Title

Dissection of E. coli RNA polymerase holoenzyme into the minimal portion sufficient for non-template strand binding

Author

Young, Brian Anthony

Publication Date

2003

Peer reviewed|Thesis/dissertation

Dissection of *E. coli* RNA polymerase holoenzyme into the minimal portion sufficient for non-template strand binding

by

Brian Anthony Young

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

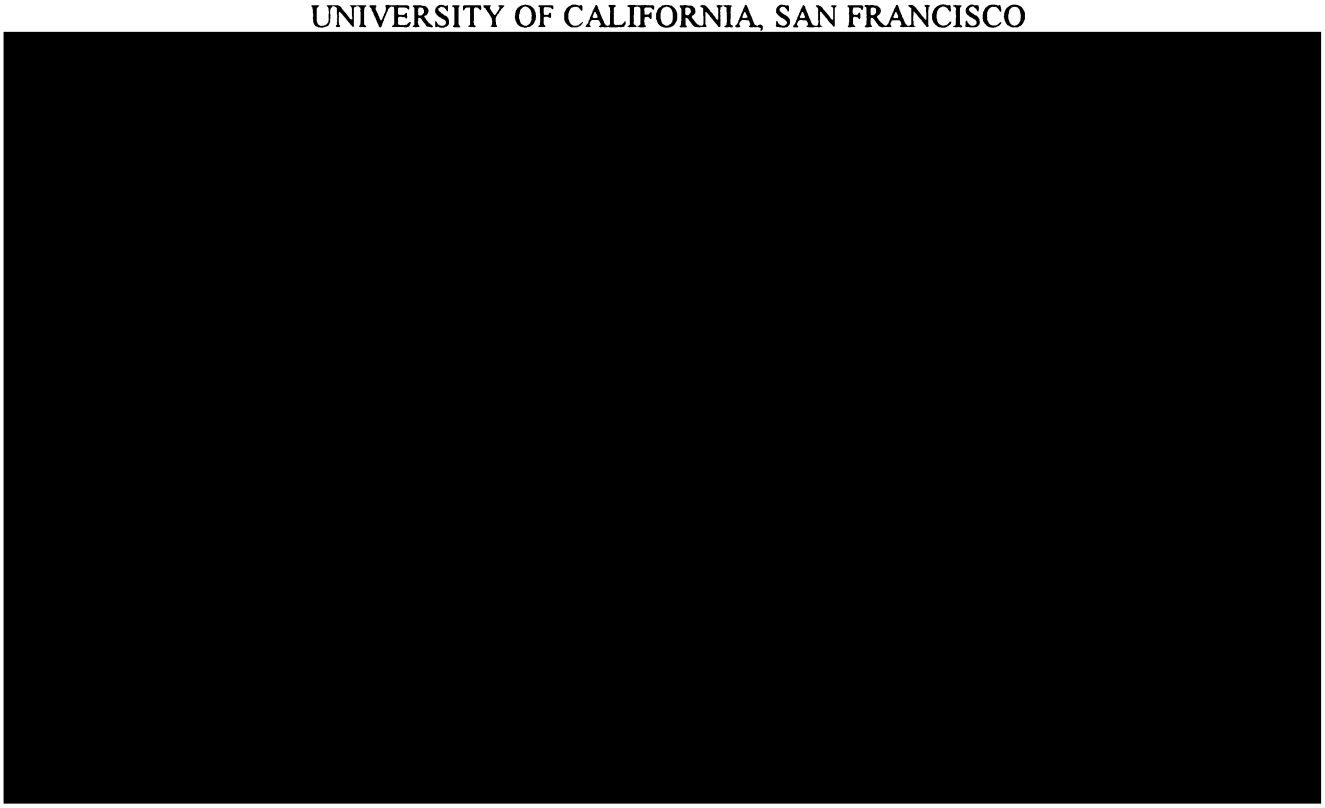
Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Acknowledgements

As I had an opportunity to express my gratitude to the long list of people who contributed to this work in my thesis talk, here I would like to thank just a special few. First, I would like to thank my thesis advisor and mentor, Carol Gross. I learned a tremendous amount from Carol, and in Carol's lab. Her door is always open to people in her lab, and I learned an enormous amount from our frequent conversations. I also really appreciated that, as much as Carol tried to steer me towards a reasonable project, she gave me the opportunity to pursue a few grandiose dream projects. Finally, one of the things I came to appreciate and respect most about Carol is that, though she wants good science to come from her lab, she really cares about her people and wants to see them success.

Next I would like to thank my other lab mentors, Tanja Gruber and Christophe Herman. I probably learned as much from each of them as from Carol, and I am very grateful for all they spent with me. Each is a remarkable scientist, and I expect great things from them in the future. Christophe really impresses me with his creativity and encyclopedic grasp of the literature; Tanja with her enormous technical gifts, determination and patience. Perhaps most of all, I am grateful for their friendship.

I would also like to thank the two fantastic technicians in the lab, Chi Zen Lu and Hong Ji Zhong. They are the glue that holds the lab together; without them the Gross Lab would be a very different and inferior place. I would especially like to thank Chi Zen Lu for her great kindness and patience.

Finally, and perhaps most of all, I would like to thank my parents. Thank you for your constant support. Despite how hard it must have been for you, you never made me feel guilty for having to be so far away. I love you.

Statement regarding Previously Published Material with Multiple Authors

Chapter Two has been previously published as (Young, B. A., Anthony, L. C., Gruber, T. M., Arthur, T. M., Heyduk, E., Lu, C. Z., Sharp, M. M., Heyduk, T., Burgess, R. R., and Gross, C. A. (2001). A coiled-coil from the RNA polymerase β' subunit allosterically induces selective non-template strand binding by σ_{70}. Cell *105*, 935-944.) and is reprinted here with permission from Elsevier Science. I performed all experiments with the exception of: 1) figure 3, which was performed by Larry Anthony, Terry Arthur, Ewa Heyduk, Tomasz Heyduk, and Richard Burgess, 2) the non-competitive coimmunoprecipitation σ/core binding experiments (not shown in the figures but mentioned in the text), which were performed by Meghan Sharp. Tanja Gruber provided me with the cloned domains constructs of core polymerase and Chi Zen Lu helped me with purification of proteins. The manuscript was written by me.

Chapter Three has been previously published as (Young, B.A., Gruber, T. M., and Gross, C. A. (2002) and is reprinted here with permission from Elsevier Science. Views of transcription initiation. Cell *109*, 417-420). This manuscript was written by me with revisions and editing by Tanja Gruber and Carol Gross.

Date: Tue, 01 Apr 2003 12:42:52 PST
From: byoung@itsa.ucsf.edu
To: permissions@elsevier.co.uk
Subject: (No subject)

To whom it may concern:

I am writing to request permission to include the following publications in my doctoral thesis dissertion:

Young, B.A., Gruber, T.M., Gross, C.A. (2002) Views of Transcription Initiation. Cell 109:417-420

Young, B.A., Anthony, L.C., Gruber, T.M., Arthur, T.M., Heyduk, E., Lu, C.Z., Sharp, M.M., Heyduk, T., Burgess, R.R., and Gross, C.A. (2001) A Coiled-coil from the RNA polymerase beta'Subunit Allosterically Induces Selective Nontemplate Strand Binding by sigma70. Cell 105: 935-944

This dissertation will be archived on microfilm by University Microfilms Incorporated, and they request permission to supply single copies of the dissertation upon demand.

Please email me or respond by FAX at ((415)-514-4080.

Sincerely,
Brian Young
Gross Laboratory
Department of Microbiology and Immunology
Genentech Hall, Room S 376
600 16th Street, Box 2200
University of California San Francisco
San Francisco, CA 94143-2200
415-476-1493

Date: Tue, 15 Apr 2003 09:54:43 +0100
From: "Clapton, Sheila (ELS)"
<S.Clapton@elsevier.co.uk>
To: "'byoung@itsa.ucsf.edu'" <byoung@itsa.ucsf.edu>
Subject: Permissions

<<...OLE_Obj...>>

15 April 2003 Our ref: HW/smc/April 2003.jl283

Mr Brian Young

byoung@itsa.ucsf.edu

Dear Mr Young

CELL, Vol 109, 2002, Pages 417-420, Young et al, "Views of Transcription Initiation"; Vol 105, 2001, Pages 935-944, Young et al, "A coiled-coil from the RNA ..."

As per your letter dated 1 April 2003, we hereby grant you permission to reprint the aforementioned material at no charge in your thesis subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

> "Reprinted from Publication title, Vol number, Author(s),Title of article, Pages No., Copyright (Year), with permission from Elsevier".

3. Reproduction of this material is confined to the purpose for which permission is hereby given.

4. This permission is granted for non-exclusive world English rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form. Should you have a specific electronic project in mind please reapply for permission.

5. This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially,please reapply for permission.

Yours sincerely
<<...OLE_Obj...>>

Helen Wilson
Rights Manager

Your future requests will be handled more quickly if you complete the online form at
www.elsevier.com/homepage/guestbook/?form=permis
<http://www.elsevier.com/homepage/guestbook/?form=permis>

Dissection of *E. coli* RNA polymerase holoenzyme into the minimal portion sufficient for non-template strand binding

by

Brian Young

Carol A. Gross, Ph.D.
Thesis Advisor
Chairperson, Thesis Committee

Abstract:

Initiation of transcription is the first step in gene expression and one of the major points at which expression is regulated. In prokaryotes transcription is accomplished by the five subunit core RNA polymerase in conjunction with the initiation specific σ factor; the complex of these factors is referred to as holoenzyme. Crucial to initiation is the recognition and melting of promoter DNA. Holoenzyme carries out several binding activities to effect this melting. Here I determine the minimal portion of holoenzyme capable of its single stranded non-template strand binding activity. This activity may serve to complete the process of promoter melting by "clamping off" or precluding the non-template strand from re-annealing with the template strand. I find that the small (~48) amino acid coiled-coil of the β' subunit is the important part of core polymerase for this activity and that it appears to function as an allosteric effector of σ conformation. I provide evidence that the allosteric binding site for the coiled coil is the most conserved sub-region of σ, region 2.2. Furthermore, I show that the β'coiled coil and Domain 2 of σ--comprising only 47 kDa, or 1/10, of the holoenzyme—are sufficient to reconstitute

this important activity. Finally, I review the current structural information concerning holoenzyme.

Table of Contents

List of Figures

Chapter One

Chapter Two

Chapter Three

Chapter One

Introduction

Initiation of transcription is the first step in gene expression and one of the major points at which this process is regulated. The enzyme that accomplishes this task, RNA polymerase, is a remarkable molecular machine. In this early phase of transcription alone it must: recognize promoter DNA from the vast excess of non-promoter DNA, pull apart the double helix at the promoter to access the template strand of DNA, and begin a nucleic acid polymer using only mononucleotides (as opposed to simply extending an existing nucleic acid polymer as do most polymerases). Finally, having begun an RNA, it must make the transition from being a sequence specific promoter binding protein to being non-specifically but tightly attached to the DNA, so that it can move processively forward (and in some cases backward) along the DNA. In prokaryotes, the catalytic core of RNA polymerase (core for short) capable of elongating transcripts is made up of five tightly associated subunits:$\beta\beta'\alpha_2\omega$. Core is highly conserved throughout bacteria and shares considerable sequence homology, and even more structural homology with its eukaryotic counterparts (Cramer et al., 2000; Ebright, 2000; Zhang et al., 1999).

Core is unable to begin transcription at a promoter alone; to do so, it requires the initiation specific σ subunit. Almost all bacteria contain several σ factor: a primary σ that transcribes most housekeeping genes and alternative σ's that predominantly transcribe genes involved in response to a variety of stress(Gross et al., 1992). In this thesis, I use the term σ synonymously with the primary σ in *E.coli*, σ^{70}. The housekeeping σ's of all bacteria contain four conserved regions that generally correspond to σ's folded domains (Campbell et al., 2002; Lonetto et al., 1992). Domain 1 is an autoinhibitory module; domains 2, 3 and 4 contain both core and promoter recognition determinants. (Note that, though for consistency sake I call "domain 1" a domain, it may not be tightly folded (Vassylyev et al., 2002).) In eukaryotes the role of σ is played by a set of general transcription factors (in mRNA synthesis about 30) sharing little homology with σ(Orphanides et al., 1996). Despite this apparent difference in initiation factors, the

high structural homology among prokaryotic and eukaryotic RNA polymerase suggests the possibility that understanding the simple case of prokaryotic initiation may give insight into what may be an analogous mechanism in eukaryotes. Even if this is not the case, we believe understanding the structural biology of prokaryotic initiation will still be interesting (and from a medical standpoint perhaps even more interesting) as it may provide targets for future anti-microbial agents.

The complex of σ and core is referred to as holoenzyme. Holoenzyme recognizes promoters containing two conserved hexamers, one at -35 and one at -10 relative to the start site of transcription (designated +1). The polymerase melts these promoters roughly between -11 to +4(Record, 1996). This thesis focuses on understanding how RNA polymerase effects this recognition and, especially, this melting.

While it is known that the formation of melted promoter complexes (also called open complexes) is a complex process requiring a number of conformational intermediates(Record, 1996), even now it is not clear exactly how these intermediates bring about DNA melting. However, as I was beginning my thesis work, we were learning that holoenzyme was capable of a number of different binding activities that may drive this process of melting. These are: binding to double stranded DNA(Gross et al., 1998); binding to the "fork junction" between double stranded and single stranded DNA located at the upstream edge of the transcriptional bubble(Guo and Gralla, 1998); and binding the non-template strand of the promoter in the –10 region in single stranded form(Callaci and Heyduk, 1998; Fedoriw et al., 1998; Heyduk and Heyduk, 1999; Huang et al., 1997; Roberts and Roberts, 1996; Severinova et al., 1996). The purpose of initial recognition of the double stranded version of promoters is readily understandable; the purpose of the other two activities is somewhat speculative. (Figure 1-1). It has been proposed that the fork junction binding activity may nucleate promoter melting (Fenton et al., 2000). Assume RNA polymerase has bound to the promoter in double stranded form. At low frequency, individual base pairs break and reform and reference therein). If a

base near what will be the upstream edge of the transcriptional bubble briefly loses pairing with its complement and through thermal motion leaves the duplex, (see (deHaseth and Helmann, 1995) and references therein) this creates a double stranded/single stranded junction. The ability of RNA polymerase to bind this "fork junction" structure, pulling on the unpaired base, may destabilize the surrounding DNA, nucleating greater "breathing" of the promoter. Once nucleation has been achieved, the single stranded non-template binding activity may then complete melting by clamping off this strand so that it can no longer re-anneal with its complement, leaving the template strand accessible for polymerization. These models are in line with previous kinetic and thermodynamic evidence indicating that melting occurs via a slow nucleation step and then a more rapid melting completion step (Liermo, 1990). One should note, however, the base flipping model of nucleation posited above is only one possibility of how nucleation may occur. Other models such as polymerase torqueing or bending the DNA are possible as well (Figure 1-2) and these models are obviously not mutually exclusive (i.e. bending might faciliate base flipping, etc)(deHaseth and Helmann, 1995).

It was clear at the beginning of my work that wherease holoenzyme is capable of all of these activities, neither σ alone nor core alone is capable of any of them. My thesis work was to determine which portions of holoenzyme were participating in these activities and the nature of their contribution. I therefore set about attempting to dissect the ~450 kDa polymerase into the minimal portions sufficient to accomplish these activities. Previous evidence had indicated that σ was critically important for all of these activities: 1) when conserved region 1 of σ, an autoinhibitory region, is removed, σ has weak double stranded promoter binding activity(Dombroski et al., 1992); 2) free σ^{54} (an alternate sigma factor which does not share sequence homology with the other sigma factors) has been demonstrated to have weak fork junction binding activity (Guo and Gralla, 1998)and 3) σ appeared to be the only portion of holoenzyme able to form zero length crosslinks specifically with non-template strand (Marr and Roberts, 1997).

Furthermore, σ was the only portion of the polymerase known to specifically interact with promoter DNA: conserved σ sub-region 2.3/2.4 was the only portion of the polymerase known to interact with the –10 hexamer and conserved σ sub-region 4.2 was the only portion of the polymerase known to interact with the –35 hexamer(Gross et al., 1998). With this in mind, I was persuaded that σ might be responsible for all the activities of holoenzyme described above, with core acting only as an allosteric effector stimulating σ to perform these activities. As σ appears to be lost at the transition to processive elongation, this is an aesthetically appealing hypothesis: thus all initiation specific activities are confined to the initiation specific factor. As small molecules consisting of only a few dozen atoms can convey allosteric effects in metabolism, I speculated that in this scenario, only very small portions of core polymerase might be necessary to act as allosteric effectors of σ conformation. If this were true it would make dissecting the polymerase to determining what was important in holoenzyme for these activities especially feasible. However, even if this weren't true the dissections might still be practical. For example, if core participated in the activities by directly contributing DNA binding determinants, as long as these binding determinants were located relatively close in the primary structure of the core subunits to σ binding motifs the dissection approach might still yield results. There was reason to believe this to be likely, as it was clear from the emerging work of Tanja Gruber, a postdoctoral fellow in the lab, that many portions of core interacted with σ (Gruber et al., 2001).

In order to divide the subunits I was able to take advantage of previous work that had suggested the stable folded domain structure of the different subunits of core(Severinov et al., 1992; Severinov et al., 1996). This work involved proteolytic cleavage and split site studies (in some organisms, what in *E. coli* is a single subunit is separated into multiple polypeptides; it had been shown that the *E. coli* subunits can be analogously split and still reconstitute into a functional enzyme). Using this information, Dr. Gruber had already cloned and purified these putative domains of core in order to

look for sites of interaction between core and σ. In addition, as I was beginning my work the high resolution structure of *Thermus aquaticus* core became available(Zhang et al., 1999). This provided me with additional domain information I might be able to use to re-engineer domain constructs as well as perhaps further divide core. The structure also provided valuable information about the secondary structure of core. If I could identify roughly which portion of the polymerase was important for an activity, even if the smaller fragments of this region were incapable of folding on their own, knowing the secondary structure of the sub-regions in the native enzyme might allow me to use previous protein engineering discoveries to find ways to assist fragment folding. For example, the Frankel lab had previously stimulated a portion of the HIV Rev protein-- known to form an α helix in the native protein-- to fold in isolation by adding charge groups and alanine residues to its termini (Tan et al., 1993). All this led me to hope that the dissection approach might be practical.

As I was beginning my work, a preliminary report suggested that the isolated β' subunit 1407 amino acids) was sufficient together with σ to reconstitute the non-template strand binding activity (Kulbachinskiy et al., 1999). Although intact β' contains almost half of the mass of core and although the report described no controls to rule out the possibility that the activity was due to trace amounts of intact core polymerase in the β' preparation, it also served to encourage me that the dissection approach might work.

In this thesis (chapter 2) I present evidence that a small (~48 amino acid) coiled coil in the β' subunit of core RNA polymerase, first discovered to be a σ binding motif by the Burgess laboratory (Arthur et al., 2000; Arthur and Burgess, 1998), allosterically induces σ to bind to the non-template strand via an interaction with the most highly conserved sub-region of σ, region 2.2. Furthermore, I show that the β'coiled coil and Domain 2 of σ--or only 47 kDa, or 1/10, of the holoenzyme are sufficient to reconstitute this important activity. I also review recent structural information about prokaryotic holoenzyme (Chapter 3). Finally I discuss the significance of my findings and

considering possible future directions that take into account my work and the recent structural data (Chapter 4).

References

Arthur, T. M., Anthony, L. C., and Burgess, R. R. (2000). Mutational analysis of beta '260-309, a sigma 70 binding site located on Escherichia coli core RNA polymerase. J Biol Chem *275*, 23113-23119.

Arthur, T. M., and Burgess, R. R. (1998). Localization of a sigma70 binding site on the N terminus of the Escherichia coli RNA polymerase beta' subunit. J Biol Chem *273*, 31381-31387.

Callaci, S., and Heyduk, T. (1998). Conformation and DNA binding properties of a single-stranded DNA binding region of sigma 70 subunit from Escherichia coli RNA polymerase are modulated by an interaction with the core enzyme. Biochemistry *37*, 3312-3320.

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell *9*, 527-539.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640-649.

deHaseth, P. L., and Helmann, J. D. (1995). Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol Microbiol *16*, 817-824.

Dombroski, A. J., Walter, W. A., Record, M. T., Siegele, D. A., and Gross, C. A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell *70*, 501-512.

Ebright, R. H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol *304*, 687-698.

Fedoriw, A. M., Liu, H., Anderson, V. E., and deHaseth, P. L. (1998). Equilibrium and kinetic parameters of the sequence-specific interaction of Escherichia coli RNA polymerase with nontemplate strand oligodeoxyribonucleotides. Biochemistry *37*, 11971-11979.

Fenton, M. S., Lee, S. J., and Gralla, J. D. (2000). Escherichia coli promoter opening and -10 recognition: mutational analysis of sigma70. Embo J *19*, 1130-1137.

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol *63*, 141-155.

Gross, C. A., Lonetto, M., and Losick, R. (1992). Bacterial sigma factors. In Transcriptional regulation, S. L. McKnight, and K. R. Yamamoto, eds. (Cold Spring Harbor, Cold Spring Harbor Laboratory Press), pp. 129-176.

Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., Artsimovitch, I., Geszvain, K. M., Arthur, T. M., Burgess, R. R., *et al.* (2001). Binding of the initiation factor σ70 to core RNA polymerase is a multi-step process. Molecular Cell *8*, 21-31.

Guo, Y., and Gralla, J. D. (1998). Promoter opening via a DNA fork junction binding activity. Proc Natl Acad Sci U S A *95*, 11655-11660.

Heyduk, E., and Heyduk, T. (1999). Architecture of a complex between the sigma70 subunit of Escherichia coli RNA polymerase and the nontemplate strand oligonucleotide. Luminescence resonance energy transfer study. J Biol Chem *274*, 3315-3322.

Huang, X., Lopez de Saro, F. J., and Helmann, J. D. (1997). sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA. Nucleic Acids Res *25*, 2603-2609.

Kulbachinskiy, A., Mustaev, A., Goldfarb, A., and Nikiforov, V. (1999). Interaction with free beta' subunit unmasks DNA-binding domain of RNA polymerase sigma subunit. FEBS Lett *454*, 71-74.

Liermo, S., Record, M.T. Jr. (1990). Structural, thermodynamic and kinetic studies of the interaction of Eσ RNA polymerase with promoter DNA. In Nucleic Acids and Molecular Biology, D. M. J. L. a. F. Eckstein, ed. (Berlin, Springer-Verlag), pp. 123-151.

Lonetto, M., Gribskov, M., and Gross, C. A. (1992). The σ^{70} family: sequence conservation and evolutionary relationships. J Bacteriol *174*, 3843-3849.

Marr, M. T., and Roberts, J. W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science *276*, 1258-1260.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

Record, M. T., Reznikoff, W., Craig M., McQuade, K., and Schlax, P. (1996). *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcriptional initiation. In *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, F. C. Neidhardt, ed. (Washington, D.C., ASM Press), pp. 792-820.

Roberts, C. W., and Roberts, J. W. (1996). Base-specific recognition of the nontemplate strand of promoter DNA by E. coli RNA polymerase. Cell *86*, 495-501.

Severinov, K., Mustaev, A., Kashlev, M., Borukhov, S., Nikiforov, V., and Goldfarb, A. (1992). Dissection of the beta subunit in the Escherichia coli RNA polymerase into domains by proteolytic cleavage. J Biol Chem *267*, 12813-12819.

Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S. A., and Goldfarb, A. (1996). Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of Escherichia coli RNA polymerase. J Biol Chem *271*, 27969-27974.

Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E. N., Roberts, J. W., Chait, B. T., and Darst, S. A. (1996). Domain organization of the Escherichia coli RNA polymerase sigma 70 subunit. J Mol Biol *263*, 637-647.

Tan, R., Chen, L., Buettner, J. A., Hudson, D., and Frankel, A. D. (1993). RNA recognition by an isolated alpha helix. Cell *73*, 1031-1040.

Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature *417*, 712-719.

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

Figure 1-1: How holoenzyme binding activities may drive melting

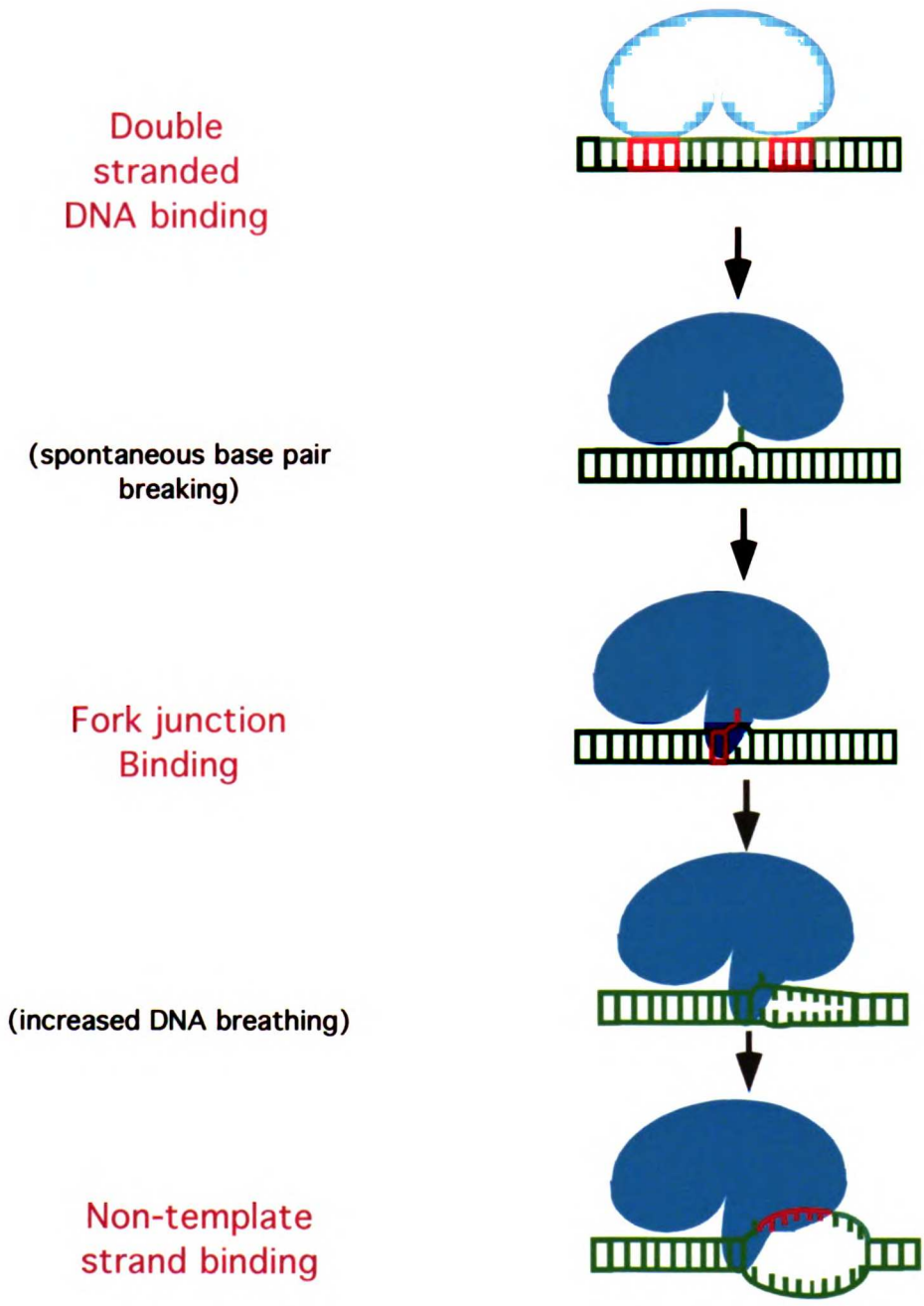
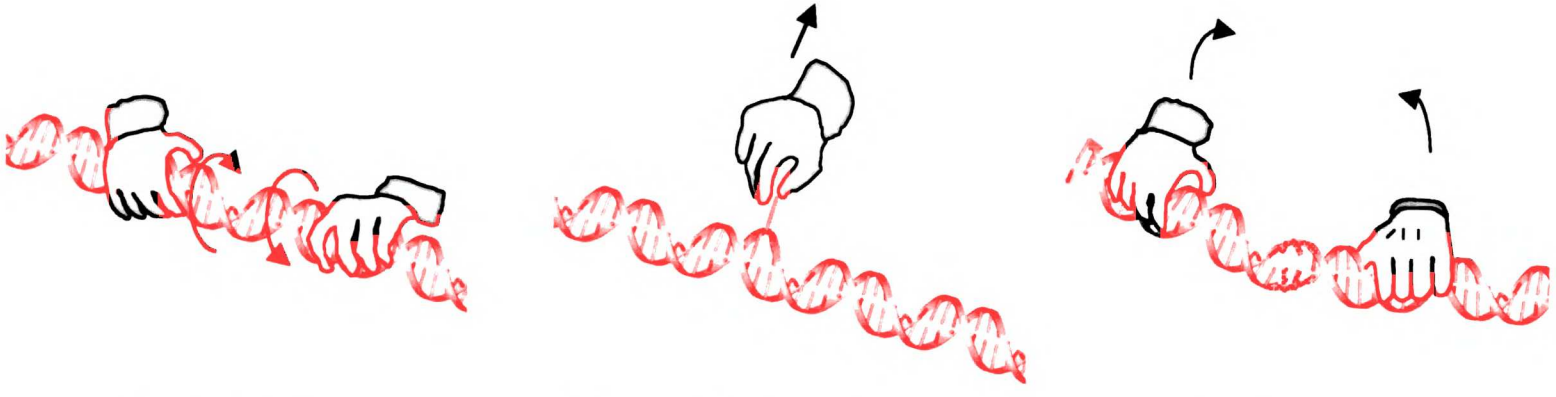


Figure 1-2



Torquing Base Flipping Bending

Chapter Two

A Coiled-Coil from the RNA Polymerase β' Subunit Allosterically Induces Selective Non-Template Strand Binding by σ^{70}

Summary

For transcription to initiate, RNA polymerase must recognize and melt promoters. Selective binding to the non-template strand of the –10 region of the promoter is central to this process. We show that a 48 amino acid (aa) coiled-coil from the β' subunit (aa262-309) induces σ^{70} to perform this function almost as efficiently as core RNA polymerase itself. We provide evidence that interaction between the β' coiled-coil and Region 2.2 of σ^{70} promotes an allosteric transition that allows σ^{70} to selectively recognize the non-template strand. As the β'262-309 peptide can function with the previously crystallized portion of σ^{70}, non-template recognition can be reconstituted with only 48 kDa, or 1/10 of holoenzyme.

Introduction

Transcription of RNA is a crucial part of gene expression in all organisms. Transcription initiation requires that RNA polymerase recognize and then locally unwind promoter DNA to gain access to the template strand. This process is carried out as a collaboration between RNA polymerase and its initiation factors, a form of RNA polymerase called holoenzyme. The multisubunit RNA polymerase is conserved throughout evolution (Young, 1991; Sentenac et al., 1992; Ebright, 2000) The five subunits of prokaryotic RNA polymerase (β',β, α_2, and ω) are all conserved in eukaryotic polymerases, though the latter contain several additional subunits (Young 1991; Sentenac et al., 1992; Ebright, 2000; Minakhin et al., 2001). In contrast, initiation factors are distinct in prokaryotic and eukaryotic organisms. A set of proteins called

general transcription factors (e.g. TFIIB, TFIID, TFIIF, etc. for RNA polymerase II) facilitate initiation in eukaryotes, whereas a single protein, called σ, serves as an initiation factor in prokaryotes (Orphanides et al., 1996; Record et al., 1996). Several alternative sigma factors are present in almost all bacteria; here we consider initiation mediated by σ^{70}, the primary or housekeeping sigma factor in *Escherichia coli* (Gross et al., 1992).

σ^{70} holoenzyme recognizes promoters characterized by two conserved hexamers centered at –10 and –35 relative to the start site of transcription, and then melts promoters roughly between –11 to +1, a process called open complex formation. Though the structural mechanism of this process is unknown, several discrete activities have been defined: binding to double stranded DNA in the promoter (Gross et al., 1998), binding to the junction between double stranded and single stranded DNA, (Guo and Gralla 1998;Matlock and Heyduk, 2000) and binding to the non-template strand of the –10 region of the promoter (Roberts and Roberts, 1996; Severinova et al., 1996; Huang et al., 1997; Marr and Roberts, 1997; Callaci and Heyduk, 1998; Fedoriw et al., 1998; Heyduk and Heyduk, 1999). Although the initiation factor σ^{70} provides most or all of the specific recognition determinants for each of these processes, free σ^{70} is unable to perform any of these activities by itself (Kenney et al., 1989; Siegele et al. 1989; Daniels et al., 1990; Waldburger et al., 1990; Dombroski et al., 1992; Dombroski et al. 1993, Severinova et al., 1996, Dombroski, 1997; Marr and Roberts, 1997; Naryshkin et al., 2000.) Rather, core RNA polymerase (core) is required and is likely to contribute at least two functions: unmasking recognition determinants in sigma and strengthening binding by providing additional non-specific DNA interactions. We would like to dissect the very large (450 kDa) holoenzyme into the minimal portions capable of reconstituting each activity to facilitate future biochemical and structural study of open complex formation.

Selective binding of holoenzyme to the non-template strand of the –10 region of the promoter is likely to capture, stabilize and extend transient strand separation. We

therefore focused our attention on determining the minimal assembly of σ^{70} and core that could carry out this activity. There is good understanding of the regions of σ^{70} required for this process. The σ^{70} family of proteins has four regions of sequence conservation which have been divided into subregions (see Figure 2-5A for a description of these conserved regions; Lonetto et al., 1992). When bound to core, the fragment of σ^{70} which has been crystallized (which extends from the middle of conserved Region 1 to the end of conserved Region 2.4) specifically recognizes the non-template strand (Severinova et al., 1996;Malhotra et al. 1996). Within this fragment, Region 2.4 has at least some of the recognition determinants for non-template strand binding as mutational alterations in 2.4 that broaden promoter specificity also broaden recognition of the non-template strand (Marr and Roberts, 1997). From the core side, it appears that the isolated β' subunit, together with σ^{70}, is sufficient for selective recognition of non-template strand (Kulbachinskiy et al., 1999).

In this report, we show that a peptide containing only 48 of the 1407 amino acids of β' confers to σ^{70} (or to a fragment of σ^{70} corresponding to that previously crystallized) the ability to selectively recognize the non-template strand. This peptide corresponds to a coiled-coil region of RNA polymerase, which was previously shown to include a sigma binding motif (Arthur and Burgess, 1998, Arthur et al. 2000, Anthony et al., 2001, submitted). Our evidence suggests that interaction of the coiled-coil of β' with Region 2.2 of σ^{70} produces an allosteric change in Region 2.4 that is required for selective binding to the non-template strand of the –10 region of the promoter.

Results

Amino Acids 262-309 of β' Reconstitute Non-Template Strand Binding with σ^{70}

We determined the minimal portion of RNA polymerase capable of promoting selective binding of sigma to the non-template strand of the –10 region of the promoter using the previously described crosslinking assay to assess binding (Kulbachinskiy et al., 1999; Marr and Roberts, 1997). Briefly, following the incubation of σ^{70} with either intact RNA polymerase or a fragment of polymerase on ice, radiolabeled oligonucleotides corresponding to the non-template or the template strand were added. Reactions were further incubated at room temperature and then irradiated with UV light to form zerolength crosslinks between sigma and bound DNA. Binding is visualized by the presence of a radiolabeled protein band following SDS polyacrylamide gel electrophoresis. Previous work had established that when holoenzyme is incubated with the oligonucleotides, crosslinking is primarily to σ^{70} and is selective for the non-template strand oligonucleotide. Quantitation of triplicate experiments indicates that holoenzyme shows approximately 7.7±1 fold greater crosslinking to the non-template strand than to the template strand control under our assay conditions (data not shown).

We tested a series of seven fragments (Severinov et al.,1992; Severinov et al., 1996; Gruber et al., 2001) covering the β' and β subunits of RNA polymerase for their ability to induce σ^{70} to bind the non-template oligonucleotide. A fragment of β' from amino acid 1- 550 was the only one of these seven capable of inducing this binding activity in σ^{70} (Figure 2-1A, lane 3 and data not shown). Little crosslinking is present in the σ^{70} alone reaction (Figure 2-1A, lane 2), and no significant crosslinking occurs in the β'1-550 alone reaction (Figure 2-1A, lane 4). The radiolabeled band must be σ^{70}, rather than the β' fragment, as varying the mobility of the β' portion does not change the mobility of the radiolabeled band (Figure 2-1A: lanes 5-12), while varying the mobility of σ^{70} (by using fragments of sigma) does change its mobility (Figure 2-5B). Binding is

specific to the non-template strand as little or no signal is observed when the oligonucleotide has the sequence of the template strand (Figure 2-1B, lanes 3 and 4). In this and all subsequent cases of positive binding, we have eliminated the possibility that RNA polymerase contamination of the fragments is responsible for binding (see Experimental Procedures).

We narrowed the region of β' required to induce non-template binding by performing deletion analysis of the 1-550 fragment. The recently published crystal structure of *Thermus aquaticus* RNA polymerase shows that aa 1-550 of β' has two solvent exposed portions separated by a hydrophobic buried stretch (Zhang et al., 1999). Amino acids 1-314, corresponding to one of these solvent exposed portions, stimulated non-template strand binding by sigma selectively (compare Figures 2-1A and 2-1B, lanes 5, 6), whereas a fragment containing β'315-550 did not (data not shown). Much of the N-terminus of β'1-550 is irrelevant for induction of non-template binding as β'1-260 cannot induce σ^{70} to bind to the non-template strand (data not shown), while a β' fragment from 237-550 can (compare Figures 2-1A and B, lanes 7, 8). Interestingly, our β'260-550 construct induced selective non-template strand binding at 0°C (compare Figure 2-1A and B, lanes 9, 10) but not at room temperature (data not shown). This increase in binding at lower temperatures does not appear to be to due to generally increased binding to the non-template at lower temperatures as neither σ^{70} alone nor holoenzyme binding to the non-template strand is significantly increased at 0°C (data not shown). Rather, the charge introduced by the N-terminal hexahistadine tag positioned adjacent to the coiled-coil α helical stretch may destabilize the helices so that they form only at low temperature. This destabilization could be due to the hexahistidine charge intensifying the helical macropole (the dipole moment formed by the hydrogen bonding of backbone elements) as this is one of the energetic barriers to formation of α helices. It has been previously shown that charge groups at the ends of peptides can affect helix formation (Tan et al., 1993). The fact that lower temperature, which favors helix

formation, allowed this construct to induce non-template binding suggested that the coiled-coil was important for function. We therefore tested whether region 262-309 of β', which contains only the 48 amino acids of β' constituting the coiled-coil, was by itself sufficient to reconstitute selective non-template strand binding with σ^{70}. This fragment was tagged at its C-terminus to avoid disrupting the coiled-coil. In this orientation, the His-tag should, if anything, lessen the macropole, thereby promoting helicity. Strikingly, this peptide, which is about ~1/70 of RNA polymerase, is able to induce binding (compare Figures 2-1A and B lanes 11, 12). We conclude that amino acids 262-309 of β' are necessary and sufficient to reconstitute selective binding by σ^{70} to the non-template strand of the –10 region of the promoter.

Previously, three substitutions in this region of β' have been shown to disrupt binding to sigma *in vivo* and *in vitro*, without disrupting assembly of RNA polymerase itself (R275Q, E295K, and A302D; Arthur et al., 2000). These same three changes disrupt the ability of β'1-550 to promote specific non-template strand binding by σ^{70} (Figure 2-2). The single change that has been introduced into β'262-309 (E295K) also disrupts specific non-template strand binding by σ^{70} (Figure 2-2). Taken together, these experiments provide evidence both that our minimal system recapitulates an interaction seen in the holoenzyme context and that the coiled-coil is the important functional element in this interaction.

Induction of Non-Template Binding of σ^{70} By a Fragment of β' is Efficient

We used a quantitative assay based on luminescence resonance energy transfer (analogous to fluorescence resonance energy transfer) to measure the efficiency with which fragments of β' convert σ^{70} to a form that selectively recognizes the non-template strand of the –10 region. In this assay, σ^{70} and the non-template oligonucleotide are each labeled with a fluorophore. When σ^{70} is bound to the oligonucleotide, excitation with a wavelength specific to the σ^{70} fluorophore results in transfer of excitation energy to the

other fluorophore, causing the energy to be emitted at a different wavelength. Thus, emission at this latter wavelength is a quantitative measure of the fraction of σ^{70} bound selectively to the non-template strand. Maximal fluorescence transfer is reached when the ratio of the β'240-309 fragment to σ^{70} is about 1:1 (Figure 2-3, closed circles). Moreover, comparison of this value to the fluorescence transfer conferred by core RNA polymerase (Figure 2-3, open square), indicates that the fragment by itself is 65% as efficient as intact RNA polymerase in inducing σ^{70} to bind to the non-template strand. Other experiments validated this assay: substitution in 240-309 (R275Q) eliminated induction; and binding was specific because it could be competed by increasing amounts of unlabeled non-template oligonucleotide but not by a non-specific oligonucleotide (data not shown). Taken together, these experiments strongly indicate that the coiled-coil alone is approximately as efficient as core RNA polymerase in converting σ^{70} to a form that exhibits selective binding to the non-template strand of the promoter.

DNA Binding by aa 262-309 of β' Does Not Appear to be Required for Selective Non-Template Recognition.

We considered the possibility that β'262-309 might exert its effect by directly contributing contacts to the DNA, rather than by functioning as an allosteric effector of the conformation of σ^{70}. Recently, a comprehensive study of contacts between RNA polymerase holoenzyme and promoter DNA in the open complexes was performed using photoactivatable chemical crosslinkers positioned between every other nucleotide (Naryshkin et al., 2000). As the non-template strand binding activity constitutes one portion of open complex interactions (Roberts and Roberts, 1996), this approach also identified potential RNA polymerase/non-template strand binding contacts. It indicated that β'1-550, (and thus potentially aa 262-309) appears to be in spatial proximity to the non-template strand only in the region between nucleotides –15 to –20. We therefore determined whether β'262-309 could still promote selective non-template binding of an

oligonucleotide spanning nucleotides –13 to +1, which lacks the potential β'1-550 DNA binding site. Contrary to the prediction of a model requiring such binding, β'262-309 still induced sigma to bind this shortened non-template strand oligonucleotide (Figure 2-4). Indeed, the fold induction for the full length and truncated oligonucleotide were about the same within the error of measurement (7.96 ±1.23, and 5.45 ±.96). We note that the magnitude of binding is somewhat smaller for the shortened oligonucleotide. Decreased efficiency of crosslinking could result from loss of non-specific contacts by either σ^{70} or β' to this region, both of which are positioned to potentially make such contacts in the open complex (Naryshkin et al., 2000). Regardless, this experiment is consistent with the idea that β'262-309 can promote specific recognition of the non-template strand by σ^{70} without contacting the non-template oligonucleotide itself.

σ^{70} Requirements for Reconstituting Non-template Strand Binding

Previous work had indicated that both Regions 1.1-2.4 and the fragment of σ^{70} previously crystallized (Region 1.2-2.4), in combination with core RNA polymerase were sufficient for selective non-template recognition (Severinova et al., 1996). These same two fragments of σ^{70} allow selective non-template recognition when activity is reconstituted with β'262-309. Full length σ^{70} (Figure 2-5B, lanes 1, 2), σ^{70} Region 1.1-2.4 (Figure 2-5B, lanes 3,4) and σ^{70} Region 1.2-2.4 (Figure 2-5B, lanes 5,6) all allow selective binding to the non-template strand oligonucleotide. Thus, non-template strand binding can be reconstituted with only 47 kDa, or approximately 1/10 of the total mass of holoenzyme (41.4 kDa sigma + 5.6 kDa β').

Mutations in several regions of σ^{70} disrupt binding to RNA polymerase (marked with a * in Figure 2-5A; Sharp et al., 1999; Gruber et al., 2001). GST pull-down assays show that β'1-550 interacts with Region 2.2 of σ^{70}(Gruber et al., 2001) as Region 2.2 mutants Q406A and E407K disrupt the β'1-550 interaction with σ^{70}. The data presented here indicate that this interaction is also necessary for the induction of non-template strand

binding as Q406A and E407K Region 2.2 mutants disrupt the ability of either β'1-550 or β'262-309 to induce binding (Figure 2-6A). Previous work showed that an additional Region 2.2 mutant, E407A, decreased interaction between β'1-550 and σ^{70} somewhat, although its effect on binding was not as severe as the other two changes (Gruber, T.M., unpublished observation). Consistent with this, the E407A mutation was defective in inducing non-template strand binding at low concentrations of σ^{70} and β' fragment (Figure 2-6B), but not at the higher concentrations we typically use in this assay (β' fragment: 900 nM; σ^{70}:450 nM; data not shown).

We believe that the defects of Region 2.2 substitutions result from weakening the binding site for the coiled-coil rather than from a structural perturbation in σ^{70}. The following observations argue against a general structural perturbation: 1) the crystal structure indicates that these residues are α-helical and surface-exposed; alanine substitutions are known to promote α-helix formation (Marqusee et al., 1989; Tan et al., 1993), and thus are unlikely to disturb even regional secondary structure; 2) non-competitive co-immunoprecipitation binding experiments indicate that approximately the same percentage of mutant and wt σ^{70} preparations bind to core RNA polymerase (see Experimental Procedures); and 3) in denaturant melt experiments, measured by circular dichroism, the denaturation curve of mutants and wild type are essentially superimposable, indicating that the mutants are as stably folded as wild type (Figure 2-6C). Taken together, we feel confident that the primary effect of the substitutions is to eliminate the direct binding site of the β' coiled-coil region.

Discussion

We show here that the coiled-coil motif in the β' subunit of RNA polymerase comprised of amino acids 262-309 allows σ^{70} to selectively recognize the non-template strand of the promoter. The coiled-coil alone is ~65% as efficient on a molar basis as

RNA polymerase itself in inducing non-template strand binding by σ^{70}. DNA binding by the β' coiled-coil does not appear to be required for this effect. Instead, this effect is mediated by binding of the β' coiled-coil to Region 2.2 of σ^{70}, which induces an allosteric change that is necessary for selective recognition of the non-template strand by σ^{70}. To our knowledge, this is the first demonstration that a single contact between RNA polymerase and an initiation factor is sufficient to reconstitute a function essential for initiation.

The knowledge that holoenzyme specifically recognizes the non-template strand of the –10 region of the promoter was a fortuitous outgrowth of long-term studies by the Roberts group to understand the mechanism of Q mediated anti-termination. A promoter-proximal pause required for this antitermination was shown to result from binding σ^{70} to a sequence in the transcript that strongly resembled the –10 region of the promoter (Ring and Roberts, 1994; Ring et al., 1996). Subsequent studies established that σ^{70} bound this same sequence (the non-template strand of the –10 region) in the open complex, indicating that this activity is part of the normal process of open complex formation (Roberts and Roberts, 1996; Marr and Roberts, 1997). Several lines of evidence indicated that σ^{70} itself was responsible for recognizing the non-template strand. Selective crosslinking occurs uniquely to σ^{70}, indicating that this protein is the one in closest proximity to the non-template strand. Region 2.4 of σ^{70}, which is implicated in recognition of the duplex –10 region (Kenney et al., 1989; Siegele et al., 1989; Daniels et al., 1990; Waldburger et al. 1990; Dombroski et al., 1992; Dombroski, 1997) is also involved in recognition of the non-template strand as Region 2.4 mutations that decrease the specificity of σ^{70} interaction with DNA also alter non-template strand recognition (Marr and Roberts, 1997). Finally, mutations in Region 2.3, known to alter melting also specifically alter cross-linking of σ^{70} to the non-template strand (Huang et al., 1997) .

Although σ^{70} selectively recognizes the non-template strand, binding to RNA polymerase is required for this activity to be manifest. By itself, σ^{70} recognizes the non-template strand quite weakly, and only slightly better than a template strand control (Callaci and Heyduk, 1998). Upon RNA polymerase binding, recognition of the template strand decreases and that of the non-template strand increases, which together result in selective recognition by holoenzyme (Callaci and Heyduk, 1998). RNA polymerase presumably induces alternate σs, which differ in their conserved –10 promoter sequences, to perform this activity as part of their initiation process. Thus one might expect RNA polymerase to induce this activity by altering σ conformation, rather than by collaborating with σ in this activity by directly providing specific DNA contacts. The finding that some residues in region 2.3 become less solvent exposed upon binding to core was consistent with the idea that RNA polymerase induced a conformational change in σ^{70} (Callaci and Heyduk, 1998). The results presented here considerably advance this idea. We show that a very small region of β' is both necessary and sufficient to efficiently promote this transition. As the coiled coil region of β' still induces non-template binding by σ^{70} even when the oligonucleotide lacks the nucleotides with which it could potentially interact (Naryshkin et al., 2000), it is reasonable to conclude that β'262-309 functions as an allosteric effector of σ^{70} function.

The studies presented here solidify the idea that Region 2.2 of σ^{70} is the binding site for this allosteric interaction. Although 2.2 is the region most highly conserved among the σ^{70} family of proteins, it had been without a functional assignment until Joo et al. genetically identified a residue in Region 2.2 of σ^{32} that was involved in core binding (Joo et al., 1997) . Then, two different genetic selections, one for mutants defective in binding to RNA polymerase and a second for mutants defective in mediating the pause required for Q antitermination, indicated that a cluster of surface-exposed residues in Region 2.2 was involved in binding RNA polymerase (Ko et al., 1998; Sharp et al., 1999). Importantly, these studies also indicated that the defect in mediating the pause

required for Q antitermination was specifically associated with a defect in this particular binding interaction between σ^{70} and RNA polymerase, rather than being a general consequence of weakened interaction with RNA polymerase (Ko et al., 1998; Sharp et al., 1999). That the genetically identified residues directly identify the σ^{70} binding site rather than generally destabilizing protein structure is indicated by our demonstration that these substitutions do not perturb σ^{70} stability.

The mechanism by which binding of β':262-309 to Region 2.2 allosterically regulates the conformation of Region 2.3-2.4 so that it selectively recognizes the non-template strand DNA remains to be determined. The early idea that a disordered acidic loop (aa ~192-212 of σ^{70}) located near region 2.3-2.4 inhibits non-template strand binding by free σ^{70} (Malhotra et al., 1996) is likely to be incorrect as region 2.3-2.4 appears solvent exposed in free sigma (Callaci and Heyduk, 1998). Further, even when this loop is absent, as in *Bacillus subtilis* σ^{43}, binding by core polymerase is still required to unmask this binding determinant (Huang et al. 1997). As first noted by Ko et al.(1998), the crystal structure indicates that the α-helix comprising Region 2.2 is immediately behind the continuous α-helix formed by Regions 2.3-2.4, indicating that changes in one helix could alter the other (Figure 2-7). Several possibilities, not all mutually exclusive, can be suggested for how interaction of 2.2 with the coiled-coil could alter the 2.3-2.4 recognition helix. Such an interaction could either stabilize the conformation of 2.2 and thereby "freeze" residues in the 2.3-2.4 helix into a conformation complementary to the non-template strand or, conversely, disrupt the binding of the helices to each other, thereby exposing additional hydrophobic residues needed for non-template strand binding. This interaction could shift the orientation of 2.2 relative to 2.3-2.4; subsequent repacking of the 2.2/2.3-2.4 interface would alter surface exposed residues. Finally, interaction could rotate 2.2, causing 2.3-2.4 to rotate in turn, in a movement akin to that of a meshed gear and thereby alter the surface exposed residues.

The interaction of the coiled-coil of β' with Region 2.2 of σ^{70} clearly has profound influences on the activity of σ^{70}. Whether this interaction also affects the activity of RNA polymerase is currently unknown. We have recently found that the helices of the β'262-309 can be covalently crosslinked and still retain the ability to bind to σ^{70} and induce its non-template strand binding activity, indicating that the coiled-coil can function as an allosteric effector without conformational change (Anthony et al, 2001, submitted). However, the coiled-coil supports the rudder, a finger-like projection that has been implicated in separating the RNA/DNA hybrid (Zhang et al., 1999; Korzheva et al., 2000). It is not inconceivable that coiled-coil interaction with σ^{70} could alter the placement of the rudder, moving it away from its normal position, thus removing steric conflicts between the rudder and promoter DNA during the initial binding process. We note that β' 262-309 shares greater than average sequence homology with its eukaryotic counterparts (25% identity for this region of *E. coli* β' and Rpb1 in both *S. cerevisiae* and humans vs. 11-12% overall identity; data not shown and Ebright, R. personal communication). Furthermore, comparison of the crystallized structures of *T. aquaticus* and yeast pol II RNA in the 262-309 region indicates that they share extensive structural conservation (see Ebright, 2000; the *T. aquaticus* β' aa 537-584 homologous to *E. coli* 262-309 comprises much of lobe 1 shown in Figure 4 of that paper). This suggests the possibility that the allosteric function of this region may be conserved in eukaryotes.

Experimental Procedures

Plasmid Construction and Mutagenesis

Fragments of β' were PCR amplified from pET15b-β'1-550 (Gruber et al., 2001) digested with restriction endonucleases, and ligated in pET15b. Mutations in the β'1-550

and β'262-309 constructs were created using the Quikchange Site-Directed Mutagenesis kit (Stratagene). All constructs were verified by sequencing.

Overproduction, purification, and quantitation of proteins

Proteins used in all experiments except Figure 2-3 were purified as follows (Figure 2-3 purifications are described in the luminescence resonance energy transfer binding assay section). β' fragments: expression of His-tagged fragments was induced with 1 mM IPTG in BL21(DE3)*pLys E. coli* at OD_{600}=.3 for 1 hour at 37°C. Cells were pelleted, frozen at –80°C, resuspended by homogenization into lysis buffer (20mM Tris pH 8.0, 500 mM NaCl, 1 mM βME, 10% glycerol, 10 mM $MgCl_2$, 6 M guanidine HCl), and heated for 1 hr at 37° C to remove residual core contamination from some fragments. Lysates were sonicated, centrifuged, batch bound to Ni-NTA resin, and washed three times in batch with 15 x the volume of resin (Wash buffer: 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM βME, 10% glycerol, 10 mM $MgCl_2$, 20 mM imidazole, 6M urea). All fragments except β'262-309 were then batch eluted (Elution buffer: 50 mM Tris pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 10 mM DTT, 10% glycerol, 200 mM imidazole, 6 M urea), renaturation buffer (50 mM Tris 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 10 mM βME, 600 mM KCl, 10% glycerol, 0.1% Triton X-100) was slowly added to dilute out the urea, and protein was dialyzed into storage buffer (0.1% Triton X-100, 10 mM Tris pH 8.0, 1 mM EDTA, 5 mM βME, 300 mM NaCl, 50% glycerol). For β'262-309 fragment, the amount of His-tagged protein overproduced was insufficient to block binding of a variety of non-specific proteins to the Ni-NTA column. Here, washed Ni-NTA beads containing the β'262-309 were loaded onto a column and eluted with a gradient of imidazole from 10 mM to 200 mM (gradient buffer: 50 mM Tris pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 10 mM βME, 10 % glycerol, 6 M Urea). Early fractions free of non-specifically bound

contaminants were pooled, renatured by step dialysis in renaturation buffer with decreasing amounts of urea, and then dialyzed into storage buffer.

Portions of all β' fragment preparations (~200 μl) were TCA precipitated, run on SDS-polyacrylamide gels and stained with Coomassie Blue next to a standard curve of RNA polymerase. The β' fragments preparations showed no detectable RNA polymerase contamination (data not shown). In these experiments, quantities of RNA polymerase contamination that would introduce only 0.5 nM RNA polymerase into our crosslinking assay can be easily detected—thus RNA polymerase contamination of our β' fragments (if present) introduces less than this amount into crosslinking assays. 0.5 nM of renatured RNA polymerase is insufficient to induce σ binding to the non-template strand in the crosslinking assay (data not shown).

His-tagged wild type and mutant σ's were overproduced and purified as previously described (Sharp et al., 1999). σ preparations were confirmed to be active using non-competitive immunoprecipitation assays. In these assays, the activity of E407A and Q406A were essentially indistinguishable from wild type; E407K was within two-fold of wild type activity. GST σ fragments were purified using the protocol described for the His-tagged protein, except that glutathione conjugated agarose beads were used instead of Ni-NTA, and glutathione was used instead of imidazole to elute protein.

Core polymerase was overexpressed and purified according to the method of Murakani, K and Darst, S., to be published elsewhere (Darst, S. personal communication.)

For crosslinking experiments, proteins were quantitated by visual inspection of protein samples on Coomassie stained SDS polyacrylamide gels next to a standard curve of BSA. In cases where more careful quantitation was necessary , i.e. σ mutants and β' fragment

mutants relative to their wild type counterparts, triplicate dilutions of each mutant were run on Coomassie stained SDS polyacrylamide next to a standard curve of wild type protein and quantitated using Alpha Innotech AlphaImager 4.0 spot densitometry. Quantitation of all samples relative to the standard curve indicated that all measurements were made within the linear range of Coomassie staining. Standard deviation of the triplicate dilutions of each mutant indicated that the error in protein quantitation was less than 20%. σ preparations for CD melts were quantified by UV absorbance at 280 nm.

Crosslinking assay for non-template strand binding

Oligonucleotides for the crosslinking assay were synthesized and HPLC or PAGE purified by Integrated DNA Technologies. Non-template and template oligonucleotide sequences were those used previously (Kulbachinskiy et al., 1999); their sequences are similar to that of *lacUV5* promoter (Non-template:ATTGCGCGTATAATGTGTGGA; Template:TCCACACATTATACGCAAT). Labeling was performed using NEB T4 polynucleotide kinase in 100 μL reaction (1X NEB buffer, .5 μM oligonucleotide, .25 μM ^{32}P-γ–ATP (6000 Ci/mmol)) for 45 minutes at 37°C. Labeling was stopped by addition of EDTA to 10 mM. Free ^{32}P-γ–ATP was removed by passing aliquots of the reactions over Boehringer Mannheim G-25 Quick Spin columns. To determine the concentration of cleaned, labeled oligonucleotide, a sample of it was run next to a known quantity of the uncleaned, labeled oligonucleotide, on a 19% 19:1 denaturing polyacrylamide gel and the radioactivity of the two oligo bands was compare using Molecular Dynamics phosphor screens and ImageQuant 1.2 software.

The crosslinking assay was performed essentially as earlier described (Marr and Roberts, 1997) except that:1) when β' fragments were dilute enough that they affected the overall buffer concentration, control reactions were supplemented with β' fragment storage

buffer to an equivalent level and 2) Figure 2-6 crosslinking experiments were performed with .05% Triton X-100 present in reaction buffer. Briefly, core polymerase or fragments of core polymerase (generally β' fragments) and σ were mixed on ice for 30 minutes in reaction buffer (20 mM Tris pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 5% glycerol, 100 ug/ml BSA, 1 mM DTT). Oligonucleotide was then added, bringing the final reaction volume to 20 μL. Concentration of oligonucleotide in reactions was 6.25 nM; core fragment and sigma concentrations were as described in the figures (typically 900 nM core fragment and 450 nM σ). Reactions were incubated for 30 minutes at room temperature and then irradiated with UV (256 nm) for 10 minutes in a Stratagene UV Stratalinker 1800. 5 μL aliquots of reactions were removed, 2X SDS loading dye was added and the samples run on 8% SDS polyacrylamide gel. The σ regions of the gels were quantified using Molecular Dynamics phosphor screens and ImageQuant 1.2 densitometry software.

Luminescence Resonance Energy Transfer Binding Assay (LRET)

For LRET experiments β'240-309 was purified according to the β'260-309 purification described earlier (Arthur and Burgess, 1998) and core and σ^{70} according to Matlock and Heyduk (2000). Interaction of σ^{70} with the non-template strand oligonucleotide was measured using sensitized acceptor signal as described (Matlock and Heyduk, 2000). Briefly, a single reactive cysteine mutant with the reactive cysteine at position 59 was labeled with the donor europium chelate probe ((Eu^{3+})-DTPA-AMCA). The 12 nt nontemplate strand oligonucleotide (TCGTATAATGTG, corresponding to positions -15 to -5 of the lacUV5 promoter) was labeled at the 5' end with the acceptor probe (Cy5) and the labeled oligonucleotide was purified as described (Heyduk and Heyduk, 1999). A mixture of 50 nM donor-labeled σ^{70} and 75 nM acceptor-labeled oligonucleotide was titrated with increasing concentrations of β' fragment. At each concentration a total

sensitized acceptor intensity was measured by quantifying sensitized Cy5 emission at 670 nm following excitation at 337 nm with nitrogen laser. The experiments were performed at 25°C in 20 mM Tris (pH 8.0), 100 mM NaCl, 10 mM $MgCl_2$, 0.1 mM DTT, and 5% glycerol.

Denaturation Melt Curves

σs were diluted to 1.5 μM (WT, E407A, Q406A) or 1.8 μM (E407K) in CD buffer (10 mM KAc pH 8.0, 100 mM KCl) supplemented with various concentration of guanidine HCl and equilibrated overnight on ice. Circular Dichroism (CD) measurements were made using a J-715 spectropolarimeter (Jasco) in a temperature controlled cuvette holder (at 15°C) and a 0.2 cm path length cuvette. The CD signal/mol protein for all mutants was within 10% of WT σ (within the error of protein quantitation). To control for minor error in protein quantitation, signals were normalized by multipying by the ratio of WT signal/mutant signal at 0M guanidine HCl in CD buffer. Guanidine concentrations were calculated using refractometer measurements.

The $\Delta G_{unfolding(H20)}$ of mutant and WT σs was calculated assuming a two state model and using the method described in Pace et al. (1989)[WT 2.84 ± .55 kcal/mol; E407K 2.45 ± .26 kcal/mol; Q406A 2.73 ± .60 kcal/mol; E407A 2.00 ± .305 kcal/mol]. Using these values, we calculate the percent σ unfolded in native buffer to be: WT .68%, E407K 1.3%, Q406A .83%, and E407A 2.9%.

Acknowledgements

We are grateful to Seth Darst and Katsuhiko Murakani for providing us with *T. aquaticus* core RNA polymerase structure coordinates and for their new method of purifying *E. coli*

core RNA polymerase, Vadim Nikiforov for sharing with us sequence alignments of β' and Rpb1 from various organisms, Richard Ebright for sharing information about sequence identity between prokaryotic β' and eukaroytic Rpb1, and to Christophe Herman and Cathy Chan for continual advice and support. We thank David Agard and Yoko Shibata for advice on performing and interpreting CD experiments, Alan Frankel for structural discussions, and Bob Landick and the rest of the Gross lab for other helpful discussions and support. This work was supported by NIH (GM30477) to CAG, NIH (GM50514) to TH, and NIH (GM28575) to RRB. BAY was supported by NSF predoctoral fellowship DGE-9616282. TMG was supported in part by the American Cancer Society-California Division, Fellowship #2-19-99.

References

Arthur, T. M., Anthony, L. C., and Burgess, R. R. (2000). Mutational analysis of beta '260-309, a sigma 70 binding site located on *Escherichia coli* core RNA polymerase. J Biol Chem *275*, 23113-23119.

Arthur, T. M., and Burgess, R. R. (1998). Localization of a sigma70 binding site on the N terminus of the *Escherichia coli* RNA polymerase beta' subunit. J Biol Chem *273*, 31381-31387.

Callaci, S., and Heyduk, T. (1998). Conformation and DNA binding properties of a single-stranded DNA binding region of sigma 70 subunit from *Escherichia coli* RNA polymerase are modulated by an interaction with the core enzyme. Biochemistry *37*, 3312-3320.

Daniels, D., Zuber, P., and Losick, R. (1990). Two amino acids in an RNA polymerase sigma factor involved in the recognition of adjacent base pairs in the -10 region of a cognate promoter. Proc Natl Acad Sci U S A *87*, 8075-8079.

Dombroski, A. J. (1997). Recognition of the -10 promoter sequence by a partial polypeptide of sigma70 in vitro. J Biol Chem *272*, 3487-3494.

Dombroski, A. J., Walter, W. A., and Gross, C. A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. Genes Dev *7*, 2446-2455.

Dombroski, A. J., Walter, W. A., Record, M. T., Siegele, D. A., and Gross, C. A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell *70*, 501-512.

Ebright, R. H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol *304*, 687-698.

Fedoriw, A. M., Liu, H., Anderson, V. E., and deHaseth, P. L. (1998). Equilibrium and kinetic parameters of the sequence-specific interaction of *Escherichia coli* RNA polymerase with nontemplate strand oligodeoxyribonucleotides. Biochemistry *37*, 11971-11979.

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol *63*, 141-155.

Gross, C. A., Lonetto, M., and Losick, R. (1992). Bacterial sigma factors. In Transcriptional regulation, S. L. McKnight and K. R. Yamamoto, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 129-176.

Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., Artsimovitch, I., Geszvain, K. M., Arthur, T. M., Burgess, R. R., Landick, R., Severinov, K., and Gross, C. A. (2001). Binding of the initiation factor σ70 to core RNA polymerase is a multi-step process. Mol. Cell *in press*.

Guo, Y., and Gralla, J. D. (1998). Promoter opening via a DNA fork junction binding activity. Proc Natl Acad Sci U S A *95*, 11655-11660.

Heyduk, E., and Heyduk, T. (1999). Architecture of a complex between the sigma70 subunit of Escherichia coli RNA polymerase and the nontemplate strand oligonucleotide. Luminescence resonance energy transfer study. J Biol Chem *274*, 3315-3322.

Huang, X., Lopez de Saro, F. J., and Helmann, J. D. (1997). Sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA. Nucleic Acids Res *25*, 2603-2609.

Joo, D. M., Ng, N., and Calendar, R. (1997). A sigma32 mutant with a single amino acid change in the highly conserved region 2.2 exhibits reduced core RNA polymerase affinity. Proc Natl Acad Sci U S A *94*, 4907-4912.

Kenney, T. J., York, K., Youngman, P., and Moran, C. P. (1989). Genetic evidence that RNA polymerase associated with sigma A factor uses a sporulation-specific promoter in Bacillus subtilis. Proc Natl Acad Sci U S A *86*, 9109-9113.

Ko, D. C., Marr, M. T., Guo, J., and Roberts, J. W. (1998). A surface of *Escherichia coli* sigma 70 required for promoter function and antitermination by phage lambda Q protein. Genes Dev *12*, 3276-3285.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). A structural model of transcription elongation. Science *289*, 619-625.

Kulbachinskiy, A., Mustaev, A., Goldfarb, A., and Nikiforov, V. (1999). Interaction with free beta' subunit unmasks DNA-binding domain of RNA polymerase sigma subunit. FEBS Lett *454*, 71-74.

Lonetto, M., Gribskov, M., and Gross, C. A. (1992). The σ^{70} family: sequence conservation and evolutionary relationships. J. Bacteriol. *174*, 3843-3849.

Malhotra, A., Severinova, E., and Darst, S. A. (1996). Crystal structure of a sigma 70 subunit fragment from E. coli RNA polymerase. Cell *87*, 127-136.

Marqusee, S., Robbins, V. H., and Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci U S A *86*, 5286-5290.

Marr, M. T., and Roberts, J. W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science *276*, 1258-1260.

Matlock, D. L., and Heyduk, T. (2000). Sequence determinants for the recognition of the fork junction DNA containing the -10 region of promoter DNA by E. coli RNA polymerase. Biochemistry *39*, 12274-12283.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci U S A *98*, 892-897.

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

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

Pace, C. N., Shirley, B.A., Thomson, J.A. (1989). Measuring the Conformational Stability of a Protein. In Protein Structure: A Practical Approach, T. E. Creighton, ed. (Oxford University Press Inc, New York: IRL press), pp. 311-330.

Record, M. T., Reznikoff, W., Craig M., McQuade, K., and Schlax, P. (1996). *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcriptional initiation. In *Escherichia coli* and *Salmonella* Cellular and Molecular Biology, F. C. Neidhardt, ed. (Washington, D.C.: ASM Press), pp. 792-820.

Ring, B. Z., and Roberts, J. W. (1994). Function of a nontranscribed DNA strand site in transcription elongation. Cell *78*, 317-324.

Ring, B. Z., Yarnell, W. S., and Roberts, J. W. (1996). Function of E. coli RNA polymerase sigma factor sigma 70 in promoter- proximal pausing. Cell *86*, 485-493.

Roberts, C. W., and Roberts, J. W. (1996). Base-specific recognition of the nontemplate strand of promoter DNA by E. coli RNA polymerase. Cell *86*, 495-501.

Sentenac, A., Riva, M., Thuriaux, P., Buhler, J.-M., Treich, I., Carles, C., Werner, M., Ruet, A., Huet, J., Mann, C., Chiannilkulchai, N., Stettler, S., and Mariotte, S. (1992). Yeast Polymerase Subunits and Genes. In Transcriptional Regulation. McKnight, S.R. and Yamamoto, K.R., ed. (New York: Cold Spring Harbor Laboratory Press), pp. 27-54.

Severinov, K., Mustaev, A., Kashlev, M., Borukhov, S., Nikiforov, V., and Goldfarb, A. (1992). Dissection of the beta subunit in the *Escherichia coli* RNA polymerase into domains by proteolytic cleavage. J Biol Chem *267*, 12813-12819.

Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S. A., and Goldfarb, A. (1996). Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of *Escherichia coli* RNA polymerase. J Biol Chem *271*, 27969-27974.

Severinova, E., Severinov, K., Fenyo, D., Marr, M., Brody, E. N., Roberts, J. W., Chait, B. T., and Darst, S. A. (1996). Domain organization of the *Escherichia coli* RNA polymerase sigma 70 subunit. J Mol Biol *263*, 637-647.

Sharp, M. M., Chan, C. L., Lu, C. Z., Marr, M. T., Nechaev, S., Merritt, E. W., Severinov, K., Roberts, J. W., and Gross, C. A. (1999). The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. Genes Dev *13*, 3015-3026.

Siegele, D. A., Hu, J. C., Walter, W. A., and Gross, C. A. (1989). Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. J Mol Biol *206*, 591-603.

Tan, R., Chen, L., Buettner, J. A., Hudson, D., and Frankel, A. D. (1993). RNA recognition by an isolated alpha helix. Cell *73*, 1031-1040.

Waldburger, C., Gardella, T., Wong, R., and Susskind, M. M. (1990). Changes in conserved region 2 of *Escherichia coli* sigma 70 affecting promoter recognition. J Mol Biol *215*, 267-276.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci U S A *98*, 892-897.

Sentenac, A., Riva, M., Thuriaux, P., Buhler, J.-M., Treich, I., Carles, C., Werner, M., Ruet, A., Huet, J., Mann, C., Chiannilkulchai, N., Stettler, S., and Mariotte, S. (1992). Yeast Polymerase Subunits and Genes. In Transcriptional Regulation. McKnight, S.R. and Yamamoto, K.R., ed. (New York: Cold Spring Harbor Laboratory Press), pp. 27-54.

Young, R. A. (1991). RNA polymerase II. Annu Rev Biochem *60*, 689-715.

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

Figure Legends

Figure 2-1. β'262-309 induces σ to bind to the non-template strand of promoter.

Fragments of β' were tested for the ability to induce σ to bind to single stranded oligonucleotides corresponding either to the non-template (A) or the template strand (B) of the -10 region of the promoter using the crosslinking assay described in the text. Presence (+) or absence (-) of core polymerase or σ is indicated above each lane. The particular β' fragment used is indicated by the amino acids comprising the fragment in the "β' fragment" box above each lane. Reactions in Lanes 9 and 10 were incubated on ice; all others were incubated at room temperature. Concentrations of proteins present in reactions (in this and all subsequent figures unless otherwise indicated) were: β' fragments:900 nM, σ:450 nM, and core:100 nM. (Note that holoenzyme controls in this figure [A and B, lane 1] are slightly underloaded in A and overloaded in B.)

Figure 2-2. Mutations in the coiled-coil region of β' known to disrupt σ binding abolish induction of σ non-template strand binding.

Wild type or mutant fragments of β' (as indicated underneath each bar) were tested for their ability to induce σ to bind the non-template strand by crosslinking assay. Bars represent the average signal intensity of the crosslinked σ band from three experiments after subtraction of the σ alone signal; error bars indicate standard deviation from the average. In this experiment and throughout the paper, the signal intensity for β'1-550+σ and β'262-309+σ were 9-fold and 8-fold, respectively, above the σ alone background (β' fragment alone background was essentially negligible).

Figure 2-3 The coiled-coil region of β' acts efficiently.

Efficiency of β'240-309 induction of non-template strand binding by σ was measured by β'240-309 titration using the luminescence resonance energy transfer assay as described in text. A single concentration of core polymerase (100 nM; square point) is provided for comparision. Cy5 labeled non-template strand and σ^{70} (59) Eu^{3+} concentrations were 75 nM and 50 nM, respectively.

Figure 2-4. β'262-309 does not appear to need to contact DNA to induce non-template binding.

β'262-309 was tested for the ability to induce σ to bind to a truncated non-template strand oligonucleotide lacking the region of the non-template strand with which β'262-309 could potentially interact (Naryshkin et al., 2000). For comparision, binding to full length (FL) non-template strand oligonucleotides is shown. Bars represent average signal intensity of crosslinked σ band from three experiments; error bars represent standard deviation from average.

Figure 2-5. Non-template strand binding can be reconstituted with only ~47 kDa of holoenzyme.

A) Schematic of σ^{70} indicating its conserved regions and subregions and their known functions. Sites of interaction with core identified by mutation are marked with a (*). The previously crystallized portion is indicated.

B) β'262-309 was tested for its ability to induce fragments of σ^{70} to bind the non-template strand using the crosslinking assay. Extent of σ^{70} fragments is indicated by the

subregions each fragment contains in the σ box above each lane; FL indicates full length control. Non-template strand binding is shown in top panel, template strand control is shown in bottom panel. Note that the mobility of the crosslinked protein band correlates with the mobility of the σ band, indicating that the crosslinked band is indeed σ. The relatively minor difference in mobility between full length and 1.1-2.4 σ^{70} is due to their different purification tags; full length σ^{70} is tagged with hexahistidine, while the 1.1-2.4 construct carries a much larger (26 kDa) GST tag (as does the 1.2-2.4 construct).

Figure 2-6. Region 2.2 of σ is a determinant for β' coiled-coil region allosteric binding.

A and B) σs carrying region 2.2 substitutions (indicated beneath bars) were tested for their ability to be induced by β' fragments to bind non-template strand oligonucleotides using the crosslinking assay. Each bar represents the average signal intensity of the crosslinked σ band from three experiments with relevant σ alone background subtracted. Error bars represent standard deviation from average. In B) protein concentrations were: β' fragment:112.5 nM, σ:75 nM.

C) Guanidine HCl denaturation melts were performed to assess the structural integrity of mutant σs.

Figure 2-7. Structure of the β' coiled-coil and the σ^{70} region involved in allosteric interaction.

The crystal structure of *E. coli* σ 1.2-2.4 is shown at top; a portion of the crystal structure of *T. aquaticus* core polymerase containing the β' coiled-coil region is shown at bottom. The coiled-coil region of ß' and region 2.2 of σ are colored yellow. Region 2.3-2.4 of σ, believed to responsible for non-template strand binding is colored green. Amino acids in

both σ (E407, Q406) and β'(R275, E295, A302) shown here to affect allosteric interaction when mutagenized are shown in red; other amino acids in σ^{70}(D403, N409) likely to be involved in interaction with coiled-coil because of their involvement in core binding and promoter proximal pausing are shown in pink (Ko et al., 1998; Sharp et al., 1999).

Figure 2-1

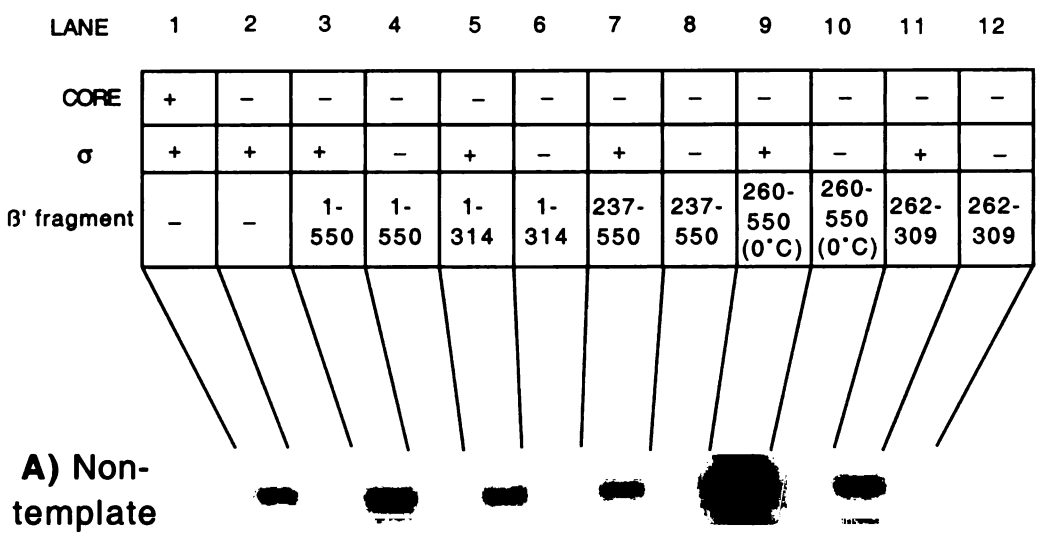

LANE 1 2 3 4 5 6 7 8 9 10 11 12
CORE + – – – – – – – – – – –
σ + + + – + – + – + – + –
ß' fragment – – 1-550 1-550 1-314 1-314 237-550 237-550 260-550 (0°C) 260-550 (0°C) 262-309 262-309
A) Non-template
B)Template

Figure 2-2

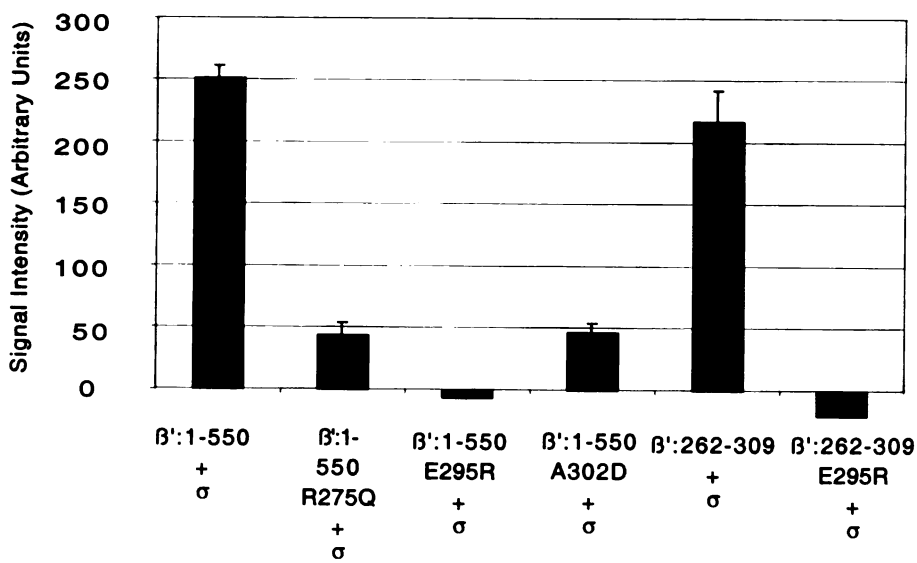

Signal Intensity (Arbitrary Units)
300
250
200
150
100
50
0
β':1-550 + σ
β':1-550 R275Q + σ
β':1-550 E295R + σ
β':1-550 A302D + σ
β':262-309 + σ
β':262-309 E295R + σ

Figure 2-3

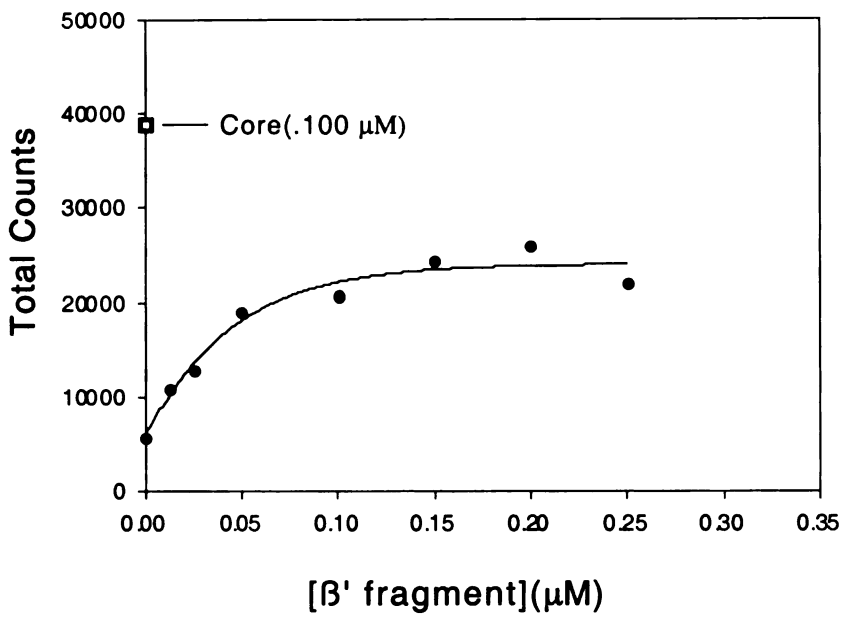

50000
40000
30000
20000
10000
0
Total Counts
Core(.100 µM)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
[ß' fragment](µM)

Figure 2-4

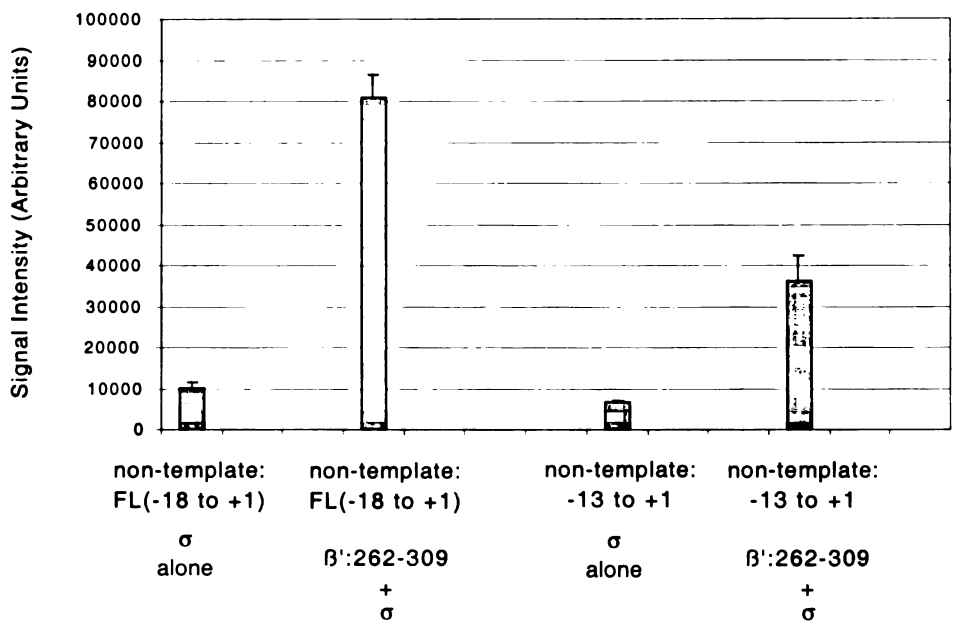

Signal Intensity (Arbitrary Units)
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
non-template: FL(-18 to +1)
σ alone
non-template: FL(-18 to +1)
ß':262-309 + σ
non-template: -13 to +1
σ alone
non-template: -13 to +1
ß':262-309 + σ

Figure 2-5

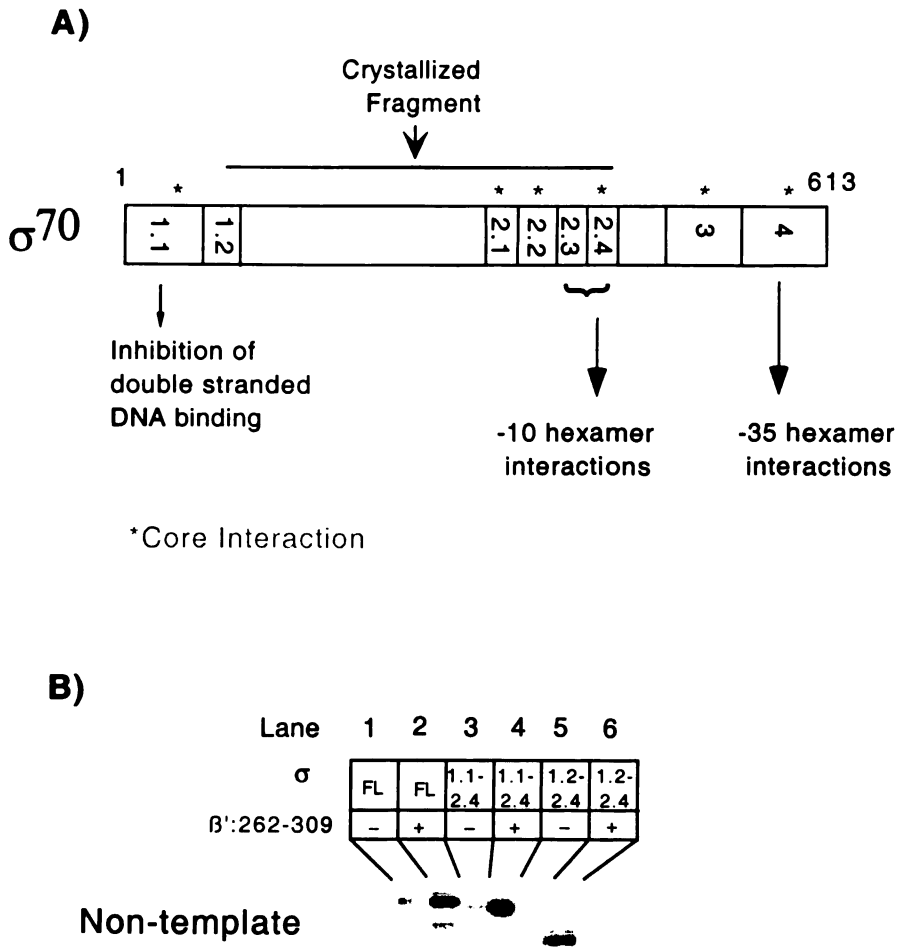
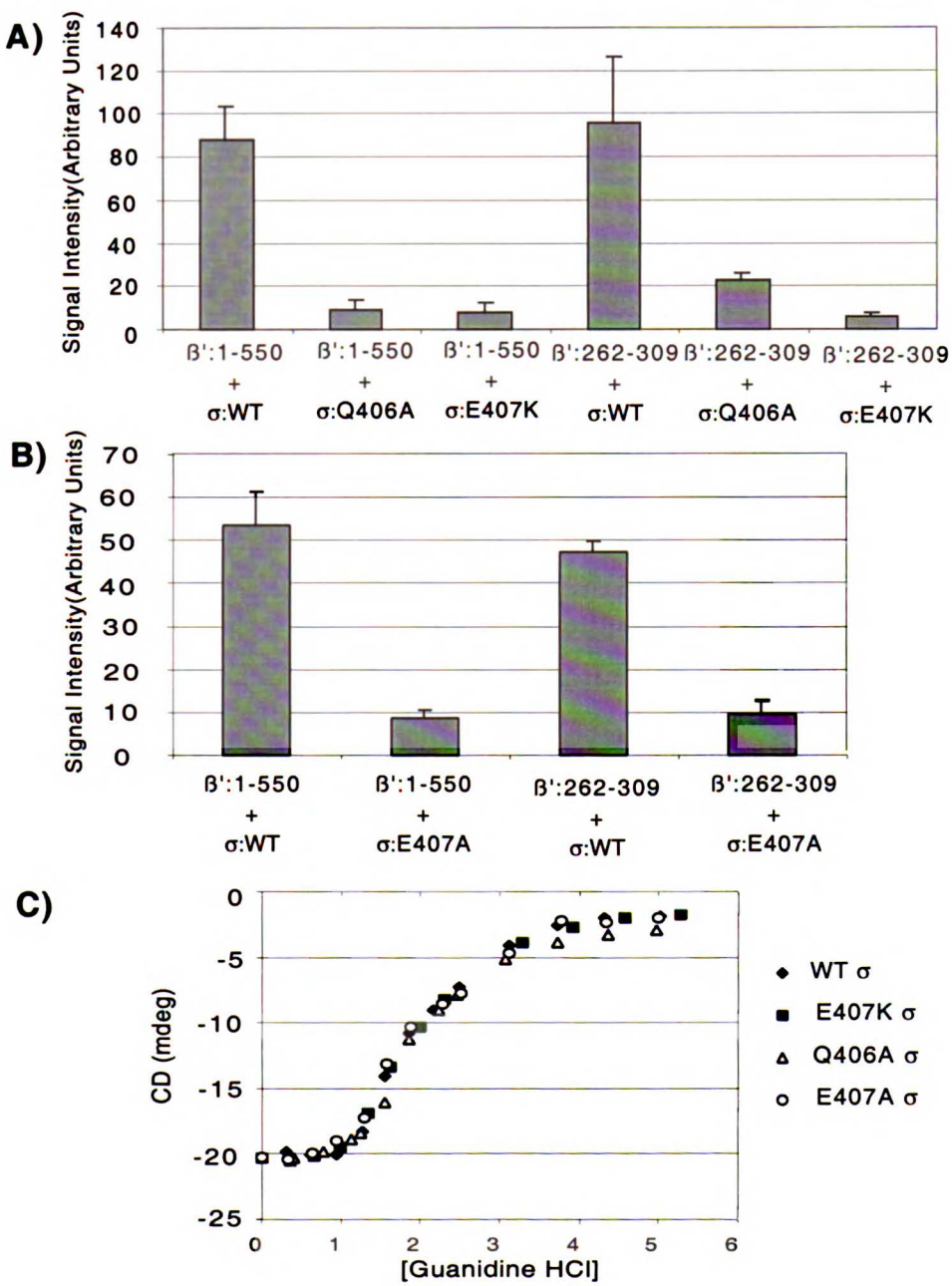

A)
Crystallized Fragment
1
613
σ70
1.1
1.2
2.1
2.2
2.3
2.4
3
4
Inhibition of double stranded DNA binding
-10 hexamer interactions
-35 hexamer interactions
*Core Interaction
B)
Lane 1 2 3 4 5 6
σ
FL FL 1.1-2.4 1.1-2.4 1.2-2.4 1.2-2.4
ß':262-309
– + – + – +
Non-template
Template

Figure 2-6


A)
Signal Intensity(Arbitrary Units)
140
120
100
80
60
40
20
0
β':1-550 + σ:WT
β':1-550 + σ:Q406A
β':1-550 + σ:E407K
β':262-309 + σ:WT
β':262-309 + σ:Q406A
β':262-309 + σ:E407K
B)
Signal Intensity(Arbitrary Units)
70
60
50
40
30
20
10
0
β':1-550 + σ:WT
β':1-550 + σ:E407A
β':262-309 + σ:WT
β':262-309 + σ:E407A
C)
CD (mdeg)
0
-5
-10
-15
-20
-25
0
1
2
3
4
5
6
[Guanidine HCl]
WT σ
E407K σ
Q406A σ
E407A σ

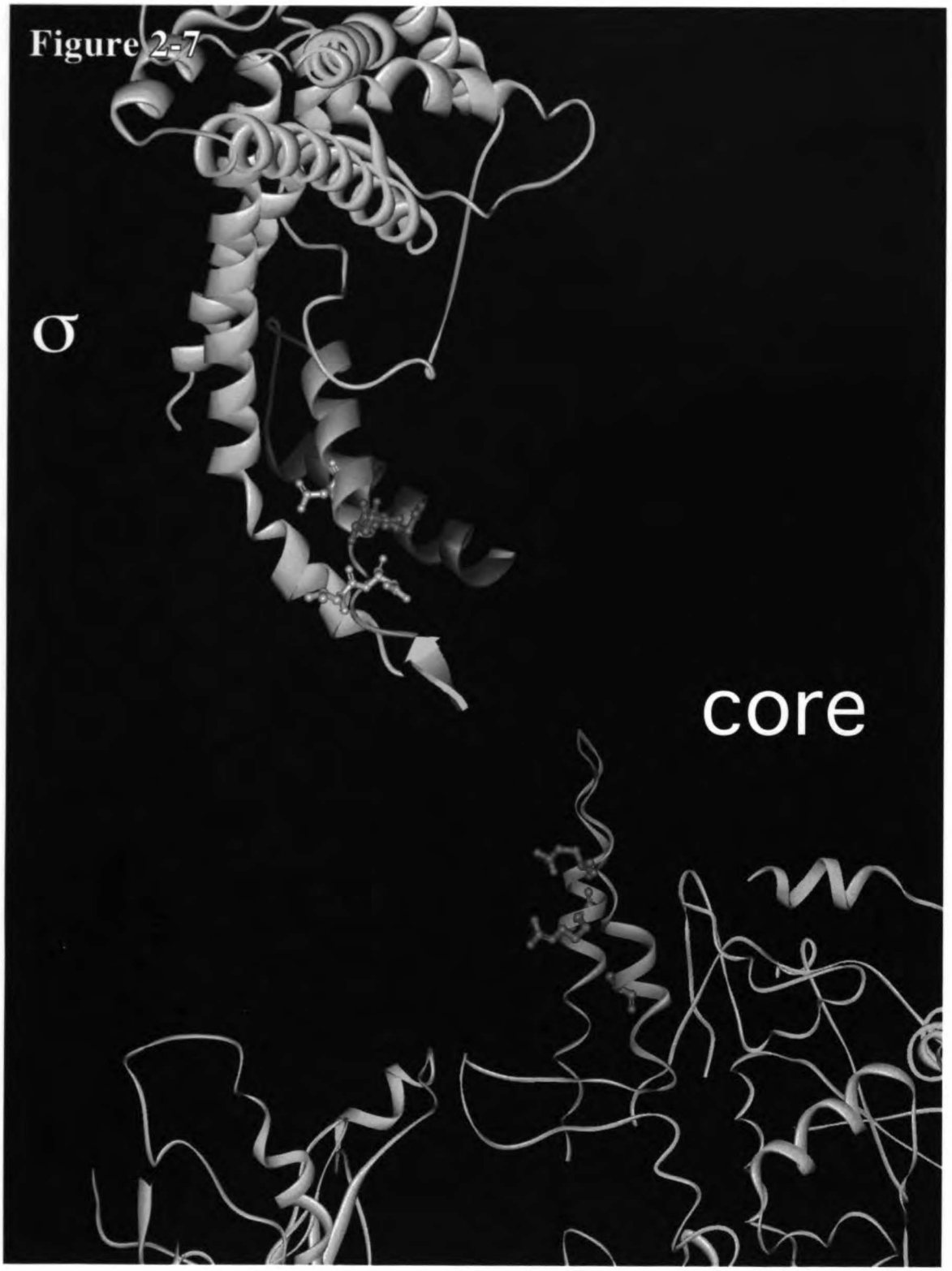


Figure 2-7
σ
core

Chapter Three

Views of Transcription Initiation

Summary

Initiation of transcription is the first step in gene expression and a major point of regulation. Recent structural studies reveal the nature of the initiating complex and suggest new ways of accomplishing the processes required for initiation.

RNA polymerase (RNAP), the enzyme that carries out transcription, is a remarkable molecular machine. During initiation, it must recognize promoter DNA from the vast excess of non-promoter DNA, separate the duplex to expose the template strand and initiate RNA synthesis using only mononucleotides. Before beginning processive elongation, it must transition to a non-sequence specific DNA binding protein that moves forward (and in some cases backward) along the DNA. Four recent reports illuminate these processes. Three use crystallography to provide structural information about the prokaryotic initiation factor σ, (Campbell et al., 2002), and the initiating form of prokaryotic RNAP without (Murakami et al., 2002b) and with promoter DNA (Murakami et al., 2002a). A fourth provides a distance constrained model of initiating RNAP and its interaction with promoter DNA based upon systematic measurements of fluorescence resonance energy transfer (FRET) of probes located throughout the initiating RNAP and in the DNA (Mekler et al., 2002). We discuss these structures and the insights and speculations they provoke about how this machine accomplishes these complex processes.

Structure

Transcription initiation in prokaryotes is carried out by holoenzyme (*holo*), comprised of core RNAP (*core*) plus the initiation specific subunit, σ. Core is an ~400 kDa complex of five subunits ($\alpha_2\beta\beta'\omega$), which shares considerable sequence and even more structural homology with its eukaryotic counterparts (e.g. RNAP II), whereas σ has little sequence homology to its counterparts, the general transcription factors. Holo first

recognizes the two conserved hexamers in the promoter, located at -10 and –35 relative to the transcription startpoint of +1, then melts the DNA from -11 to +4 to form the *open complex*, and then begins synthesizing the nascent RNA. The three sections below summarize the structures of σ, holo and the open complex.

σ subunit

All bacteria have one primary σ factor, which directs the majority of its transcription. σs have four conserved regions, which mediate binding to core and to DNA (Fig. 3-1). As the structure of only one σ domain (σ_2: Region 1.2-2.4 from *E. coli*) had been described (Malhotra et al., 1996), Campbell et al. (2002) tried to crystallized σ^A, the primary σ of the thermophile *Thermus aquaticus (Taq)*. This proved impossible but serendipitous protease contamination produced crystallizable fragments diffracting to ~2Å.

σ^A has three stably folded domains, σ_2, σ_3 and σ_4, connected by flexible linkers. Each domain is predicted to bind both core and DNA (Fig. 3-1). σ_2 is essentially identical to *E. coli* σ_2, with an exposed region 2.2 helix predicted to form a primary interface with core and the region 2.3-2.4 helix, which recognizes the –10 element and contains aromatic residues important for melting and recognition of the non-template strand of the -10 element. Both σ_3 and σ_4 are comprised of three helices. One helix in σ_3 is responsible for recognizing two conserved bases located upstream of the –10 region, present in "extended -10 promoters", which do not need a –35 promoter element. Two helices in σ_4 form a HTH motif; one of these helices recognizes the –35 region of the promoter. Campbell et al. were also able to obtain the structure of σ_4 complexed with a -35 element, allowing the first high resolution view of promoter recognition. This pivotal work defined the domain structure of σ, provided assurance that the genetic inferences about how σ recognized the –35 element were generally correct and produced high resolution structures that allowed definitive placement of σ on the holo structure.

Holoenzyme

Two studies provide our first glimpse of the structure of an initiation competent multi-subunit RNAP. A not quite atomic resolution (4 Å) electron density map for *Taq* holo, combined with structures of core (Zhang et al., 1999) and portions of σ^A (Campbell et al., 2002) allowed the Darst group to provide an excellent reconstruction of holo and a first view of a (nearly) intact σ (Murakami et al., 2002b). The distance constrained model based on FRET experiments measuring distances between probes in σ^{70} and core allowed the Ebright group to also place σ^{70} on core (Mekler et al., 2002). These studies are complementary. The FRET analysis provides information on region 1.1, which was absent in the *Taq* holo crystals, and the crystallographic work gives high resolution structural information. Overall, both models agree with one another and with previous biochemical and genetic evidence (Gross et al., 1998). When describing these models, we use the term *upstream* to refer to DNA before the start site (-1 to -∞) and *downstream* to refer to DNA after the start site (+1 to +∞).

To put these structures in context, we first revisit the model of elongating prokaryotic RNAP (Fig. 3-2), derived by combining the structure of core with crosslinking studies that place nucleic acids on the structure (Korzheva et al., 2000; Zhang et al., 1999). RNAP can be crudely described as a crab claw whose active site is positioned at the base of its two pincers. Downstream DNA, located in an internal channel formed between the pincers (also called the *jaws*), separates into its two strands near the active site. The strands turn upward (relative to the plane of the page in Fig.3-2), taking different paths through the polymerase and reanneal to form the upstream duplex, which is at a right angle to the downstream DNA. Nascent RNA follows the template strand for about 9 bases and then exits the polymerase underneath a flap that juts out from the bottom of the pincers. Studies of multiple crystals of polymerase indicate that the pincers and the flap are mobile. The particularly flexible top pincer, called the *clamp,* is

derived primarily from a portion of the β' subunit. The bottom pincer (with two independently mobile modules, β1 and β2) and the flap are derived from the β subunit. Flexibility is presumably required for the conformational changes necessary to accommodate steps in transcription.

The holo structure provides evidence of the importance of these mobile features (Fig. 3-3). The three domains of σ are spread out across one face of core, each interacting with and altering the position of a mobile domain of ββ' relative to its position in core. Additionally, since σ_2 interacts with the β' clamp in the upper pincer and σ_3 interacts with β1 in the lower pincer, these interactions can modulate opening and closing of the downstream DNA channel. Likewise, interaction of σ_4 with the β flap can alter the RNA exit channel. Additionally, several regions disordered in core become ordered upon interaction with σ, including structures homologous to the zipper and the lid of RNAP II, which may be important in guiding nucleic acids through RNAP (Gnatt et al., 2001). If the zipper and lid structures are maintained in elongating bacterial RNAP, as they are in elongating RNAP II, they will have different interacting partners than they do in the holoenzyme or in the initiation complex and may therefore be key focal points for the transition between initiating and elongating RNAP.

σ is properly positioned in holo to bind promoter DNA. Its DNA binding determinants are solvent exposed, with a spacing roughly consistent with that expected from the position of the DNA elements to which they bind. For example, σ_2 (–10 recognition) and σ_4 (–35 recognition) are separated by about 76Å, roughly the distance between the middle of –10 region and the –35 region in B-form DNA (Murakami et al., 2002b). One remarkable feature of σ that allows such separation is the σ_3–σ_4 linker. Its 33 amino acids, derived mostly from region 3.2, traverse the 45 Å that separates σ_3 and σ_4, skirting the active site and passing through cores's RNA exit channel before connecting up with σ_4.

Interestingly, FRET analysis locates σ region 1.1 in the downstream DNA channel (Mekler et al., 2002); a similar conclusion is reached by the Darst group more indirectly (Murakami et al., 2002b). Region 1.1 must move from this position to form an open complex, and the next set of structures described indicates that this is the case.

Clearly, the interface between σ and core is crucial for reconfiguring both partners for initiation; some notion of its importance comes from the realization that ~8500 Å^2 are buried, almost twice that of the largest reversible protein/protein interface known, and actually more comparable to interaction surfaces in oligomeric proteins (~10000 Å^2) (Murakami et al., 2002b).

Open complex

The same two groups investigate the structure of the open complex. The Ebright group used the FRET analysis to determine the positions of σ and downstream DNA in the open complex (Mekler et al., 2002). The Darst group (Murakami et al., 2002a) opted to crystallize holo bound to a well-characterized open complex mimic—the fork junction, a fragment of upstream promoter DNA that is double stranded from -41 to -12, with the non-template strand continuing in single stranded form to –7. Absence of the template strand downstream from -12 prevents formation of the multiple conformational intermediates in the process of open complex formation that would occur at the low temperatures employed in crystallization if intact open complexes were used.

With the exception of region 1.1, which moves from the inside to the outside of the active site channel (as determined by the FRET analysis), the positions of σ are similar in holo and the open complex. Fork junction DNA lies across the upstream face of holo, with essentially all sequence specific recognition carried out by σ (Fig. 3-4). Predicted contacts in the –10 region validate previous genetic assignments: the –12 position is very close to the very region 2.4 amino acids believed to contact this base-pair; the non-template strand is draped over the region 2.3 aromatic residues implicated in

their recognition; and the extended –10 recognition determinants are very close to the Region 3.0 amino acids believed to contact them. In contrast, most likely as a result of crystal packing artifacts, σ_4 is slightly mispositioned from the -35 region placement that is predicted genetically and observed in the structure of the σ_4–35 promoter region complex (Campbell et al., 2002).

Both groups use the current information and previous data (Gnatt et al., 2001; Korzheva et al., 2000), to generate similar models of RNAP open complexes (Fig. 3-4). The upstream DNA drapes over the domains of σ. At -11 the template and non-template strands separate. The template strand turns sharply down into the active site, completely enclosed in a positively charged protein tunnel formed of portions of σ, β and β', and framed by universally conserved basic residues of σ. The non-template stand continues to interact with σ until about –7 and then bends down between two lobes of the β subunit (β1 and β2). The strands reanneal at about +5 and this downstream duplex continues through the main channel in the tunnel formed by the jaws, just as in the structure of the elongation complex. Overall, a sharp bend is introduced between the upstream DNA and the downstream DNA.

Function

Promoter Recognition and Spacer Accommodation

The work presented here goes some way towards explaining how binding to core relieves region 1.1 autoinhibition of DNA binding in free σ. In holo, region 1.1 is removed from its location in σ and placed in the active site channel of polymerase, possibly because its high negative charge allows it to act as a downstream DNA mimic. This same idea could explain autoinhibition: region 1.1 might bind DNA recognition determinants in free σ because of its negative charge, thereby out competing promoter DNA.

The structures also suggest a mechanism by which RNAP binds promoters with as few as 16 or as many as 18 nucleotides in the spacer region between the -35 and -10 elements. (These variations can change the distance between these elements by as much as 10 Å.) σ_4 (–35 recognition), is perched on the end of the flexible flap of β. Shifting the angle at which this flexible flap juts out of the core enzyme alters the distance between σ_4, and σ_2 (–10 recognition) somewhat. Larger lengths of DNA can be accommodated by stretching (or "kinking") them over a bulge in β' that intervenes between the domains in σ that recognize the -10 and -35 hexamers. The enhanced DNaseI hypersensitivity in the spacer region of promoters with longer spacers is consistent with this explanation, as such "kinked" DNA would be expected to be more susceptible to DNase I cleavage (Murakami et al., 2002a).

Region 1.1 and open complex formation

Region 1.1 also plays a positive role in transcription initiation: holo having σ^{70} lacking region 1.1 forms open complexes very slowly at several promoters (Gross et al., 1998 and references therein). The jaws are closed in the holo structure (which was obtained with σ^{A} lacking region 1.1) but must be open in wild type holo to permit downstream DNA to enter. Thus, the Darst group proposes that opening the jaws is a rate limiting step at some promoters, and that region 1.1 accelerates this opening by binding between them, thereby accelerating open complex formation. At one weak promoter, holo lacking region 1.1 forms melted complexes more readily than intact holo (Vuthoori et al., 2001), perhaps as suggested by the Ebright group, because the rate limiting step at this promoter is ejecting region 1.1 from the jaws.

A great deal about the relationship between σ and jaw opening remains to be worked out. Region 1.1 is conserved only in primary or housekeeping σ factors; how are the jaws of the polymerase opened during initiation with alternate σ factors, which lack region 1.1? Additionally, the notion that σ opens the jaws of polymerase currently lacks

experimental backing. The most recent analysis of *E. coli* core RNAP indicates that its jaws are wide-open (Darst et al., 2002). Although this could be artifactual, it is worthwhile recalling that σ (and TFIIF in eukaryotes) decreases binding to non-promoter DNA. Do they do this by partially closing the jaws? Even after solution measurements of the placement of the jaws in core and holo resolve this particular question, additional work is clearly needed to understand how downstream DNA is efficiently placed in the jaws during open complex formation.

De Novo RNA synthesis

In contrast to DNA polymerases, which can only extend existing nucleic acid chains, RNAP is able to create a nucleic acid polymer *de novo*--using only mononucleotides. The initiation step is difficult: both incoming nucleotide and the initiating nucleotide, which attacks the incoming NTP, must be stabilized in the correct geometry. This is especially difficult because RNAP prefers to begin RNA chains with an ATP (which base pairs with the template strand more weakly than would CTP or GTP).

How does RNAP accomplish *de novo* synthesis? Specific interactions with the initiating nucleotide must hold it rigidly in place, facilitating chemical attack on the incoming nucleotide. Such specific interactions would explain why polymerase prefers to start transcripts with ATP (followed by GTP, UTP and then CTP). Indeed, a sub-complex of core polymerase, $\alpha_2\beta$, and possibly even the isolated β subunit has a site for the initiating nucleotide (Naryshkina et al., 2001). Darst now suggests that a disordered loop of σ near the beginning of the σ_3-σ_4 linker, pointing towards the active site, assists in binding the initiating nucleotide. They tested this idea, using an extended –10 promoter, which does not use –35 recognition determinants and can therefore be transcribed by holo ending at σ_3, which lacks the disordered loop. Holo with this truncated σ requires a much higher concentration of initiating nucleotide to reach maximal levels of transcription than does holo with full length σ (Campbell et al., 2002). Using σ to provide specificity is

appealing. A marked preference for a particular nucleotide in the attacking site might have disagreeable side effects when elongating a transcript. If σ performed this function, the selectivity required for *de novo* synthesis would be present only at initiation.

We suggest another possible role for this disordered loop of σ--stabilization of the melted state of the promoter by binding to the template strand near the start site in single stranded form (thus keeping it from reannealing). This idea comes from a consideration of the ribosomal RNA (rRNA) promoters. rRNA promoters cannot form stable open complexes, they require high levels of initiating nucleotide to stabilize the melted state required for efficient transcription. This requirement is an important regulation mechanism in the cell, and is mediated in part by the presence of a stretch of CG bases, called the discriminator region, near the start site of rRNA promoters. The unique nucleotides in the discriminator may prevent the stabilizing interactions between the disordered loop in σ and the template strand, thereby preventing stable open complex formation. In this context, an alternate single strand specific interaction may be necessary to achieve a stable open complex. The bridge created between the template strand and RNAP by the initiating nucleotide may provide this interaction-- thus the requirement for high initiating nucleotides. Likewise, removal of the disordered loop in σ should prevent stable open complex formation and create a requirement for high initiating nucleotides.

Promoter Clearance and Abortive Initiation

All RNAP reiteratively synthesize and release short RNA transcripts called abortives, (~2-9 nucleotides in length). Based on the structure of elongating eukaroytic RNAP II, Kornberg proposed that shorter nascent RNAs dissociate because they make fewer contacts with polymerase than do longer RNAs (Gnatt et al., 2001). Another explanation for these transcripts has emerged here. Region 3.2 occupies the RNA exit channel, leading to the speculation that nascent RNAs must successfully compete with region 3.2 to be retained in elongating polymerase. When RNA transcripts lose the

competition, they are ejected as abortive transcripts; when they win, region 3.2 is ejected and the transcript is successfully elongated. Consistent with this idea, holo having σ lacking region 3.2 produced fewer abortives relative to full length transcript than does holo with wt σ (Campbell et al., 2002). If this is a universal explanation for abortive transcription, a general transcription factor must play this role in eukaryotic initiation.

Why might polymerase place region 3.2 of σ right where the RNA must go, thus wasting valuable NTP energy synthesizing abortive RNAs? Maybe the competition between σ and RNA is an important part of the promoter clearance process. During promoter clearance the polymerase must extricate itself from promoter specific contacts so it can processively elongate trancripts. Interestingly, promoter clearance tends to coincide with the end of abortive synthesis and could be set in motion by release of region 3.2 from core by the successfully elongating RNA chain. Release of region 3.2 could cause promoter clearance because it weakens the σ/core interface thus allowing core to dissociate from σ. Alternatively, either movement of region 3.2 out of the channel or of RNA into the channel could alter the position of σ_4 (which is perched on the flap surrounding the channel) making correct interactions with -35 element impossible and promoting promoter dissociation. In either case, the NTP energy utilized during abortive initiation may be a small price to pay in order to switch from specific interaction with the promoter to processive elongation.

Conclusion

These first glimpses of the initiation competent polymerase provide extraordinary insight into the functions the machine performs. Although these structures do not provide final answers for how these processes are accomplished, they do allow us to conceptualize concrete models. No doubt these first peeks into the structure of initiating RNAP will motivate an enormous number of future experiments to test these ideas.

Selected Reading

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002). Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. Mol. Cell *9*, 527-39.

Darst, S. A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G., and Wriggers, W. (2002). Conformational flexibility of bacterial RNA polymerase. Proc. Natl. Acad. Sci. USA *99*, 4296-301.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science *292*, 1876-82.

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb. Symp. Quant. Biol. *63*, 141-55.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). A structural model of transcription elongation. Science *289*, 619-25.

Malhotra, A., Severinova, E., and Darst, S. A. (1996). Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. Cell *87*, 127-36.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R., and Ebright, R. H. (2002). Structural organization

of bacterial RNA polymerase holo and the RNA polymerase-promoter open complex. Cell *108*, 599-614.

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002a). Structural basis of transcription initiation: an RNA polymerase holo/DNA complex at 6.5 Å resolution. Science *in press*.

Murakami, K. S., Masuda, S., and Darst, S. A. (2002b). Structural basis of transcrition initiation: *T. aquaticus* RNA polymerase holo at 4 Å resolution. Science *in press*.

Naryshkina, T., Mustaev, A., Darst, S. A., and Severinov, K. (2001). The β' subunit of *E. coli* RNA polymerase is not required for interaction with initiating nucleotide but is necessary for interaction with rifampicin. J. Biol. Chem. *276*, 13308-13.

Vuthoori, S., Bowers, C. W., McCracken, A., Dombroski, A. J., and Hinton, D. M. (2001). Domain 1.1 of the σ^{70} subunit of *E. coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. J. Mol. Biol. *309*, 561-72.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of *T. aquaticus* core RNA polymerase at 3.3 Å resolution. Cell *98*, 811-24.

Figure 3-1

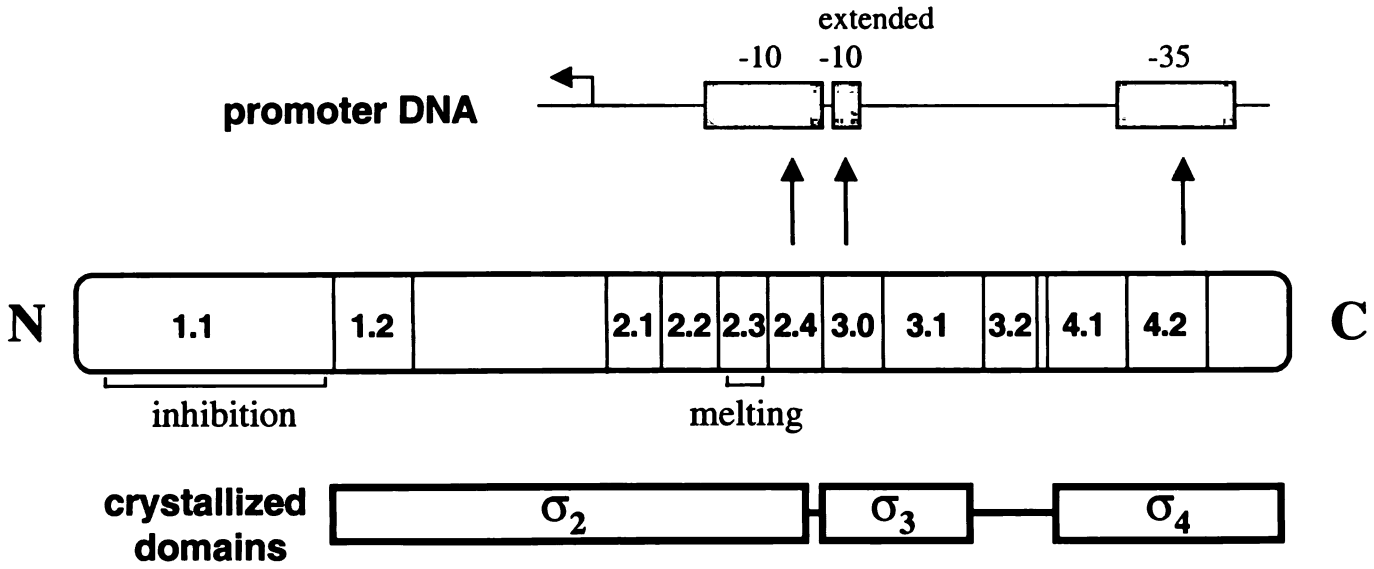
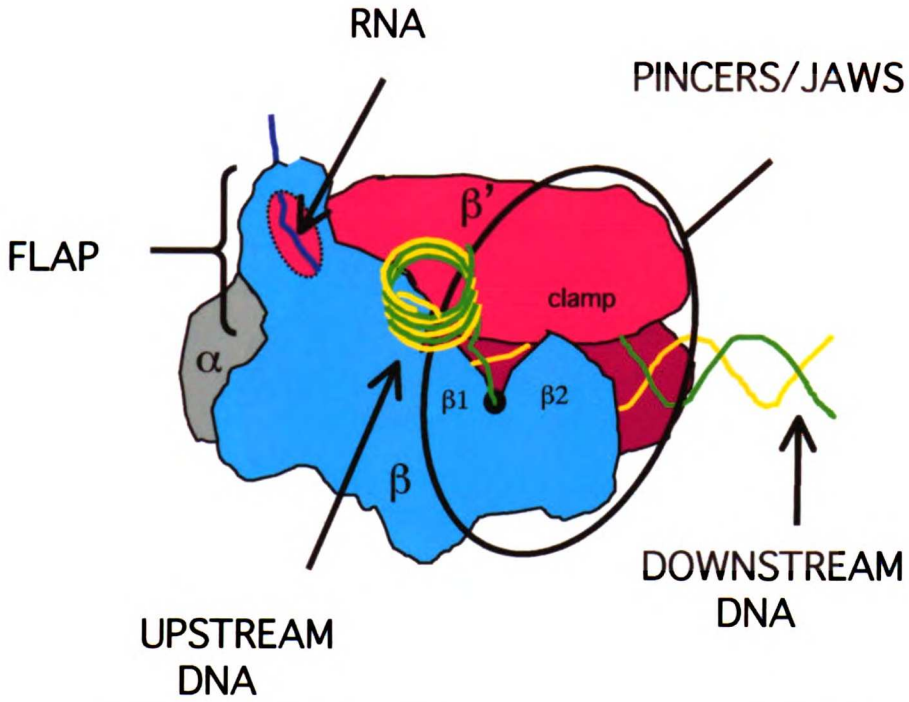
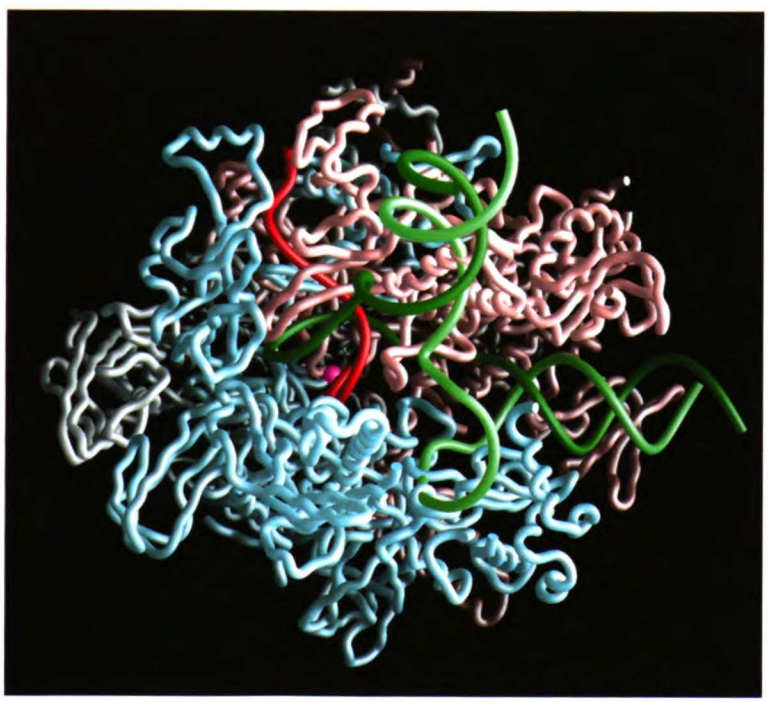

extended
-10
-10
-35
promoter DNA
N
1.1
1.2
2.1
2.2
2.3
2.4
3.0
3.1
3.2
4.1
4.2
C
inhibition
melting
crystallized domains
σ2
σ3
σ4

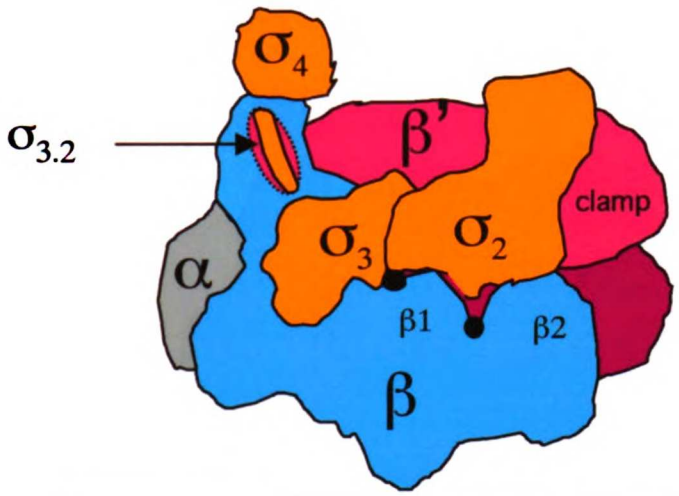
Figure 3-2

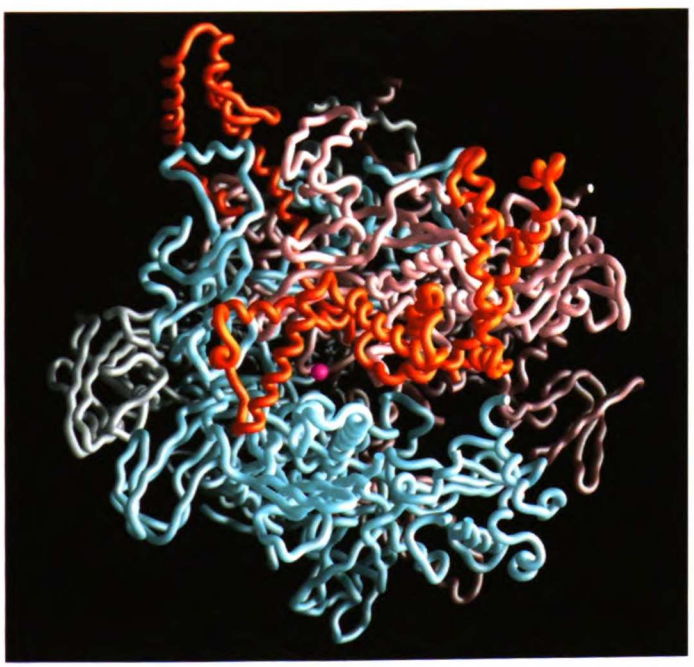




Courtesy: Seth Darst

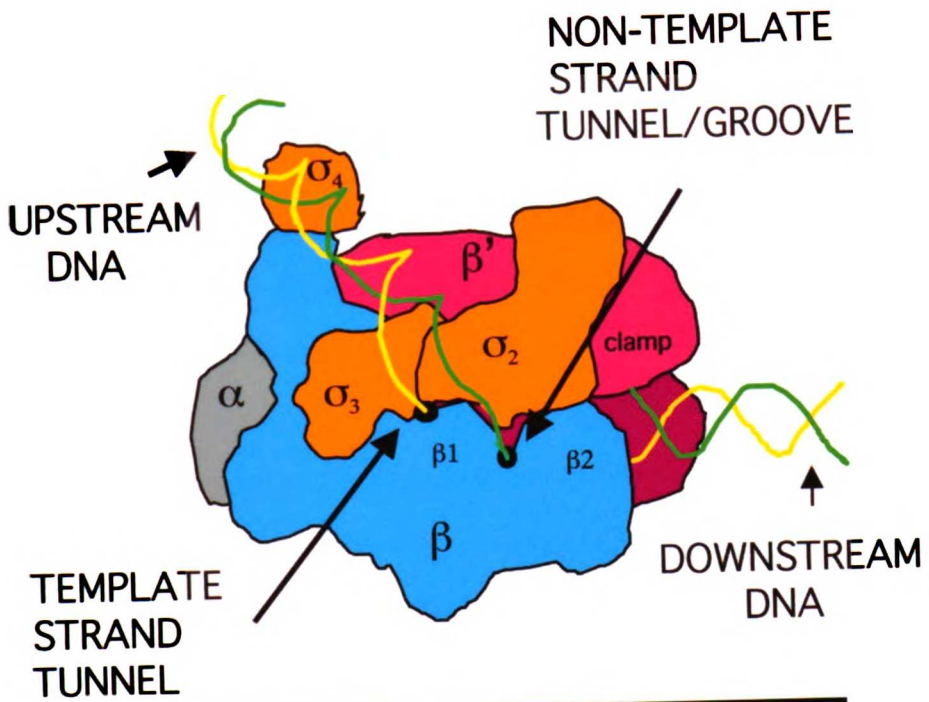
Figure 3-3

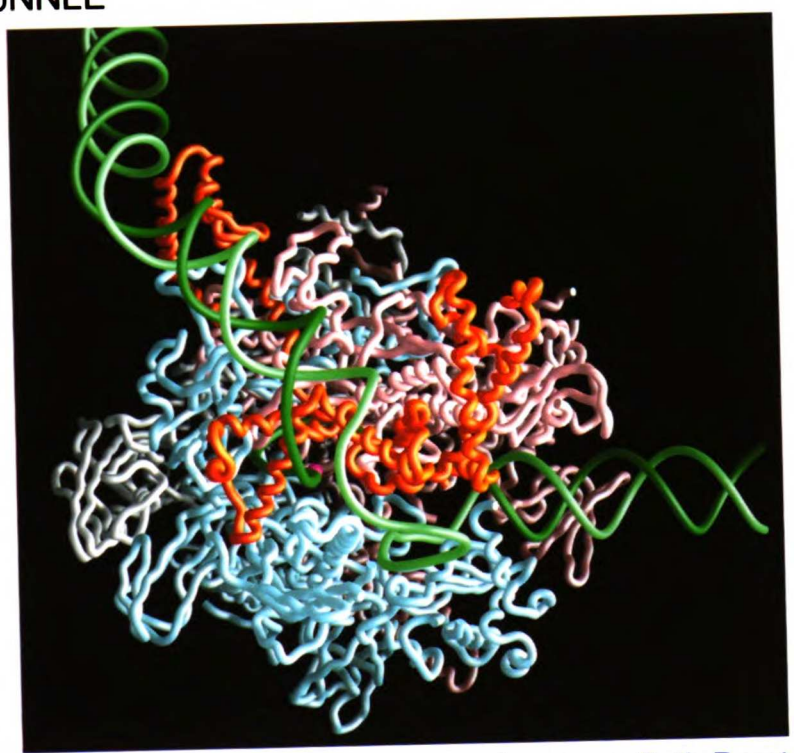




Courtesy: Seth Darst

Figure 3-4





Courtesy: Seth Darst

Chapter Four

Conclusions and Future Directions

By determining the minimal portion of holoenzyme capable of the non-template strand binding or "clamping open" activity, this thesis demonstrates the feasibility of dissecting RNA polymerase into minimal functional assemblies. My finding that the small (48 a.a.) coiled coil of the β' subunit is the important part of core polymerase for this activity and that it appears to act as an allosteric effector of σ conformation, demonstrates the utility of the approach. Below I consider how similar dissections might therefore be worthwhile for the other known promoter melting activities, as well as perhaps for melting as a whole. This work also established, in conjunction with the work of Dr. Gruber, that the β' coiled coil binds to sub-region 2.2 of σ (Chapter 2, Gruber et al., 2001)). Together with experiments from the Roberts lab, it helped explain the function of this the most conserved sub-region in σ: an allosteric binding site(Ko et al., 1998). That the coiled coil binds region 2.2 has been confirmed by later crystal structures of bacterial holoenzymes (Murakami et al., 2002; Vassylyev et al., 2002). Still unclear and interesting, however, is the nature of the allostery that promotes non-template binding; I discuss this below.

This thesis also showed that a discrete portion of a large protein/protein interface can be responsible for allosterically affecting the function of one of the partner proteins. This is an especially exciting insight from a technological standpoint. If small portions of large interfaces can be responsible for allostery, it may then be possible to block allosteric function with small molecule mimics. There are currently efforts to find such mimics capable of blocking the allosteric binding of the coiled coil to σ(Bergendahl et al., 2003). As non-template binding appears a vital function in an essential process, such small molecules could potentially serve as future anti-microbial agents.

Another exciting possibility raised by this thesis involves the greater than average sequence identity and considerable structural homology between the *E. coli* coiled coil and its counterpart regions in eukaryotic RNA polymerases. This suggests that this region in eukaryotes may act allosterically in interacting with one of the general

transcription factors. Potentially, this eukaryotic factor might be performing an analogous non-template strand binding activity as part of promoter melting in eukaryotes—revealing that the mechanism of initiation in prokaryotes and eukaryotes may be more similar than previously thought. Below I consider experiments testing this potential conservation of function. (Of course, such functional conservation could be problematic for the drug potential of small molecules targeting the prokaryotic coiled coil, noted above: they might be toxic to humans as well as to bacteria because of the conservation between coiled coils. However, even if the function of the coiled coil is conserved, drugs specific to the prokaryotic coiled coil may still be found as the homology is not overwhelming.)

Finally, in Chapter 3 of this thesis, as part of my consideration of the recent structural evidence about holoenzyme, I proposed the possibility that σ might perform an important single stranded template binding activity downstream of the –10 hexamer. I conclude this chapter by explaining this hypothesis and how it might be addressed in more detail.

Dissecting holoenzyme into minimal assemblies sufficient for other melting activities

a. Initial Promoter Recognition.

It is still not entirely clear how binding of core polymerase allows σ to initially recognize the promoter. It is known that region 1 of σ inhibits DNA binding determinants in σ regions 2/3 and 4, preventing promoter binding by free σ(Dombroski et al., 1992). However, even if region 1 is removed, σ alone binds promoters extremely weakly: unlike holoenzyme, it is incapable of footprinting or gel shifting promoter DNA (Suh, W.C. and Gross, C. unpublished results). Apparently, core must be doing more in the process than simply relieving autoinhibition.

One way in which core may do so is indicated by the recent structures of holoenzyme in conjunction with previous work from the Heyduk lab (Callaci et al., 1998; Murakami et al., 2002; Vassylyev et al., 2002). Because portions of core widely separated in the structure make interactions with the different domains of σ, core appears to move σ's DNA binding determinants so that they are better positioned to simultaneously make their respective promoter contacts. This allostery by global positioning of domains may also help to explain how autoinhibition is relieved. Because region 1 binds a portion of core at a distance from the DNA binding determinants of σ, core may move it away from these determinants, preventing region 1 from occluding DNA binding.

My work indicates that discrete portions of the σ/core interface can functionally reorganize the conformation of an individual folded domain of sigma. This sort of "micro-allostery" could potentially be important in initial promoter recognition, as well. For example, the β flap or the β pincer may by themselves allosterically stimulate promoter recognition by σ domains 4 (containing region 4) or 3 (containing σ sub-region 3.0—involved in extended -10 interaction—see chapter 3) respectively. "Micro-allostery" may also be involved in alleviating auto-inhibition: interaction between domain 1 of σ and the downstream DNA channel of core might change domain 1's conformation stimulating it to release domain 4, which it otherwise appears to bind (Dombroski et al., 1993). One argument for why "micro-allostery" could be helpful for the polymerase comes from a consideration of promoter clearance. It is known that when σ and core dissociate after initiation, σ does not remain associated with the promoter. (If it did so this would tend to target core—when it reassociated with σ— to sites of pre-existing gene expression, rather than allowing free holoenzyme to distribute to promoters based on the current cellular conditions.) If core "micro-allosterically" stimulates σ promoter binding, then loss of this "micro-allostery" could facilitate release of σ from the promoter once σ and core are no longer in contact. This wouldn't seem to be the case if global

positioning allostery was the only mechanism utilized by core to stimulate promoter recognition. In that case, even after core had moved downstream, the DNA binding determinants of σ would be likely remain bound to their respective elements.

There is a third possible way for core to stimulate initial promoter binding: by directly contributing contacts to the DNA. The recent structures of holoenzyme and the low resolution structure of holoenzyme complexed with fork junction DNA, provide only one strong candidate for such a contact: a zinc finger domain at the very N-terminus of β'. Currently, there has been no functional test of the importance of this region in polymerase in initial promoter recognition.

The approach of testing minimal assemblies of holoenzyme provides a way of directly determining whether known discrete σ/core contacts may act "microallosterically" to stimulate promoter recognition. Furthermore, looking into the ability of fragments of β' with or without the Zinc finger motif to stimulate promoter recognition may be able to provide evidence that this region directly contacts DNA. Much of this can readily be done with constructs that are already present in the lab. Furthermore, work from the Meares laboratory provides for an excellent assay with which to perform these experiments: FeBABE conjugated σ DNA footprinting(Owens et al., 1998)(Bown et al., 1999). This assay is sensitive, non-perturbing, and includes a built in binding specificity control. All this may make it relatively simple to carry out these experiments.

b. Fork Junction Recognition

The recent structure of holoenzyme complexed with fork junctions provides some picture of how holoenzyme performs this activity. However, because of the quite low resolution (~6.0 Å), most of the important details are not present. No doubt, a higher resolution structure would provide more information; one way such a structure might be made possible is by dissecting the polymerase into the minimal assembly responsible for

this activity. Eliminating the extraneous portions of the polymerase --some of which may be less ordered and thus interfere with crystallization—may greatly increase the chances of obtaining crystals that diffract to high resolution. Unfortunately, the reconstitution of fork junction binding is greatly complicated by the absence of a robust assay for this activity. A battery of gel shift controls was necessary to argue for the presence of this activity when it was first discovered (Guo and Gralla, 1998). Repeating all of these assays to test the functionality of various minimal assemblies of holoenzyme is likely to be a long and complicated process. Furthermore, during the course of this thesis, the Heyduk lab has presented evidence calling the fork junction activity of holoenzyme into question (Matlock and Heyduk, 2000). The original fork junction paper presented evidence that holoenzyme binds promoter fragments containing a double stranded/ single stranded DNA junction in addition to the single stranded non-template strand much more tightly than the non-template strand alone. Matlock et al. found that fork junction fragment binding was at best only marginally stronger than non-template strand binding alone.) Considering both these facts, other experiments may be a higher priority at present.

c. Overall Melting

Obviously, an exciting prospect is determining the minimal portion of holoenzyme capable of achieving melting. Extrapolating from my results, one simple model for how melting may occur would be that, once initial promoter recognition is achieved, the coiled coil allosterically stimulates σ to perform *all* the activities necessary for melting. One might imagine this occurring either via an initial base flipping nucleation step and followed by non-template strand binding or simply by non-template strand binding on its own. (Note that though the evidence for a fork junction binding activity may prove incorrect and thus unable to support a base flipping, no evidence has currently been advanced to *disprove* the possibility of a base flipping nucleation step.) I am currently in

the process of trying to test this model. The coiled coil and full length σ do not appear capable of achieving melting at general promoters. This is not surprising. As discussed above, initial promoter recognition is likely to minimally require both global repositioning of region 4 of σ to allow –35 recognition, and removing autoinhibition. However, I currently have preliminary evidence that a fragment of β' containing the coiled coil, together with a fragment of σ, lacking region 1, can achieve melting at a special class of promoter lacking the –35 element (the extended –10 promoters mentioned in chapter 3). While more experiments remain to be done, and while non-specific binding of core to DNA near the start site likely stabilizes the melt, this exciting evidence tends to support the simple model described above.

Nature of non-template binding allostery

In the discussion of Chapter 2, I proposed several mechanisms for how coiled coil binding to σ sub-region 2.2 might increase the non-template strand binding functionality of σ sub-regions 2.3/2.4. All mechanisms involved transmitting a conformational change across the interface between helix 2.2 and helix 2.3/2.4: rotation of helix 2.2 causing helix 2.3/2.4 to rotate and expose residues; additional/decreased ordering of helix 2.2 leading to additional/decreased ordering of helix 2.3/2.4; and movement of helix 2.2 disrupting its interaction with helix 2.3/2.4 thereby exposing residues of 2.3/2.4. However, the structure of region 2.2/2.3/2.4 in holoenzyme (*Thermus thermophilus*) is essentially superimposible with the same region in a domain of free σ ((*T. aquaticus*). Moreover, the B-factors (or thermal motion factors) of the atoms do not appear to change significantly. Thus there is little support for the above models. However, these results could also arise from crystallization artifacts. Possibly free σ is in equilibrium between two forms: a form incapable of non-template binding (that would greatly predominate in

solution) and a form capable of non-template binding (the form of σ stimulated by core). If the form capable of non-template binding were to be favored under crystallization conditions–perhaps because it is more ordered—it would be obtained even when free σ is crystallized. Alternatively, holoenzyme could be in equilibrium between two forms: one in which σ is non-template strand binding capable, and another where it is incapable. In this case perhaps the incapable form is more readily crystallized. NMR solution structures of free σ and coiled coil bound σ (perhaps using a small alternative σ) might be able to test these possibilities.

Nonetheless, other models for the allosteric mechanism need to be considered. For all the reasons presented in the discussion of chapter 2, I still disfavor the early proposition that an acidic disordered loop in σ inhibits non-template strand binding, and core binding to 2.2 removes this inhibition. Still, it may be worthwhile to directly test this hypothesis by removing the loop in *E. coli* σ and checking for non-template binding in the absence of core. Currently I am interested in a possible allosteric mechanism involving a different portion of σ's second domain, sub-region 1.2. While regions 2.2-2.4 are little changed in the holoenzyme structure, the position of the sub-region 1.2 helix does move. In fact, based on the low resolution fork junction structure, several solvent exposed residues in region 1.2 (in *E. coli* numbering Threonine 110 Arginine 111 and Glutamate 112) move toward the non-template strand. Potentially, core binding causes allostery by allowing these residues in sub-region 1.2 to make direct contacts with the non-template strand they are incapable of in free σ. Consistent with this, site directed mutagenesis of region 2 has found only a few residues affecting non-template strand binding; additional contacts may be needed(Huang et al., 1997; Tomsic et al., 2001).

Previously, experiments in *B. subtilus* had shown when a σ containing a non-template strand binding defective mutation was expressed in the absence of wild type σ it grew poorly (Tomsic et al., 2001). As a preliminary genetic experiment to test if region 1.2 was critical for non-template strand binding, I tested complementation of σ's bearing

alanine mutations at the solvent exposed region 1.2 residues listed above. Discouragingly, no growth phenotype was observed. However, *B. subtillus* holoenzyme appears to interact less tightly with melted promoter complexes than does the *E. coli* enzyme (*B. subtillus* holoenzyme can be competed off melted promoter complexes with the polyanion heparin, while the *E.coli* enzyme is much more resistant to such competition). Thus a growth phenotype might be difficult to detect in *E. coli*, even for a mutant strongly affecting non-template strand binding. It may be worthwhile to test the effect of the region 1.2 mutants in *B. subtillus*; alternatively, the *E. coli* mutants might be purified and an effect on non-template strand binding directly measured *in vitro*.

Interaction of the eukaryotic coiled coils and the general transcription factors? A eukaryotic non-template binding activity?

If the coiled coil in eukaryotic polymerases indeed functions as an allosteric effector, one can imagine several potential targets among the general transcription factors. Among the pol II general transcription factors TFIIF, TFIIB and TBP, like σ, interact both with the core polymerase and DNA (Orphanides et al., 1996). Of these, TFIIF is the most tantalizing, as its RAP30 subunit has previously been found to have some sequence similarity to σ region 2 and 4 and because, like σ, it contains an N-terminal domain that autoinhibits DNA binding. TFIIF has also been shown to compete with σ for binding to prokaryotic core polymerase (reviewed in Orphanides et al., 1996). Moreover, TFIIF appears to be poised, in melted DNA complexes, at the upstream edge of the transcriptional bubble, like σ region 2 (Kim et al., 2000).

Of course, in order for the pol II coiled coil to allosterically affect a general transcription factor, it would have to bind it. Preliminary pulldown experiments have thus been performed looking for interaction between pol II coiled coil and intact TFIIF, TBP, TFIIB, and the isolated subunits of TFIIF, RAP30 and RAP74 (Wember, K.,

Gruber T., Young, B.A., unpublished data). Of these, only TBP and isolated RAP30 showed significant interaction with the coiled coil. For both, unfortunately, significant interactions also occurred with GST alone (in fact, for RAP30, interaction with GST coiled coil and GST alone was about equivalent) – suggesting that binding might simple to due to general "stickiness" of the factors. Obviously, this fails to convincingly support an allosteric interaction between a eukaryotic coiled coil and a general transcription factor; note however, that the results are quite preliminary. For example, additional experiments indicate that the particular GST tagged eukaryotic coiled coil construct co-purifies with significant amounts of the chaperone Dna K (Young, B.A. unpublished observation). This suggests that lack of significant interaction could be due to failure of this coiled coil construct to fold (perhaps folding of the eukaryotic version is less robust). It might be profitable to repeat these experiments with larger constructs, where surrounding structure might assist folding of the coiled coil. In addition, another simple preliminary experiment that might be profitable in this area: a direct test of whether pol II is capable of inducing TFIIF (or potentially the other factors) to bind to single stranded eukaryotic non-template DNA, either by gel shift or by the crosslinking assay I utilized in chapter 2.

Is the σ hairpin loop involved in a single stranded template binding activity important for melting?

There is a particularly fascinating class of promoters in *E. coli* exemplified by rrnB P1. rrnB P1 is involved in growth rate control of ribosome levels: cells with plenty of nutrients available synthesize abundant quantities of ribosomes so they can make more protein and thus rapidly grow and divide. When nutrients are scarce cells downregulate ribosome synthesis, as they have less need to create more protein. The rrnB P1 promoter

transcribes ribosomal RNA, the functional constituent of the ribosome. It turns out that levels of ATP, the primary energy currency of the cell and the nucleotide with which the rrnB P1 promoter initiates, directly affect the ability of holoenzyme to begin transcription at this promoter. They do so because holoenzyme readily recognizes and binds the rrnB P1 promoter, but has a hard time melting the promoter and therefore starting transcription. High levels of ATP (as are present in an energy rich cells) stabilize the melted complexes at the rrnB P1 promoter, leading to more transcription from the promoter and thus more ribosomal RNA (Gourse Science paper). Why rrnB P1 and promoters like it are difficult for the polymerase to melt, and how this difficulty is reversed by the initiating nucleotide, is unclear, but it seems to involve the presence of a stretch of CG bases near the start site known as the discriminator (Barker and Gourse, 2001; Gaal et al., 1997).

Recently, Campbell et. al. found that when holoenzyme was reconstituted with a fragment of σ, which lacked—among other things— a portion of region 3, it could transcribe only in the presence of high levels of initiating nucleotide(Campbell et al., 2002). The low resolution structure of holoenzyme suggested that the loop of region 3 (which was missing in this fragment) pointed in toward the active site of the core enzyme (Murakami et al., 2002). As discussed in Chapter 3, this led Campbell et al. to propose that this portion of σ was involved in stabilizing the initiating nucleotide in transcription—i.e. the first nucleotide in a trancript. Naryshkina et al., of course, had earlier provided evidence that the isolated β subunit contained the initiating nucleotide binding site; perhaps the loop of σ region 3, however, further supplemented the binding site in holoenzyme for the nucleotide(Campbell et al., 2002; Naryshkina et al., 2001). In recent high resolution *Thermus Thermophilius* structure, however, though the loop of region 3 (renamed the σ hairpin loop) does indeed point toward the active site it appear to be a considerable distance from where the initiating nucleotide is expected to be (~10 Å away) (Vassylyev et al., 2002 and Borukhov, S., personal communication).

In Chapter 3, I provided an alternate model to explain the Campbell et al. observation: perhaps in the absence of the hairpin loop, the polymerase has trouble melting the promoter. This could occur if this hairpin loop region of σ normally stabilizes melted complexes by binding to the template strand of the promoter near the start site in single stranded form. It is possible that ribosomal promoters lack the template region normally stabilized by this hairpin loop. Perhaps the discriminator sequence of rrnB P1, unlike most promoter's template strands near the start site, is recognized poorly by the σ hairpin loop, thus accounting for their inability to form stable open complexes. Whether the proposed binding interaction is eliminated by removing the hairpin loop or the appropriate template sequence, this weakened melted complex might be stabilized by an initiating nucleotide. By binding both the template strand and the polymerase near the active site (the β site identified by Naryshkina et al.) it would provide a bridge that allows the single stranded template strand and the polymerase to interact (Naryshkina et al., 2001). This would stabilize the melted form of the promoter and stimulate transcription. That holoenzyme might have a single stranded template binding activity may seem counter-intuitive. In Chapter 2, I showed lack of template strand binding both by holoenzyme and my minimal complexes as a control demonstrating that they bound the non-template specifically (rather than being able to bind any single stranded DNA). There are several possible explanations for this failure to observe template strand binding that are consistent with the model above. For example, perhaps at most promoters specific template strand binding is much weaker than non-template strand binding and thus is difficult to observe. Alternatively, the hairpin loop may make non-specific interaction with the template strand, but the bulky nature of the discriminator sequence in rrnB P1 may clash with the hairpin loop, thus blocking such interactions. Either form of binding, while important for melting, might only be strong enough to occur when the template strand is presented to the hairpin loop in the proper geometry by the rest of σ holding the promoter in place. Consistent with these

possibilities, the Gralla lab has found that if holoenzyme is presented with a fragment of a promoter DNA that is double stranded from –35 to –11 with one additional single stranded template base overhang (a fork junction) it is recognized much more poorly than a similar probe that included the single stranded template up to the start site (Guo and Gralla, 1998). Of course, this observation may simply be due to non-specific binding of core to the template strand near the active site. On the other hand, it may indicate the binding of the σ hairpin loop to the template strand posited above.

In preliminary experiments, I have determined that promoter melting of holoenzyme reconstituted with the Campbell fragment of σ (lacking the hairpin loop) is indeed weaker than holoenzyme reconstituted with full length σ. It is not clear at this point, though, if increased initiating nucleotide concentrations result in more robust melting as would be expected if my model is correct. Furthermore, though these experiments may help indicate that the model is correct, they are somewhat artificial.. They would be more interpretable if they were performed with a σ construct where the only change present was the loss of the hairpin loop (perhaps by substitution of a flexible linker domain) rather than a fragment of σ also lacking additional regions. Moreover, more direct tests for a σ hairpin loop interaction with the template strand, and whether the discriminator sequence eliminates this interaction can be imagined. For example: does loss of the hairpin loop decrease the ability of holoenzymeto bind the fork junction/long single stranded template overhang probe noted above? If the discriminator sequence is introduced into the fork junction/long single stranded template overhang probe, does this decrease binding by intact holoenzyme? These and other experiments may help determine if the model is correct.

References

Barker, M. M., and Gourse, R. L. (2001). Regulation of rRNA transcription correlates with nucleoside triphosphate sensing. J Bacteriol *183*, 6315-6323.

Bergendahl, V., Heyduk, T., and Burgess, R. R. (2003). Luminescence resonance energy transfer-based high-throughput screening assay for inhibitors of essential protein-protein interactions in bacterial RNA polymerase. Appl Environ Microbiol *69*, 1492-1498.

Bown, J. A., Owens, J. T., Meares, C. F., Fujita, N., Ishihama, A., Busby, S. J., and Minchin, S. D. (1999). Organization of open complexes at Escherichia coli promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. J Biol Chem *274*, 2263-2270.

Callaci, S., Heyduk, E., and Heyduk, T. (1998). Conformational changes of Escherichia coli RNA polymerase sigma70 factor induced by binding to the core enzyme. J Biol Chem *273*, 32995-33001.

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002). Structure of the Bacterial RNA Polymerase Promoter Specificity sigma Subunit. Mol Cell *9*, 527-539.

Dombroski, A. J., Walter, W. A., and Gross, C. A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. Genes Dev *7*, 2446-2455.

Dombroski, A. J., Walter, W. A., Record, M. T., Siegele, D. A., and Gross, C. A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell *70*, 501-512.

Gaal, T., Bartlett, M. S., Ross, W., Turnbough, C. L., Jr., and Gourse, R. L. (1997). Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. Science *278*, 2092-2097.

Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., Artsimovitch, I., Geszvain, K. M., Arthur, T. M., Burgess, R. R., *et al.* (2001). Binding of the initiation factor σ70 to core RNA polymerase is a multi-step process. Molecular Cell *8*, 21-31.

Guo, Y., and Gralla, J. D. (1998). Promoter opening via a DNA fork junction binding activity. Proc Natl Acad Sci U S A *95*, 11655-11660.

Huang, X., Lopez de Saro, F. J., and Helmann, J. D. (1997). sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA. Nucleic Acids Res *25*, 2603-2609.

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

Ko, D. C., Marr, M. T., Guo, J., and Roberts, J. W. (1998). A surface of Escherichia coli sigma 70 required for promoter function and antitermination by phage lambda Q protein. Genes Dev *12*, 3276-3285.

Matlock, D. L., and Heyduk, T. (2000). Sequence determinants for the recognition of the fork junction DNA containing the -10 region of promoter DNA by E. coli RNA polymerase. Biochemistry *39*, 12274-12283.

Murakami, K. S., Masuda, S., and Darst, S. A. (2002). Structural basis of transcrition initiation: *T. aquaticus* RNA polymerase holoenzyme at 4 Å resolution. Science *in press*.

Naryshkina, T., Mustaev, A., Darst, S. A., and Severinov, K. (2001). The beta ' subunit of Escherichia coli RNA polymerase is not required for interaction with initiating nucleotide but is necessary for interaction with rifampicin. J Biol Chem *276*, 13308-13313.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes Dev *10*, 2657-2683.

Owens, J. T., Chmura, A. J., Murakami, K., Fujita, N., Ishihama, A., and Meares, C. F. (1998). Mapping the promoter DNA sites proximal to conserved regions of sigma 70 in an Escherichia coli RNA polymerase-lacUV5 open promoter complex. Biochemistry *37*, 7670-7675.

Tomsic, M., Tsujikawa, L., Panaghie, G., Wang, Y., Azok, J., and deHaseth, P. L. (2001). Different roles for basic and aromatic amino acids in conserved region 2 of Escherichia coli sigma(70) in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. J Biol Chem *276*, 31891-31896.

Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature *417*, 712-719.