

## An Additional Basal Promoter Element Recognized by Free RNA Polymerase $\sigma$ Subunit Determines Promoter Recognition by RNA Polymerase Holoenzyme

Andrey Feklistov,<sup>1,2,3</sup> Nataliya Barinova,<sup>1</sup> Anastasiya Sevostyanova,<sup>1</sup> Ewa Heyduk,<sup>4</sup> Irina Bass,<sup>1</sup> Irina Vvedenskaya,<sup>5</sup> Konstantin Kuznedelov,<sup>5</sup> Eglė Merkienė,<sup>6</sup> Elena Stavrovskaya,<sup>7</sup> Saulius Klimašauskas,<sup>6</sup> Vadim Nikiforov,<sup>1,3</sup> Tomasz Heyduk,<sup>4</sup> Konstantin Severinov,<sup>1,5,\*</sup> and Andrey Kulbachinskiy<sup>1,5,\*\*</sup>

<sup>1</sup>Institute of Molecular Genetics

Moscow 123182

Russia

<sup>2</sup>Department of Molecular Biology

Moscow State University

Moscow 119992

Russia

<sup>3</sup>Public Health Research Institute

Newark, New Jersey 07103

<sup>4</sup>St. Louis University School of Medicine

St. Louis, Missouri

<sup>5</sup>Waksman Institute

Rutgers University

Piscataway, New Jersey 08854

<sup>6</sup>Institute of Biotechnology

Vilnius 2028

Lithuania

<sup>7</sup>Institute for Information Transmission Problems

Moscow

Russia

\*Correspondence: severik@waksman.rutgers.edu (K.S.)

\*\*Correspondence: akulb@img.ras.ru (A.K.)

### Summary

During transcription initiation by bacterial RNA polymerase, the  $\sigma$  subunit recognizes the  $-35$  and  $-10$  promoter elements; free  $\sigma$ , however, does not bind DNA. We selected ssDNA aptamers that strongly and specifically bound free  $\sigma^A$  from *Thermus aquaticus*. A consensus sequence, GTA(C/T)AATGGGA, was required for aptamer binding to  $\sigma^A$ , with the TA(C/T)AAT segment making interactions similar to those made by the  $-10$  promoter element (consensus sequence TATAAT) in the context of RNA polymerase

holoenzyme. When in dsDNA form, the aptamers function as strong promoters for the *T. aquaticus* RNA polymerase  $\sigma^A$  holoenzyme. Recognition of the aptamer-based promoters depends on the downstream GGG motif from the aptamers' common sequence, which is contacted by  $\sigma^A$  region 1.2 and directs transcription initiation even in the absence of the  $-35$  promoter element. Thus, recognition of bacterial promoters is controlled by independent interactions of  $\sigma$  with multiple basal promoter elements.

### Introduction

Initiation of transcription in bacteria is carried out by the RNA polymerase (RNAP) holoenzyme, a complex of the core enzyme (subunit composition  $\alpha_2\beta\beta'\omega$ ) and a specificity subunit,  $\sigma$ . Bacteria contain several  $\sigma$  subunits, each responsible for transcription from different types of promoters. Most promoters of house-keeping genes are recognized by holoenzyme containing the primary  $\sigma$  subunit, called  $\sigma^{70}$  in *Escherichia coli*. Analysis of promoter sequences from *E. coli* and other bacteria identified two elements with consensus sequences TTGACA and TATAAT that are centered, respectively, 35 and 10 base pairs upstream of the transcription start site. The so-called "extended  $-10$ " promoters contain an additional TG motif located one base pair upstream of the  $-10$  element; these promoters often lack the  $-35$  element (Gross et al., 1998).

Primary  $\sigma$  subunits contain four conserved regions, which can be further subdivided (Gross et al., 1998). In the RNAP-promoter complex, the  $-10$ , the extended  $-10$ , and the  $-35$  elements are recognized by  $\sigma$  conserved regions 2.4, 2.5, and 4.2, respectively. RNAP interactions with the  $-10$  element play a crucial role in RNAP-induced promoter melting (Fenton et al., 2000; Gross et al., 1998; Juang and Helmann, 1994; Tomsic et al., 2001). RNAP recognizes the  $-10$  element through base-specific interactions of  $\sigma$  with the nontemplate DNA strand (Roberts and Roberts, 1996). The holoenzyme, but not free  $\sigma$ , is also able to recognize the nontemplate strand of the  $-10$  element in single-stranded oligonucleotides (nontemplate oligos) (Kulbachinskiy et al., 1999; Marr and Roberts, 1997; Savinkova et al., 1988; Young et al., 2001).

Although  $\sigma$  plays the main role in promoter recognition by RNAP, free  $\sigma$  does not recognize promoters or nontemplate oligos (Burgess et al., 1969; Callaci and Heyduk, 1998; Gross et al., 1998). The intrinsic ability of  $\sigma$  to recognize DNA is revealed upon the holoenzyme formation and is due to structural rearrangement(s) in  $\sigma$  induced by the core binding (Callaci et al., 1998, 1999; Callaci and Heyduk, 1998; Gross et al., 1998; Marr and Roberts, 1997). Crystal structures of RNAP holoenzymes from *T. aquaticus* (*Taq*) and *T. thermophilus* revealed that  $\sigma$  contains three domains connected by flexible linkers (Murakami et al., 2002b; Vassilyev et al., 2002). The relative positions of these domains in free  $\sigma$  and in the holoenzyme differ. In particular, binding of  $\sigma^{70}$  to RNAP core is accompanied by a large increase in the interdomain distance between  $\sigma$  regions 2 and 4 (Callaci et al., 1998). A short fragment of the  $\beta'$  subunit corresponding to the major  $\sigma$  binding site on the core enzyme induces a similar change in  $\sigma^{70}$ , allowing it to recognize the  $-10$  element and initiate promoter melting (Young et al., 2001, 2004). However, the detailed nature of these conformational changes remains unknown.

Here, we used systematic evolution of ligands by exponential enrichment (SELEX) (Gold et al., 1995) to study DNA binding by free *Taq*  $\sigma^A$ . We obtained ssDNA aptamers that bound  $\sigma^A$  with high affinity and exquisite specificity. The aptamers contain a  $-10$  element-like sequence followed by an additional, previously unidentified four-nucleotide motif that is contacted by  $\sigma^A$  region 1.2 and is required for binding. Binding of aptamers causes a significant increase in the interdomain distance between  $\sigma^A$  regions 2 and 4. Thus, the interaction of free  $\sigma^A$  with aptamers simulates two crucial processes in promoter recognition, i.e., conformational changes in  $\sigma$  and specific interactions with DNA. We further demonstrate that the additional aptamer motif functions as a basal promoter element and allows promoter recognition by *Taq* RNAP holoenzyme in the absence of the  $-35$  promoter element. Similar promoter elements may govern promoter recognition in other bacteria.

## Results

### Structural Features of Aptamers Specific to Free *Taq* $\sigma^A$

ssDNA aptamers interacting with *Taq*  $\sigma^A$  were selected from a library of 75 nt-long ssDNA containing a 32 nt random central region (Figure 1Afig1). Like *E. coli*  $\sigma^{70}$ , free *Taq*  $\sigma^A$  does not interact with double-stranded promoters or nontemplate oligos (data not shown).

Therefore, as expected, the initial ssDNA pool exhibited poor  $\sigma^A$  binding ( $K_d > 10 \mu\text{M}$ ). Selection of aptamers was performed at  $25^\circ\text{C}$ ; in each round of selection, the library was allowed to bind to immobilized  $\sigma^A$ , and bound DNA was eluted, amplified, and used in the next round of selection. After 12 rounds, the affinity of the enriched library significantly increased (apparent  $K_d \sim 10 \text{ nM}$ ). The enriched library was cloned, and sequences of 33 clones were determined. Representative sequences are shown in Figure 1B. Individual aptamers bound *Taq*  $\sigma^A$  with high affinity ( $K_d \sim 7\text{--}14 \text{ nM}$ ) but did not bind *E. coli*  $\sigma^{70}$  or  $\sigma^A$  from *Deinococcus radiodurans* ( $K_d > 1 \mu\text{M}$ ). We therefore named these aptamers sTaps (for sigma *Taq* aptamers).

All sequenced sTaps contain a 11 nt long conserved region with a consensus sequence, GTA(C/T)AATGGGA (Figures 1B and 1C). In most aptamers, this sequence is located close to the center of the randomized region of the initial library. The underlined sequence is similar to the primary  $\sigma -10$  promoter element consensus sequence (TATAAT) derived from the analysis of housekeeping bacterial promoters (Gross et al., 1998). The G immediately upstream of the  $-10$ -like sequence and the GGA tetranucleotide located downstream of it do not correspond to known promoter elements. The presence of the  $-10$  promoter motif in the aptamers suggests that *Taq*  $\sigma^A$  recognizes the same  $-10$  element consensus sequence as the *E. coli*  $\sigma^{70}$ . The distribution of individual nucleotides within the consensus sTap sequence is shown in Figure 1C. Nucleotides at the first, second, and last positions of the TA(C/T)AAT motif are identical in all aptamers. Other positions are less conserved. In particular, more than half of sTaps contain a C instead of a T found at the third position of the  $-10$ - consensus element. Interestingly, similar preference for a C at this position was observed among *E. coli* RNAP binding sequences selected from a pool of promoter DNA with the randomized  $-10$  element sequence (Xu et al., 2001).

One of the aptamers, sTap1, was studied further. Several shortened sTap1 variants were designed and tested for interaction with  $\sigma^A$ . The shortest sTap1 variant that bound  $\sigma^A$  with high affinity was 35 nt long and contained the entire 11 nt conserved region. A putative secondary structure of the minimal sTap1—a short imperfect hairpin with a 4 nt loop—is shown in Figure 1D. The first four

nucleotides of the  $-10$ -like element are single stranded, whereas the last two nucleotides base pair with the 3' part of the aptamer. Other aptamers can be folded into similar structures (data not shown). The unpaired nucleotides within the hairpin are essential for  $\sigma^A$  binding because substitutions that increase complementarity within the hairpin impair  $\sigma$  binding (data not shown).

Probing sTap1 with  $\text{KMnO}_4$ , an agent specific for single-stranded thymines, supported the proposed sTap1 structure (Figure 1D and Figure S1 available in the Supplemental Data with this article online). When the aptamer was incubated with  $\sigma^A$  at  $20^\circ\text{C}$ , a condition when efficient binding occurs, the first two thymines from the conserved region became protected from  $\text{KMnO}_4$ , indicating formation of close DNA-protein contacts. In contrast, a thymine in the loop became hyperreactive, apparently as a result of structural distortions caused by  $\sigma^A$  binding (Figure 1D and Figure S1). No changes in the aptamer structure were observed when the experiment was repeated at  $70^\circ\text{C}$ , indicating that  $\sigma^A$  does not bind the aptamer at high temperature.

To determine the sites of sTap1 that are in close contact with  $\sigma^A$ , we performed hydroxyl radical footprinting. Almost every position of the minimal sTap1 aptamer was protected, with two adjacent adenines from the  $-10$ -like element and several 3' end-proximal nucleotides being notable exceptions. The strongest protection was observed for the second A and the last T of the  $-10$ -like element, for the GGG A element, and for several nucleotides in the 3' end-proximal half of the aptamer (Figure 1D and Figure S2). These nucleotides likely form very tight contacts with  $\sigma^A$ .

### The $-10$ -like Element in Aptamers Is Recognized by Region 2 of $\sigma^A$

We investigated the role of individual elements within the conserved sTap region in aptamer binding by studying several variants of minimal sTap1. In the first variant, the upstream G nucleotide was replaced with C, in the second, the  $-10$ -like TATAAT sequence was replaced with ATCTCG (an "anticonsensus" sequence; [Marr and Roberts, 1997]), and in the third, the GGG A motif was replaced with CCCT. None of the mutant aptamers bound  $\sigma^A$  ( $K_d > 4 \mu\text{M}$ ), and therefore, all nucleotides from the conserved region are important for aptamer recognition by  $\sigma^A$ .

Sequence similarity between the sTap  $-10$ -like element and the  $-10$  promoter consensus element suggests that they make identical contacts with  $\sigma$  in binary  $\sigma$ -sTap complex

and the open RNAP-promoter complex, respectively. A nontemplate adenine at the second position of the  $-10$  promoter element is critical for promoter opening, and substitution of this base with 2-aminopurine inactivates a model promoter (Lim et al., 2001). The  $\sigma^A$  affinity of sTap1 mutant harboring a corresponding substitution was drastically reduced ( $K_d > 4 \mu\text{M}$ ). Thus, adenine at the second position of the  $-10$ -like element is essential for efficient aptamer recognition by  $\sigma^A$ .

The primary role of *E. coli*  $\sigma^{70}$  region 2.4 in recognition of the  $-10$  element was initially revealed in genetic screens for  $\sigma^{70}$  mutants that suppressed substitutions in the  $-10$  element (Gross et al., 1998). Specifically, it was shown that a holoenzyme carrying  $\sigma^{70}$  with a Gln to His change at position 437 efficiently transcribed from a promoter with a T to C downmutation at position  $-12$  (Waldburger et al., 1990). We generated a homologous Gln  $\rightarrow$  His substitution in  $\sigma^A$  position 260 (Q260H), which corresponds to the  $\sigma^{70}$  position 437. Control experiments demonstrated that the mutation suppressed a T to C substitution at the first position of the  $-10$  promoter element during transcription with the *Taq*  $\sigma^A$  RNAP holoenzyme (data not shown). We then tested the ability of the mutant  $\sigma^A$  to recognize an sTap1 variant bearing a T  $\rightarrow$  C substitution at the first position of the  $-10$ -like element (Figure 2fig2). The Q260H  $\sigma^A$  bound the wild-type sTap1 with the same affinity as the wild-type  $\sigma^A$ . At the same time, its affinity for the mutant aptamer was higher ( $K_d \sim 100 \text{ nM}$ ) than the wild-type  $\sigma^A$  affinity ( $K_d > 1 \mu\text{M}$ ). The effect of the Q260H substitution was position specific because it did not suppress changes at other positions of sTap1 (data not shown). Thus, Q260H is a "relaxed-specificity" mutation that allows efficient recognition of various nucleotides at the first position of the  $-10$ -like element of sTap. The identical effects of Q260H substitution on the recognition of mutant promoters and sTaps argue strongly that the  $-10$ -like element of sTaps and the  $-10$  element of promoters interact with  $\sigma$  in the same way.

### The GGG A Motif in Aptamers Is Recognized by Region 1.2 of $\sigma^A$

To identify  $\sigma^A$  regions that interact with the GGG A motif of aptamers, we performed crosslinking with sTap1 bearing a photoactivatable crosslinkable 6-

thio-guanine at the second position of the GGGA motif. Control experiments demonstrated that the modification did not interfere with  $\sigma^A$  binding. Irradiation of the  $\sigma^A$ -aptamer complex with uv light (365 nm) resulted in formation of a crosslink between the aptamer and  $\sigma^A$  with high (2%) efficiency (Figure 3Afig3, lane 1). The crosslink was specific because no crosslink was formed in the presence of an excess of unmodified sTap1 (lane 2), whereas a control oligo of the same length did not affect the crosslink (lane 3).

Available models of open promoter complex (Murakami et al., 2002a) indicate that nontemplate DNA downstream of the  $-10$  promoter element may be contacted by conserved region 1.2 (amino acids 88–125) of  $\sigma^A$  (see Figure 7 and also the Discussion). To test whether this region is involved in the recognition of the GGGA motif, we created two  $\sigma^A$  derivatives. The first derivative lacked N-terminal amino acids 1–91 (corresponding to conserved region 1.1); the second, S132C, contained a substitution of serine at position 132 by cysteine (note that there are no cysteines in the wild-type  $\sigma^A$  sequence).

The 92–438  $\sigma^A$  fragment bound the aptamer with the same affinity ( $K_d$  5 nM) and crosslinked to derivatized sTap1 as efficiently as the wild-type  $\sigma^A$  (Figure 3A, lane 4 and data not shown). Thus, region 1.1 of  $\sigma^A$  is not required for aptamer recognition, and the crosslink must be located C-terminal of  $\sigma^A$  amino acid 92. The S132C mutant also efficiently crosslinked to derivitized sTap1 (Figure 3A, lane 5). Treatment of crosslinked complexes with NTCBA, a chemical protease that specifically cleaves polypeptide chains at cysteines, resulted in the appearance of an additional radioactive band on SDS gels, which migrated faster than the marker  $\sigma^A$  fragments (amino acids 92–438 and 1–390) (Figure 3B). This indicated that the labeled band corresponds to N-terminal cleavage product (calculated molecular weight 28 kDa) and that the crosslink is located N-terminal of  $\sigma^A$  residue 132. Thus, the data position the crosslink between amino acids 92 and 132 of  $\sigma^A$ . We therefore conclude that the GGGA motif is contacted by  $\sigma^A$  region 1.2.

### Aptamers Induce a Conformational Change in $\sigma^A$

Distance measurements using luminescence resonance energy transfer (LRET) technique revealed large conformational changes in *E. coli*  $\sigma^{70}$  upon binding to RNAP core (Callaci et al., 1999). Presumably, these changes are required for promoter binding by  $\sigma$ . Because sTaps bind  $\sigma^A$  with high affinity in the absence of the core, we tested whether sTaps could induce conformational

changes in  $\sigma^A$ . To this end, a  $\sigma^A$  mutant containing a pair of cysteines in regions 2 and 4 was created. The resultant double mutant  $\sigma^A$  (N263C, G406C) was fully functional (data not shown). It was double labeled with luminescence donor and acceptor probes and used for LRET measurements. The results are shown in Figure 4fig4 and Figure S3. The apparent interdomain distance between regions 2 and 4 in free  $\sigma^A$  was similar to that measured in free  $\sigma^{70}$  (Callaci et al., 1999). The addition of sTap1 increased the apparent distance between regions 2 and 4. However, the increase was less than that caused by the core binding (Figure 4). A control nontemplate oligo containing the TATAAT sequence but lacking other specific aptamer sequences did not cause any changes in the interdomain distance (data not shown). Thus, the binding of sTap to  $\sigma^A$  induces a conformational rearrangement of the protein, which may unmask the DNA binding sites of  $\sigma$  (see Discussion).

The distance between the two domains of  $\sigma^A$  measured by LRET in the presence of core (53 Å) was somewhat shorter than the distance between the corresponding residues of  $\sigma$  observed in the *Taq* holoenzyme structure (68.9 Å, see Figure 7) (Murakami et al., 2002a). This distance most likely corresponds to the distance of the closest approach of donor and acceptor fluorophores in the course of LRET measurement, which is likely to be shorter than the distance observed in the structure due to local and segmental motions of the protein in the solution (Callaci et al., 1999).

### Double-Stranded DNA Fragments Based on $\sigma^A$ Aptamers Function as Promoters for $\sigma^A$ Holoenzyme

The apparent similarity of interactions between  $\sigma^A$  and the  $-10$  consensus element in aptamers and natural promoters suggested that sTap sequences, when present in double-stranded form, may function as promoters. To test this hypothesis, we prepared double-stranded DNA fragments containing sequences of several aptamers and determined their ability to support transcription in vitro (Figure 5Afig5). As can be seen, every sTap-based DNA fragment tested supported robust synthesis of a single major RNA product by *Taq* holoenzyme (Figure 5B). Primer extension indicated that, in each case, transcription initiated six to seven nucleotides downstream of the aptamer's  $-10$  element-like element, demonstrating

that this element in fact functions as a bona fide  $-10$  element (data not shown). However, double-stranded DNA sTap fragments were completely inactive when the *E. coli*  $\sigma^{70}$  RNAP holoenzyme was used instead of the *Taq*  $\sigma^A$  holoenzyme. This is an unexpected result, because heretofore no differences in promoter recognition by the *E. coli*  $\sigma^{70}$  and *Taq*  $\sigma^A$  RNAP holoenzymes have been reported (Minakhin et al., 2001; Xue et al., 2000).

The sTap-based promoters contain a  $-10$  element but no recognizable  $-35$  element. The absence of apparent homology between sequences surrounding the 11 nt-conserved aptamer region in sTap-based promoters (see Figures 1A and 5A) suggested that this region alone was responsible for promoter function. Several derivatives of one sTap-based promoter, sTap2, were created to determine (1) elements responsible for efficient recognition by *Taq*  $\sigma^A$  holoenzyme and (2) elements that need to be introduced to allow recognition by the *E. coli*  $\sigma^{70}$  holoenzyme. The sTap2 promoter contains a TG-dinucleotide upstream of the  $-10$  element that is similar to the extended  $-10$  promoter motif but is not separated from the  $-10$  element by an additional nucleotide. In an sTap2 derivative called “-TG,” the TG dinucleotide was substituted by AC (Figure 5A). In the “-GGGA” derivative, the downstream GGGA motif was substituted by CCCT. The sTap2 derivative named “+35” had a consensus  $-35$  element introduced 18 bp upstream of the  $-10$  element. In the “+35 -TG” derivative, the  $-35$  element was present but the TG dinucleotide was substituted by AC. Finally, the “+35 -GGGA” promoter was a derivative of the +35 template with a CCCT instead of the GGGA motif. Equal amounts of sTap2 promoter fragment or its derivatives were used in a multiple-round transcription assay with *Taq*  $\sigma^A$  or *E. coli*  $\sigma^{70}$  holoenzymes. The substitution for the TG motif significantly decreased transcription by *Taq* RNAP (Figure 5C, top, lanes 1 and 2). This indicates that TG can be specifically recognized by *Taq* RNAP despite its position being different from the position in the extended  $-10$  promoters. The substitution for the GGGA motif abolished activity of the sTap promoter (Figure 5C, top). *E. coli* RNAP was inactive at all three promoters (Figure 5C, bottom, lanes 1–3). The introduction of the  $-35$  element led only to a small stimulation of transcription by the *Taq* enzyme but strongly stimulated transcription by *E. coli* RNAP (Figure 5C, lane 4). When the  $-35$  promoter element was present, removal of the TG motif had little effect on transcription by either enzyme. Removal of the GGGA motif in the presence of the  $-35$  element had a slight stimulatory effect on *E. coli* RNAP but inhibited transcription by the *Taq* enzyme. Thus, our results demonstrate that the GGGA motif located downstream of the  $-10$  element allows promoter recognition and/or open

promoter complex formation by *Taq* RNAP when both the  $-35$  element and the TG motif are absent. In the case of *E. coli* RNAP transcription, the  $-35$  element is strictly required for promoter utilization, whereas GGGA plays no role or may be slightly inhibitory.

KMnO<sub>4</sub> probing revealed that the lack of transcription from sTap2 or its derivatives was due to the absence of open promoter complex formation (Supplemental Data and Figure S4A). At the same time, gel retardation experiments demonstrated that removal of the GGGA motif in the -GGGA promoter did not impair closed complex formation by *Taq*  $\sigma^A$  holoenzyme (Supplemental Data and Figure S4B). It is therefore likely that this motif is not essential for initial promoter binding but is important at later stages of the open complex formation.

To further understand the role of the GGGA motif, additional point mutations were created and tested in an in vitro transcription assay with *Taq* RNAP (Figure 5D and Figure S5). The GGGA motif was not sufficient for promoter utilization in the absence of the  $-10$  element, because changing the  $-10$  element to the anticonsensus sequence (in the “anti-10” derivative) abolished promoter function even when the GGGA motif was intact. Within the GGGA motif, every position was important, because substitution of any base pair with its complement (in the “-6C”, “-5C”, “-4C”, and “-3T” derivatives) strongly decreased transcription. However, the effect of the downstream A:T base pair substitution was less drastic. As the point mutations did not change the G/C content of the GGGA region, the effect of the mutations was likely due to changes in specific RNAP-DNA contacts rather than to changes in stability of the DNA duplex.

To reveal the functional role of the GGGA element, we compared stability of promoter complexes formed by *Taq* RNAP on sTap2, the +35, and +35 -GGGA promoters. Promoter complexes of *Taq* RNAP rapidly dissociate in the presence of DNA competitors such as heparin, making direct measurements of open complex half-life impossible (Fabry et al. [1976], Kulbachinskiy et al. [2004a], Kuznedelov et al. [2003], Schroeder and deHaseth [2005], and data not shown). We therefore tested the stability of promoter complexes by performing transcription at different ionic strengths. At low ionic strength (40 mM KCl), all

three promoters were efficiently utilized by *Taq* RNAP, as expected (Figure 5E, top). However, in the presence of 130 mM KCl, only the +35 promoter was active (Figure 5E, bottom). Thus, neither the GGGGA motif nor the –35 element on their own is sufficient for transcription at elevated salt concentrations. However, when both elements are present, efficient promoter utilization in high salt becomes possible.

To test whether the sTap-based promoters can be recognized by RNAPs from other bacteria, we studied transcription on the sTap2 promoter by RNAPs from *B. subtilis*, *T. thermophilus*, and *D. radiodurans*. *B. subtilis* RNAP was active on the +35 promoter but was inactive on the wild-type sTap2 promoter (Figure 5F). At the same time, RNAPs from *T. thermophilus* and *D. radiodurans* were active on both promoters (Figure 5F and data not shown). The fact that the sTap2 promoter is efficiently recognized by RNAPs from *T. aquaticus*, *T. thermophilus*, and *D. radiodurans*, which is the closest mesophilic relative of *Thermus*, indicates that the ability to utilize this type of promoters is not related to thermal adaptation but may be specific for *Thermus/Deinococcus* group of bacteria.

### Natural *Taq* Promoters Contain the GGGGA Motif

The question arises whether the GGGGA motif discovered through in vitro analysis is relevant for understanding of promoter utilization in *Thermus* and related bacteria. Analysis of annotated *Thermus* promoters reveals that some of them contain sequences similar to the GGGGA motif immediately downstream of –10 elements (Figure 6Afig6 and Supplemental Data). Preliminary analysis of intergenic sequences from the genome of *Thermus thermophilus* also suggests that sequences immediately downstream of –10 elements of predicted promoters are enriched with guanines in the nontemplate strand (see Supplemental Data for details). To show that in natural *Thermus* promoters the GGGGA motif has a function similar to that in sTap-based promoters, we analyzed transcription from *T. thermophilus dnaK* promoter that contains the –35 element and the GGGGA motif (Osipiuk and Joachimiak, 1997). The *dnaK* promoter was efficiently utilized by *Taq*  $\sigma^A$  holoenzyme as judged by its higher activity compared to the activity of a control T7 A1 promoter (Figure 6B, lanes 1 and 5). In contrast, the *E. coli*  $\sigma^{70}$  holoenzyme was poorly active on *dnaK* but was highly active on T7 A1 (Figure 6B, lanes 6 and 10). Several derivatives of *dnaK* similar to the sTap2 promoter derivatives described above were created and tested for their ability to support transcription by both holoenzymes (Figure 6B and Figure S5). As can be seen, substitution of the GGGGA motif with CCCT in the –GGGA

derivative dramatically decreased the efficiency of promoter utilization by *Taq* RNAP, and substitution of both –35 and GGGGA elements (“–GGGA –35”) completely inactivated the promoter. At the same time, the mutant lacking the –35 element (“–35”) was transcriptionally active, though the level of activity was strongly reduced (Figure 6B, lanes 2–4). Thus, the GGGGA motif of the *dnaK* promoter can direct transcription initiation by *Taq* RNAP even in the absence of the –35 element. Substitution of the GGGGA motif did not cause any changes in the efficiency of transcription by *E. coli* RNAP, whereas substitution of the –35 element decreased the promoter activity to undetectable levels.

To test whether the GGGGA motif can increase the efficiency of transcription by *Taq*  $\sigma^A$  holoenzyme on other promoters, we created two variants of the T7 A1 promoter in which four nucleotides immediately downstream of the –10 element (TACT) were replaced with GGGGA or CCCT (Figure S5). These substitutions had opposite effects on promoter utilization by *Taq* RNAP. Whereas the presence of GGGGA significantly stimulated RNAP activity, the CCCT substitution had an inhibitory effect on transcription (Figure 6C, lanes 1–3). Both mutations had only minor effect on the activity of *E. coli* RNAP (Figure 6C, lanes 4–6).

### Discussion

In this work, we show for the first time, to our knowledge, that a primary RNA polymerase  $\sigma$  subunit can bind DNA with high affinity and specificity. Analysis of the mechanism of aptamer binding demonstrated that *Taq*  $\sigma^A$  recognizes a continuous 11 nt long DNA region containing a –10 element sequence flanked by additional conserved motifs. Several lines of evidence suggest that recognition of the –10-like sequence in the context of aptamers by free  $\sigma$  is very similar to recognition of the –10 element by RNAP holoenzyme and that, therefore,  $\sigma$  on its own specifically recognizes all six nucleotides of the –10 element. First, the frequency profile of individual bases within the sTap –10-like element corresponds to that observed in natural promoters of *E. coli* and *T. thermophilus*, with nucleotides located at positions corresponding to the –10 element positions –12, –11, and –7 being the most

conserved (Harley and Reynolds, 1987; Maseda and Hoshino, 1995). Second, mutations in the  $-10$  element have a dramatic effect on both the recognition of promoters by RNAP holoenzyme and the interaction of aptamers with  $\sigma$ . Third, a T  $\rightarrow$  C mutation at the first position of the  $-10$ -like element of sTap1 is suppressed by the Q260H substitution of *Taq*  $\sigma^A$ . An identical substitution at the corresponding position of *E. coli*  $\sigma^{70}$  (Q437H) allows RNAP holoenzyme to recognize promoters and single-stranded nontemplate oligos containing a T  $\rightarrow$  C mutation at the first position of the  $-10$  element. Therefore, the same amino acid residue of  $\sigma$  interacts with the first nucleotide of the  $-10$  element in promoters, nontemplate oligos, and sTaps. Thus, the simple  $\sigma$ -aptamer complex is a good model for functional and structural understanding of promoter recognition.

Previous studies demonstrated that a conformational change in  $\sigma$  induced by the core polymerase is required for efficient recognition of promoters or nontemplate oligos. As aptamers obtained in this work are specifically recognized by free  $\sigma^A$ , we propose that they induce conformational changes in  $\sigma$  that may be in some aspect similar to those induced by RNAP core and thus allow specific interaction with DNA. Indeed, direct measurements of interdomain distances demonstrated that aptamer binding increases the distance between  $\sigma^A$  regions 2 and 4, although to a lesser extent than core RNAP. This indicates that region 4 may mask DNA binding sites in free  $\sigma$  and that it must be moved away from region 2 to allow DNA recognition.

The principal result of this work is the demonstration that a tetranucleotide motif, GGGGA, downstream of the  $-10$  element plays an essential role in promoter recognition by *Taq*  $\sigma^A$  holoenzyme. The GGGGA motif (together with the  $-10$  element) allows transcription even in the absence of the  $-35$  element or the extended  $-10$  motif. Analysis of known *Thermus* promoters revealed that some of them contain such a motif and that it has an important role in promoter function. In particular, a substitution of the GGGGA motif in the *dnaK* promoter leads to a dramatic decrease in promoter activity. Conversely, introduction of the GGGGA motif in the T7 A1 promoter strongly stimulates transcription by *Taq* RNAP but has only a small effect on the activity of *E. coli* RNAP.

Using site-specific DNA-protein crosslinking, we demonstrated that, in a complex of the sTap1 aptamer with  $\sigma^A$ , nucleotides from the GGGGA motif interact with  $\sigma^A$  region 1.2. Apparent homology of recognition of the  $-10$  element in aptamers by free  $\sigma$  and in promoters by RNAP holoenzyme suggests that recognition of the GGGGA motif in promoters also occurs through specific interactions between region 1.2 of  $\sigma^A$  and the nontemplate DNA strand.

Indeed, DNA-protein crosslinking experiments revealed close contacts between  $\sigma^{70}$  and nontemplate DNA between the  $-10$  element and the transcription start point in the *lacUV5* promoter open complex (Brodolin et al., 2000; Naryshkin et al., 2000; Simpson, 1979). Analysis of the structure of *Taq* RNAP  $\sigma^A$  holoenzyme complex with a fork-junction promoter DNA template also suggests that nontemplate strands downstream of the  $-10$  element (not present in the structure) may interact with  $\sigma$  conserved region 1.2 (Murakami et al., 2002a). In a model of open promoter complex based on this structure, the closest distance between this DNA segment and  $\sigma$  region 1.2 is less than 4 Å (Figure 7fig7) (Murakami et al., 2002a). Interestingly, in a model of open promoter complex proposed by Artsimovitch and coauthors (Artsimovitch et al., 2004), the corresponding segment of the nontemplate strand is placed too far from  $\sigma$  to make direct contacts. Thus, the  $\sigma$ -aptamer complex and RNAP-fork-junction complex may correspond to an intermediate of the open complex formation pathway, and the position of the nontemplate strand downstream of the  $-10$  element may change as RNAP contacts with DNA downstream of the transcription start point develop.

The position of the GGGA motif coincides with the position of G/C rich “discriminator” found in some *E. coli* promoters that are regulated by stringent response (Travers, 1980). However, whereas mutations in the discriminator sequence were shown to increase stability of open promoter complex (Gourse et al., 1998; Lamond and Travers, 1985; Pemberton et al., 2000; Haugen et al., 2006), mutations in the GGGA motif have an opposite effect and destabilize the complex. It was proposed that discriminator acts by increasing the stability of the DNA duplex and thus destabilizing the open complex (Gourse et al., 1998; Jung and Lee, 1997; Lamond and Travers, 1985; Pemberton et al., 2000). Our data suggest that the overall stability of the open complex may also depend on direct contacts of  $\sigma$  with DNA between the  $-10$  element and the transcription start point. In this case, the destabilizing effect of discriminator may be explained by the absence of favorable  $\sigma^{70}$ -DNA contacts that facilitate promoter melting. In fact, Gourse and colleagues (Haugen et al., 2006) have demonstrated that a base on the nontemplate strand two nucleotides downstream of the  $-10$  hexamer crosslinks to *E. coli*  $\sigma^{70}$  region 1.2; the identity of this base affects the lifetime of the promoter

complex with RNAP in vitro and regulation of rRNA promoter activity in vivo.

Interestingly, mutations downstream of a  $-10$ -like element sequence that induces a  $\sigma^{70}$ -dependent pause during transcription elongation were shown to decrease pausing efficiency (Ring and Roberts, 1994; Ring et al., 1996). The finding suggests that the  $\sigma^{70}$  subunit might also specifically recognize sequences downstream of the  $-10$  element during transcription elongation. Downstream elements of varying sequence may be present in promoters and regulatory pause sequences from different bacteria and may control promoter complex formation and promoter-proximal pausing through specific interaction with  $\sigma$  region 1.2. Selection of aptamers specifically interacting with  $\sigma$  subunits from different bacteria, coupled with bioinformatic searches of bacterial genome sequences, could be a powerful approach for uncovering such elements.

## Experimental Procedures

### Proteins

Wild-type and mutant *Taq*  $\sigma^A$  and *E. coli*  $\sigma^{70}$ , and *Taq* and *E. coli* holoenzyme RNAPs were obtained by standard procedures, as described in the Supplemental Data.

### Selection of Aptamers to *Taq* $\sigma^A$

The selection was performed at 25°C essentially as described previously (Kulbachinskiy et al., 2004b). Determination of equilibrium  $K_d$  values for binding of aptamers to RNAP was done either by using the nitrocellulose binding method or by measuring fluorescence quenching of a 5'-fluorescein-labeled sTap1 aptamer. Detailed description of the procedures can be found in the Supplemental Data.

### Crosslinking Experiments

Modified sTap1 containing 6-thio-guanine at the second position of the GGGA element was purchased from Oligos Etc. (Wilsonville, OR). Crosslinking experiments and cysteine-specific cleavage of  $\sigma$  were performed as described in the Supplemental Data.

### LRET Distance Measurements

The double-cysteine mutant of  $\sigma^A$  was labeled with europium chelate and Cy5 and purified as previously described for  $\sigma^{70}$  (Callaci et al., 1999). Distance measurements were conducted as described in the Supplemental Data (Heyduk and Heyduk, 2002).

## In Vitro Transcription

Double-stranded DNA fragments containing promoter sequences were prepared by PCR as described in the Supplemental Data. The sequences of all studied promoters are shown in Figure S5. Transcription was performed in a buffer containing 40 mM HEPES (pH 7.9), 40 mM KCl, and 10 mM MgCl<sub>2</sub>. Holoenzymes were prepared by incubating the core enzyme (100 nM) with  $\sigma$  (500 nM) for 5 min at room temperature. Promoter DNA fragment was added to 20 nM, and the samples were incubated for 10 min at either 65°C (for *Taq* RNAP) or 37°C (for *E. coli*, *D. radiodurans*, and *B. subtilis* RNAPs). Samples were supplemented with 25  $\mu$ M of each rNTPs (in run-off transcription experiments) or CpA primer (10  $\mu$ M) and UTP (25  $\mu$ M) (in abortive transcription assays) and incubated for 10 min at the same temperature. In both cases, RNA products were labeled by the addition of 2.5  $\mu$ Ci of  $\alpha$ -[<sup>32</sup>P]-UTP (3000 Ci/mmol, NEN). The reactions were terminated by the addition of 10  $\mu$ l of 8 M urea, 20 mM EDTA, 2 $\times$  TBE, and the products were separated by 15% denaturing PAGE and analyzed by PhosphorImager (Amersham BioSciences).

## Supplemental Data

Supplemental Data include Supplemental Results, Supplemental Experimental Procedures, Supplemental References, and six figures and can be found at the end of this article.

## Acknowledgments

This work was supported by the Russian Foundation for Basic Research grant 05-04-49405, National Institutes of Health (NIH) R01 grant GM64530 (to K.S.), NIH grant GM30717 (to A. Goldfarb), NIH grant GM50514 (to T.H.), by the grant of the President of Russian Federation MK-952.2005.4 and the Busch Fellowship from the Waksman Institute (to A.K.), and an International research fellowship grant from the Howard Hughes Medical Institute (to S.K.). A.K.'s work at the Institute of Biotechnology in Vilnius was supported by a travel grant from the U.S. National Academy of Sciences with partial support from the Howard Hughes Medical Institute. We thank S. Borukhov for help with mapping of  $\sigma$ -aptamer contacts, N. Zenkin and A. Mustaev for help with crosslinking experiments, L. Minakhin for proteins,

S. Darst for coordinates of the open promoter complex model, M. Gelfand for help with bioinformatic analysis, V.K. Potapov for oligonucleotide synthesis, S. Nechaev, I. Artsimovitch, A. Hochschild, and P. Geiduschek for reading the manuscript, and R. Gourse for sharing his results prior to publication. A.K. is grateful to all members of the S. Klimašauskas laboratory for their help during the work.

Received: January 14, 2006

Revised: April 11, 2006

Accepted: June 15, 2006

Published online: June 22, 2006

## References

- Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M.N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassylyev, D.G. (2004). Structural basis for transcription regulation by alarmone ppGpp. *Cell* 117, 299–310.
- Brodolin, K., Mustaev, A., Severinov, K., and Nikiforov, V. (2000). Identification of RNA polymerase beta' subunit segment contacting the melted region of the *lacUV5* promoter. *J. Biol. Chem.* 275, 3661–3666.
- Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* 221, 43–46.
- 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.
- 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.
- Callaci, S., Heyduk, E., and Heyduk, T. (1999). Core RNA polymerase from *E. coli* induces a major change in the domain arrangement of the sigma 70 subunit. *Mol. Cell* 3, 229–238.
- Fabry, M., Sumegi, J., and Venetianer, P. (1976). Purification and properties of the RNA polymerase of an extremely thermophilic bacterium: *Thermus aquaticus* T2. *Biochim. Biophys. Acta* 435, 228–235.
- 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.
- Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995). Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* 64, 763–797.
- Gourse, R.L., Gaal, T., Aiyar, S.E., Barker, M.M., Estrem, S.T., Hirvonen, C.A., and Ross, W. (1998). Strength and regulation without transcription factors: lessons from bacterial rRNA promoters. *Cold Spring Harb. Symp. Quant. Biol.* 63, 131–139.
- 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.
- Harley, C.B., and Reynolds, R.P. (1987). Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* 15, 2343–2361.
- Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006). rRNA promoter regulation by nonoptimal binding of  $\sigma$  region 1.2: an additional recognition element for RNA polymerase. *Cell*, in press.
- Heyduk, E., and Heyduk, T. (2002). Conformation of fork junction DNA in a complex with *E. coli* RNA polymerase. *Biochemistry* 41, 2876–2883.
- Juang, Y.L., and Helmann, J.D. (1994). A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids. *J. Mol. Biol.* 235, 1470–1488.
- Jung, Y.H., and Lee, Y. (1997). *Escherichia coli* rnpB promoter mutants altered in stringent response. *Biochem. Biophys. Res. Commun.* 230, 582–586.
- 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.
- Kulbachinskiy, A., Bass, I., Bogdanova, E., Goldfarb, A., and Nikiforov, V. (2004a). Cold sensitivity of thermophilic and mesophilic RNA polymerases. *J. Bacteriol.* 186, 7818–7820.
- Kulbachinskiy, A., Feklistov, A., Krashennikov, I., Goldfarb, A., and Nikiforov, V. (2004b). Aptamers to *Escherichia coli* core RNA polymerase that sense its interaction with rifampicin, sigma-subunit and GreB. *Eur. J. Biochem.* 271, 4921–4931.
- Kuznedelov, K., Minakhin, L., and Severinov, K. (2003). Preparation and characterization of recombinant *Thermus aquaticus* RNA polymerase. *Methods Enzymol.* 370, 94–108.

- Lamond, A.I., and Travers, A.A. (1985). Genetically separable functional elements mediate the optimal expression and stringent regulation of a bacterial tRNA gene. *Cell* 40, 319–326.
- Leontiadou, F., Triantafyllidou, D., and Choli-Papadopoulou, T. (2001). On the characterization of the putative S20-thx operon of *Thermus thermophilus*. *Biol. Chem.* 382, 1001–1006.
- Lim, H.M., Lee, H.J., Roy, S., and Adhya, S. (2001). A “master” in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc. Natl. Acad. Sci. USA* 98, 14849–14852.
- Marr, M.T., and Roberts, J.W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* 276, 1258–1260.
- Maseda, H., and Hoshino, T. (1995). Screening and analysis of DNA fragments that show promoter activities in *Thermus thermophilus*. *FEMS Microbiol. Lett.* 128, 127–134.
- Minakhin, L., Nechaev, S., Campbell, E.A., and Severinov, K. (2001). Recombinant *Thermus aquaticus* RNA polymerase, a new tool for structure- based analysis of transcription. *J. Bacteriol.* 183, 71–76.
- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296, 1285–1290.
- Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296, 1280–1284.
- 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.
- Osipiuk, J., and Joachimiak, A. (1997). Cloning, sequencing, and expression of dnaK-operon proteins from the thermophilic bacterium *Thermus thermophilus*. *Biochim. Biophys. Acta* 1353, 253–265.
- Pemberton, I.K., Muskhelishvili, G., Travers, A.A., and Buckle, M. (2000). The G+C-rich discriminator region of the tyrT promoter antagonises the formation of stable preinitiation complexes. *J. Mol. Biol.* 299, 859–864.
- 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.
- Sato, S., Nakada, Y., Kanaya, S., and Tanaka, T. (1988). Molecular cloning and nucleotide sequence of *Thermus thermophilus* HB8 trpE and trpG. *Biochim. Biophys. Acta* 950, 303–312.
- Savinkova, L.K., Baranova, L.V., Knorre, V.L., and Salganik, R.I. (1988). Binding of RNA-polymerase from *E. coli* with oligodeoxyribonucleotides homologous to transcribed and non-transcribed DNA stands in the -10-promoter region of bacterial genes. *Mol. Biol. (Mosk.)* 22, 807–812.
- Schroeder, L.A., and deHaseth, P.L. (2005). Mechanistic differences in promoter DNA melting by *Thermus aquaticus* and *Escherichia coli* RNA polymerases. *J. Biol. Chem.* 280, 17422–17429.
- Simpson, R.B. (1979). The molecular topography of RNA polymerase-promoter interaction. *Cell* 18, 277–285.
- 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.
- Travers, A.A. (1980). Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J. Bacteriol.* 141, 973–976.
- Vassilyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417, 712–719.
- 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.

Xu, J., McCabe, B.C., and Koudelka, G.B. (2001). Function-based selection and characterization of base-pair polymorphisms in a promoter of *Escherichia coli* RNA polymerase-sigma(70). *J. Bacteriol.* *183*, 2866–2873.

Xue, Y., Hogan, B.P., and Erie, D.A. (2000). Purification and initial characterization of RNA polymerase from *Thermus thermophilus* strain HB8. *Biochemistry* *39*, 14356–14362.

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 sigma(70). *Cell* *105*, 935–944.

Young, B.A., Gruber, T.M., and Gross, C.A. (2004). Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. *Science* *303*, 1382–1384.

A

GGGAGCTCAGAATAAACGCTCAA-32N-TTCGACATGAGGCCCGGATC

B

```

sTap1  CACGGCCGGAGTGTATAATGGGAGCGGTATCG
sTap2  GGCCACGTGTAAACTGGGAGCACCTACGGATG
sTap3  ACCGCGCCGAGAGTACAATGGGAGCTGAACAG
sTap4  TGTACAATGGGATGCACTCGAGAATCGTAGTG
sTap5  GGAGGGCGGAAGGTACAATGGGAGCCGACTGG
sTap6  TGCCGCATGTACAATGGGAGGCCCCCTCGGTGG
sTap7  GGGGCGAGGGTACAATGGGACGCCTTCACGG
sTap8  GGGCTGCAGTGTACAATGGGTGGCTGTTCAAG
sTap9  CGTGGGCGAAGGTACAATGGGATGCCGCAAGG
sTap10 GGAGGGCGAGTGTACCATGGGAGGCTGCAGAG
sTap11 GTCGGGGCGAGGGTATAATGGGAGGCGGTTGG
sTap12 GGAGGCGCCGATGTACAATGGGATGCCAAAGG
sTap13 GGCGGGGCTGCGTGTATAATGGGAGGCTTTGG
sTap14 GTGCGGGGCGGGGTGTATAATGGGAGCTTTAG

```

C

	G	T	A	c/t	A	A	T	G	G	G	A
G	33	0	0	0	0	0	0	33	33	33	0
A	0	0	33	3	32	31	0	0	0	0	32
T	0	33	0	10	0	0	33	0	0	0	1
C	0	0	0	20	1	2	0	0	0	0	0

D

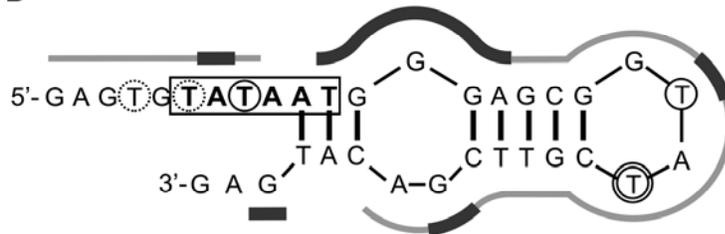


Figure 1. Single-Stranded DNA Aptamers against *Taq* RNAP  $\sigma^A$

(A) Random ssDNA library used in the selection experiment.

(B) Sequences of several representative aptamers. Only central 32 nt long regions of the aptamers are shown. The conserved aptamer motif is gray.

(C) Frequency of bases at individual positions of the conserved motif in 33 cloned aptamers. The consensus sequence is shown on the top.

(D) The proposed secondary structure of sTap1. The structure was predicted by using the RNA Structure 3.7 software. The  $-10$ -like element is boxed. The regions of sTap1 protected from hydroxyl radicals in the  $\sigma^A$ -sTap1 complex are shown by a gray line; the thickness of the line corresponds to the degree of protection (see Figure S2). Thymine residues modified by  $\text{KMnO}_4$  are encircled (Figure S1). Two thymines protected from modification in the  $\sigma^A$ -sTap1 complex are marked with dashed circles; the residue that becomes hyperreactive in the presence of  $\sigma^A$  is shown by a double circle.

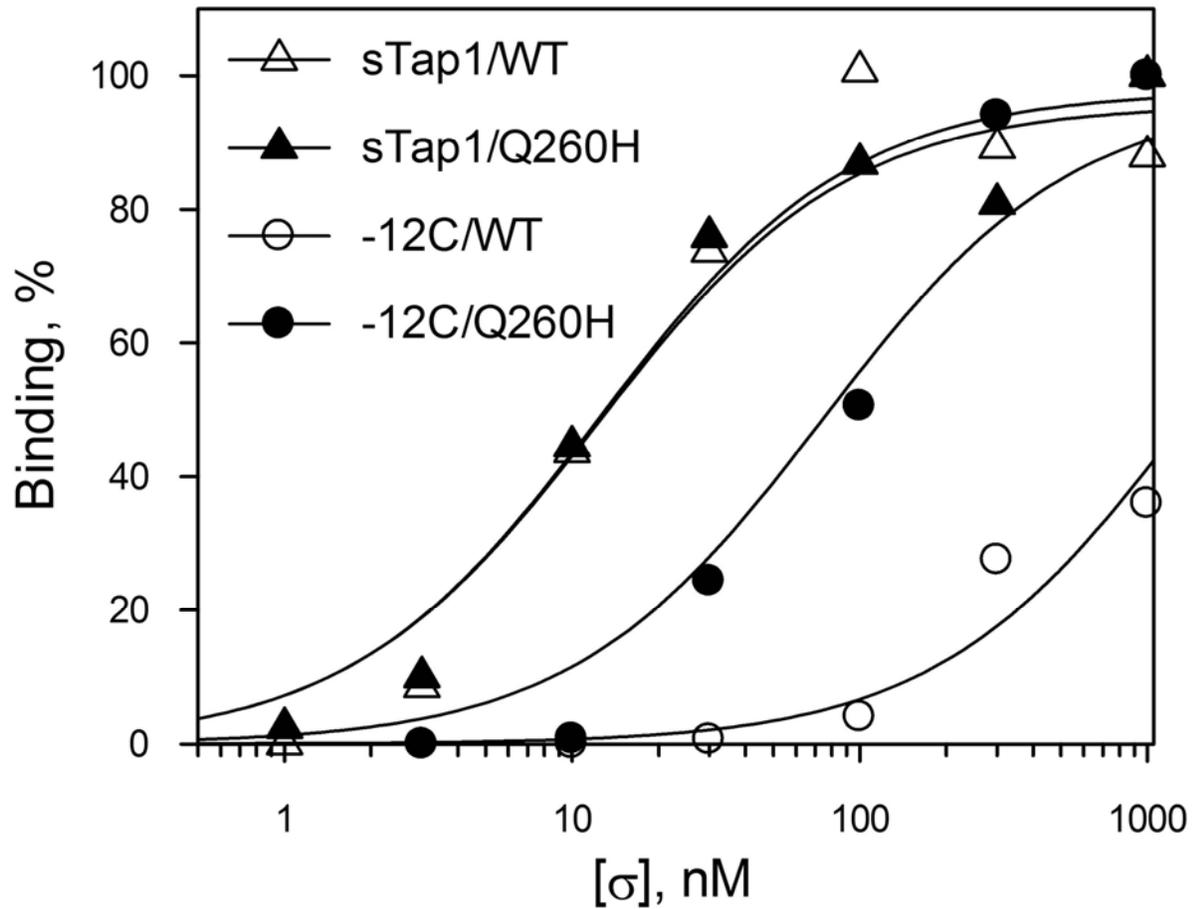


Figure 2. Recognition of the -12C Mutant sTap1 by the Wild-Type and the Mutant Q260H  $\sigma^A$  Subunits

Wild-type or mutant sTap1 was incubated in the presence of increasing concentrations of  $\sigma$ , and the binding was measured by nitrocellulose binding method (see Experimental Procedures). For each aptamer, binding is shown as a percentage of the maximum binding observed for Q260H  $\sigma^A$ .

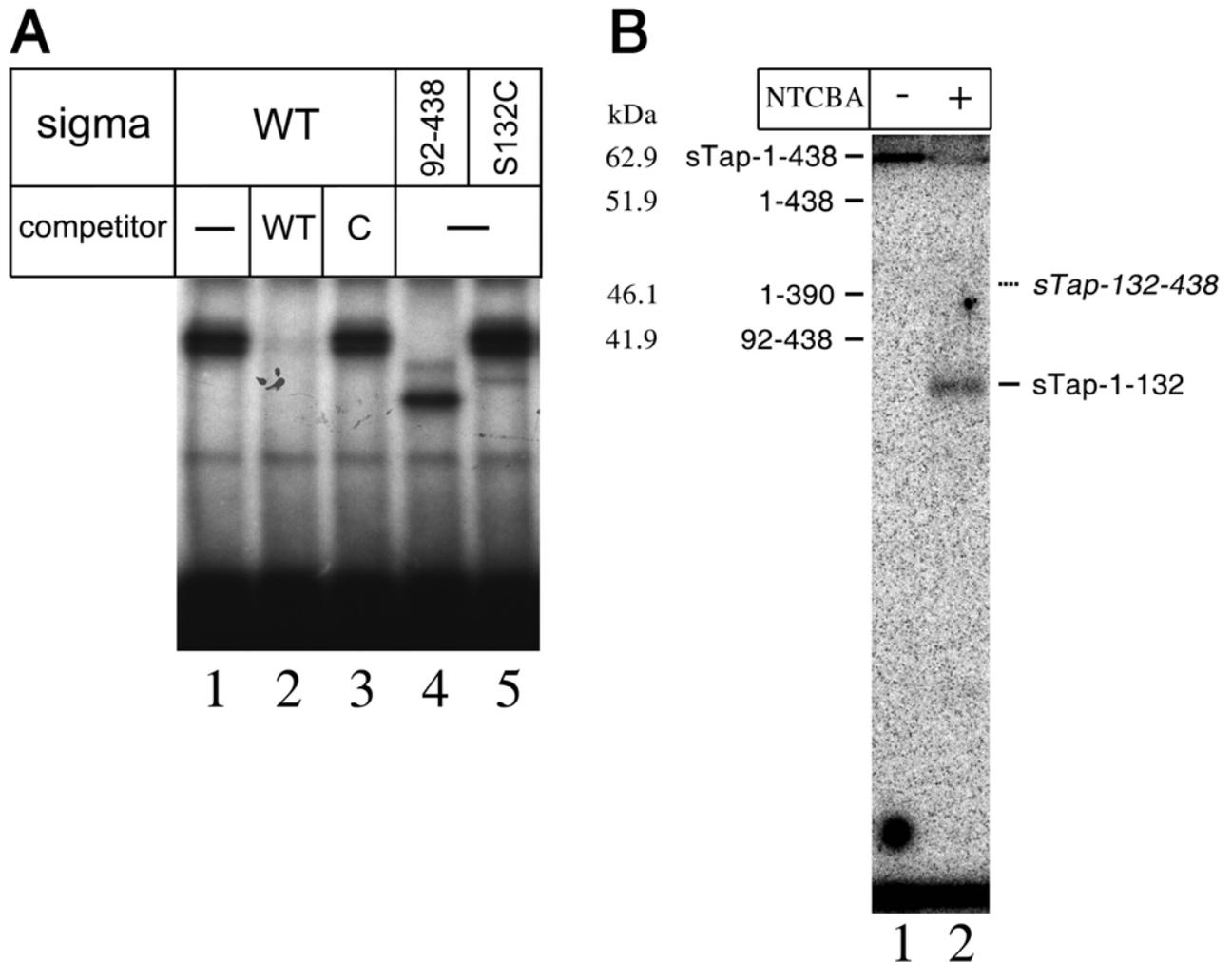


Figure 3. Site-Specific Crosslinking of sTap1 with the  $\sigma^A$  Subunit

(A) Crosslinking of 6-thio-guanine-modified sTap1 with wild-type  $\sigma^A$  (lanes 1–3), the 92–438 fragment (lane 4), and the S132C mutant (lane 5). 5'-labeled-modified sTap1 and the  $\sigma$  subunits were present at 0.1 and 1  $\mu$ M, respectively. In the control samples, crosslinking was performed in the presence of either unmodified sTap 1 (lane 2) or a control nonspecific oligonucleotide (lane 3); both oligonucleotides were present at 1  $\mu$ M.

(B) Cysteine-specific cleavage of S132C  $\sigma$  in the complex with radioactively labeled 6-thio-guanine-sTap1 (lane 2). The sample on lane 1 is a control without NTCBA cleavage. Positions and molecular weights of markers used for the analysis (full-length  $\sigma^A$ , 92–438, and 1–390  $\sigma^A$  fragments and the crosslinked  $\sigma^A$ -sTap1 complex) are indicated on the left of the gel. The position of the radiolabeled N-terminal cleavage product is shown on the right of the figure. Predicted position of the C-terminal 132–438 peptide in the complex with sTap1 (calculated Mr 46.2 kDa) is shown on the right by a dashed line.

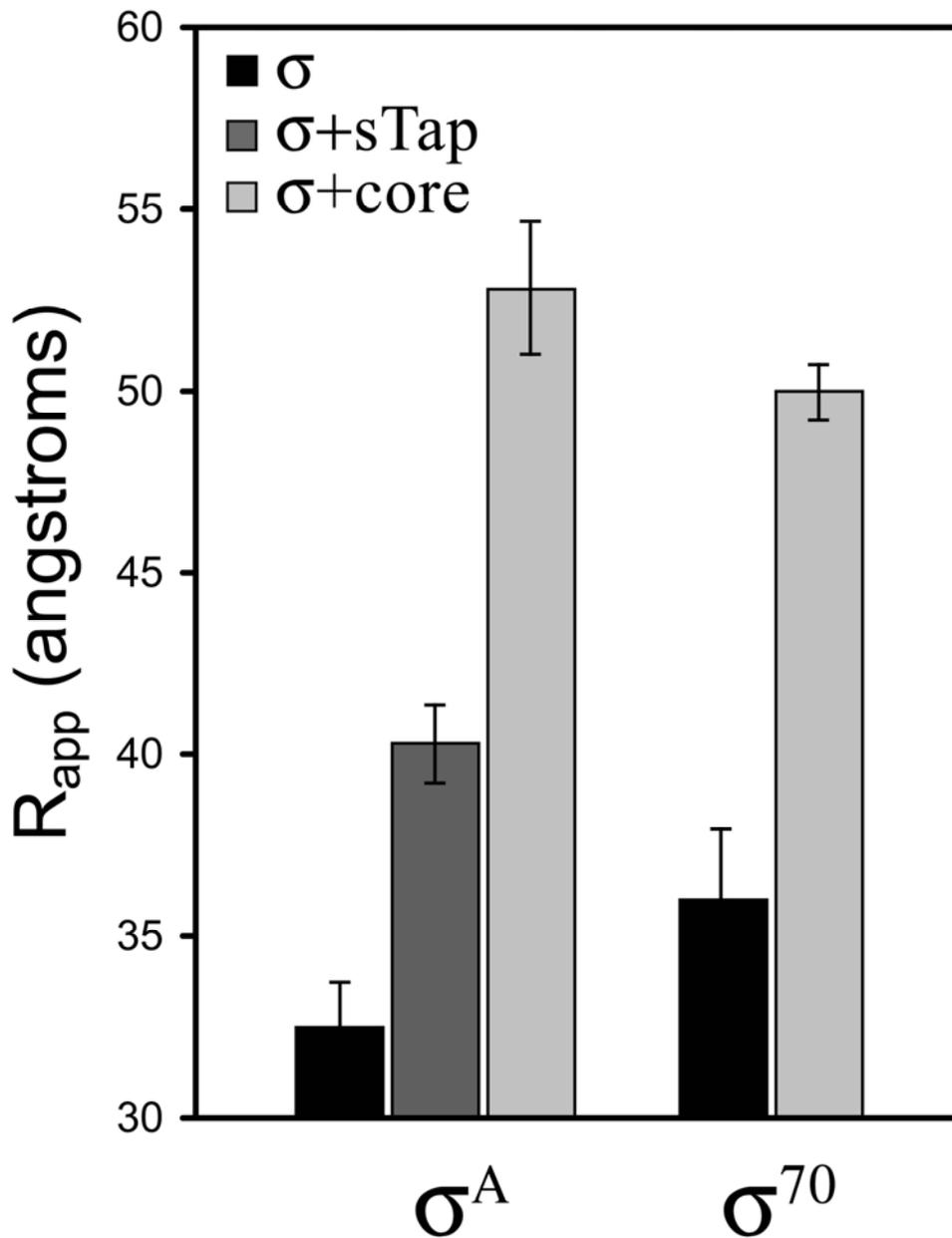


Figure 4. Apparent Interdomain Distances in the  $\sigma^A$  Subunit Measured by LRET

Apparent distances between the donor and acceptor fluorochromes introduced to cysteine residues at positions 263 and 406 of the  $\sigma^A$  subunit were measured as described in the Experimental Procedures. Averages and standard deviations from three to five experiments are shown. Data for apparent interdomain distances between regions 2 and 4 in the *E. coli*  $\sigma^{70}$  subunit are taken from Callaci et al. (1999).

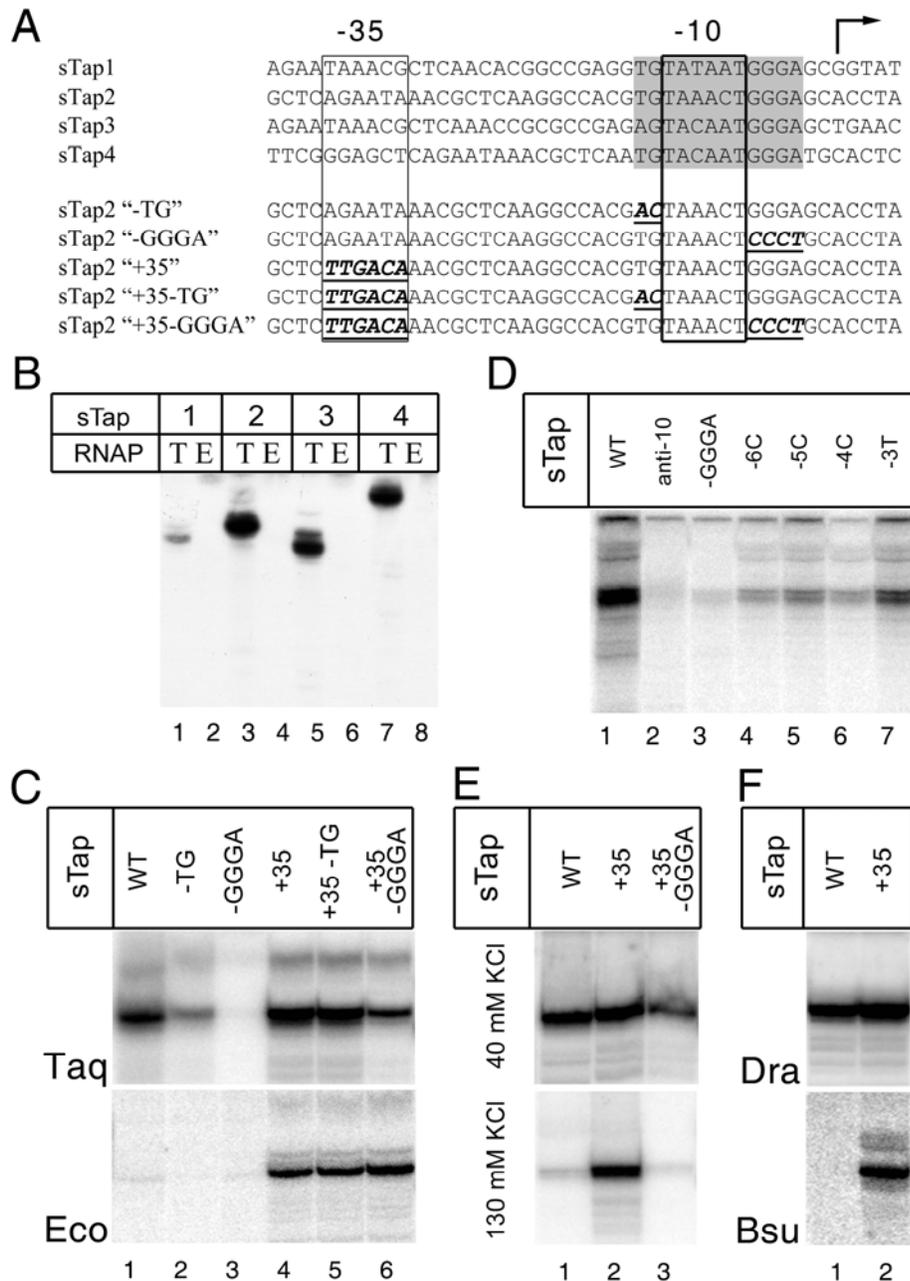


Figure 5. Transcription Properties of Double-Stranded DNA Fragments Containing Aptamer Sequences

Transcription was performed at 65°C for *Taq* RNAP and at 37°C for *E. coli*, *D. radiodurans*, and *B. subtilis* RNAPs. Transcription buffer contained 40 mM KCl unless otherwise indicated.

(A) Sequences of sTap promoters used in transcription experiments. Shown are nucleotides from –40 to +5 relative to the starting point of transcription; the full-length sequences are presented in Figure S5. The numbering of sTap clones corresponds to Figure 1B. The conserved aptamer motif is shown in gray. Nucleotides corresponding to the –10 and –35 promoter elements are boxed. The starting point of transcription is indicated by an arrow. Nucleotides changed in the mutant derivatives of sTap2 are bold underlined italics.

- (B) Synthesis of run-off RNA transcripts on several aptamer-based promoters. Abbreviation: T, *Taq* RNAP; E, *E. coli* RNAP.
- (C) Transcription from sTap2 derivatives by *Taq* (top) and *E. coli* RNAPs (bottom).
- (D) Effect of mutations in the conserved aptamer sequence on the utilization of sTap2-based promoter by *Taq* RNAP.
- (E) Salt sensitivity of transcription by *Taq* RNAP from the sTap2 promoter. Transcription buffer contained either 40 mM (top) or 130 mM of KCl (bottom).
- (F) Transcription on the wild-type (lane 1) and “+35” (lane 2) sTap2 promoters by RNAPs from *B. subtilis* (*Bsu*) and *D. radiodurans* (*Dra*).

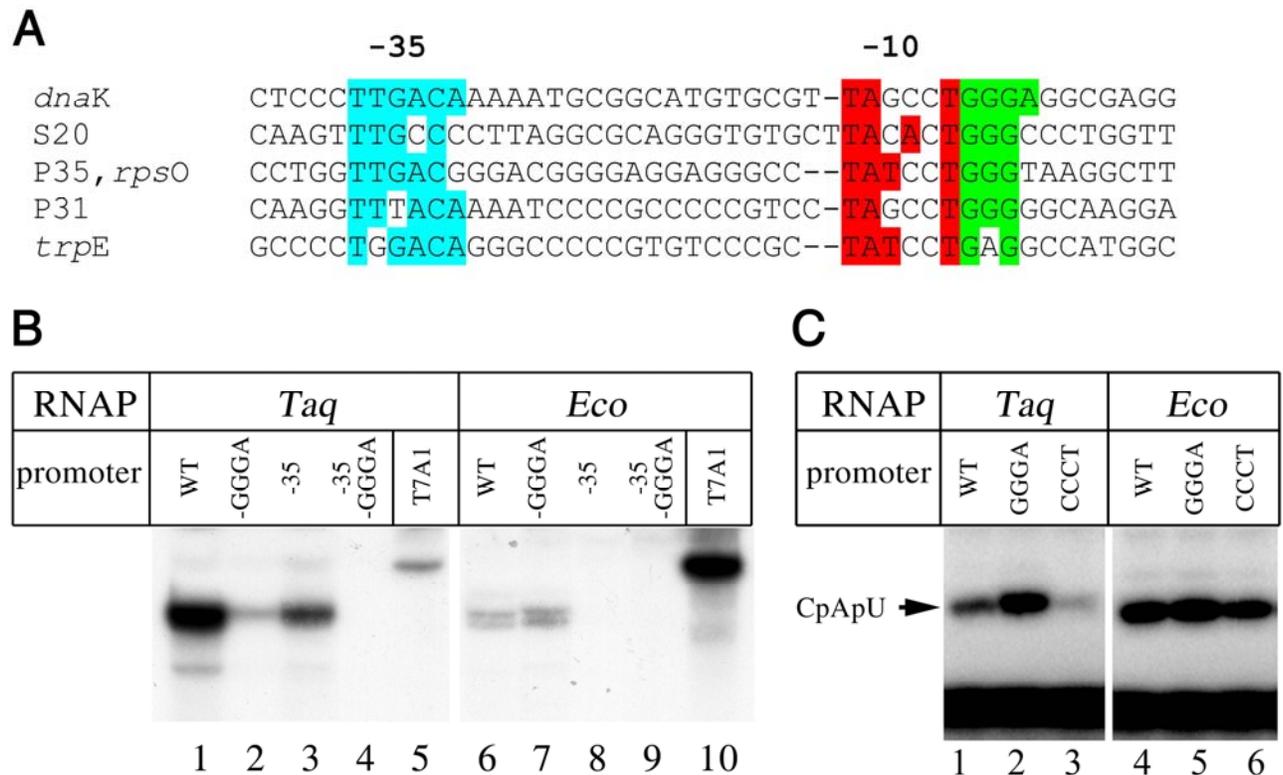


Figure 6. Role of the GGGA Motif in the Recognition of Natural Promoters

(A) Sequences of annotated *T. thermophilus* promoters containing GGGA-like motifs. The data are from Leontiadou et al. (2001), Maseda and Hoshino (1995), Osipiuk and Joachimiak (1997), and Sato et al. (1988) (see Figure S6 for the complete list of promoters). The nucleotides identical to the -10 promoter consensus element are shown in red, the nucleotides of the GGGA-like motif are green, and the nucleotides of the -35 element are blue.

(B) Transcription from the *dnaK* promoter variants by *Taq* and *E. coli* RNAPs. The T7 A1 promoter was used as a control template.

(C) Effect of the GGGA motif on the efficiency of abortive synthesis by *Taq* and *E. coli* RNAP on the T7 A1 promoter. Position of abortive product CpApU is indicated by an arrow. Transcription was performed with the wild-type promoter and two derivatives containing either GGGA or CCCT downstream of the -10 promoter element. The temperature of the reaction was 65°C in the case of *Taq* RNAP and 37°C in the case of *E. coli* RNAP.

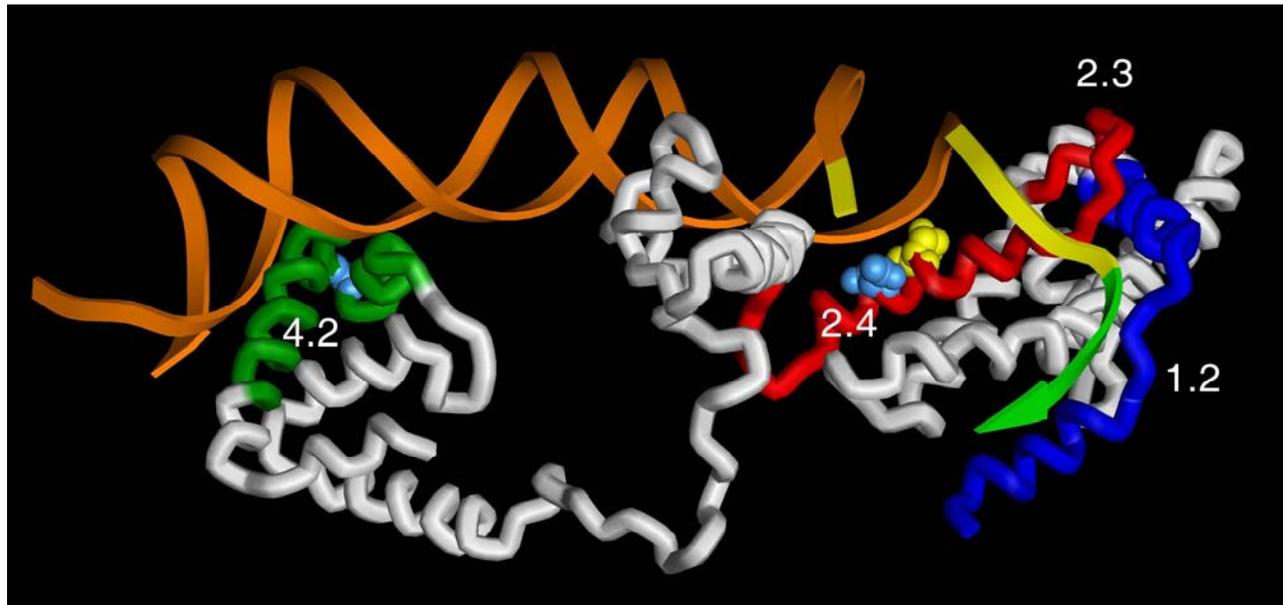


Figure 7. DNA Recognition Regions of the  $\sigma$  Subunit

The model illustrates contacts of the *Taq*  $\sigma^A$  subunit with a fork-junction DNA fragment containing the  $-35$  and  $-10$  promoter elements. The figure is based on the model of the open promoter complex of *Taq* RNAP (Murakami et al., 2002a). The double-stranded part of DNA shown on the figure includes the  $-35$  element, the spacer between the  $-10$  and  $-35$  elements, and the first T of the  $-10$  element (shown in yellow); the single-stranded 3' extension of the nontemplate strand corresponds to the remaining five nucleotides from the  $-10$  element (yellow) and four downstream nucleotides corresponding to the GGGA motif (light green). The  $\sigma^A$  regions involved in the recognition of the promoter elements are colored as follows: region 4.2 (amino acids 397–424), dark green; regions 2.3 and 2.4 (amino acids 242–257 and 258–276), red; and region 1.2 (amino acids 92–125), dark blue. In the model, the distance between the second nucleotide of the GGGA motif and Leu108 from  $\sigma$  region 1.2 is  $\leq 4\text{\AA}$ . The position of the Q260H mutation is shown in yellow. Amino acids changed in the double-cysteine mutant of  $\sigma$  used in LRET measurements are light blue (N263C and G406C). The figure was generated with the ViewerLite 4.2 software (Accelrys Inc.).

## Supplemental Data

### An Additional Basal Promoter Element Recognized

### by Free RNA Polymerase $\sigma$ Subunit Determines

### Promoter Recognition by RNA Polymerase Holoenzyme

Andrey Feklistov,<sup>1,2,3</sup> Nataliya Barinova,<sup>1</sup> Anastasiya Sevostyanova,<sup>1</sup> Ewa Heyduk,<sup>4</sup> Irina Bass,<sup>1</sup> Irina Vvedenskaya,<sup>5</sup> Konstantin Kuznedelov,<sup>5</sup> Eglė Merkienė,<sup>6</sup> Elena Stavrovskaya,<sup>7</sup> Saulius Klimašauskas,<sup>6</sup> Vadim Nikiforov,<sup>1,3</sup> Tomasz Heyduk,<sup>4</sup> Konstantin Severinov,<sup>1,5,\*</sup> and Andrey Kulbachinskiy<sup>1,5,\*\*</sup>

## Supplemental Results

### Formation of open and closed complexes by *Taq* RNAP on the sTap2 promoter

To study the role of the GGA motif in open complex formation, we performed KMnO<sub>4</sub> probing of promoter complexes formed by *Taq* RNAP on sTap2 and its mutant derivatives at 65 °C (Figure S4A). As can be seen, *Taq* RNAP melted the wild type and “+35-GGA” promoters with similar efficiencies (Figure S4A, lanes 2 and 5). The most efficient modification was observed in the case of the “+35” promoter indicating that the presence of both the GGA and -35 motifs further stimulated open complex formation (lane 4). In contrast, no open complexes were formed when both the -35 and the GGA motifs were absent (lane 3). This indicated that the absence of activity of *Taq* RNAP on the “-GGA” promoter could be explained by its inability to form open complex on this template.

To test whether the GGA motif could be important for formation of a closed promoter complex on sTap2, gel retardation experiments with the wild type and “-GGA” sTap2 promoters were performed at 4 and 20 °C. As was shown previously, *Taq* RNAP does not form open promoter complex at these temperatures (Kuznedelov et al., 2002; Kulbachinskiy et al., 2004a). No significant difference in the complex formation was observed between the wild type and mutant promoters when the experiment was performed at 4 °C (Figure S4B, lanes 2 and 4). In contrast, the mutation of the GGA motif impaired promoter complex formation when the experiment was repeated at 20 °C. In particular, the efficiency of the complex formation was about 10 times lower on the mutant promoter than on the wild type sTap2 (Figure S4B, lanes 6 and 8). We therefore speculate that the complexes formed at 4 °C correspond to the closed

promoter complex while those formed at 20 °C may represent intermediate promoter complexes containing partially bent/distorted DNA (Saecker et al., 2002). Thus, removal of the GGGA motif impairs formation of the open and, presumably, intermediate promoter complexes, but does not affect closed complex formation by *Taq* RNAP. This finding is in agreement with a structural model of the closed complex in which DNA segment downstream of the -10 element, corresponding to the GGGA motif, is not contacted by RNAP (Murakami et al., 2002). In the intermediate complex formed at 20 °C, promoter DNA may be bent by RNAP thus allowing specific interactions of the GGGA motif with  $\sigma$  region 1.2 (Saecker et al., 2002; Scavi et al., 2005). Further experiments are required to establish the complete pathway of open complex formation by *Taq* RNAP.

### Analysis of promoter regions in the *T. thermophilus* genome

Analysis of literature and public databases reveals a limited number of experimentally defined promoter sequences from *Thermus* bacteria almost all of which come from *T. thermophilus*. A compilation of published sequences of *T. thermophilus* -10/-35 promoters is presented on Figure S6. As can be seen from the figure, two of the promoters contain the full GGGA-motif found in the  $\sigma^A$  aptamers and a considerable fraction (6 out of 22) contains the GGG-sequence downstream of the -10 element. In contrast, only one of the promoters contain the complementary CCC-sequence. Although the limited number of analyzed promoters does not allow to make statistically significant conclusions, this may indicate that the GGGA-like sequences play an important role in recognition of natural *Thermus* promoters. To further analyze the occurrence of the GGGA-like motifs in *Thermus* promoters, we performed a preliminary bioinformatic analysis of promoters in the genome of *T. thermophilus*. To predict promoter sequences, we created recognition profiles for the -35 and -10 promoter consensus elements using a starting set of 22 annotated promoters of *T. aquaticus* and *T. thermophilus* that was compiled from the literature (see Figure S6). The recognition profiles (positional weight matrices, PWMs) were created using SignalX software (Mironov et al., 2000). The resulting PWMs were used to search two samples of intergenic fragments extracted from the whole genome sequence of *T. thermophilus* HB27 (gene bank accession number NC\_005835). The first sample contained regions between divergently transcribed genes (divergons), whereas the second sample contained regions between convergently transcribed genes (convergons). Divergons are expected to contain divergently oriented promoters, whereas convergons are not expected to contain

promoters and thus serve as a control. In total, 308 divergon fragments with an average length of 135 nucleotides, and 404 convergon fragments with an average length of 133 nucleotides were included in the analysis.

In each region, the highest scoring candidate promoter was identified using PWMs. We next filtered out those putative promoters whose scores were less than the worst score observed in the starting promoter set. The filtering procedure reduced the total number of putative promoters to 115 candidates in divergons (37% of the total number of divergons) and 63 candidate promoters in convergons (16% of the total number of convergons). The relative enrichment of candidate promoters in divergons demonstrates that the constructed PWMs have some, although not absolute, predictive power, comparable to that of other promoter recognition tools (Robison et al., 1998).

We next searched the set of predicted promoters for the presence of a motif consisting of three Gs immediately downstream of the -10 element (we did not search for the whole GGGA-motif as it was demonstrated that the last A from the motif is less important for promoter recognition than the first three Gs, see Figure 5C). For each candidate promoter we computed the number of cases when at least two out of three positions immediately downstream of putative -10 promoter elements were Gs. As a control, we also computed the number of cases with at least two Cs in the same positions. The results showed that 33% of divergon promoters contained two or more Gs immediately after predicted -10 promoter elements, while 19% contained two or more Cs. In contrast, in the control set of convergon promoters, 24% contained two or more Gs and 22% contained two or more Cs. Thus, predicted promoters in divergons of *T. thermophilus* appear to be enriched with stretches of Gs immediately after the -10 elements compared to control sets. Similar analysis of 197 *E. coli* promoters from the *dpinteract* database (Robison et al., 1998) revealed that only 7% of them contained stretches of Gs and 11% contained Cs, confirming that the observed preference for Gs in this region is specific for the genome of *T. thermophilus*.

## Supplemental Experimental Procedures

### Plasmids and proteins

pET28TaD encoding the *Taq*  $\sigma^A$  subunit with a His<sub>6</sub> tag in the N-terminus was described previously (Minakhin et al., 2001). Genes encoding two fragments of the  $\sigma^A$  subunit (92-438 and 1-390) were obtained by PCR from the wild type *Taq rpoD* gene. Plasmid pET28TaD(Q260H) encoding the mutant Q260H  $\sigma^A$  was obtained by two-step PCR

mutagenesis of the wild type *rpoD* gene. Plasmids encoding the mutant S132C, H278S and E281A  $\sigma^A$  subunits and the double-mutant  $\sigma^A$  containing two cysteine residues at positions 263 and 406 (N263C, G406C) were also prepared by site-directed mutagenesis. Wild-type and mutant  $\sigma$ s were overexpressed in the BL21 (DE3) *E. coli* strain and purified as described (Minakhin et al., 2001). Core enzymes of *E. coli* and *Taq* RNAPs were purified as described (Kulbachinskiy et al., 2004b).

### Selection of aptamers to *Taq* $\sigma^A$ subunit

ssDNA library used in the selection experiments was purchased from Operon technologies (Figure 1A). The selection was performed at 25 °C essentially as described in (Kulbachinskiy et al., 2004b). The amounts of ssDNA and *Taq*  $\sigma^A$  subunit varied from 3 nmol and 500 pmol, respectively, in the first round of selection to 100 and 10 pmol in subsequent rounds. DNA was diluted in 1 ml of binding buffer (20 mM Tris-HCl, pH 7.9, 10 mM MgCl<sub>2</sub>, 240 mM NaCl, 60 mM KCl), heated for 3 minutes at 95°C and chilled on ice. The solution was passed through a 50  $\mu$ l Ni<sup>2+</sup>-NTA-agarose (Qiagen) micro-column pre-equilibrated with binding buffer to remove the sequences that nonspecifically interact with the sorbent. The  $\sigma^A$  subunit was added to the solution and the mixture was incubated for 20 minutes. 30  $\mu$ l of Ni-NTA-agarose was added and the incubation continued for another 20 minutes with occasional shaking. The solution containing unbound DNA was removed and the sorbent was washed 2-4 times with 1 ml of binding buffer (for a total time of 30-60 minutes).  $\sigma$ -DNA complexes were eluted with 300  $\mu$ l of binding buffer containing 200 mM imidazole. The solution was treated with 300  $\mu$ l of phenol and with 300  $\mu$ l of chloroform. DNA was ethanol-precipitated, dissolved in water and amplified using Deep Vent DNA polymerase (New England BioLabs) and primers corresponding to fixed regions of the library (5'-GGGAGCTCAGAATAAACGCTCAA and BBB-5'-GATCCGGGCCTCATGTTCGAA, where B is a biotin residue). Two DNA strands were separated by size on 10% denaturing PAAG, the nonbiotinylated strand was eluted and used for the next round of selection. After ten rounds of selection, the enriched library was amplified with primers containing EcoRI and HindIII sites and cloned into the pUC19 plasmid. The sequences of inserts in 33 individual clones were determined using standard sequencing protocols. Individual ssDNA aptamers were obtained by PCR with the primers corresponding to aptamer flanks, followed by strand separation. Shortened variants of aptamers were purchased from Syntol (Moscow). 5'-fluorescein labeled aptamers were kindly provided by V.K. Potapov.

### Quantitation of the binding of aptamers to the $\sigma^A$ subunit

All binding experiments were performed at 20-25 °C. Determination of equilibrium  $K_d$  values for binding of oligonucleotides to RNAP was done either using the nitrocellulose binding method (Carey et al., 1983) or by measuring fluorescence quenching of a 5'-fluorescein labeled sTap1 aptamer. In the former case, oligonucleotides were 5'-end labeled with  $\gamma$ -[ $^{32}$ P]-ATP (6000 Ci/mmol, NEN) and T4 polynucleotide kinase (New England BioLabs) and purified by PAGE. Each aptamer (0.1 nM) was incubated with a series of dilutions of the  $\sigma$  subunit (from 1 nM to 10  $\mu$ M) in 50  $\mu$ l of binding buffer for 30 minutes at room temperature; the samples were then filtered through 0.45  $\mu$ m nitrocellulose filters (HAWP, Millipore) prewetted in the same buffer. The filters were washed with 5 ml of the buffer and quantified with PhosphorImager (Amersham Biosciences). The interaction of 5'-fluorescein labeled sTap1 with  $\sigma^A$  was monitored in a spectrofluorometer (Perkin-Elmer LS-50B luminescence spectrometer). The excitation wavelength was set at 496 nm; the emission was measured at 517 nm; slit widths were 5 and 10 nm, respectively. The binding of  $\sigma^A$  resulted in strong quenching of the sTap1 fluorescence. For each sample, the fluorescence intensity was measured every second for 20 seconds and then averaged. To calculate equilibrium dissociation constants, the data were fit to the hyperbolic equation

$$B = \frac{B_{\max}[\sigma]}{[\sigma] + K_d}$$

where B is a percentage of DNA bound,  $B_{\max}$  is the maximum binding at infinite concentration of  $\sigma$  and  $K_d$  is the observed dissociation constant (Connolly et al., 2001).

### Hydroxyl radical footprinting

*Taq*  $\sigma^A$  subunit (500 nM) was mixed with 5'-radioactively-labeled 75 nt long sTap1 (5 nM) in 85  $\mu$ l of binding buffer and incubated for 30 minutes at 25 °C. Footprinting reaction was initiated by adding 5  $\mu$ l of 0.6% H<sub>2</sub>O<sub>2</sub>, 5  $\mu$ l of 20 mM ascorbic acid, 2.5  $\mu$ l of 0.4 mM Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>·6H<sub>2</sub>O and 2.5  $\mu$ l of 0.8 mM EDTA. The reaction was stopped after 2 minutes with 10  $\mu$ l of 0.1 M thiourea, DNA was ethanol precipitated and dissolved in 5  $\mu$ l of a formamide containing sample buffer. DNA fragments were separated on 10% urea PAAG and analyzed with PhosphorImager.

### KMnO<sub>4</sub> footprinting of the sTap1 aptamer

The 5'-labeled minimal sTap1 aptamer (10 nM) was incubated at 20 °C or 70 °C in 10 µl of the binding buffer. The  $\sigma^A$  subunit, when present, was added to 500 nM. The samples were treated with 5 mM KMnO<sub>4</sub> for 30 seconds and processed as described in (Minakhin and Severinov, 2003). The cleavage products were separated on 10% PAAG and analyzed with PhosphorImager.

### Cross-linking experiments

The 6-thioguanine-modified sTap1 was 5'-labeled with  $\gamma$ -[<sup>32</sup>P]-ATP (6000 Ci/mmol, NEN) and T4 polynucleotide kinase (New England BioLabs) and purified by PAGE. The  $\sigma^A$  subunit or its mutant derivatives (1 µM) was mixed with the modified aptamer (0.1 µM) in 20 µl of the binding buffer. The samples were incubated for 20 minutes at 25 °C and irradiated for 10 minutes with a 365 nm ultraviolet lamp (4W, Spectroline) placed on the top of open eppendorf tubes. The samples were supplemented with SDS-containing loading buffer and loaded separated on 8% SDS-PAGE. In the competition experiments, the labeled modified aptamer was mixed with an excess of either unlabeled sTap1 or a control oligonucleotide (1 µM) prior to  $\sigma$  addition. The sequence of the control oligonucleotide was 5'-

GGGAGCTCAGAATAAACGCTCAATTCGACATGAGGCCCGGATC.

In the NTCBA cleavage experiments, the samples obtained after irradiation were supplemented with 100 µl of the binding buffer and 10 µl of Ni<sup>2+</sup>-NTA-agarose (Qiagen) and incubated for 20 minutes at 4 °C with occasional shaking. Supernatant containing unbound sTap1 was removed and the Ni<sup>2+</sup>-NTA-agarose pellet was washed three times with 1 ml of the binding buffer and then with 1 ml of cleavage buffer (20 mM Tris-HCl, pH 8.4, 8 M urea, 10 mM  $\beta$ -mercaptoethanol). The  $\sigma$ -aptamer complexes were eluted with 30 µl of the cleavage buffer containing 30 mM EDTA and incubated for 10 minutes at 65 °C to allow protein denaturation. Half of the sample (15 µl) was supplemented with the SDS-loading buffer and served as a control on the absence of nonspecific protein degradation. To initiate cysteine-specific cleavage, 1.5 µl of 0.5 M 2-nitro-5-thiocyanobenzoic-acid (Sigma, USA) was added to 15 µl sample and the reaction was incubated for 30 minutes at 37 °C. The sample was supplemented with 1 µl of 1 M NaOH and incubation continued for 16 hours at 37 °C. The products of the cleavage were analyzed by 11% SDS-PAGE.

### LRET distance measurements

The double cysteine mutant of  $\sigma^A$  was labeled with europium chelate and Cy5 and purified as previously described for  $\sigma^{70}$  (Callaci et al., 1999). This allowed to measure the interdomain distances in  $\sigma^A$  using luminescence resonance energy transfer (LRET). Distance measurements were conducted at 25 °C as described in detail in (Heyduk and Heyduk, 2002). Briefly, decay curves of 25 nM labeled  $\sigma^A$  in 120  $\mu$ l of buffer (50 mM HEPES (pH 7.9), 100 mM NaCl, 10  $\mu$ M EDTA, 5% glycerol, 5 mM MgCl<sub>2</sub>, 0.1 mg/ml BSA, and 0.1 mM DTT) were collected using LN300 nitrogen laser (Laser Photonics, Salt Lake City, UT) as the excitation source on a laboratory-built two-channel spectrofluorometer (Heyduk and Heyduk, 2002). Oligonucleotides and the *Taq* core polymerase were added to 100 nM and 60 nM, when present, respectively. The sequences of the control nontemplate oligonucleotide was 5'-GGCGTTTATAATGG. Decay curves were analyzed by fitting to a sum of exponentials by nonlinear regression using SCIENTIST (Micromath Scientific Software, Salt Lake City, UT). Apparent donor-acceptor distances were calculated from sensitized acceptor decay curves as described (Heyduk and Heyduk, 2002; Heyduk and Heyduk, 2001). Luminescence lifetimes of sensitized acceptor ( $\tau_{da}$  values) are shown in Supplementary Figure S3. All sensitized acceptor decays contained a minor (10-20 % of the total amplitude) fast decaying component ( $\sim 30$   $\mu$ sec) and a major slower decaying component. The nature of this fast component is unclear (it could be due to protein aggregation). All distances were calculated using average amplitude-weighted lifetime (Heyduk and Heyduk, 2002; Heyduk and Heyduk, 2001) and should thus be considered as apparent distances. If only the major decay component was used for distance calculations, obtained values were only slightly different and the relative effects of aptamers and core enzyme remained the same.

### *In vitro* transcription

Double stranded DNA fragments containing the aptamer sequences were prepared by PCR from the aptamer pUC19 clones using standard sequencing primers. The sequence of the *dnaK* promoter was obtained by PCR from a synthetic oligonucleotide template corresponding to positions -56 to +38 of the *dnaK* promoter, cloned into the pUC19 plasmid and amplified with standard primers. Mutant derivatives of the sTap2, *dnaK* and T7A1 promoters were obtained in the same way. The PCR products were separated from primers by agarose gel electrophoresis and purified with QiaQuick gel extraction kit (Qiagen). The sequences of all studied promoters are presented in Supplemental Figure S5.

### KMnO<sub>4</sub> footprinting of the sTap<sub>2</sub>-based promoter

Double stranded DNA fragments containing the sequence of either the wild type sTap<sub>2</sub> promoter or its mutant derivatives were digested with the EcoRI restriction endonuclease (the recognition site for this endonuclease was introduced in the PL primers used for amplification, Figure S5) and labeled with 1 unit of the Klenow enzyme (Amersham Biosciences) and  $\alpha$ -[<sup>32</sup>P]-dATP for 15 minutes at 25 °C in 20  $\mu$ l of reaction buffer. The samples were then supplemented with 1  $\mu$ l of 10 mM dNTP mixture and incubated for 5 minutes at 25 °C. Labeled DNA fragments were purified by 5% nondenaturing PAGE.

Holoenzyme RNAP was prepared by mixing the core enzyme (100 nM final concentration) and the  $\sigma$  subunit (500 nM) in 10  $\mu$ l of the transcription buffer. Labeled DNA fragment was added to 50 nM and the samples were incubated for 10 minutes at 65 °C to allow open complex formation. KMnO<sub>4</sub> was added to 2 mM at 65 °C and the reaction was stopped after 15 seconds by the addition of 10  $\mu$ l of a buffer containing 1 M  $\beta$ -mercaptoethanol and 1 M sodium acetate (pH 4.8). The samples were processed as described in (Minakhin and Severinov, 2003) and analyzed by 10% urea PAGE.

### Gel retardation experiments

Radioactively labeled double stranded DNA fragments containing sequences of either the sTap<sub>2</sub> promoter or its “-GGGA” derivative were obtained as described above. Holoenzyme RNAP was prepared by mixing the core enzyme (10 nM final concentration) and the  $\sigma$  subunit (100 nM) in 10  $\mu$ l of the transcription buffer containing 50  $\mu$ g/ml BSA. Labeled DNA fragment was added to 50 nM and the samples were incubated for 30 minutes at either 4 or 20 °C. The samples were separated on 4% nondenaturing PAGE at either 4 or 20 °C and analyzed by phosphorimaging.

## Supplemental References

Callaci, S., Heyduk, E., and Heyduk, T. (1999). Core RNA polymerase from *E. coli* induces a major change in the domain arrangement of the sigma 70 subunit. *Mol. Cell* 3, 229-238.

- Carey, J., Cameron, V., de Haseth, P. L., and Uhlenbeck, O. C. (1983). Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry* 22, 2601-2610.
- Connolly, B. A., Liu, H.-H., Parry, D., Engler, L. E., Kurpiewski, M. R., and Jeu-Jacobson, L. (2001). Assay of restriction endonucleases using oligonucleotides. In *DNA-protein interactions*, T. Moss, ed. (Totowa, NJ, Humana Press), pp. 465-490.
- Croft, J. E., Love, D. R., and Bergquist, P. L. (1987). Expression of leucine genes from an extremely thermophilic bacterium in *Escherichia coli*. *Mol. Gen. Genet.* 210, 490-497.
- Faraldo, M. M., de Pedro, M. A., and Berenguer, J. (1992). Sequence of the S-layer gene of *Thermus thermophilus* HB8 and functionality of its promoter in *Escherichia coli*. *J. Bacteriol.* 174, 7458-7462.
- Hartmann, R. K., and Erdmann, V. A. (1989). *Thermus thermophilus* 16S rRNA is transcribed from an isolated transcription unit. *J. Bacteriol.* 171, 2933-2941.
- Hartmann, R. K., Ulbrich, N., and Erdmann, V. A. (1987). An unusual rRNA operon constellation: in *Thermus thermophilus* HB8 the 23S/5S rRNA operon is a separate entity from the 16S rRNA operon. *Biochimie* 69, 1097-1104.
- Heyduk, E., and Heyduk, T. (2002). Conformation of fork junction DNA in a complex with *E. coli* RNA polymerase. *Biochemistry* 41, 2876-2883.
- Heyduk, T., and Heyduk, E. (2001). Luminescence energy transfer with lanthanide chelates: interpretation of sensitized acceptor decay amplitudes. *Analytical Biochemistry* 289, 60-67.
- Koyama, Y., and Furukawa, K. (1990). Cloning and sequence analysis of tryptophan synthetase genes of an extreme thermophile, *Thermus thermophilus* HB27: plasmid transfer from replica-plated *Escherichia coli* recombinant colonies to competent *T. thermophilus* cells. *J. Bacteriol.* 172, 3490-3495.
- Kulbachinskiy, A., Bass, I., Bogdanova, E., Goldfarb, A., and Nikiforov, V. (2004a). Cold sensitivity of thermophilic and mesophilic RNA polymerases. *J. Bacteriol.* 186, 7818-7820.
- Kulbachinskiy, A., Feklistov, A., Krashennnikov, I., Goldfarb, A., and Nikiforov, V. (2004b). Aptamers to *Escherichia coli* core RNA polymerase that sense its interaction with rifampicin, sigma-subunit and GreB. *Eur. J. Biochem.* 271, 4921-4931.

- Kuznedelov, K., Korzheva, N., Mustaev, A., and Severinov, K. (2002). Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J.* *21*, 1369-1378.
- Leontiadou, F., Triantafyllidou, D., and Choli-Papadopoulou, T. (2001). On the characterization of the putative S20-thx operon of *Thermus thermophilus*. *Biol. Chem.* *382*, 1001-1006.
- Maseda, H., and Hoshino, T. (1995). Screening and analysis of DNA fragments that show promoter activities in *Thermus thermophilus*. *FEMS Microbiol Lett.* *128*, 127-134.
- Minakhin, L., Nechaev, S., Campbell, E. A., and Severinov, K. (2001). Recombinant *Thermus aquaticus* RNA polymerase, a new tool for structure- based analysis of transcription. *J. Bacteriol.* *183*, 71-76.
- Minakhin, L., and Severinov, K. (2003). On the role of the *Escherichia coli* RNA polymerase sigma 70 region 4.2 and alpha-subunit C-terminal domains in promoter complex formation on the extended -10 galP1 promoter. *J. Biol. Chem.* *278*, 29710-29718.
- Mironov, A. A., Vinokurova, N. P., and Gelfand, M. S. (2000). GenomeExplorer: software for analysis of complete bacterial genomes. *Mol. Biol. (Mosk.)* *34*, 222-231.
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* *296*, 1285-1290.
- Nardmann, J., and Messer, W. (2000). Identification and characterization of the dnaA upstream region of *Thermus thermophilus*. *Gene* *261*, 299-303.
- Osipiuk, J., and Joachimiak, A. (1997). Cloning, sequencing, and expression of dnaK-operon proteins from the thermophilic bacterium *Thermus thermophilus*. *Biochim. Biophys. Acta* *1353*, 253-265.
- Robison, K., McGuire, A. M., and Church, G. M. (1998). A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.* *284*, 241-254.
- Saecker, R. M., Tsodikov, O. V., McQuade, K. L., Schlax, P. E., Jr., Capp, M. W., and Record, M. T., Jr. (2002). Kinetic studies and structural models of the association of *E. coli* sigma(70) RNA polymerase with the lambdaP(R) promoter: large scale conformational changes in forming the kinetically significant intermediates. *J. Mol. Biol.* *319*, 649-671.

- Sanchez, R., Roovers, M., and Glansdorff, N. (2000). Organization and expression of a *Thermus thermophilus* arginine cluster: presence of unidentified open reading frames and absence of a Shine-Dalgarno sequence. *J. Bacteriol.* *182*, 5911-5915.
- Sato, S., Nakada, Y., Kanaya, S., and Tanaka, T. (1988). Molecular cloning and nucleotide sequence of *Thermus thermophilus* HB8 trpE and trpG. *Biochim. Biophys. Acta* *950*, 303-312.
- Sclavi, B., Zaychikov, E., Rogozina, A., Walther, F., Buckle, M., and Heumann, H. (2005). Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 4706-4711.
- Serganov, A., Rak, A., Garber, M., Reinbolt, J., Ehresmann, B., Ehresmann, C., Grunberg-Manago, M., and Portier, C. (1997). Ribosomal protein S15 from *Thermus thermophilus*--cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein. *Eur. J. Biochem.* *246*, 291-300.
- Struck, J. C., Toschka, H. Y., and Erdmann, V. A. (1988). Nucleotide sequence of the 4.5S RNA gene from *Thermus thermophilus* HB8. *Nucleic Acids Res.* *16*, 9042.
- Van de Casteele, M., Chen, P., Roovers, M., Legrain, C., and Glansdorff, N. (1997). Structure and expression of a pyrimidine gene cluster from the extreme thermophile *Thermus* strain ZO5. *J. Bacteriol.* *179*, 3470-3481.

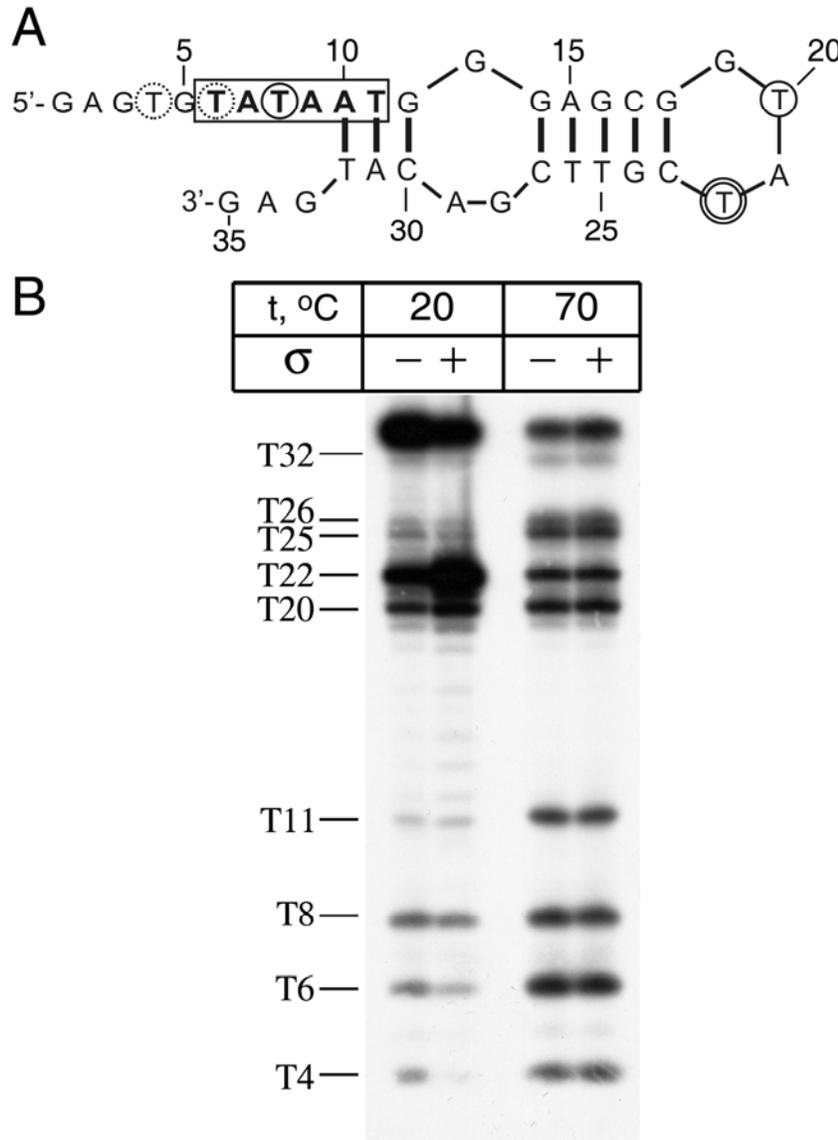


Figure S1. Probing of the secondary structure of sTap1 with  $\text{KMnO}_4$

(A) The proposed secondary structure of sTap1. Aptamer bases are numbered starting from the 5'-end. The -10-like element is boxed. Thymine residues modified by  $\text{KMnO}_4$  are encircled; thymines protected from modification in the  $\sigma^A$ -sTap1 complex are marked with dashed circles; the residue that becomes hyper reactive in the presence of  $\sigma^A$  is shown by a double-circle.

(B)  $\text{KMnO}_4$  probing was performed with minimal sTap1 at 20°C (lanes 1, 2) or 70 °C (lanes 3, 4) as described in Supplemental Experimental Procedures. Samples in lanes 2 and 4 contained *Taq*  $\sigma^A$  subunit. Positions of thymine residues starting from the 5'-end of the aptamer are shown on the left.

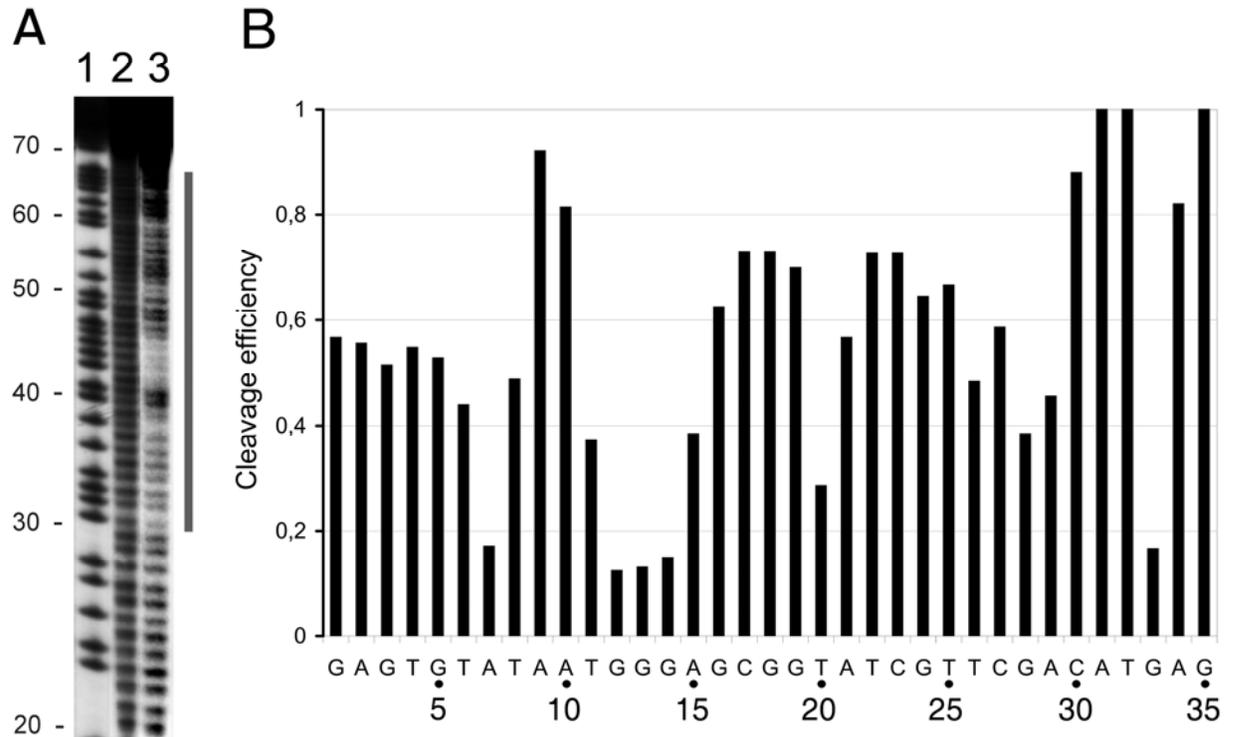


Figure S2. Hydroxyl radical footprinting of the  $\sigma^A$ -sTap1 complex

(A) Electrophoretic analysis of DNA cleavage products. Lane 1 – A+G sequencing marker, lane 2 – cleavage of the 5'-end labeled full length sTap1 aptamer in the absence of  $\sigma^A$ , lane 3 – cleavage of sTap1 in the complex with  $\sigma^A$ .

Numbers on the left indicate nucleotide positions starting from the 5'-end of the aptamer. Gray bar on the right indicates footprint area corresponding to the minimal 35 nt long sTap1.

(B) Efficiencies of nucleotide cleavage at different aptamer positions in the  $\sigma^A$ -sTap1 complex relative to free sTap1. The aptamer sequence is shown below the figure; the bases are numbered starting from the 5'-end of the aptamer.

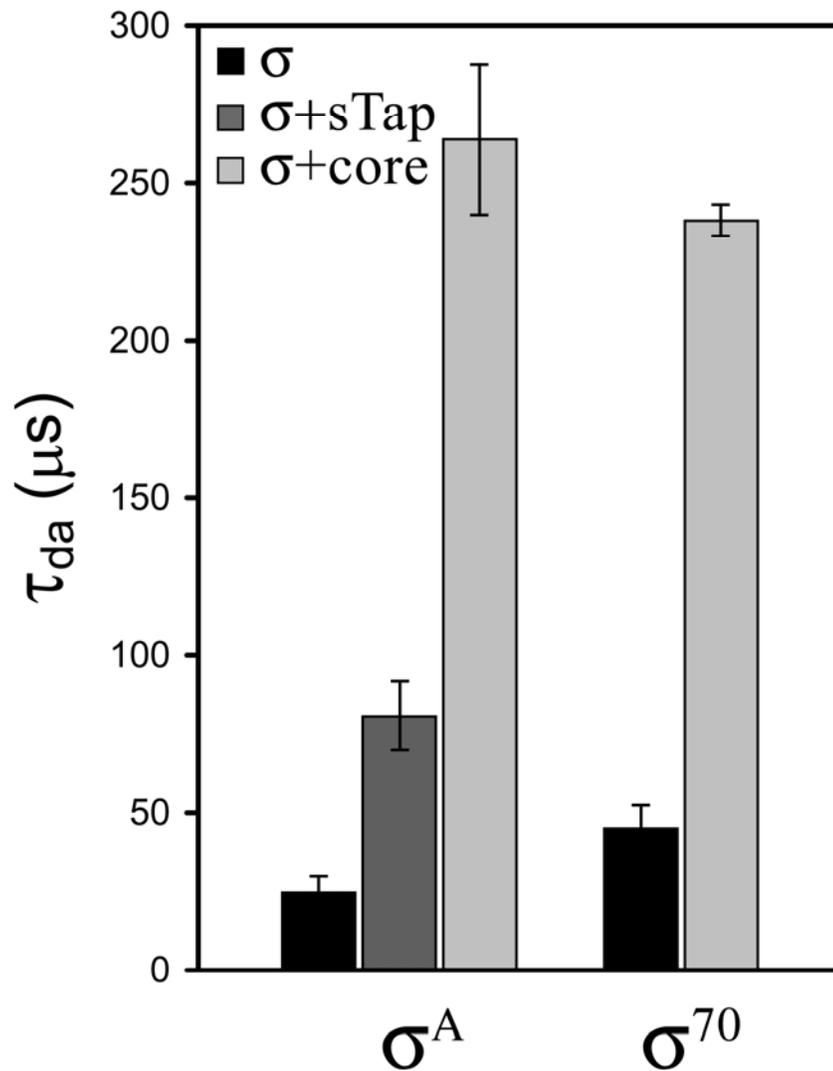


Figure S3. LRET measurements of interdomains distances in the  $\sigma^A$  subunit

The figure shows luminescence lifetimes of sensitized acceptor ( $\tau_{da}$ ) measured by LRET (see (Heyduk and Heyduk, 2002; Heyduk and Heyduk, 2001) for details). Averages and standard deviations from 3-5 experiments are shown. The unquenched donor decay had a lifetime of 600 microsec. Data for *E. coli*  $\sigma^{70}$  subunit are taken from (Callaci et al., 1999).

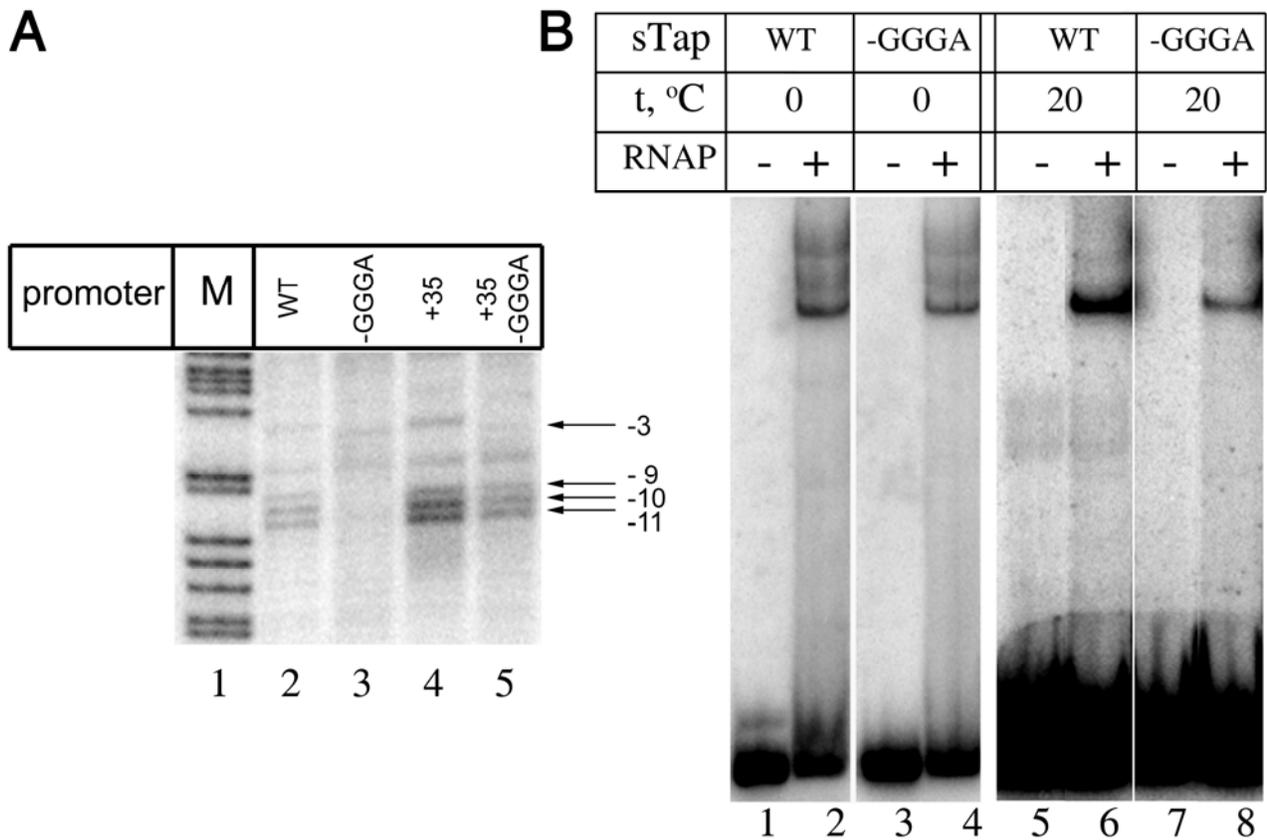


Figure S4. Effect of mutations in the sTap2-based promoter on the open and closed complex formation by *Taq* RNAP (A) Permanganate footprinting of the complexes of *Taq* RNAP with different derivatives of the sTap2 promoter. The experiment was done at 65 °C. Lane 1 contains an A+G cleavage marker. The positions of modified thymine residues relative to the starting point of transcription are indicated on the right of the figure (see Figure S5 for the sTap2 sequence).

(B) Gel retardation experiment with *Taq* RNAP and either the wild type (WT) or “-GGGA” sTap2 promoters at 4 and 20 °C (see Supplemental Experimental Procedures for details). Control samples on lanes 1, 3, 5 and 7 contained no RNAP.

**A**

\*

sTap1 GGGAGCTCAGAATAAACGCTCAACACGGCCGAGG**TGTATAATGGGA**GCGGTATCGTTCGACATGAGGCCCGGATC

sTap2 GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGTAAACTGGGA**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap3 GGGAGCTCAGAATAAACGCTCAAACCGCGCCGAG**AGTACAATGGGA**GCTGAACAGTTCGACATGAGGCCCGGATC

sTap4 GGGAGCTCAGAATAAACGCTCA**ATGTACAATGGGA**TGCACTCGAGAATCGTAGTGTTTCGACATGAGGCCCGGATC

sTap2 “-TG” GGGAGCTCAGAATAAACGCTCAAGGCCACG**ACTTAAACTGGGA**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “-GGGA” GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGTAAACTCCCT**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “+35” GGGAGCTC**TTGACA**AACGCTCAAGGCCACG**TGTAAACTGGGA**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “+35-TG” GGGAGCTC**TTGACA**AACGCTCAAGGCCACG**ACTTAAACTGGGA**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “+35-GGGA” GGGAGCTC**TTGACA**AACGCTCAAGGCCACG**TGTAAACTCCCT**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “anti-10” GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGATCTCGGGG**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “-6G” GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGTAAACTCGG**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “-5G” GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGTAAACTCGG**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “-4G” GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGTAAACTCGG**GCACCTACGGATGGTTCGACATGAGGCCCGGATC

sTap2 “-3A” GGGAGCTCAGAATAAACGCTCAAGGCCACG**TGTAAACTGGG**TGCACCTACGGATGGTTCGACATGAGGCCCGGATC

PL1 CAGTGAATTCGGGAGCTCAGAATAAACGCTCAAGGCC

PL2 CAGTGAATTCGGGAGCTC**TTGACA**AACGCTCAAGGCC

PR CCGAAGCTTGATCCGGGCCTCATGTGAACCATCCG

**B**

T7A1 GAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAG**GATACTTACAGCCATCGAGAGGGACACGGCGAATAG**

T7A1 “GGGA” GAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAG**GATACTGGGA**GCCATCGAGAGGGACACGGCGAATAG

T7A1 “CCCT” GAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAG**GATACTCCCT**GCCATCGAGAGGGACACGGCGAATAG

T7A1L CAGTGAATTCATTTGGATCCAGATCCCGAAAATTTATCAAAAAGAG

T7A1R CCGAAGCTTCCCCGGTGTGCGATTGGGATGGCTATTCGCCGTGTCCC

**C**

*dnaK* TCATACTCAACTCCC**TTGACA**AAAAATGCGGCATGTGCGT**TAGCCT**GGGAGGCGAGGTGAAGACGTATGGCCAAGGCAGT

*dnaK* “-35” TCATACTCAACTCCC**GGTCAC**AAAAATGCGGCATGTGCGT**TAGCCT**GGGAGGCGAGGTGAAGACGTATGGCCAAGGCAGT

*dnaK* “GGGA” TCATACTCAACTCCC**TTGACA**AAAAATGCGGCATGTGCGT**TAGCCTCCCT**GGGAGGCGAGGTGAAGACGTATGGCCAAGGCAGT

*dnaK* “-35-GGGA” TCATACTCAACTCCC**GGTCAC**AAAAATGCGGCATGTGCGT**TAGCCTCCCT**GGGAGGCGAGGTGAAGACGTATGGCCAAGGCAGT

*dnaKL* CAGTGAATTCACAAAATCATACTCAACTCCC

*dnaKR* CCGAAGCTTTCAATGCCCACTGCCTTGGCCATAC

Figure S5. Sequences of double stranded promoter DNA fragments used in this study

(A) Sequences of the sTap2 derivatives. Mutated nucleotides are bold underlined italics. The conserved aptamer motif is shown in red, the -35 promoter element is blue. The transcription start site is marked with an asterisk. Primers used in PCR reaction are shown below the sequences. Primer PL2 was used to obtain sTap2 derivatives bearing the -35 promoter consensus sequence. EcoRI and HindIII sites used for cloning are underlined.

(B) Sequences of the T7A1 derivatives. The -10 promoter element is shown in red. Primers used for PCR amplification are shown below the sequences.

(C) Sequences of the derivatives of *dnaK* promoter. The -10 element is shown in red, the -35 element is blue. Primers used for amplification are shown below the sequences.

All sequences were cloned into the pUC19 plasmid and amplified with standard M13 primers (5'-

AGCGGATAACAATTTACACAGGA and 5'-GTTTTCCCAGTCACGAC).

Name	Sequence	Reference
<b>A</b>		
<b>dnaK</b>	TCCCTTGACAAAAATGCGGCATGTGCGT--TAGCCTGGGAGGCGAGGTGAA	(Osipiuk and Joachimiak, 1997)
<b>trpAB</b>	CGAGTTTACCGGGAGGCCCTCCGGGG--TAGGATGGGAGTTGTCTTGGC	(Koyama and Furukawa, 1990; Faraldo et al., 1992)
<b>s20</b>	AAGTTTGCCCTTAGGCGCAGGGTGTGCTTACTACTGGGCCCTGGTTTGCC	(Leontiadou et al., 2001)
<b>rpsO</b>	CTGGTTGACGGGACGGGGAGGAGGGCC--TATCCTGGGTAAGGCTTGCCG	(Serganov et al., 1997)
<b>dnaA</b>	TCCCATCTCACTCTAGCGGGTCCCGGGGTAAAAATGGGGACCGGGACGA	(Nardmann and Messer, 2000)
<b>trpE</b>	CCCCTGACAGGGCCCCCGTGTCCCGC--TATCCTGAGGCCATGGCCCTT	(Sato et al., 1988)
<b>16S RNA</b>	AGCCTTGACAAAAAGGAGGGGGATTGA--TAGCATGGCTTTTCTGCGCGG	(Hartmann and Erdmann, 1989)
<b>pyr</b>	TCCCTTGCCGGGACGGGGCAGGGGGTG--TAAGGTAGGGGTGGCCTTAA	(Van de Castele et al., 1997)
<b>arg</b>	GCCCTTGACATAAGTTTGCGGGCACGGGG--TATGCTTAAGGCCTCATGGGG	(Sanchez et al., 2000)
<b>ORF4 (arg)</b>	CCACTTGACAGCTTTTGTATTCTGAGTC--TATCCTCTATTTCGGGGAGCGT	(Sanchez et al., 2000)
<b>slp</b>	CCGCTTGACAAGGGCGCGTGAGGTTTT--TACGATAGCGCCGGATGCGGG	(Faraldo et al., 1992)
<b>leu</b>	AAGCTTGACCCCGCAGGCCTCGAGGGCT--TACCTTAGGGGCAATGCGGC	(Croft et al., 1987)
<b>4.5S RNA</b>	CCTCTAGCCTCAGGGCTTCCATGGGTGC--TATACTACCCGAGCCCCCGGT	(Struck et al., 1988)
<b>23S RNA</b>	GCCCTTGACAAAGGCCATGCCTCCTTGG--TATCTTCCCTTTTGCCTGCGC	(Hartmann et al., 1987)
<b>B</b>		
<b>P35</b>	CTGGTTGACGGGACGGGGAGGAGGGCC--TATCCTGGGTAAGGCTTGCCG	(Maseda and Hoshino, 1995)
<b>P31</b>	AAGGTTTACAAAAATCCCCGCCCCCGTCC--TAGCCTGGGGGCAAGGAGGTT	(Maseda and Hoshino, 1995)
<b>P39</b>	GCCCTTGACGGGGAGGAGGCAACGGGG--TAAAAACAGGGGCGAGAGCGGC	(Maseda and Hoshino, 1995)
<b>P211</b>	TAGAATTCGGACACCCCTGTGC--TAAGCTGAGGCCGTGAACCTG	(Maseda and Hoshino, 1995)
<b>P215</b>	GCACTTGACATCATAAAGTGCTTGAGG--TATCATCCGACCTGGGCGCAA	(Maseda and Hoshino, 1995)
<b>P43</b>	GGTCTTGTCAAGTAAGCTTAGCTATGG--TAACATAGACCTGGGAGGTAA	(Maseda and Hoshino, 1995)
<b>P7</b>	GGGGTTGCCAGATGAAGAAAAAGCGT--TATGCTAAGCCCTAGGGATTC	(Maseda and Hoshino, 1995)
<b>P37</b>	GGCGTTGACCATCTTCTCCTTGGCCT--TATCCTTAGGGTGCCTCCGCC	(Maseda and Hoshino, 1995)
<b>P214</b>	GGGCTTGCCAATCCGCCCTTAGAGTG--TACCATAGCGATTGCCCGAGG	(Maseda and Hoshino, 1995)

Figure S6. Published sequences of *T. thermophilus* promoters

Nucleotides identical to the -10 promoter consensus element are shown in red, the nucleotides of the -35 element are blue and the nucleotides of the GGGGA-like motif are green.

(A) Promoters of *T. thermophilus* found in regulatory regions of known genes.

(B) Promoters identified in a genetic screen for promoter activities in *T. thermophilus* (Maseda and Hoshino, 1995).

Note, that P35 is identical to the rpsO promoter characterized by (Serganov et al., 1997).