT strand and did not generate significant signal with the exception of the -9 position (9 C and data not shown).

pBpa cross-linking as an assay for detecting the state of the RNAP clamp?

At 65 °C and in the presence of rATP, rCTP and rGTP, additional cross-links to downstream positions develop on the template strand from +1 to +8 (compare Figure 3. 12A and 12B with Figure 3. 10A and 10B). There are two possible explanations for this: 1) the TFB1 B-reader follows the RNAP as the active site moves from -1/+1 to +5/+6; or 2) the DNA cross-linked to TFB1 *p*-Bpa mutants becomes protected from nuclease digestion because of formation of a protective pocket by RNAP and causing a larger stub to appear, extending the signal detection range for the assay further downstream as a result.

To test this, we performed the same cross-linking experiment as in Figure 3. 11, but instead of digesting with micrococcal nuclease and DNase I under non-

denaturing conditions, we digested with DNase I followed by denaturation and digestion with S1 nuclease. The downstream, nucleotide dependent, cross-links disappear upon digestion with S1 nuclease whereas the upstream cross-linking signal (-6 and -2 on the T strand) displays a lower molecular weight, presumably because the extended DNA stub has been digested since the protection afforded by RNAP under non-denaturing conditions would be expected to disappear after denaturation and treatment with S1 (Figure 3. 14, lanes 5 and 8).

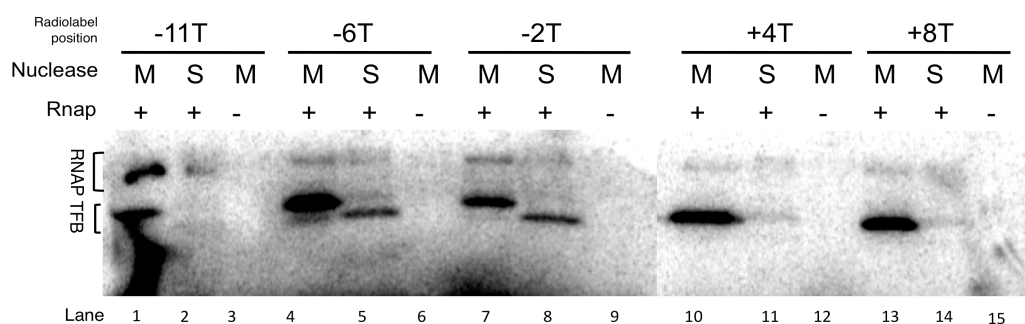


Figure 3.14. The effect of S1 nuclease treatment on pBpa cross-linking signal in RFP_{init} at 65°C. Reactions were assembled as described in the methods and the legend to Figure 3.3 and contained R52pBpa. Following cross-linking, the reactions labeled M were digested with Micrococcal nuclease and DNase I whereas the reactions labeled S were treated with S1 nuclease and DNase I.

Discussion

Role of TFB in transcription: post-assembly events leading to promoter escape

Over the past two decades, biochemical, genetic and crystallographic data have elucidated in molecular detail the assembly of transcription initiation complexes in both Archaea and Eukarya (13, 33, 34). It is now clear that after the recruitment of RNAP, the DNA duplex unwinds, forming the transcription bubble, in which transcribed and non-transcribed strands follow separate paths. The template strand is

loaded into the active site cleft of the polymerase where it base-pairs with incoming nucleotides while the non-template strand is directed away from the T-strand, rejoining the template strand at the edge of the transcription bubble following separation of the growing RNA transcript from the template DNA strand.

We do not yet know the details of the protein-DNA interactions mediating the formation of the transcription bubble, the determinants of the transcription start site selection, of abortive transcription, or the events that lead to the formation of the stable elongation complex following initiation. Several lines of evidence, including data from our own laboratory, have suggested the TFB NTD plays a crucial role in these transitions (5, 15-17, 21, 35-37). Domain swaps between TFB1 and TFB2 have shown that the N-terminus of TFB2 (which lacks the conserved B-reader sequence) is responsible for its lower transcriptional activity and that unlike TFB1, TFB2 is deficient in promoter opening (5).

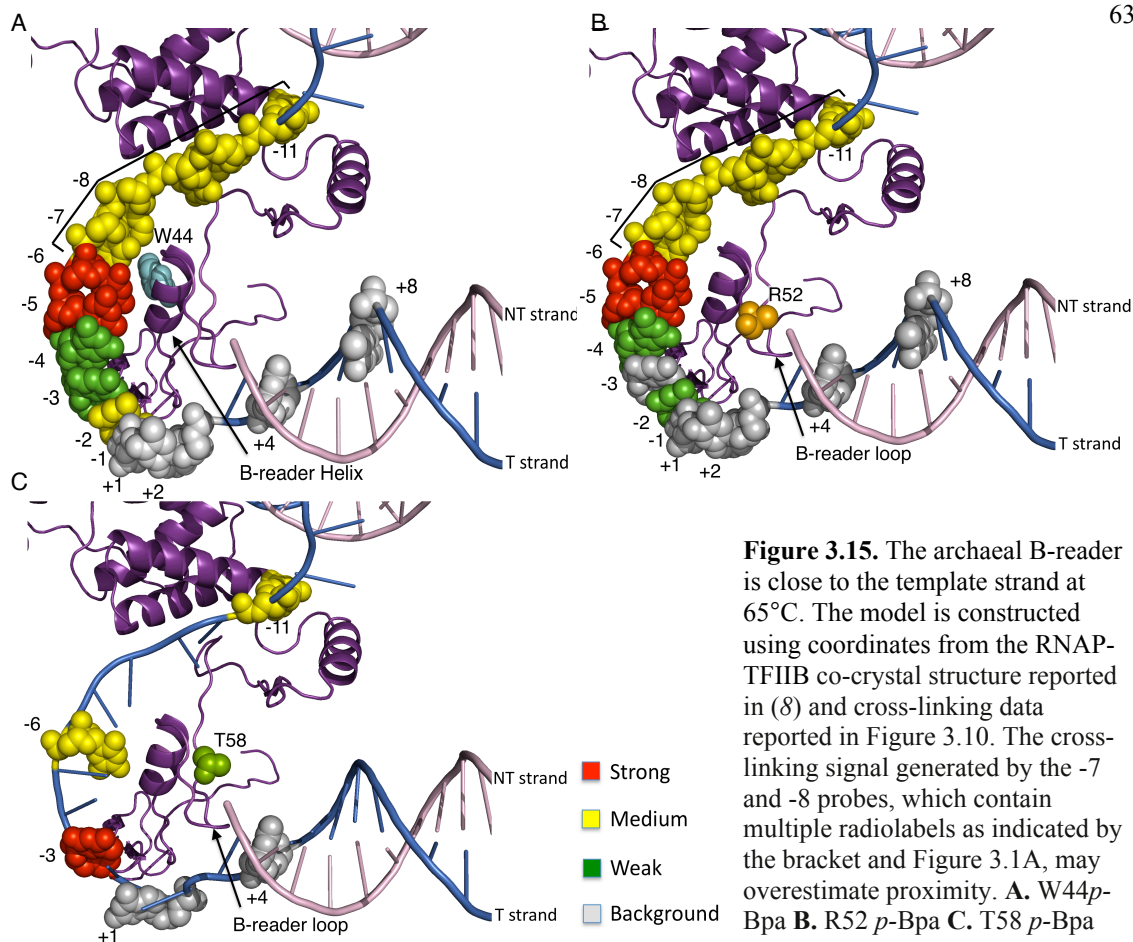
A novel method for mapping protein-DNA interacting surfaces

To begin to answer these questions we have focused on the NTD B-reader and have determined its orientation and position relative to promoter DNA. We have established that the *p*-Bpa cross-linking method combined with radioactive probes encompassing a wide range of positions spanning the transcription bubble and flanking regions is well suited for detecting protein proximity to specific DNA positions, with close to amino acid to nucleotide resolution. With the large set of promoter variants reported here, we have gained an overview of the general position

of the B-reader within the archaeal transcription initiation complex, thus complementing the existing models based on crystal structures with functional information.

The B-reader: a role for transcribed strand placement

In one of two recently published co-crystal structures (8, 22) Cramer and colleagues were able to resolve distinct structures within the TFIIB linker region previously assigned to the B-finger. The crystal structures were formed in the absence of nucleic acids, and so the position of the conserved regions during transcription initiation is not clear. The archaeal TFB1 N-terminus contains sequences similar to those that make up the eukaryotic B-finger/B-reader structures, as well as some regions that were not well resolved in the electron density maps, presumably because of multiple conformations or mobility. We have chosen to model a subset of our cross-linking results using this structure since it provides the most detailed information regarding position and secondary structures within the TFIIB N-terminus (Figure 3.15).



Our data derived from the highly homologous archaeal transcription system supports the differential positioning of the B-reader helix and B-reader loop while providing additional positional information collected in the presence of nucleic acids which supplements the x-ray crystal structure models with functional data. Our data show that the archaeal B-reader sequence is located in close proximity to the template strand (Figures 10 and 15) and that it is not close to the non-template strand (Figures 8 and 9). Furthermore, the data places the B-reader closer to the positions immediately upstream of the transcription start site and further away from downstream positions, since little or no signal is observed from template strand positions such as +4 or +1

(Figure 3.10 and 15). Our data thus supports a structural arrangement of the initiation complex in which the B-reader helix is poised to interact with the template strand in the transcription bubble, upstream of the transcription start site. Thus the B-reader may assist in strand-separation, guiding the template DNA towards the active site, and/or in preventing its re-association with the non-template strand.

The position of TFB1 during the initiation of transcription

Passage through the transcription initiation cycle by archaeal RNAP is likely to involve kinetically distinct steps (see Figure 3.1) that reflect structural transitions, each of which may be influenced by transcription factors. Knowing the position of TFB1 in the transcription complex at each step could suggest potential roles for the factor and its subregions in the transitions between these structural species. Therefore, we examined the positioning of the TFB subregions at 4 different steps in the transcription cycle: FP, RFP_C, RFP_O and RFP_{init}.

One important reaction parameter that is expected to affect the state of the transcription complex is temperature. While *in vitro* transcription assays performed at 65°C generate significant transcript on a number of promoters, little or no transcription is observed when the reaction temperature is below 55 °C (data not shown). Since transcription experiments only monitor the accumulation of run-off transcript, a number of explanations exist for the lack of transcription at lower temperatures. One possibility is that promoter melting cannot occur at the lower temperature, a view supported by the low KMnO₄ reactivity observed at low temperature (Grunberg & Thomm, personal communication). It is also possible that

the open complex transiently forms at lower temperatures, but the transition to elongation (which requires the RNAP to break contacts with the promoter and transcription factors) may not occur at 50°C or may only occur very slowly. One proposed mechanism for promoter opening involves the interaction of B-reader residues with template strand positions. This model posits such interactions trap the template strand away from the non-template strand and thus facilitate promoter opening. If promoter opening can still occur at low temperature, we would expect to detect cross-linking to the template strand around the transcription start site but not to the corresponding non-template strand positions.

The data reported here indicate differences in the orientation of the B-reader between 50 and 65 °C. (compare Figure 3.10 to 11 and Figure 3.15 to Figure 3.16). At 65 °C both the B-reader helix and loop are repositioned closer to upstream positions between -8T and -4T. At 50 °C the B-reader loop is in close proximity to downstream positions (+2 to +4) unlike the B-reader helix (Figures 11 and 16). Thus when the B-reader changes its relative position to the promoter as the transcription complex undergoes the conformational changes that occur between the closed and open states, the helix and loop may interact with different regions of the DNA, possibly because these structures may play different roles in this process (8).

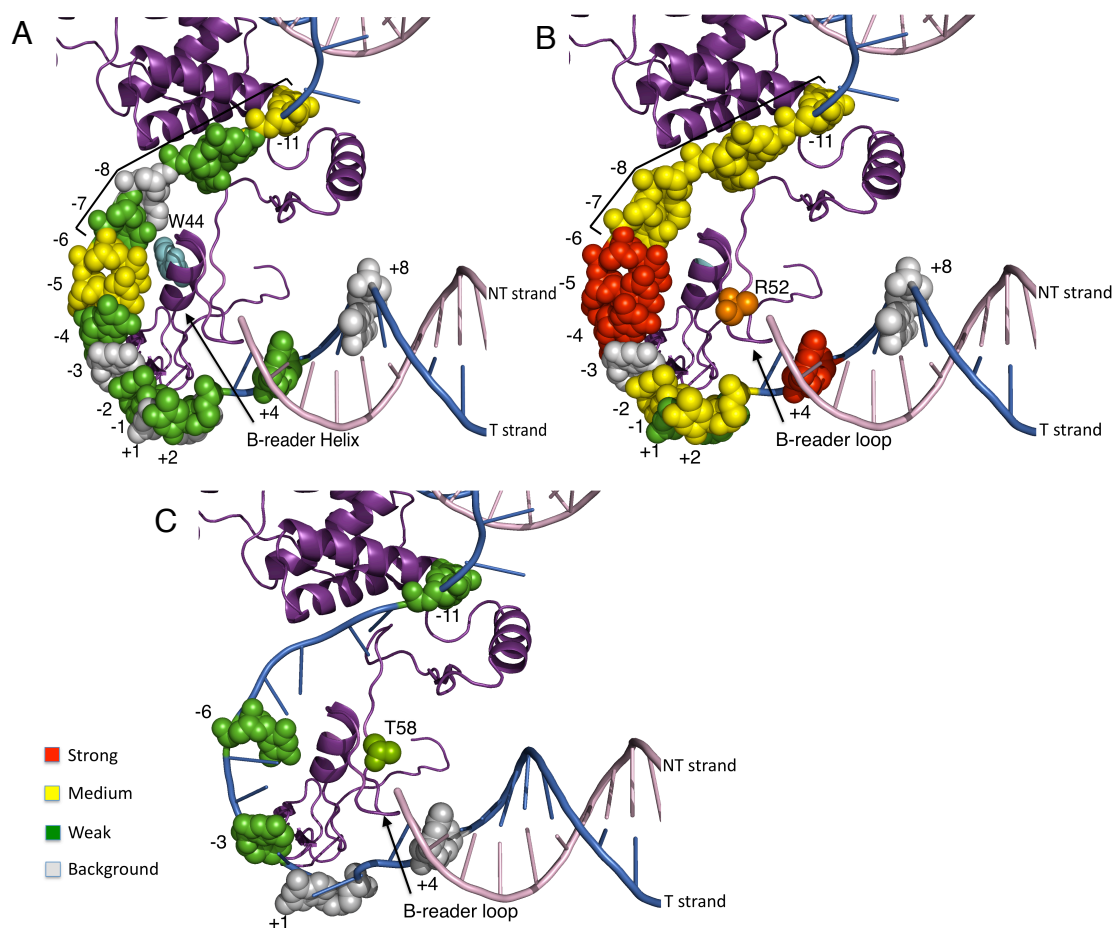


Figure 3.16. The archaeal B-reader during transcription initiation. A. Model based on cross-linking results reported in Figure 3.11A (W44p-Bpa and 50 °C) B. Model based on cross-linking results reported in Figure 3.11B (R52p-Bpa and 50 °C). C. Model based on cross-linking results reported in Figure 3.11C (T58p-Bpa and 50 °C). All three panels utilize the RNAP-TFIIB co-crystal structure reported in (8). The cross-linking signal generated by the -7 and -8 probes, which contain multiple radiolabels as indicated by the bracket and Figure 3.1A, may overestimate proximity.

In addition, our results support the trapping of the template strand by the B-reader, since the intensity of template strand cross-links increase under reaction conditions designed to favor RFP_O.

Another reaction parameter that is expected to affect the stability of transcription complexes is the inclusion of rATP, rCTP and rGTP. Complexes assembled on the *gdh* promoter in the presence of these rNTPs are expected to

synthesize a 5 nt transcript before the first rUTP (absent from the reaction) must be incorporated and would represent an intermediate similar to abortive initiation complexes (RFP_{init}) and the escape commitment intermediate proposed for RNAP II (38). If the B-reader is involved in this transition, its position might change between the two states. In addition, the presence of rATP, rCTP and rGTP would cause movement of the RNAP active site relative to the T-strand DNA, and this should be reflected by a different set of cross-links as the B-reader is exposed to downstream template positions after the 5-mer is synthesized. Our data shows that in the presence of NTPs additional downstream cross-links develop for all three *p*-Bpa mutants (Figure 3.12 and 17) and that, in addition, the molecular weight of the labeled TFBs increases slightly (Figure 3.12).

It is possible the presence of a short transcript induces a conformational change that stabilizes the alignment of the template strand relative to the active site. This may lead to an increase in template strand protection from digestion by Mnase but not from the more stringent S1 treatment. When we tested this in complexes expected to contain a 5nt transcript (RFP_{init}), higher molecular weight cross-linked TFB species were observed, that disappeared following denaturing nuclease treatment (Figure 3.14). We hypothesize that the synthesis of a short transcript protected T-strand DNA from digestion, because of formation of a more stable complex, perhaps involving closing the downstream clamp (previously observed in X-ray crystal structures of RNAP II elongation complexes).

Experiments with both W44pBpa and R52pBpa conducted at 50°C show that in the absence of nucleotides, downstream cross-links occur similar to those observed at 65 °C in the presence of nucleotides (compare Figure 3.11 to 12 and 16 to 17) with a similar shift in the motility of labeled TFB1. This may indicate the DNA is protected at 50 °C even in the absence of nucleotides and that less protection occurs at 65 °C in the absence of nucleotides. The question is still open as to whether the downstream 50 °C cross-links disappear upon S1 treatment and even to what extent the upstream cross-links can withstand this treatment. It is possible that this potentially clamp-closed state at 50 °C is responsible for most of the observed cross-links at this temperature and may be responsible for the lack of transcription seen at low temperatures. The clamp may transiently open to allow template strand loading at higher temperatures and then close upon the addition of initiating nucleotides.

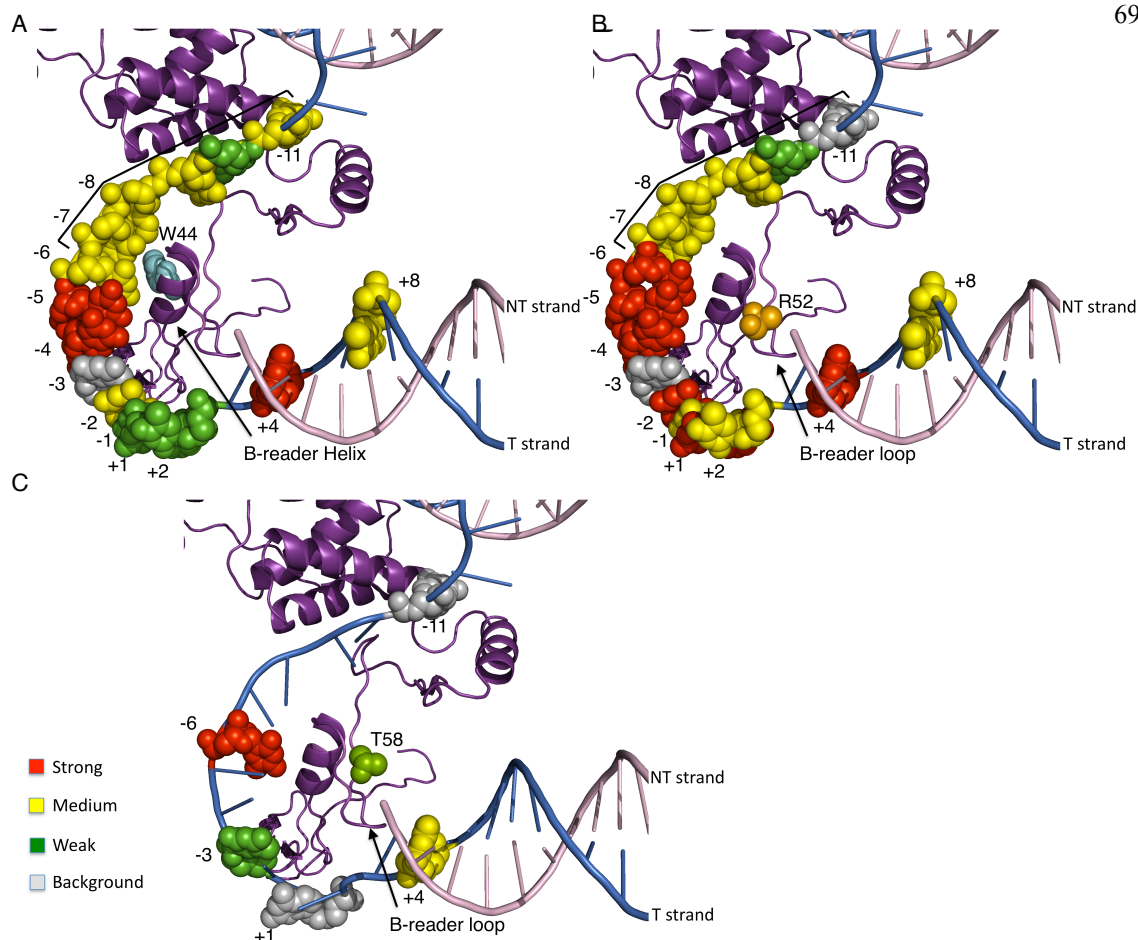


Figure 3.17. The archaeal B-reader during transcription initiation in the presence of initiating nucleotides at 65°C. **A.** W44*p*-Bpa **B.** R52*p*-Bpa **C.** T58*p*-Bpa All three panels utilize the RNAP-TFIIB co-crystal structure reported in (8). The cross-linking signal generated by the -7 and -8 probes, which contain multiple radiolabels as indicated by the bracket and Figure 3.1A, may overestimate proximity.

Our data can be used to further characterize the position of the B-reader helix and loop in the context of the transcription initiation states. Before RNAP recruitment (the FP state, Figure 3.1), the B-reader helix and loop are located further away from promoter DNA than after recruitment. In the absence of RNAP, we only observed cross-links to the -9 position on the non-template strand from R52*p*-Bpa (Figure 3.9A) and to a lesser extent from W44*p*-Bpa, indicating that unlike in post-recruitment

complexes (RFP_C , RFP_O and RFP_{init}) the B-reader (especially the B-reader loop) is positioned closer to the upstream edge of the transcription bubble and the NT-strand. However, we observed that the cross-linking signal generated by -9NT disappears in the presence of RNAP (Figure 3.9C and 9D), indicating that a significant B-reader rearrangement occurs upon RNAP recruitment. RFP_C complexes generate cross-links from T-strand positions that span the transcription bubble region and the molecular weight of labeled TFB1 is slightly greater than in RFP_O complexes, perhaps because RNAP closure protects labeled DNA from the nuclease treatment. The B-reader (especially the B-reader loop) is thus expected to be repositioned close to the T-strand as transcription initiation complexes transition from the FP to the RFP_C state. Transient clamp opening and the loading of the template strand into the active site are expected to define RFP_O complexes, as evidenced by the lower molecular weight TFB1 species (Figure 3.10) and repositioning of the B-reader helix closer to upstream T-strand positions than in RFP_C (compare Figures 10A and 11A). In RFP_{init} complexes we observe that labeled TFB1 is once again higher in molecular weight, indicating a possible clamp closure event upon the initiation of RNA synthesis. In addition, the B-reader helix and loop maintain their proximity to upstream positions on template strand, indicating that as a short RNA transcript is synthesized, the position of the template strand relative to the B-reader may remain fixed until an eventual steric clash with the growing RNA transcript is expected to lead to promoter clearance and the RFP_{elong} state (8, 21).

Experimental Procedures

Mutagenesis and protein purification

TFB1 variants containing site-specific placement of *p*-benzoyl phenylalanine were made using an engineered tRNA-aminoacyl tRNA synthetase system. The genetic system makes use of a modified tRNA and cognate aaRS gene derived from *Methanococcus jannaschii*. The tRNA gene has been modified to specifically recognize an amber codon while the aa RS mutant charges the modified tRNA with the *p*-benzoyl-L-phenylalanine, a cross-linkable amino acid analogue derived from phenylalanine. Both engineered genes are carried on the pSup-BpaRS-6TRN plasmid (24). Mutant versions of histidine-tagged TFB1 DNA with amber codons substituted at the desired positions, W44, R52, T58 or F192 were created using mutagenic primer extension followed by Dpn digestion of templates followed by sequencing. The primer sequences were as follows: F192pBpa forward: 5'-AAGGAAATTGGAAGAAGTTACAGATAGATTGCGAGAAATCTCAATTTAAC TC-3' F192pBpa reverse: 5'GAGTTAAATTGAGATTTCTCGCAATCTATCTGTA ACTTCTTCCAATTTCCCTT-3'; R52pBpa forward: 5'CGTGCTTTTGATGCTTCTC AATAGGAACGCAGGTCTAGAAC R52pBpa reverse: 5'GTTCTAGACCTGCGT TCCTATTGAGAAGCATCAAAAGCACG-3'; W44pBpa forward: 5'CATAATTGA TATGGGTCCTGAGTAGCGTGCTTTTGATG W44pBpa reverse: 5'CATCAAAAG CACGCTACTCAGGACCCATATCAATTATG-3'; T58pBpa forward: 5'CTTCTCA AAGGGAACGCAGGTCTAGATAGGGTGCACCAGAAAG-3' T58pBpa reverse: 5'CTTTCTGGTGCACCCTATCTAGACCTGCGTTCCCTTTGAGAAG-3'.

The resulting plasmid DNA was transformed into a BL21 expression strain carrying the modified tRNA and aaRS genes on the pSup-BpaRS-6TRN plasmid. The mutant TFBs were overexpressed in media supplemented with 1mM *p*-benzoyl-L-phenylalanine and purified using standard his-tag purification protocols. Following elution from the Ni⁺⁺ column, proteins were concentrated and buffers swapped to storage buffer (500 mM NaCl, 20 mM Tris-Cl pH 8, 1mM EDTA) using microcon 10 kD cutoff filtration units. In order to reduce the amount of termination products we then used a microfiltration column with a 30 kD cutoff, which selectively enriched our preps for the full-length TFB variants. The purity of TFB following purification was estimated to be greater than 75% by densitometry. TFBs were quantified on a SDS-PAGE gel by comparison against a BSA standard after staining with Coomassie Blue.

Standard transcription assays

Transcription reactions were performed essentially as described previously (15). The 12.5- μ l reaction mixtures contained 40 mM Na-HEPES, pH 7.3, 250 mM NaCl, 2.5 mM MgCl₂, 0.1 mM EDTA, 5 mM beta-mercaptoethanol, 5% (v/v) glycerol, and 0.1 μ g μ l⁻¹ bovine serum albumin; 10 nM promoter DNA was combined with 60 nM TBP, 60 nM TFB1, or TFB2 and 20 nM RNAP. Increasing the TFB concentration to values greater than 60 nM resulted in no increase in transcription, indicating that 60 nM is saturating for TFB under these conditions. The reaction mixture was overlaid with mineral oil and incubated at 65°C for 40 min. Heparin was added (to a concentration of (to 50 μ g ml⁻¹), followed 30 s later by ribonucleotide triphosphates

(500 μ M GTP, CTP and ATP, 10 μ M [α - 32 P]UTP at \sim 40 Ci mmol $^{-1}$) to initiate transcription. Reactions were stopped after 20 min by addition of 80 μ l of stop solution (20 mM EDTA containing a radiolabeled DNA recovery marker at a known concentration). Nucleic acids were purified by phenol-chloroform treatment, followed by ethanol precipitation. The transcripts were resolved by gel electrophoresis and analyzed by phosphorimaging, essentially as described previously (39, 40)

Protein-DNA Cross-linking

Transcription initiation complexes were assembled with *gdh* DNA template (1-2 nM) containing radioactively labeled positions prepared as previously described (15). The reactions contained 60nM TFB1, W44p-Bpa, R52p-Bpa, F192p-Bpa as specified, 60nM TBP and 20nM RNAP. Upon assembly photocross-linking was induced at 65 °C by exposing the reactions to UV light (350 nm) for 60 minutes with a hand-held lamp. Immediately following cross-linking, the reactions were treated with Micrococcal and DNaseI nuclease, or S1 and DNase I for 20 minutes, to digest promoter DNA not protected by the transcription complexes. The crosslinked complexes thus formed only become radioactively labeled if the Bpa residue is located in close proximity to the labeled probe position. In this manner a full range of promoter positions can be examined for specific interactions with the modified B-finger residues by visualizing the crosslinked complexes with SDS-PAGE followed by phosphorimaging.

References

1. T. J. Darcy *et al.*, *J Bacteriol* **181**, 4424 (Jul, 1999).
2. C. Hethke, A. C. Geerling, W. Hausner, W. M. de Vos, M. Thomm, *Nucleic Acids Res* **24**, 2369 (Jun 15, 1996).
3. D. Langer, J. Hain, P. Thuriaux, W. Zillig, *Proc Natl Acad Sci U S A* **92**, 5768 (Jun 20, 1995).
4. T. L. Marsh, C. I. Reich, R. B. Whitelock, G. J. Olsen, *Proc Natl Acad Sci U S A* **91**, 4180 (May 10, 1994).
5. M. Micorescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
6. S. A. Qureshi, B. Khoo, P. Baumann, S. P. Jackson, *Proc Natl Acad Sci U S A* **92**, 6077 (Jun 20, 1995).
7. T. Rowlands, P. Baumann, S. P. Jackson, *Science* **264**, 1326 (May 27, 1994).
8. D. Kostrewa *et al.*, *Nature* **462**, 323 (Nov 19, 2009).
9. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
10. C. Woese, O. Kandler, M. Wheelis, *Proc Natl Acad Sci U S A* **87**, 4576 (Jun, 1990).
11. D. Alexander *et al.*, *J Biol Chem* **279**, 32401 (Jul 30, 2004).
12. G. Kassavetis, C. Bardeleben, A. Kumar, E. Ramirez, E. Geiduschek, *Mol Cell Biol* **17**, 5299 (Sep, 1997).
13. M. S. Bartlett, *Curr Opin Microbiol* **8**, 677 (Dec, 2005).
14. M. B. Renfrow *et al.*, *J Biol Chem* **279**, 2825 (Jan 23, 2004).
15. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
16. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).
17. I. Pinto, W. H. Wu, J. G. Na, M. Hampsey, *J Biol Chem* **269**, 30569 (Dec 2, 1994).
18. W. Zhu *et al.*, *Nat Struct Biol* **3**, 122 (Feb, 1996).
19. H. T. Chen, S. Hahn, *Mol Cell* **12**, 437 (Aug, 2003).
20. S. D. Bell, C. P. Magill, S. P. Jackson, *Biochem Soc Trans* **29**, 392 (Aug, 2001).
21. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
22. X. Liu, D. Bushnell, D. Wang, G. Calero, R. Kornberg, *Science* (Nov 12, 2009).
23. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
24. Y. Ryu, P. G. Schultz, *Nat Methods* **3**, 263 (Apr, 2006).
25. L. Wang, A. Brock, B. Herberich, P. Schultz, *Science* **292**, 498 (Apr 20, 2001).
26. L. Glatzer, D. Rincon, A. Eisenstark, *Biochim Biophys Acta* **418**, 137 (Jan 19, 1976).

27. G. Sun, C. Fecko, R. Nicewonger, W. Webb, T. Begley, *Org Lett* **8**, 681 (Feb 16, 2006).
28. K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, *Science*. **296**, 1285 (2002).
29. N. Korzheva *et al.*, *Science*. **289**, 619 (2000).
30. L. M. Hsu, *Biochim Biophys Acta* **1577**, 191 (Sep 13, 2002).
31. L. M. Hsu, *Methods* **47**, 25 (Jan, 2009).
32. K. S. Murakami, S. A. Darst, *Curr Opin Struct Biol* **13**, 31 (Feb, 2003).
33. Y. Korkhin *et al.*, *PLoS Biol* **7**, e102 (May 5, 2009).
34. E. P. Geiduschek, M. Ouhammouch, *Mol Microbiol* **56**, 1397 (Jun, 2005).
35. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
36. H. T. Chen, S. Hahn, *Cell* **119**, 169 (Oct 15, 2004).
37. H. T. Chen, L. Warfield, S. Hahn, *Nat Struct Mol Biol* **14**, 696 (Aug, 2007).
38. J. F. Kugel, J. A. Goodrich, *J Biol Chem* **275**, 40483 (Dec 22, 2000).
39. M. S. Bartlett, T. Gaal, W. Ross, R. L. Gourse, *J Bacteriol* **182**, 1969 (Apr, 2000).
40. G. Vierke, A. Engelmann, C. Hebbeln, M. Thomm, *J Biol Chem* **278**, 18 (Jan 3, 2003).

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The data presented in this thesis elucidates the positional and functional importance of the TFB B-reader region in the transcription initiation process. Chapter II presents data that shows TFB2, an archaeal TFB paralog that lacks the conserved B-reader sequence, has impaired transcriptional activity across a number of promoter templates. In order to probe the difference in activity between the TFB1 and TFB2, we have created hybrid proteins by swapping their N and C terminal domains as well as their B-reader regions. In addition to examining the activity of the hybrid TFBs *in vitro*, the stimulatory activity of TFE was also examined using this system. The results show the N-terminal domain of TFB2 is responsible for its lower activity and that the conserved B-reader motif may play a role in promoter opening. The data also show that TFE can stimulate transcriptional activity of TFB by compensating for the absent B-reader (1).

We have thus established that TFB2, although it is able to direct transcription initiation from a number of promoter templates only displays at most 20% of the activity seen with TFB1. Furthermore, KMnO₄ footprinting experiments also show TFB2 is less active because it is less efficient in opening the promoter DNA (1). Domain swapping experiments presented in chapter II show the TFB2 N-terminal domain confers low activity on the TFB1 C-terminal domain (hybrid 1::2), and the

TFB1 N-terminal domain confers much higher activity on the TFB2 C-terminal domain (hybrid 2::1).

We also swapped the B-reader and linker regions between TFB1 and TFB2 in order to establish whether the low activity associated with the TFB2 N-terminus was solely attributable to the lack of a B-reader sequence. Both the 1Bf and 2Bf hybrids displayed low activity, suggesting that neither the TFB1 B-reader nor the Zn-ribbon alone is sufficient for high activity and that an interaction between the two motifs may be required for full activity.

In chapter III we examined the orientation of the TFB B-reader region in archaeal transcription initiation complexes by tracking the proximity of residues within the B-reader helix and B-reader loop to radiolabeled promoter DNA via protein-DNA cross-linking. The data shows the B-reader is in close proximity to the template strand but not to the non-template strand under reaction conditions known to result in the formation of open transcription initiation complexes (RFP_O). We also examined the position and orientation of the B-reader under conditions designed to favor RFP_C (50°C) and in the presence of initiating nucleotides. The data suggests that the B-reader loop and the B-reader helix may be involved in early initiation events that lead to the separation of the template strand away from the non-template strand and to the formation of a stable transcription bubble.

Transcription factor E has been shown to stimulate transcription up to three-fold at some promoters, and in the presence of TFB2 and to stabilize the RFP_O while the absence of a B-reader enhances TFE activation (1, 2). I have examined the role of

TFE in promoter binding and transcription complex assembly (Appendix). I confirmed that TFE stimulates transcription and that TFE activation is enhanced at suboptimal TBP concentrations at weak promoters. I also address the function of TFE by using gel-shift assays for assembly of transcription complexes, but results were complicated by non-specific RNAP binding to radiolabeled promoter DNA. The data presented shows that although the absence of the TFB B-reader increases TFE activation it does not appear to greatly affect the formation of initiation complexes. Taken together, the results support a model that suggests TFE activation takes place at least in part at steps following the formation of RFP_C.

Unanswered questions and future experiments

Does the archaeal B-reader play a role in the formation of the initial phosphodiester bond ?

We have observed the lack of a B-reader sequencer leads to decreased transcriptional activity and have confirmed the B-reader is located in close proximity to the transcription start site. It is thus possible that in addition to roles in open complex formation and transcription start site determination, the B-reader may assist in the formation of the initial phosphodiester bond between the first rNTP base-paired at +1 and the incoming rNTP. The glutamate dehydrogenase promoter can be used to examine this question, since by omitting rUTP and rATP transcription ceases after the synthesis of a 3-mer with the sequence 5'-GCC-3'. It is thus expected that transcription reactions assembled with TFBs lacking a B-reader (TFB2 or the 2Bf

hybrid) should produce a comparatively lower amount of 5'-GCC-3' than reactions assembled with TFBs containing a B-reader sequence (TFB1 or the 1Bf hybrid). The 3-nucleotide transcript can be easily radioactively labeled and visualized with high percentage denaturing PAGE.

Does the archaeal B-reader play a role in abortive transcription?

After the formation of RFP_O and the initiation of RNA synthesis, the B-reader is predicted to sterically clash with the growing RNA transcript after the addition of the 8th or 9th nucleotide. It has been hypothesized that at this point in the transcription cycle, an equilibrium exists between the formation of the elongation complex (RFP_{elong}) and the formation of a short abortive product mediated by the steric clash, which is necessary for the detachment of the RNAP from the TBP/TFB/DNA complex. Our established *in vitro* system can be used to study the contribution of the B-reader to promoter clearance and abortive transcription because short transcripts can be easily visualized and quantified in the absence (TFB2 or 2Bf) or presence (TFB1 or 1Bf) of a B-reader.

Is the positioning of the TFB B-reader promoter-dependent?

I have used *p*-Bpa TFB variants and protein-DNA cross-linking to establish the positioning of the B-reader with respect to the strong glutamate dehydrogenase promoter. However, the position and orientation of the B-reader helix and loop may vary in a promoter-sequence dependent fashion. It is possible that differences in

promoter strength and in overall transcriptional activity may be, at least in part, correlated with promoter-specific differences in B-reader positioning. Protein-DNA cross-linking studies with the *p*-Bpa B-reader variants described so far can be used to determine the role, if any, of the promoter sequence in the position and orientation of the B-reader.

The location of other parts of the NTD: where is the B-linker?

The recently published RNAP-TFIIB co-crystal structure suggests that other parts of the TF(II)B N-terminus play a role in transcription initiation events. The TFB linker connects the B-reader region with the TFB C-terminal domain and is also located in the active cleft of RNAP. In the new structure, a B-linker helix (residues 71-76 in *P.furiosus* TFB) and a B-linker strand were resolved and predicted to be in close proximity to the non-template strand while mutations in the archaeal TFB B-linker helix result in a defect in promoter opening (3). It is thus likely that while the B-reader is poised to interact with the template strand, the B-linker helix and/or B-linker strand may interact with the non-template strand, perhaps assisted by TFE under certain conditions. The *p*-Bpa cross-linking approach described in chapter III can be used to determine the orientation and position of the TFB B-linker helix and B-linker strand with respect to specific positions on the template and non-template strands and can thus serve to complement the recent structural and genetic data with functional information.

What are the determinants of promoter opening?

It has been established that TFB2 is less active in transcription and that this decreased activity is due, at least in part, to a defect in promoter opening (Chapter II and (1)) while Chapter III and Kostrewa et al. (4) show that the B-reader and B-linker are in close proximity to the upstream edge of the transcription bubble. However, the specific residues and mechanism responsible for nucleating the formation of the transcription bubble remain unknown. An extensive mutagenesis study of TFB B-reader and B-linker residues coupled with potassium-permanganate foot-printing experiments to monitor the extent of promoter opening can confirm the importance of specific residues in this process. The position and orientation of key residues relative to either the template or the non-template strand could then be determined with *p*-Bpa cross-linking.

What is the contribution of other TFB NTD regions to TFB function?

In chapter II I have shown that swapping in the TFB1 NTD restored the transcriptional activity of TFB2 to near TFB1 levels while solely swapping in the TFB1 B-reader/B-linker region did not have a similar effect. It is possible that the NTD Zn-ribbon motif contributes significantly to TFB transcriptional activity. Additional, more specific, swaps could further characterize the contribution of the NTD to TFB function. Swapping the TFB2 Zn-ribbon sequence with its TFB1 counterpart or the swapping of specific portions of the B-reader and/or B-linker

regions could thus reveal the contribution of specific TFB NTD structures to the overall difference in transcriptional activity observed between TFB1 and TFB2.

How does TFE contribute to promoter opening?

I have investigated TFE activation and showed in chapter II and the Appendix that TFE can stimulate transcription in a promoter dependent fashion, especially under suboptimal TBP concentrations and in the absence of a B-reader sequence. It is likely that in addition to enhancing TBP binding and/or stabilizing the initiation complex, TFE plays a role in post recruitment events as it has been shown that TFE plays a role in promoter opening at the upstream edge of the transcription bubble and that it stabilizes the initiation complex in an open conformation (5). However, the specific TFE surfaces and residues mediating activation are yet to be determined. Residues important for TFE activation could first be identified using mutagenesis and the *in vitro* transcription conditions already established (Chapter II and Appendix) in combination with potassium-permanganate footprinting. Protein-DNA cross-linking could then be used to further characterize the mechanism of TFE activation by determining the position of specific TFE residues relative to promoter DNA. This could be done by constructing *p*-Bpa TFE variants similar to the ones described for TFB in chapter III and by performing analogous cross-linking experiments.

References

1. M. Micorescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
2. S. Grunberg, M. S. Bartlett, S. Naji, M. Thomm, *J Biol Chem* **282**, 35482 (Dec 7, 2007).
3. D. Kostrewa *et al.*, *Nature* (Oct 9, 2009).
4. D. Kostrewa *et al.*, *Nature* **462**, 323 (Nov 19, 2009).
5. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).

Terminal References

Chapter I

1. C. Woese, O. Kandler, M. Wheelis, *Proc Natl Acad Sci U S A* **87**, 4576 (Jun, 1990).
2. L. Aravind, E. V. Koonin, *Nucleic Acids Res* **27**, 4658 (Dec 1, 1999).
3. E. Perez-Rueda, J. Collado-Vides, *J Mol Evol* **53**, 172 (Sep, 2001).
4. Y. Korkhin *et al.*, *PLoS Biol* **7**, e102 (May 5, 2009).
5. A. Hirata, K. Murakami, *Curr Opin Struct Biol* **19**, 724 (Dec, 2009).
6. T. J. Darcy *et al.*, *J Bacteriol* **181**, 4424 (Jul, 1999).
7. D. Langer, J. Hain, P. Thuriaux, W. Zillig, *Proc Natl Acad Sci U S A* **92**, 5768 (Jun 20, 1995).
8. D. D. Leipe, L. Aravind, E. V. Koonin, *Nucleic Acids Res* **27**, 3389 (Sep 1, 1999).
9. N. A. Woychik, M. Hampsey, *Cell* **108**, 453 (Feb 22, 2002).
10. F. J. Asturias, *Curr Opin Struct Biol* **14**, 121 (Apr, 2004).
11. S. Buratowski, H. Zhou, *Proc Natl Acad Sci U S A* **90**, 5633 (Jun 15, 1993).
12. S. Malik, D. Lee, R. Roeder, *Mol Cell Biol* **13**, 6253 (Oct, 1993).
13. T. Gibson, J. Thompson, A. Blocker, T. Kouzarides, *Nucleic Acids Res* **22**, 946 (Mar 25, 1994).
14. D. B. Nikolov *et al.*, *Nature* **377**, 119 (Sep 14, 1995).
15. F. T. Tsai, P. B. Sigler, *Embo J* **19**, 25 (Jan 4, 2000).
16. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
17. S. A. Qureshi, S. P. Jackson, *Mol Cell* **1**, 389 (Feb, 1998).
18. W. Deng, S. G. Roberts, *Genes Dev* **19**, 2418 (Oct 15, 2005).
19. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
20. W. Zhu *et al.*, *Nat Struct Biol* **3**, 122 (Feb, 1996).
21. H. T. Chen, P. Legault, J. Glushka, J. G. Omichinski, R. A. Scott, *Protein Sci* **9**, 1743 (Sep, 2000).
22. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).
23. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
24. D. Kostrewa *et al.*, *Nature* **462**, 323 (Nov 19, 2009).
25. X. Liu, D. Bushnell, D. Wang, G. Calero, R. Kornberg, *Science* (Nov 12, 2009).
26. G. Kassavetis, G. Letts, E. Geiduschek, *EMBO J* **20**, 2823 (Jun 1, 2001).
27. G. Kassavetis, S. Han, S. Naji, E. Geiduschek, *J Biol Chem* **278**, 17912 (May 16, 2003).
28. I. Pinto, D. E. Ware, M. Hampsey, *Cell* **68**, 977 (Mar 6, 1992).

29. I. Pinto, W. H. Wu, J. G. Na, M. Hampsey, *J Biol Chem* **269**, 30569 (Dec 2, 1994).
30. C. S. Bangur, T. S. Pardee, A. S. Ponticelli, *Mol Cell Biol* **17**, 6784 (Dec, 1997).
31. N. A. Hawkes, R. Evans, S. G. Roberts, *Curr Biol* **10**, 273 (Mar 9, 2000).
32. N. A. Hawkes, S. G. Roberts, *J Biol Chem* **274**, 14337 (May 14, 1999).
33. M. B. Renfrow *et al.*, *J Biol Chem* **279**, 2825 (Jan 23, 2004).
34. H. T. Chen, S. Hahn, *Cell* **119**, 169 (Oct 15, 2004).
35. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
36. S. D. Bell, A. B. Brinkman, J. van der Oost, S. P. Jackson, *EMBO Rep* **2**, 133 (Feb, 2001).
37. A. Meinhart, J. Blobel, P. Cramer, *J Biol Chem* **278**, 48267 (Nov 28, 2003).
38. B. L. Hanzelka, T. J. Darcy, J. N. Reeve, *J Bacteriol* **183**, 1813 (Mar, 2001).
39. M. Ouhammouch, F. Werner, R. O. Weinzierl, E. P. Geiduschek, *J Biol Chem* **279**, 51719 (Dec 10, 2004).
40. D. Forget, M. F. Langelier, C. Therien, V. Trinh, B. Coulombe, *Mol Cell Biol* **24**, 1122 (Feb, 2004).
41. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).
42. K. Yokomori, C. P. Verrijzer, R. Tjian, *Proc Natl Acad Sci U S A* **95**, 6722 (Jun 9, 1998).
43. T. I. Lee, R. A. Young, *Annu Rev Genet* **34**, 77 (2000).
44. Y. Ohkuma, S. Hashimoto, C. K. Wang, M. Horikoshi, R. G. Roeder, *Mol Cell Biol* **15**, 4856 (Sep, 1995).
45. Y. Ohkuma, R. G. Roeder, *Nature* **368**, 160 (Mar 10, 1994).
46. H. Serizawa, J. Conaway, R. Conaway, *J Biol Chem* **269**, 20750 (Aug 12, 1994).
47. M. Thomm, G. Wich, *Nucleic Acids Res* **16**, 151 (Jan 11, 1988).
48. M. S. Bartlett, *Curr Opin Microbiol* **8**, 677 (Dec, 2005).
49. S. D. Bell, P. L. Kosa, P. B. Sigler, S. P. Jackson, *Proc Natl Acad Sci U S A* **96**, 13662 (Nov 23, 1999).
50. C. S. Bangur, S. L. Faitar, J. P. Folster, A. S. Ponticelli, *J Biol Chem* **274**, 23203 (Aug 13, 1999).
51. T. S. Pardee, C. S. Bangur, A. S. Ponticelli, *J Biol Chem* **273**, 17859 (Jul 10, 1998).
52. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
53. M. Micoreescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
54. K. R. Shockley *et al.*, *Appl Environ Microbiol* **69**, 2365 (Apr, 2003).
55. M. Micoreescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).

Chapter II

1. T. J. Darcy *et al.*, *J Bacteriol* **181**, 4424 (Jul, 1999).

2. D. Langer, J. Hain, P. Thuriaux, W. Zillig, *Proc Natl Acad Sci U S A* **92**, 5768 (Jun 20, 1995).
3. S. D. Bell, A. B. Brinkman, J. van der Oost, S. P. Jackson, *EMBO Rep* **2**, 133 (Feb, 2001).
4. B. L. Hanzelka, T. J. Darcy, J. N. Reeve, *J Bacteriol* **183**, 1813 (Mar, 2001).
5. W. Hausner, J. Wettach, C. Hethke, M. Thomm, *J Biol Chem* **271**, 30144 (Nov 22, 1996).
6. T. L. Marsh, C. I. Reich, R. B. Whitelock, G. J. Olsen, *Proc Natl Acad Sci U S A* **91**, 4180 (May 10, 1994).
7. S. A. Qureshi, P. Baumann, T. Rowlands, B. Khoo, S. P. Jackson, *Nucleic Acids Res* **23**, 1775 (May 25, 1995).
8. T. Rowlands, P. Baumann, S. P. Jackson, *Science* **264**, 1326 (May 27, 1994).
9. J. R. Palmer, C. J. Daniels, *J Bacteriol* **177**, 1844 (Apr, 1995).
10. S. A. Qureshi, S. P. Jackson, *Mol Cell* **1**, 389 (Feb, 1998).
11. M. Thomm, G. Wich, *Nucleic Acids Res* **16**, 151 (Jan 11, 1988).
12. S. D. Bell, P. L. Kosa, P. B. Sigler, S. P. Jackson, *Proc Natl Acad Sci U S A* **96**, 13662 (Nov 23, 1999).
13. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
14. M. Micoreescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
15. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).
16. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
17. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
18. M. B. Renfrow *et al.*, *J Biol Chem* **279**, 2825 (Jan 23, 2004).
19. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 31624 (Oct 13, 2000).
20. I. Pinto, W. H. Wu, J. G. Na, M. Hampsey, *J Biol Chem* **269**, 30569 (Dec 2, 1994).
21. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
22. H. T. Chen, S. Hahn, *Mol Cell* **12**, 437 (Aug, 2003).
23. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
24. K. R. Shockley *et al.*, *Appl Environ Microbiol* **69**, 2365 (Apr, 2003).
25. T. Santangelo, L. Cubonova, C. James, J. Reeve, *J Mol Biol* **367**, 344 (Mar 23, 2007).
26. M. T. Facciotti *et al.*, *Proc Natl Acad Sci U S A* **104**, 4630 (Mar 13, 2007).
27. S. Karlin, J. Mrazek, *J Bacteriol* **182**, 5238 (Sep, 2000).
28. S. Karlin, J. Mrazek, J. Ma, L. Brocchieri, *Proc Natl Acad Sci U S A* **102**, 7303 (May 17, 2005).
29. G. J. Schut, S. D. Brehm, S. Datta, M. W. Adams, *Journal of bacteriology* **185**, 3935 (Jul, 2003).
30. G. J. Schut, J. Zhou, M. W. Adams, *J Bacteriol* **183**, 7027 (Dec, 2001).
31. J. DiRuggiero, L. A. Achenbach, S. H. Brown, R. M. Kelly, F. T. Robb, *FEMS microbiology letters* **111**, 159 (Aug 1, 1993).

32. J. Rollins, I. Veras, S. Cabarcas, I. Willis, L. Schramm, *Int J Biol Sci* **3**, 292 (2007).
33. A. Saxena, B. Ma, L. Schramm, N. Hernandez, *Mol Cell Biol* **25**, 9406 (Nov, 2005).
34. D. Alexander *et al.*, *J Biol Chem* **279**, 32401 (Jul 30, 2004).
35. X. Zhao, L. Schramm, N. Hernandez, W. Herr, *Mol Cell* **11**, 151 (Jan, 2003).
36. Y. Huang, E. McGillicuddy, M. Weindel, S. Dong, R. Maraia, *Nucleic Acids Res* **31**, 2108 (Apr 15, 2003).
37. T. J. Santangelo, L. Cubonova, C. L. James, J. N. Reeve, *J Mol Biol* **367**, 344 (Mar 23, 2007).
38. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).
39. D. Forget, M. Langelier, C. Therien, V. Trinh, B. Coulombe, *Mol Cell Biol* **24**, 1122 (Feb, 2004).
40. S. Grunberg, M. S. Bartlett, S. Naji, M. Thomm, *J Biol Chem* **282**, 35482 (Dec 7, 2007).
41. N. S. Baliga *et al.*, *Mol Microbiol* **36**, 1184 (Jun, 2000).
42. P. E. Rouviere *et al.*, *Embo J* **14**, 1032 (Mar 1, 1995).
43. C. Hethke, A. C. Geerling, W. Hausner, W. M. de Vos, M. Thomm, *Nucleic Acids Res* **24**, 2369 (Jun 15, 1996).
44. G. A. Kassavetis, A. Kumar, E. Ramirez, E. P. Geiduschek, *Mol Cell Biol* **18**, 5587 (Sep, 1998).
45. Y. Korkhin, O. Littlefield, P. J. Nelson, S. D. Bell, P. B. Sigler, *Methods Enzymol* **334**, 227 (2001).
46. F. T. Robb *et al.*, *Methods Enzymol* **330**, 134 (2001).
47. G. Vierke, A. Engelmann, C. Hebbeln, M. Thomm, *J Biol Chem* **278**, 18 (Jan 3, 2003).
48. M. S. Bartlett, T. Gaal, W. Ross, R. L. Gourse, *J Bacteriol* **182**, 1969 (Apr, 2000).

Chapter III

1. T. J. Darcy *et al.*, *J Bacteriol* **181**, 4424 (Jul, 1999).
2. C. Hethke, A. C. Geerling, W. Hausner, W. M. de Vos, M. Thomm, *Nucleic Acids Res* **24**, 2369 (Jun 15, 1996).
3. D. Langer, J. Hain, P. Thuriaux, W. Zillig, *Proc Natl Acad Sci U S A* **92**, 5768 (Jun 20, 1995).
4. T. L. Marsh, C. I. Reich, R. B. Whitelock, G. J. Olsen, *Proc Natl Acad Sci U S A* **91**, 4180 (May 10, 1994).
5. M. Micoreescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
6. S. A. Qureshi, B. Khoo, P. Baumann, S. P. Jackson, *Proc Natl Acad Sci U S A* **92**, 6077 (Jun 20, 1995).
7. T. Rowlands, P. Baumann, S. P. Jackson, *Science* **264**, 1326 (May 27, 1994).
8. D. Kostrewa *et al.*, *Nature* **462**, 323 (Nov 19, 2009).

9. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
10. C. Woese, O. Kandler, M. Wheelis, *Proc Natl Acad Sci U S A* **87**, 4576 (Jun, 1990).
11. D. Alexander *et al.*, *J Biol Chem* **279**, 32401 (Jul 30, 2004).
12. G. Kassavetis, C. Bardeleben, A. Kumar, E. Ramirez, E. Geiduschek, *Mol Cell Biol* **17**, 5299 (Sep, 1997).
13. M. S. Bartlett, *Curr Opin Microbiol* **8**, 677 (Dec, 2005).
14. M. B. Renfrow *et al.*, *J Biol Chem* **279**, 2825 (Jan 23, 2004).
15. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
16. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).
17. I. Pinto, W. H. Wu, J. G. Na, M. Hampsey, *J Biol Chem* **269**, 30569 (Dec 2, 1994).
18. W. Zhu *et al.*, *Nat Struct Biol* **3**, 122 (Feb, 1996).
19. H. T. Chen, S. Hahn, *Mol Cell* **12**, 437 (Aug, 2003).
20. S. D. Bell, C. P. Magill, S. P. Jackson, *Biochem Soc Trans* **29**, 392 (Aug, 2001).
21. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
22. X. Liu, D. Bushnell, D. Wang, G. Calero, R. Kornberg, *Science* (Nov 12, 2009).
23. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
24. Y. Ryu, P. G. Schultz, *Nat Methods* **3**, 263 (Apr, 2006).
25. L. Wang, A. Brock, B. Herberich, P. Schultz, *Science* **292**, 498 (Apr 20, 2001).
26. L. Glatzer, D. Rincon, A. Eisenstark, *Biochim Biophys Acta* **418**, 137 (Jan 19, 1976).
27. G. Sun, C. Fecko, R. Nicewonger, W. Webb, T. Begley, *Org Lett* **8**, 681 (Feb 16, 2006).
28. K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, *Science* **296**, 1285 (2002).
29. N. Korzheva *et al.*, *Science* **289**, 619 (2000).
30. L. M. Hsu, *Biochim Biophys Acta* **1577**, 191 (Sep 13, 2002).
31. L. M. Hsu, *Methods* **47**, 25 (Jan, 2009).
32. K. S. Murakami, S. A. Darst, *Curr Opin Struct Biol* **13**, 31 (Feb, 2003).
33. Y. Korkhin *et al.*, *PLoS Biol* **7**, e102 (May 5, 2009).
34. E. P. Geiduschek, M. Ouhammouch, *Mol Microbiol* **56**, 1397 (Jun, 2005).
35. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
36. H. T. Chen, S. Hahn, *Cell* **119**, 169 (Oct 15, 2004).
37. H. T. Chen, L. Warfield, S. Hahn, *Nat Struct Mol Biol* **14**, 696 (Aug, 2007).
38. J. F. Kugel, J. A. Goodrich, *J Biol Chem* **275**, 40483 (Dec 22, 2000).
39. M. S. Bartlett, T. Gaal, W. Ross, R. L. Gourse, *J Bacteriol* **182**, 1969 (Apr, 2000).

40. G. Vierke, A. Engelmann, C. Hebbeln, M. Thomm, *J Biol Chem* **278**, 18 (Jan 3, 2003).

Chapter IV

1. M. Micorescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
2. S. Grunberg, M. S. Bartlett, S. Naji, M. Thomm, *J Biol Chem* **282**, 35482 (Dec 7, 2007).
3. D. Kostrewa *et al.*, *Nature* (Oct 9, 2009).
4. D. Kostrewa *et al.*, *Nature* **462**, 323 (Nov 19, 2009).
5. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).

APPENDIX

AN INVESTIGATION INTO THE ROLE OF TFE IN TRANSCRIPTION

COMPLEX ASSEMBLY

Transcription factor E (TFE) stimulates archaeal transcription in a TBP and promoter dependent fashion while the presence of the TFB B-reader obviates this activity. In this chapter I present data that suggests TFE activation takes place both during and after the formation of transcription initiation complexes and provide an explanation for how this observation may be reconciled with the dependence on a low TBP concentration.

Introduction

The basic components of archaeal and RNAP II transcriptional machinery have evolved from common ancestors (1) and even though archaeal species encode a simpler version of the eukaryotic system the structure and function of the basal apparatus has been conserved (2-5). In both domains of life promoter recognition is mediated by a saddle shaped TATA binding protein (TBP), which recognizes and binds the minor groove of a conserved TATA box sequence located approximately 30 bp upstream of the transcription start site. TBP binding significantly bends the promoter region and allows transcription factor B (TFB in Archaea and TFIIB/Brf in the RNAP II/RNAP III systems) to bind the TBP-promoter complex via sequence specific contacts between a helix-turn-helix motif in its C-terminal domain motif and

the TFB recognition element (BRE), a guanine-cytosine rich sequence located immediately upstream of the TATA box (6, 7). The TFB N-terminal domain then recruits RNAP to the promoter by binding to the RNAP dock domain (8, 9).

Upon RNAP recruitment, conformational changes within the pre-initiation complex lead to the unwinding of the DNA double-helix and the loading of the template strand into the active site of RNAP, a complex process that involves not only RNAP but also the TFB N-terminus and transcription factor E (TFE), a third archaeal transcription factor with distinct homology to the N-terminal domain of the TFIIE α subunit (10-13). The function of the α subunit of eukaryotic TFIIE has been shown to play a role in stabilizing the unwinding of the promoter double helix in conjunction with TFIIF, a process that occurs before the template strand can be loaded into the RNAP active site (14). Archaeal TFE has been less well characterized but two studies have shown that although not absolutely required for transcription, TFE stimulates transcription up to three-fold at weak promoters and under less than optimal TBP concentrations (10, 11) while more recently Naji et al. have shown TFE plays a role in promoter opening at the upstream edge of the transcription bubble and in stabilizing open complexes (12).

Several lines of evidence suggest that the TFB and TFE roles in initiation are mediated by interactions between the factors and DNA. X-ray crystallography and Fe-BABE cleavage studies in yeast suggest the TFIIB N-terminus is in close proximity to the RNAP active site and that the TFIIB B-reader may interact with the template strand based on a modeled DNA duplex (15-18). Our own protein-DNA cross-linking

studies with archaeal TFB show the TFB N-terminus is in close proximity to the template strand around the transcription start site while separate cross-linking studies indicate TFE is close to the non-template strand at the upstream edge of the transcription bubble (19). In addition, we have shown that transcription initiation complexes are subject to greater activation by TFE in the absence of the TFB B-reader (20). We thus hypothesized that the TFB N-terminus and TFE may act in concert to facilitate strand separation during the formation of the open complex and suggest that the role of TFE in transcription initiation extends beyond stabilizing TBP binding, as was suggested by the earlier studies. The experiments in this chapter test the hypothesis that TFE activation is greatest under suboptimal TBP concentrations and provide support for TFE activation after the assembly of the archaeal transcription initiation complex. In addition, the data presented here extends the characterization of TFE to additional promoters and shows that while weak promoters benefit most from TFE activation, TFE does not activate transcription by enhancing complex formation.

Results

TFE can activate archaeal transcription, but it is not yet clear whether TFE activation is due to increased transcription initiation complex formation or if TFE acts by stimulating events that take place after the recruitment of RNAP, such as promoter opening. A better understanding of the interplay between transcription factors and the mechanisms of TFE activation provide an opportunity for identifying the step(s) affected. If TFE affects the assembly of the transcription initiation complex, this

would be observed as an increase in the amount of initiation complexes formed in its presence. We first asked if activation of transcription by purified *Pyrococcus furiosus* TFE is promoter sequence-dependent by comparing the transcriptional activity of complexes assembled on weak promoters with the activity of complexes transcribing from promoters known to be highly active *in vitro*. We compared TFE transcriptional activity using two previously characterized promoters. The glutamate dehydrogenase promoter is capable of directing a high level of transcription *in vitro* while the tRNA_{asn} (GTT) promoter is comparatively weak (Chapter II and Figure A.1).

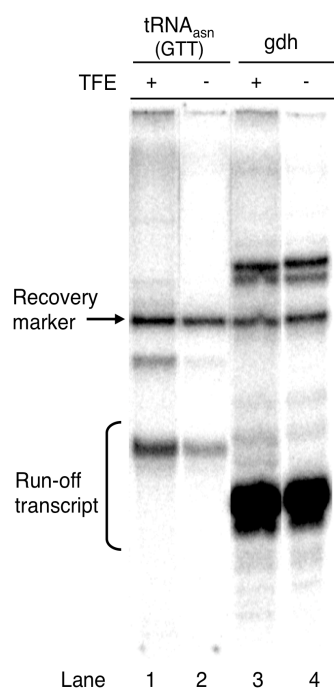
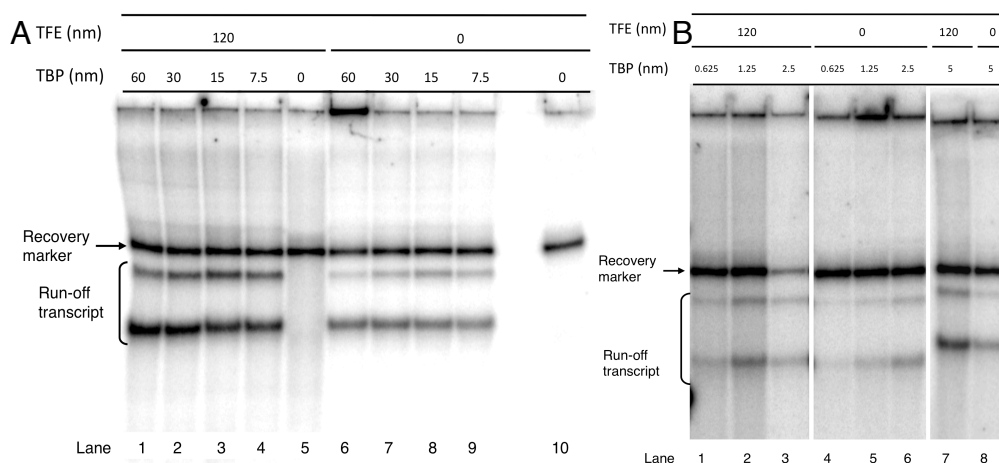


Figure A.1. TFE activates transcription from a weak promoter more than from a strong promoter. Multiple-round transcriptions were performed as described in the legend to Figure A.5 in the absence or presence of TFE (120 nM). The positions of the transcript and recovery marker are indicated.

We assembled standard transcription reactions as described in the methods section in the absence or presence 120 nM purified TFE, a concentration determined to be saturating under our experimental conditions (data not shown). As previously observed (20), TFE stimulated gdhP directed transcriptional activity only slightly

(compare lanes 3 and 4 in Figure A.1) but increased transcription from the tRNA_{asn} (GTT) approximately 2.3 fold. Thus, the effect of *P.furiosus* TFE on transcription is promoter-dependent, consistent with a previous report on *Methanococcus* TFE activity (11).

To further investigate the determinants of TFE activation, we assembled transcription reactions using *gdhP* or tRNA_{asn} (GTT) in the presence or absence of TFE over a range of TBP concentrations. As previously observed, Figure A.2 shows that under normal TBP concentrations (60 nM) there is only a slight activation by TFE. However, under suboptimal TBP concentrations (below 10 nM), TFE activation exceeds three-fold for tRNA_{asn} (GTT). It is possible that TFE enhances TBP binding at suboptimal concentrations or that the effect of TFE on promoter melting at low TBP concentrations compensates for the slow rate of TBP binding while at high TBP concentrations this effect is no longer required for optimal activity.



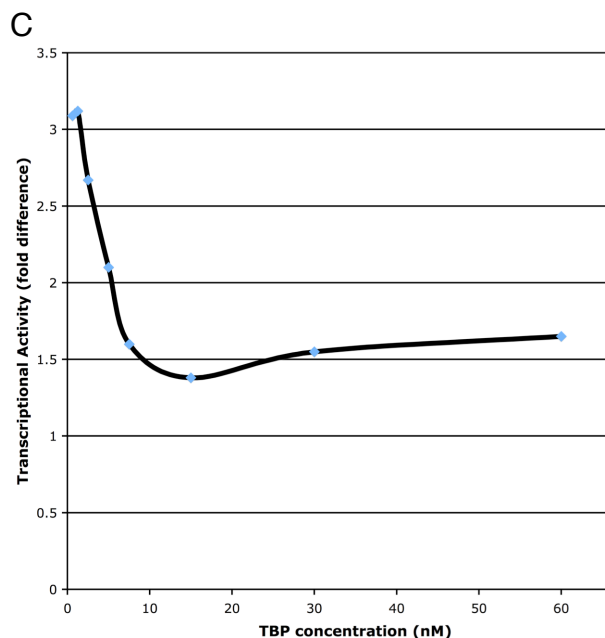


Figure A.2 – TFE activation is enhanced under suboptimal TBP concentrations for tRNA_{asn} (GTT). Transcription reactions were assembled as describe in the methods section in the presence or absence of 120 nM purified TFE. **A.** TFE activation between 60 – 7.5 nM TBP **B.** TFE activation between 5 – 0.625 nM TBP **C.** TFE activation was quantified by calculating the fold difference between the transcriptional activity achieved in its presence and absence of TFE at a range of TBP concentrations (between 0 and 60 nM).

We next wanted to determine whether TFE stimulates transcription at the level of pre-initiation complex formation, whether it influences steps in initiation following complex assembly or both. We first attempted to establish whether differences in complex assembly could be observed in a promoter dependent fashion. Complex formation was compared using an electrophoretic mobility shift assay (EMSA). Transcription factor/RNAP complexes were assembled using radiolabeled *gdh* or tRNA_{asn}(GTT) or tRNA_{lys}(TTT) promoter templates. Figure A.3 shows that TFE is able to bind archaeal transcription initiation complexes under our experimental conditions on both promoter templates, as indicated by the higher molecular weight shifts marked by arrows (4.3A, lanes 2 and 8 and 4.3B, lane 5).

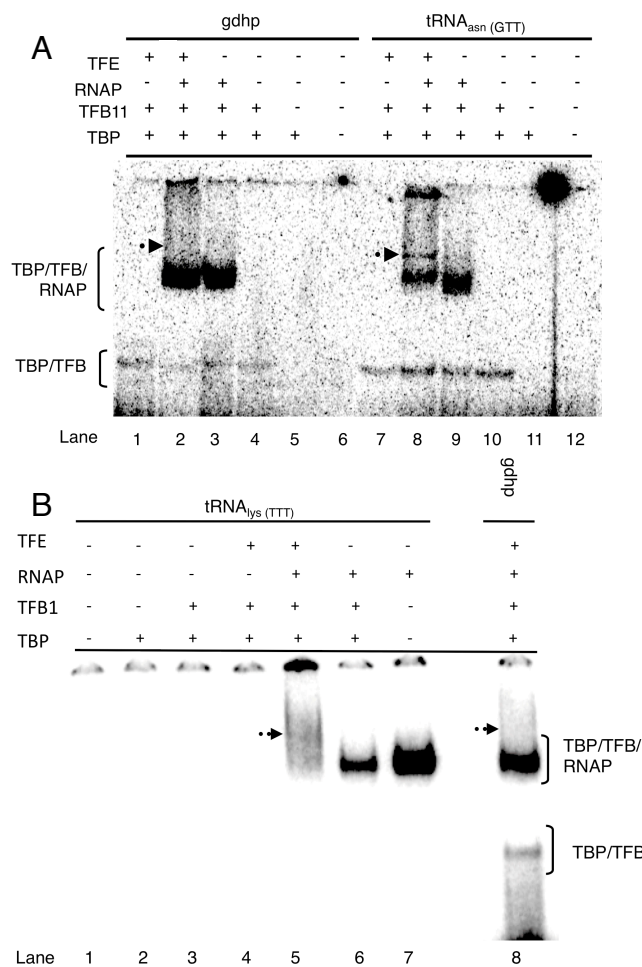


Figure A.3. TFE binds transcription initiation complexes in a promoter dependent fashion. Radiolabeled *gdh* and *tRNA_{asn}(GTT)* (panel A) or *tRNA_{lys}(TTT)* (panel B) promoter DNA was incubated with the proteins indicated above the wells as described in the methods section. The bands marked by the arrows indicate transcription complexes supershifted by TFE binding.

The data shows TFE binding is RNAP dependent since higher molecular weight complexes containing TFE are not observed in the absence of RNAP. In addition, TFE did not significantly increase the amount of DNA-TBP-TFB complexes assembled on either promoter (compare lanes 1 to 4 and 7 to 10 in 4.3A), indicating that TFE activation occurs at steps following complex assembly. However, TFE containing complexes assembled on the weak *tRNA_{asn} (GTT)* promoter were more stable (Figure AA, compare lanes 2 and 8). As such, the promoter dependence of TFE activation at

high TBP concentrations could be explained by optimal complex assembly in the absence of TFE on the strong *gdh* promoter which may inherently support a higher rate of TBP binding or promoter opening and poor complex assembly or promoter opening on the weaker the tRNA_{asn} (GTT) promoter.

We next wanted to determine if the presence or absence of a TFB B-reader sequence affects the ability of TFE to bind transcription initiation complexes, since we have shown previously that TFE can partially rescue the impaired transcriptional activity observed with TFB2 (chapter II and (20)). We examined complex assembly in the presence of TFB2 under the same experimental conditions employed for TFB1 directed assembly (Figure A.2). Figure A.3 shows that TFE directed complex assembly on *gdhP* in the absence of a B-reader is not markedly different than in its presence (lanes 3,4,9 and 10, compare with lanes 2,3,8 and 9 in Figure A.2). This suggests that the B-reader dependent differences in TFE activation observed previously (20) can occur without a significant increase in complex formation.

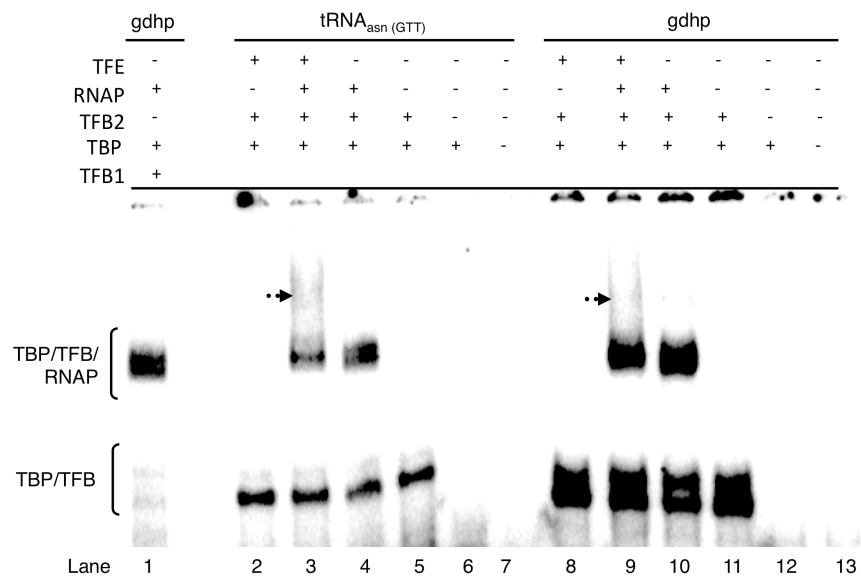


Figure A.4–TFE does not greatly affect transcription complex assembly in the absence of the TFB B-reader. Radiolabeled *gdh* or tRNA_{asn}(GTT) promoter DNA was incubated with the proteins indicated above the wells as described in the methods section. The bands marked by the arrows indicate transcription complexes supershifted by TFE binding.

Discussion

Transcription by archaeal RNAP can be activated by TFE *in vitro*, although the magnitude of the effect varies as a function of promoter-sequence. (10, 11) The mechanism of transcription activation by TFE remains undetermined. We have shown that the absence of a TFB B-reader sequence increases TFE activation in an *in vitro* transcription system, suggesting that TFE may play a role in promoter opening (Chapter II and (20)). The data in this chapter supports our previous observation that TFE stimulates transcription initiation by affecting post-recruitment steps. We have also determined that under standard *in vitro* conditions TFE activation is more

pronounced in the presence of the weak tRNA_{asn} (GTT) promoter than in the presence of the strong *gdh* promoter (Figure A.1) under suboptimal TBP concentrations.

Two separate but not mutually exclusive explanations may account for the dependence of TFE activation on promoter sequence. It is possible that the two promoter sequences differ in their ability to bind TBP, causing the observed difference in transcription efficiency. If TFE stabilizes TBP binding, it may preferentially enhance the formation of complexes at the weaker promoter, since at a strong promoter this step is not expected to be rate-limiting (21). Alternatively, TFE may play a role in the formation of the transcription bubble or in the stability of the bubble after it has formed, perhaps enhancing strand separation by interacting with the non-template strand. Like in previous studies (10, 11), we have shown that TBP concentration directly affects TFE activation at the weak promoter (Figure A.2), with suboptimal concentrations (10 nM and below) eliciting levels of activation up to three-fold while the strong promoter remains unresponsive to TFE activation. However, the higher rate of TBP binding and closed complex accumulation at a strong promoter may obscure the enhanced rate of TFE-assisted promoter opening. If the transcription initiation complex formation proceeds at a high rate under high TBP concentrations, a relatively large amount of closed complexes will form. In the absence of TFE some fraction of these complexes will undergo promoter opening and enter productive transcription. Low TBP concentrations result in lower levels of run-off transcription, suggesting initiation occurs at a lower rate (10, 11). A smaller relative number of closed complexes thus form per unit time but the same fraction undergo promoter

opening and enter productive transcription. The net result is a significant difference in runoff transcript at low TBP concentrations, as observed in our experiments. TFE addition under high TBP concentrations may result in an increase in the rate of open complex formation but this effect may be masked by the already high rate of closed complex accumulation, such that only a small difference in run-off transcription is observed (Figure A.1, compare lanes 3 and 4). However, at low TBP concentrations, the increase in the rate of open complex formation elicited by TFE addition can be observed since the rate of closed complex accumulation is expected to be lower (Figure A.2). We attempted to determine if TFE assists in transcription complex formation with EMSA and both a strong *gdh* promoter and the weaker *tRNA_{asn}* (GTT) promoter. Although we have shown TFE is able to bind transcription initiation complexes in a RNAP dependent fashion (Figure A.3), we saw no significant difference in the levels of initiation complexes between the two promoters, in spite of the increased transcription elicited by TFE at the weak *tRNA* promoter compared to the strong *gdh* promoter. Taken together these data indicate suboptimal TBP concentrations and/or suboptimal promoter-sequences prime the transcription machinery for TFE activation, and suggest that TFE activation may also affect post-recruitment steps in the initiation cycle.

We also wanted to determine if the lack of detectable TFE-dependent complex formation may be due to the presence of a B-reader motif, since our earlier results (see chapter II) indicated the B-reader might render TFE activation superfluous. To this end the experiment shown in Figure A.4 was performed under identical conditions to

those described for Figure A.3 but TFB2, which lacks a conserved B-reader sequence, was substituted for TFB1. We did not observe a significant difference in the accumulation of initiation complexes between the TFB2 and TFB1 experiments (compare Figure A.s 4.3 and 4.4) on either the *gdh* or *tRNA_{asn}* (GTT) promoters, indicating that TFE activation in the absence of a B-reader occurs after initiation complex assembly. However, the interpretation of the gel-shift results is complicated by non-specific RNAP binding, which occurred under these conditions in the absence of transcription factors with the *tRNA_{lys}* (TTT) promoter (Figure A.3B, lane 7). As such, in the absence of additional controls, it is difficult to confidently determine the fraction of TBP/TFB/RNAP complexes from non-specifically bound RNAP.

Experimental Procedures

Gene cloning and protein purification.

Recombinant *P. furiosus* TBP was prepared as described previously (22). *P. furiosus* TFB genes (Pf1377 for TFB1 and Pf0687 for TFB2) were amplified by PCR and cloned into the vector pET21b-H6-Nco (23), creating constructs that encoded proteins with a six-histidine tag at the N terminus. The overexpressed proteins were purified to near homogeneity (estimated by gel electrophoresis and Coomassie staining) by Ni^{++} ion chromatography. Native RNAP used for the experiments whose results are shown in Figure A.2, 5b, 6, and 7 was purified from *P. furiosus* cells as described by Hethke et al. (22), while the native RNAP used for the experiments

whose results are shown in Figure A.3, 5a, 8, and 9 was purified by the method described by Korkhin et al. (24).

Promoter DNA templates.

Predicted promoters were amplified from *P. furiosus* genomic DNA by PCR, as follows (sequence positions according to the annotation of Robb et al. (25)): for *gdhp* (Pf1602), positions 1494929 to 1495025; for *tRNAAsn(GTT)*, positions 1287414 to 1287519; for *tRNALys(TTT)*, positions 508388 to 508493; for *Pfr001* (16S rRNA), positions 136562 to 136687; for Pf1974, positions 1823493 to 1823598; and for Pf1377, positions 1292896 to 1292998 (primer sequences are available upon request). The promoters for Pf1882, Pf1883, and Pf1790 were described previously (26).

Standard transcription assays.

Transcription reactions were performed essentially as described previously (27). The 12.5- μ l reaction mixtures contained 40 mM Na-HEPES, pH 7.3, 250 mM NaCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 5 mM beta-mercaptoethanol, 5% (v/v) glycerol, and 0.1 μ g μ l⁻¹ bovine serum albumin; 10 nM promoter DNA was combined with 60 nM TBP, 60 nM TFB1, or TFB2 and 20 nM RNAP. Increasing the TFB concentration to values greater than 60 nM resulted in no increase in transcription, indicating that 60 nM is saturating for TFB under these conditions. The reaction mixture was overlaid with mineral oil and incubated at 65°C for 40 min. Heparin was added (to a concentration of 50 μ g ml⁻¹), followed 30 s later by ribonucleotide triphosphates (500 μ M GTP, CTP and ATP, 10 μ M [α -32P]UTP at ~40 Ci mmol⁻¹) to initiate

transcription. Reactions were stopped after 20 min by addition of 80 μ l of stop solution (20 mM EDTA containing a radiolabeled DNA recovery marker at a known concentration). Nucleic acids were purified by phenol-chloroform treatment, followed by ethanol precipitation. The transcripts were resolved by gel electrophoresis and analyzed by phosphorimaging, essentially as described previously (26, 28).

Electromotility Shift Assays

Transcription complexes were assembled as described above for the standard transcription assay but contained 0.2 nM end-labeled promoter DNA and no nucleotides were added. Upon incubation at 65°C for 30 min, heparin was added (to a concentration of 50 μ g ml⁻¹) and the reactions were immediately loaded on a 4% polyacrylamide gel (5mM MgCl₂, 0.5mM beta-mercaptoethanol, 5% (v/v) glycerol and gel-shift buffer [250 mM Tris pH 8.5, 1.9 M glycine, 10 mM EDTA]).

References

1. C. Woese, O. Kandler, M. Wheelis, *Proc Natl Acad Sci U S A* **87**, 4576 (Jun, 1990).
2. L. Aravind, E. V. Koonin, *Nucleic Acids Res* **27**, 4658 (Dec 1, 1999).
3. A. Hirata, *Seikagaku* **81**, 377 (May, 2009).
4. Y. Korkhin *et al.*, *PLoS Biol* **7**, e102 (May 5, 2009).
5. E. Perez-Rueda, J. Collado-Vides, *J Mol Evol* **53**, 172 (Sep, 2001).
6. S. D. Bell, C. P. Magill, S. P. Jackson, *Biochem Soc Trans* **29**, 392 (Aug, 2001).
7. O. Littlefield, Y. Korkhin, P. B. Sigler, *Proc Natl Acad Sci U S A* **96**, 13668 (Nov 23, 1999).
8. S. D. Bell, S. P. Jackson, *J Biol Chem* **275**, 12934 (Apr 28, 2000).

9. C. P. Magill, S. P. Jackson, S. D. Bell, *J Biol Chem* **276**, 46693 (Dec 14, 2001).
10. S. D. Bell, A. B. Brinkman, J. van der Oost, S. P. Jackson, *EMBO Rep* **2**, 133 (Feb, 2001).
11. B. L. Hanzelka, T. J. Darcy, J. N. Reeve, *J Bacteriol* **183**, 1813 (Mar, 2001).
12. S. Naji, S. Grunberg, M. Thomm, *J Biol Chem* **282**, 11047 (Apr 13, 2007).
13. F. Werner, R. O. Weinzierl, *Mol Cell Biol* **25**, 8344 (Sep, 2005).
14. Y. Ohkuma, *J Biochem* **122**, 481 (Sep, 1997).
15. D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **303**, 983 (Feb 13, 2004).
16. D. Kostrewa *et al.*, *Nature* (Oct 9, 2009).
17. X. Liu, D. Bushnell, D. Wang, G. Calero, R. Kornberg, *Science* (Nov 12, 2009).
18. G. Miller, S. Hahn, *Nat Struct Mol Biol* **13**, 603 (Jul, 2006).
19. S. Grünberg, M. S. Bartlett, S. Naji, M. Thomm, *J Biol Chem* . (Oct 5, 2007).
20. M. Micorescu *et al.*, *J Bacteriol* **190**, 157 (Jan, 2008).
21. J. F. Kugel, J. A. Goodrich, *J Biol Chem* **275**, 40483 (Dec 22, 2000).
22. C. Hethke, A. C. Geerling, W. Hausner, W. M. de Vos, M. Thomm, *Nucleic Acids Res* **24**, 2369 (Jun 15, 1996).
23. G. A. Kassavetis, A. Kumar, E. Ramirez, E. P. Geiduschek, *Mol Cell Biol* **18**, 5587 (Sep, 1998).
24. Y. Korkhin, O. Littlefield, P. J. Nelson, S. D. Bell, P. B. Sigler, *Methods Enzymol* **334**, 227 (2001).
25. F. T. Robb *et al.*, *Methods Enzymol* **330**, 134 (2001).
26. G. Vierke, A. Engelmann, C. Hebbeln, M. Thomm, *J Biol Chem* **278**, 18 (Jan 3, 2003).
27. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *J Biol Chem* **279**, 5894 (Feb 13, 2004).
28. M. S. Bartlett, T. Gaal, W. Ross, R. L. Gourse, *J Bacteriol* **182**, 1969 (Apr, 2000).