The pathway to transcriptionally active

Escherichia coli RNAP−*T7A1* promoter complex formation:

Positioning of RNAP at the promoter using X−ray

hydroxyl radical footprinting.

Anastasia Rogozina

aus

Svirsk, Russland

2009

Contents

1.   Introduction.

Transcription, the DNA−directed synthesis of RNA, is a highly regulated cellular process catalyzed by a large multisubunit protein, called RNA polymerase (RNAP). In eukaryotic species, three distinct multisubunit RNAPs are found within the cell nucleus. RNAP I synthesizes rRNA, RNAP II synthesizes mRNA and some small nuclear RNAs, RNAP III synthesizes tRNA, 5S rRNA and some small nuclear RNAs. In eubacteria and archaea, a single multisubunit RNAP is responsible for transcription of the major classes of genes including mRNA, tRNA and rRNA.

The bacterial RNAP exists in two forms: core and holoenzyme. In *Escherichia coli*, the core RNAP consists of two large subunits, named β (1342 amino acid residues, 150.6 kDa) and β' (1407 residues, 155.2 kDa), and two smaller α subunits (each 329 residues, 36.5 kDa) [Darst et al. 1998]. The smallest 91−residue ω polypeptide was also identified as part of the enzyme, but no direct role in transcription could be attributed to this subunit [Hampsey 2001].

The transcription cycle in bacterial cells can be divided into three major phases: initiation, RNA transcript elongation, RNA transcript termination and release. Although core $\alpha_2\beta\beta'\omega$ RNAP is catalytically active, it is incapable of accurate initiation. For this, it must bind an initiation factor, σ, to form the holoenzyme that can recognize a specific DNA sequence at the beginning of a gene, the promoter. Upon binding to the promoter, the RNAP holoenzyme and the bound DNA undergo a series of conformational changes from the closed to the open promoter complex, in which the DNA duplex is partially opened at the promoter region such that one DNA strand becomes accessible as a template for synthesis of the complementary RNA sequence. In the presence of ribonucleoside triphosphates (rNTPs), open promoter complex is competent to initiate RNA synthesis leading to the formation of RNA chain completing thus the initial phase of transcription.

Each of these steps is a possible target for regulation of the transcription. Given the central role of RNAP in prokaryotic gene expression, it is very important to elucidate how interactions of holoenzyme with promoter DNA lead to productive transcription initiation and how these interactions are regulated by the binding of other proteins (specific transcription activators and repressors, nucleoid proteins), cofactors or specific metabolites [Browning and Busby 2004].

Extensive studies have been done on the structure of RNAP−DNA complexes at equilibrium resulting in the characterization of the interaction of each of the enzyme's subunits with specific promoter elements [Naryshkin et al. 2000, Mekler et al. 2002, Campbell et al. 2002, Murakami et al. 2002b]. However an analysis of the dynamics of the interaction at each step of the transcription initiation is still lacking.

At some promoters the *Escherichia coli* RNAP holoenzyme is able to recognize and bind specifically to the promoter forming a functional open complex and initiate RNA synthesis in the absence of additional transcription factors. Kinetic studies of RNAP binding to these promoters demonstrated the presence of a series of isomerisation events leading to open complex formation. Furthermore, by decreasing the isomerisation rates at lower temperatures, one or more transient intermediates in the pathway from the initial closed complex to the final open complex could be trapped and characterized [Schickor et al. 1990, Craig et al. 1998, Li et al. 1998, Buckle et al. 1999, Saecker et al. 2002]. However, the large, and sometimes nonlinear, temperature dependence of some of the steps in this pathway [Johnson and Chester 1998, Saecker et al. 2002] and the occurrence of intermediates at low temperature that were not detected in the normal kinetic pathway [Li et al. 1998, Buckle et al. 1999] suggest that the structure of intermediates and / or the mechanism of promoter recognition and DNA melting could differ at low temperatures compared with normal physiological conditions. In contrast, coupling the kinetic studies with techniques providing the structural signatures of the intermediates allows a more direct, real−time, characterization of these short−lived complexes.

Synchrotron X−ray footprinting combined with stopped−flow technique provides information on time−resolved structural changes of nucleoprotein complexes and nucleic acid polymer conformation with single nucleotide resolution. This approach has been successfully used to monitor the folding of the multiple domains of complex RNA molecules [Sclavi et al, 1998; Maleknia et al., 2001]. For the first time, we applied this approach to directly characterize and compare the structural intermediates present in the pathways of final open complex formation on the *T7A1* promoter under different conditions in order to determine how temperature and DNA sequence in the -10 region may affect the structure of the intermediates.

2. Structure of the RNAP.

2.1. Structure of the RNAP core enzyme.

2.1.1. Overall structure.

From a genetic, biochemical and functional point of view, *Escherichia coli* (*E.coli*) RNAP is the best characterized bacterial RNAP. However, the structure of *E.coli* RNAP core enzyme determined by cryo−electron microscopy (cryo−EM) has a low resolution of ~ 15 Å [Darst et al. 1998]. In contrast, the structure of RNAP core enzyme from the thermophilic organism *Thermus aquaticus* (*Taq*) obtained by X−ray crystallography has the 3.3 Å resolution [Zhang et al. 1999]. *Taq* RNAP has been used as basis for further modeling studies on *E.coli* RNAP including our footprinting results. This is justified, since the subunits of *E.coli* and *Taq* RNAPs exhibit high sequence homology and are functionally similar. Moreover, there is a high structural homology as indicated by superposition of the high−resolution structure of *Taq* RNAP core enzyme and the low−resolution structure of *E.coli* RNAP core enzyme [Zhang et al. 1999].

Data from literature show that *E.coli* and *Taq* RNAPs share the similar crab claw−like shape (Figure 1). The two "pincer" of the "claw" define a cleft (the internal space of the polymerase between the pincers). Major part of one pincer is formed by β subunit, while the major part of the other pincer is formed by β' subunit.
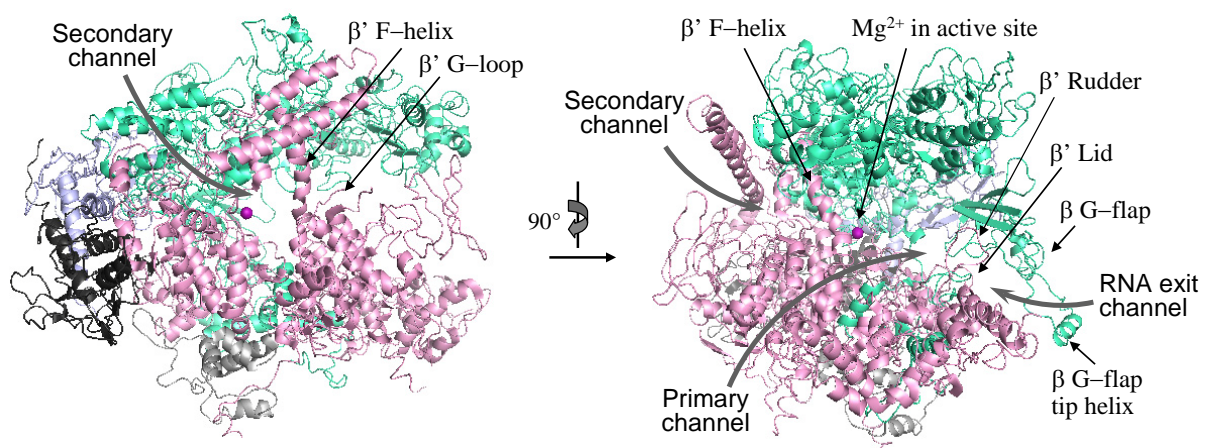
Figure 1. High−resolution crystal structure of *Taq* RNAP core enzyme [Zhang et al. 1999]. The structure is shown as cartoons using PyMOL program. The RNAP subunits are color coded as follows: α^I, light blue; α^{II}, dark gray; β, green−cyan; β', pink; ω, light gray. Magenta ball indicates the position of catalytic Mg^{2+} ion. Left panel is the secondary channel view of RNAP. Right panel showing the major channel view is obtained by rotating the left view 90° clockwise about the vertical axis.

The pincers are joined at the back by α−subunit dimer. One α subunit interacts with β (this α subunit is designated $α^I$). The other α subunit interacts with β' (this α subunit is designated $α^{II}$). Each α subunit contains two domains connected by a flexible linker. The amino–terminal domain (αNTD; residues 8–235 in *E.coli*; 26 kDa) plays a key role in RNAP assembly, providing the contact surface for dimerization of α subunits and interaction with either β or β' subunit, whereas the carboxy–terminal domain (αCTD; residues 249–329 in *E.coli*; 9 kDa) carries determinants for interaction with promoter DNA elements and with certain transcription factors. The flexible linker allows αCTD to occupy different positions relative to the remainder of RNAP in different promoter contexts [Blatter et al. 1994]. The αCTD is not resolved in the crystallographic structure of RNAP. The ω subunit is located near the base of the pincer formed mostly by β' subunit and interacts only with β'. It has been reported that although ω subunit is not required for transcription, it can assist the folding of the β' subunit [Hampsey 2001].

Both β and β' subunits consist of a number of relatively distinct, highly conserved regions (Figure 2). The β regions F, G, H and I contact the $α^I$NTD. The β' regions C, D, G and H contact the $α^{II}$NTD. Furthermore, β and β' subunits make extensive interaction with each other. A major interface between these RNAP subunits occurs at the base of the cleft.
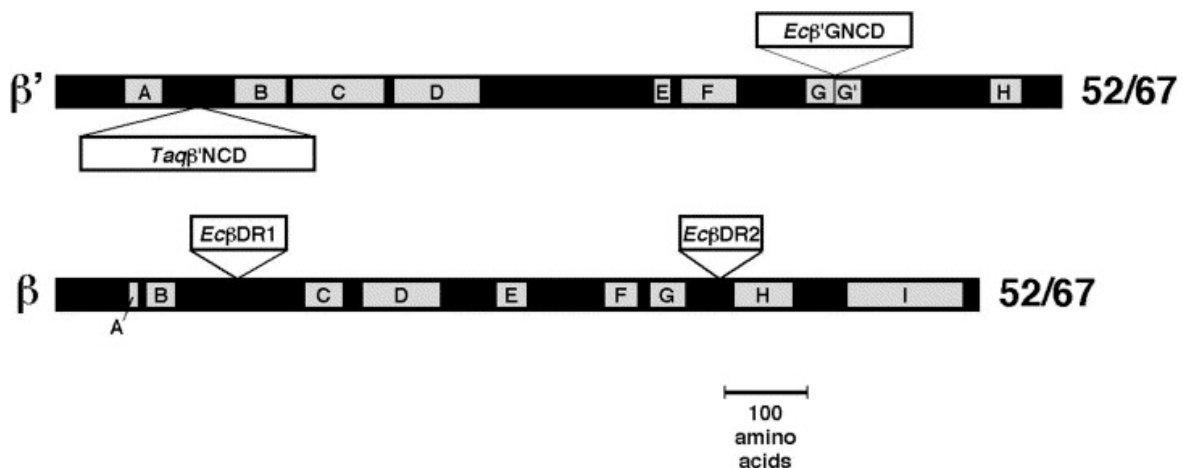
Figure 2. Sequence architecture of *E.coli* and *Taq* RNAP large subunits. The black bars represent the primary sequences of the RNAP subunits β' and β. The gray boxes indicate evolutionarily conserved regions among all prokaryotic, chloroplast, archaebacterial, and eukaryotic sequences. These are labeled A–H for β' and A–I for β [Zhang et al. 1999]. Comparing *E.coli* and *Taq* β' and β, sequence insertions larger than 15 amino acid residues are shown as white bars above (for insertions in the *E*.coli subunits) or below (for insertions in the *Taq* subunits). To the right of each subunit, the sequence identity (%)/sequence similarity (%) between the *E.coli* and *Taq* subunit is shown, calculated by ignoring the large insertions [Darst et al. 2002].

Particularly critical are interactions between β regions H and I and β' region D, which position the catalytic triad of β'$_D$ Asp residues for holding essential Mg^{2+}. β' regions C, G and H also participate in the formation of RNAP active site [Zhang et al. 1999].

Although the names of the conserved sequence regions of two large RNAP subunits, A–H for β' and A–I for β, remain in use, descriptive names such as the "clamp" or "G–flap" are used to identify structural motifs (see below) [Zhang et al. 1999].

2.1.2. Mobile domains. Conformational flexibility.

On the basis of comparisons of available crystal structures of RNAP from multiple organisms, the structural organization of RNAP is described as an immobile core module connected with four other modules able to move relative to it [Darst et al. 2002 and references therein, Murakami et al. 2002a]. The core module consists of the αNTDs, ω and portions of β and β' surrounding the active site. The mobile modules include: the clamp (the upstream half of the β' pincer) comprising the N–terminus of β' (residues 1-624 of *Taq* β' subunit) and the C–terminus of β (residues 1054-1115 of *Taq* β subunit); the two β N–terminal modules β1 (residues 22-130 and 336-392 of *Taq* β subunit) and β2 (residues 142-324 of *Taq* β subunit) that make up the top of β pincer; and the β G–flap module (residues 705-828 of *Taq* β subunit).

These mobile modules give considerable flexibility to RNAP. For instance, the flexibility of RNAP was demonstrated by comparison of the structures of *E.coli* and *Taq* core enzymes (Figure 3) [Darst et al. 2002].
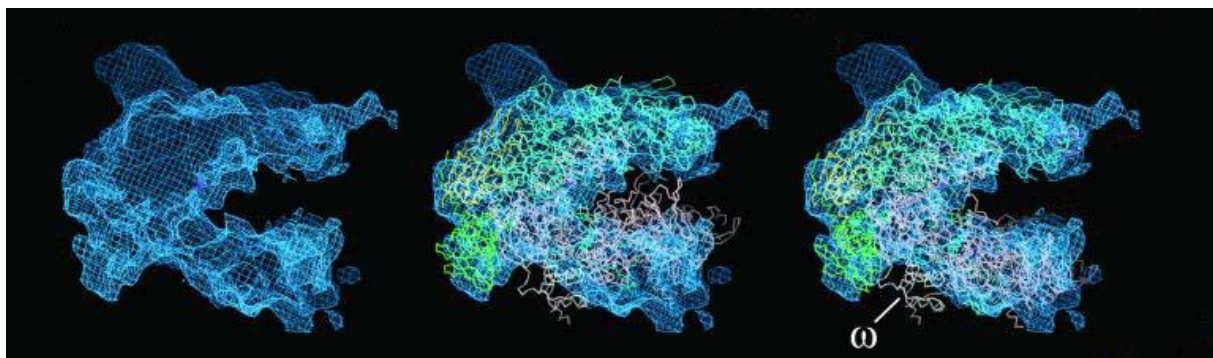
Figure 3. One view of a single *E. coli* core RNAP molecule extracted from the cryo−EM map, with (*Left*) the cryo−EM map alone (blue net), (*Center*) the original (not flexed) *Taq* core RNAP X–ray structure (α^I, yellow; α^{II}, green; β, green−cyan; β', pink; ω, white) [Zhang et al. 1999] superimposed, showing the less than ideal fit of the β' subunit, and (*Right*) the flexed *Taq* X–ray structure superimposed [Darst et al. 2002].

It has been shown that the swinging motion of the clamp, β1 and β2 modules can result in the opening of the claws by ~ 25 Å. Such flexibility is presumably required for the conformational changes necessary in the different transcription steps. In fact, the initial opening of the claws seems to be important during transcription initiation, when DNA must enter the cleft. The subsequent movement of the mobile modules resulting in closing of the claws may help RNAP to hold DNA−RNA hybrid in position during elongation and may be important for the efficient transcription of long genes (processivity) [Murakami et al. 2003 and references therein].

2.1.3. Channels.

The cleft is intersected by three channels: the major channel, named also primary channel, and two minor channels branching off from the major channel to form the "RNA exit channel" and substrate−accessible "secondary channel".

The primary channel is often subdivided into two parts: the active site channel and the downstream DNA channel. The active site channel includes the structural elements essential for catalysis and maintaining the nucleic acid scaffold. The active center is marked by a Mg^{2+} ion chelated at the base of the channel by three aspartate residues from the universally conserved NA**DFDGD** motif of β' region D. The upstream edge of the active site channel is formed by β flexible G−flap domain and the β' lid (β' region B) and zipper domains [Zhang et al. 1999]. The downstream DNA channel is formed mostly by β2 domain (also named downstream lobe) of β pincer and the downstream half of the β' pincer (named β' downstream jaw). This channel accommodates the downstream double−stranded DNA. The walls of RNA exit channel are made of the upstream portions of β and β' pincers including the β' rudder (β' region C), lid and the N−terminal Zinc−finger element (designated β'ZBD), and the β fork loop and G−flap. The wall of the secondary channel is formed by both the β' F-bridge α−helix, that connects the two pincers near the active site, and β' G−loop element, that extends into the cleft (see Figure 1).

2.1.4. Non−conserved domains.

Despite their overall similarity, the *E.coli* and *Taq* RNAPs also have some structural differences. For example, within the β subunit, a 26−residue segment between conserved
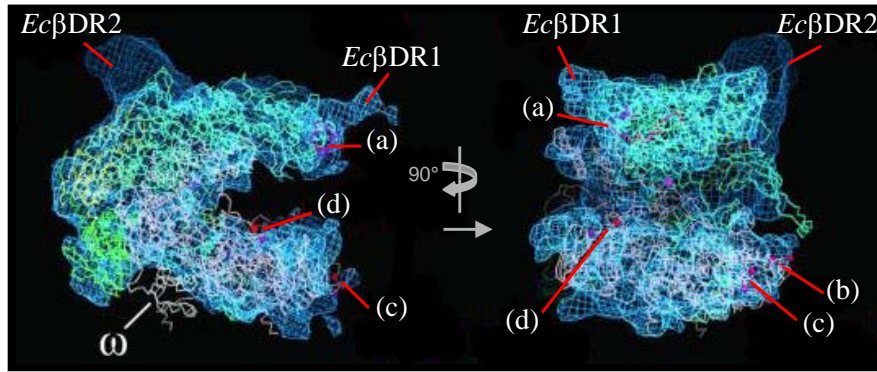
Figure 4. Differences between the *E.coli* core RNAP cryo−EM map (blue net) and the flexed *Taq* core RNAP X−ray structure (α^I, yellow; α^{II}, green; β, green−cyan; β', pink; ω, white). These are *E.coli* βDR1 and βDR2, located near their insertion points with respect to *Taq* β208–233 (colored red and labeled (a)) and *Taq* β803–806. The red atoms labeled (b) denote a gap in the *Taq* β' chain (from *Taq* β'32–68) that includes the β'ZBD universally conserved among prokaryotes. The red atoms labeled (c) denote the gap in the *Taq* β' chain (*Taq* β'156–451) caused by *Taq*β'NCD. The red atoms labeled (d) denote a gap in the *Taq* β' chain (*Taq* β'1242–1249) where *E. coli* β'GNCD is inserted (see Figure 2) [Darst et al. 2002].

regions B and C of the *Taq* subunit is replaced with a 141−residue segment in *E.coli*, a difference of 115 residues, and a 4−residue segment between conserved regions G and H of the *Taq* subunit is replaced with a 103−residue segment in *E.coli*, a difference of 99 residues. These two large insertions in *E.coli* β subunit are named Dispensable Region 1 (*Ec*βDR1) and Dispensable Region 2 (*Ec*βDR2), respectively (see Figure 2). The location of βDR1 and βDR2 in *E.coli* core RNAP was determined by flexible fitting of the high−resolution structure of *Taq* RNAP core enzyme into the low−resolution cryo−EM map of *E.coli* RNAP core enzyme (Figure 4) [Darst et al. 2002]. Both regions comprise separate, isolated domains that protrude from the RNAP surface. It has been reported that the large deletions (more than 200 amino acid residues in some cases) in *Ec*βDR1 and *Ec*βDR2 do not affect RNAP assembly and basic transcription activity *in vitro* [reviewed in Darst et al. 2002]. However, these regions could play a regulatory role in transcription. For instance, *Ec*βDR1 is targeted by the bacteriophage T4 termination factor Alc, which selectively induces premature termination of *E.coli* RNAP transcription on *E.coli* DNA during infection [reviewed in Darst et al. 2002].

Within the β' subunit, a 283−residue segment between conserved regions A and B is present in *Taq* (termed *Taq*β' Non−Conserved Domain, or *Taq*β'NCD), but it is absent in *E.coli*, while a 188−residue segment is inserted in the conserved G−loop element (between region G and G') of the *E.coli* subunit (termed *Ec*β'GNCD). *Ec*β'GNCD consists of two domains termed Sandwich−Barrel Hybrid Motif a (SBHMa) and Sandwich−Barrel Hybrid

Motif b (SBHMb) [Chlenov et al. 2005]. *Ec*β'GNCD is not visible in the cryo−EM map of *E.coli* RNAP core enzyme [Darst et al. 2002], apparently because both the N and the C termini of *Ec*β'GNCD are tethered to the enzyme via long (~ 13 residues) unstructured and flexible linkers that may give a high degree of freedom of motion to the *Ec*β'GNCD [Chlenov et al. 2005]. However, data of crosslinking studies indicated that in the ternary elongation complex (TEC, containing core RNAP, DNA template and RNA product) the SBHMa domain of *Ec*β'GNCD faces the entrance to the secondary channel (allowing it to be reached by RNA backtracked by 10−14 nucleotides) [reviewed in Chlenov et al. 2005]. Supporting crosslinking data, the structural modeling of the *Ec*β'GNCD in the context of TEC revealed that at least five Lys residues of SBHMa are exposed towards the entrance to the secondary channel and lie ~ 40−50 Å from the 3'−OH of the RNA (Figure 5 (a)). The modeled position of SBHMa domain of *Ec*β'GNCD is compatible with the binding mode of the transcript cleavage factors GreA and GreB, suggesting that *Ec*β'GNCD may interact with bound Gre factors and may influence RNAP's propensity to backtrack, affecting its pausing and arresting [Chlenov et al. 2005 and references therein].
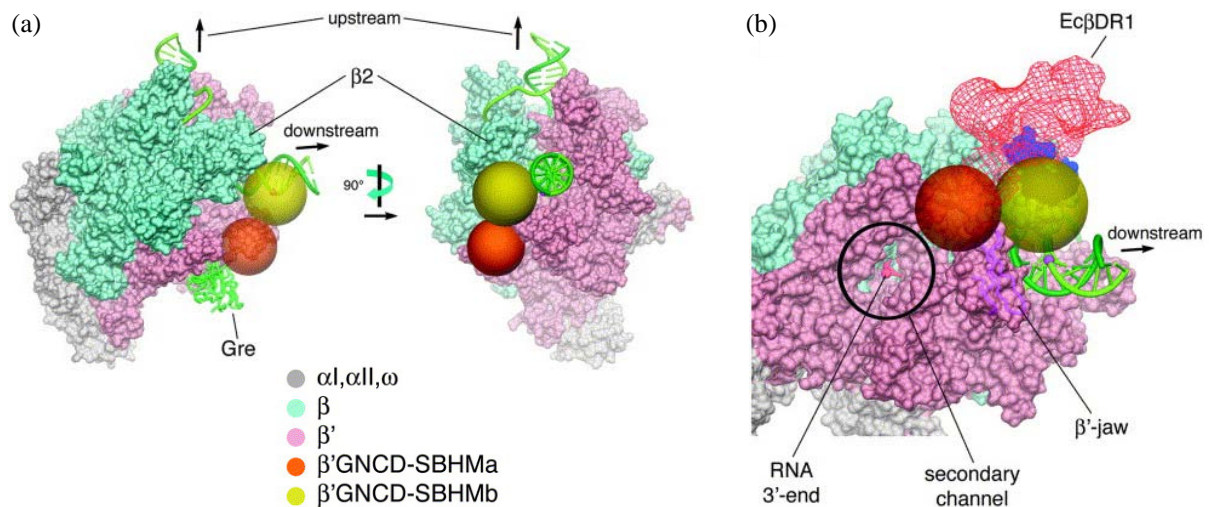
Figure 5. Ternary elongation complex model with *Ec*β'GNCD [Chlenov et al. 2005]. (**a**) Two orthogonal views of the TEC model. The RNAP is shown as a molecular surface. Subunits are color−coded as indicated at the bottom. The DNA is shown as phosphate backbone worms, with DNA bases denoted schematically as bars in duplex regions only. The template strand is colored dark green. The non−template strand is light green. The modeled position of the *Ec*β'GNCD is shown, with SBHMa and SBHMb represented as partially transparent orange and yellow spheres, respectively (the volume of the spheres corresponds to the molecular masses of the domains). In the left view only, the Gre factor interacting with the TEC is shown as a green α−carbon backbone worm. (**b**) View of the TEC model and *Ec*β'GNCD down the secondary channel. The RNA transcript 3'−end (red) is seen. The backtracked RNA 3'−end would extend out the secondary channel and could contact the SBHMa domain (orange). The backbone worm of the β'−jaw domain (magenta) is shown. The cryo−EM derived density corresponding to the *Ec*βDR1 is represented as a red mesh; the insertion point of the *Ec*βDR1 in the β2 domain is colored blue.

The modeled position of *Ec*β'GNCD SBHMb suggests that this domain, together with other parts of β and β' subunits, forms a channel that accommodates the downstream double−stranded DNA. Moreover, the close proximity of the modeled position of *Ec*β'GNCD SBHMb to the cryo−EM derived density corresponding to the *Ec*βDR1 suggests the possibility of functional interactions between these domains of RNAP (Figure 5 (b)) [Chlenov et al. 2005]. A number of evidence indicates that *Ec*β'GNCD plays a significant role in termination. Indeed, it has been found that phosphorylation of Thr[1068] in SBHMb domain by T7 Gp0.7 affects termination *in vitro* and *in vivo*. Furthermore, mutation analyses revealed 13 amino acid residues within *Ec*β'GNCD, which substitutions alter termination [reviewed in Chlenov et al. 2005].

2.2.   Structure of the RNAP holoenzyme.

2.2.1.   σ factor.

In *E.coli*, seven different species of σ subunit have been identified, each responsible for recognition of a specific set of promoters, so that regulation of binding of a σ factor to the core RNAP is a mechanism for altering the pattern of gene expression [reviewed in Ishihama 2000]. Most promoters are recognized by a holoenzyme containing the σ^{70} subunit. These promoters are characterized by two conserved hexamers near nucleotide positions -35 and -10 relative to the transcription start site (+1). The consensus sequences of these two hexamers as read on the non−template strand are, respectively, TTGACA and TATAAT.

Limited proteolysis studies have revealed that the σ factor consists of four distinct structural domains connected by flexible linkers, σ_1, σ_2, σ_3 and σ_4 [reviewed in Campbell et al. 2002]. Amino acid sequence comparison the *E.coli* σ^{70} and σ^{70}–like factors of other bacteria detected within these domains several distinct regions of sequence homology designated as 1.1., 1.2 and 2.1 to 2.4, 3.0 to 3.1, and 4.1 to 4.2, respectively [reviewed in Murakami et al. 2003]. The linker connected σ_3 and σ_4 domains is often named linker domain (LD) and comprises mostly the conserved region 3.2. The conserved regions of the σ^{70} factor are shown in Figure 6.
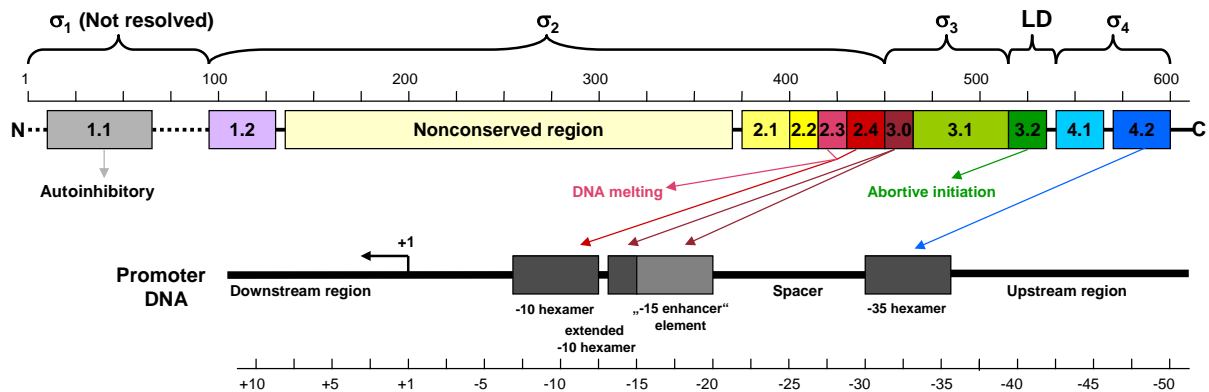
Figure 6. Structural and functional organization of the *E.coli* σ^{70} factor. Top diagram is a linear representation of σ^{70} showing structural domains and conserved regions (numbered and color−coded boxes). Bottom diagram shows DNA promoter regions and interactions made by σ binding regions.

In 1996 Malhotra and Severinova described the structure of only one domain (σ_2 domain) of the *E.coli* σ^{70} [Malhotra and Severinova 1996]. Several years later, Campbell and co−workers tried to crystallize an intact σ^A, the primary σ factor of the thermophile *Thermus aquaticus* (*Taq*) [Campbell et al. 2002]. This turned out to be impossible but *in situ* degradation of σ^A by unknown contaminating protease produced crystallizable fragments diffracting to ~ 2 Å. The crystallographically resolved portion of σ^A consists of three stably folded domains, σ_2, σ_3 and σ_4. Each domain is shown to interact with both RNAP core enzyme and DNA (see Figure 6).

2.2.2. *σ−core RNAP interactions.*

In the high−resolution *Taq* and *Thermus thermophillus* (*Tth*) holoenzyme structures, the σ subunit is visible as a V−shaped structure partially wedged between two pincers of core on the upstream face of the enzyme (Figure 7) [Murakami et al. 2002a, Vassylyev et al. 2002]. Each of the σ domains, as well as the linkers connecting them, makes interactions with RNAP core enzyme. It has been reported that the simultaneous, but independent binding of discrete domains of σ to different parts of RNAP core enzyme results in high−affinity binding (K_d ~ 10^{-9} M), without any one interaction between an individual σ domain and the core being particularly strong [Sharp et al. 1999, Vassylyev et al. 2002, Murakami et al. 2003]. This finding suggests the mechanism by which the σ domains are gradually, one by one, dissociate from core enzyme when RNAP enters the elongation phase of transcription, resulting in the eventual release of the σ factor (for details see below).

As outlined above, most contacts in the σ–core RNAP interface are relatively weak and distributed over a wide area (~ 8500 $\mathring{A}^{2}$). For the most part, these contacts are limited to the β and β' subunits of RNAP core enzyme. The strongest interaction is observed between σ_2 and β' coiled–coil domain on the upper edge of the clamp. Less strong interactions are observed between σ_4 and β G–flap, and between σ_3 and $\beta 1$ [Murakami et al. 2002a, Vassylyev et al. 2002].

2.2.3. *Conformational changes upon holoenzyme formation.*

Upon holoenzyme formation, both the core RNAP and the σ factor undergo conformational changes. In fact, some regions (β'ZBD, β' zipper, β' lid domains and *Taq*β'NCD) that were disordered in the core RNAP structure become ordered in the

holoenzyme, whereas other structural modules of core subunits move so that their positions change by 2 to 12 Å [Murakami et al. 2002a, Vassylyev et al. 2002]. For instance, in the *Taq* RNAP, rotation of the clamp domain of β' pincer and β1 domain of β pincer towards the active–site channel upon core–to–holoenzyme conversion leads to the closing of the claws by ~ 10 Å. The interaction of σ_2 and σ_3 domains with the mobile clamp and β1, respectively, suggests that these σ domains could play a role in opening and closing of the RNAP claws during different stages of transcription initiation. The interaction with σ_4 domain shifts the β G–flap by ~ 5–6 Å relative to its position in the core RNAP [Murakami et al. 2002a]. It has been shown that in the holoenzyme the clamp domain, and σ_2 bound to it, form a rigid mobile module (clamp–σ_2), whereas the β G–flap domain, and σ_4 bound to it, form yet another rigid mobile module (flap–σ_4) [Murakami et al. 2002b]. It has been proposed that the independent movement of the flap–σ_4 and the clamp–σ_2 modules allows to modulate the distance between σ_4 and σ_2 which recognize the -35 and -10 promoter hexamers, respectively (see below). Such plasticity is likely to be essential for the ability of RNAP to accommodate promoters containing variably spaced -35 and -10 hexamers.

The conformational changes occurring in σ factor upon its binding to core RNAP are mentioned below (Chapter: **3.1. Structure and role of distinct σ regions in transcription initiation.** *3.1.7. Region 1.1.*).

3.   RNAP– promoter interactions.

Two crystal structures provided information on how RNAP recognizes and binds promoter DNA: the 2.4–Å–resolution structure of σ_4 domain of $Taq\sigma^A$ in complex with -35 hexamer DNA (from position -37 to -26) [Campbell et al. 2002], and the 6.5–Å–resolution structure of *Taq* RNAP holoenzyme in complex with fork–junction promoter DNA [Murakami et al. 2002b]. The fork–junction DNA contained the double–stranded DNA from position -41 through the -35 hexamer up to the first base pair of the -10 hexamer (the -12 bp), and only single–stranded non–template DNA from -11 to -7, and no downstream DNA.

Together with available genetic and biochemical data, these structural data showed that all sequence–specific contacts with core promoter elements are mostly mediated by the

conserved regions of the σ factor and they defined the role of individual conserved regions of σ during different steps of transcription initiation.

3.1. Structure and role of distinct σ regions in transcription initiation.

3.1.1. Regions 4.2 and 4.1.

The σ_4 domain, which includes conserved regions 4.1 and 4.2, contains four α–helixes, which are arranged as a pair of helix–turn–helix motifs. Overall, the σ_4 domain is C–shaped, with concave pocket coated almost totally with hydrophobic residues of region 4.1 [Murakami et al. 2003]. It was identified that substitutions for four closely spaced residues (Glu[555], Arg[562], Phe[563], and Ile[565]) in region 4.1 and two residues (Ile[590] and Leu[598]) in region 4.2 of *E.coli* σ^{70} hinder the ability of σ to bind core RNAP [Sharp et al. 1999]. Most of these mutants occur in or around the edge of the hydrophobic pocket. Therefore, it was suggested that σ_4 domain latches onto the β flap domain of core RNAP through this hydrophobic pocket [Campbell et al. 2002].

A large number of evidence indicates that σ_4 domain determines interactions with -35 hexamer [Siegele et al. 1989, Moyle et al. 1989, Campbell et al. 2002]. These interactions occur mostly through several conserved residues of the helix–turn–helix motif of *E.coli* σ^{70} region 4.2. Among these, four key residues (Arg[584], Glu[585], Arg[586], and Gln[589]) are responsible for base–specific DNA recognition. On the template strand, the side chain of Arg[584] interacts with -31G and -30T through hydrogen bonds and van der Waal's contacts, respectively. Glu[585] interacts directly with -33C and makes a water–mediated hydrogen bond with -34A. On the non–template strand, Arg[586] and Gln[589] establish van der Waal's contacts and hydrogen bond with -35T. Additionally, several residues of both conserved regions 4.1 and 4.2 provide nonspecific interactions with the ribose and phosphate backbone of the non–template strand from -35 to -38 and the template strand from -31 to -33. Among these, Arg[588] makes water–mediated interactions with the phosphate backbone of the template strand at positions -32 and -33 and makes van der Waal's contact with -32T. Furthermore, Arg[588] appears to be a key in positioning universally conserved Glu[585].

In addition, it was revealed that the σ_4 domain frequently serves as a target for transcription activators which bind at or near the -35 hexamer. Mutation analyses identified

two clusters of σ_4 amino acid residues implicated in the interactions with activators [Campbell et al. 2002 and references therein].

3.1.2. *Region 3.2.*

The linker domain LD, comprising primarily σ region 3.2, intervenes between the σ_3 and σ_4 domains, and has mostly an extended, unfolded conformation [Murakami et al. 2002a, Vassylyev et al. 2002]. The LD contains several conserved acidic amino acid residues, giving an overall charge of -8 to -9 among σ^{70}-like factors. Roughly at its midpoint, the LD forms a hairpin loop that protrudes into the RNAP active–site channel, between the β' lid and rudder. The rest of the LD is located within the RNA exit channel, with the negatively charged LD apparently serving as a molecular mimic or molecular placeholder for RNA.

It has been observed that a *Taq* holoenzyme with C–terminally truncated variant of σ^A, lacking both the region 3.2 and the σ_4 domain, retains weak transcription activity on extended -10 promoters. Activity can be increased to a level comparable with wild–type *Taq* holoenzyme by increasing the concentration of the initiating dinucleotide, suggesting that the absence of σ regions 3.2 – 4.2 substantially decreases the apparent K_m for the initiating substrate [Campbell et al. 2002]. The proximity of the region 3.2 to the active site proposes that it may directly or indirectly participate in binding the initiating nucleotide.

The LD occupies the same space as the exiting RNA transcript of the elongation complex. Therefore, it has been hypothesized that, in the initiating complex, the LD must be displaced from the RNA exit channel upon synthesis of a ≥ 9–11nt RNA product [Mekler et al. 2002]. Competition between the LD and growing RNA transcript for binding site in the RNAP would hinder the initiating process and destabilize the transcripts, leading to abortive initiation. Several lines of evidence support the important role of LD in abortive initiating [Murakami et al. 2002a], however it was noted that such structural impediment model can only account for a basal level of abortive initiating that probably occurs in a promoter sequence–independent but transcript length–dependent manner, but it can not account for the widely different patterns of abortive initiation observed on distinct promoters [Hsu et al. 2003, Vo et al. 2003].

Moreover, the mechanism of σ dissociation from RNAP, when the enzyme enters the elongation phase of transcription, was suggested based on LD location. A steric clash of the

growing RNA transcript with LD would eventually lead to complete displacement of LD from the RNA exit channel. Once the RNA transcript grows past 16–17 nt, it clashes with σ_4 and finally causes disruption of the interactions between β G–flap and σ_4 domains. Once the contacts with LD and σ_4 are lost, the interactions with the σ_3 and σ_2 are lost slowly and stochastically [Mekler et al. 2002, Murakami et al. 2002a, Mooney et al. 2005].

3.1.3. Region 3.0 (first named as region 2.5).

Various studies showed that the σ conserved regions 2.2 to 3.0 are implicated in the interactions with the -10 hexamer and adjacent DNA sequences. In *E.coli* σ^{70} region 3.0, amino acid residues C–terminal to position 454 are involved in recognition of nucleotides from -14 to -20 [Barne et al. 1997, Bown et al. 1999]. It has been reported that the extended -10 element (consensus sequence -15 5'-TGnTATAAT-3' -7) is recognized by two residues, His^{455} and Glu^{458} of *E.coli* σ^{70}, and that Glu^{458} is critical for recognition of the 5'-TG-3' motif at extended -10 sequence, whereas His^{455} appears to play a nonspecific DNA binding role [Barne et al. 1997]. Furthermore, it has been suggested that the amino acid residues of σ region 3.0 may be involved in the interactions in the major groove of the "-15 enhancer" element (-17/-12 segment) [Liu et al. 2004].

In the crystal structure of *Taq* RNAP holoenzyme complexed with a fork–junction promoter DNA fragment, His^{278} and Glu^{281} of *Taq* σ^{A}, corresponding to His^{455} and Glu^{458} of *E.coli* σ^{70}, are exposed on the surface of the σ region 3.0 α–helix, facing the major groove of the extended -10 DNA [Murakami et al. 2002b]. Glu^{281} may make base-specific interactions with non–template strand T at position -15, whereas His^{278} may interact nonspecifically with the phosphate backbone of the non–template strand at positions -17/-18.

3.1.4. Region 2.4.

Suppression analyses have implicated amino acid residues Gln^{437}, Thr^{440} and Arg^{441} of *E.coli* σ^{70} region 2.4 in the recognition of nucleotides -13 and -12 [reviewed in Barne et al. 1997, Siegele et al. 1989]. Additionally, analysis of the effects of serine substitution for almost all hydrophilic amino acids within the region between Arg^{436} and Ile^{452} showed that three arginine residues (Arg^{436}, Arg^{441}, and Arg^{451}) appear to be involved exclusively in duplex contacts from -12 upstream and are aided by nearby residues (Asp^{445}, Gln^{446}, and

Arg[448]). Moreover, it has been observed that not only residues 445, 446 and 448 may contact duplex DNA, but this region of RNAP most probably assists the reorganization of the protein–DNA complex during opening [Fenton et al. 2002].

In the structure of the *Taq* holoenzyme– fork-junction DNA complex, Gln[260] and Asn[263], corresponding to Gln[437] and Thr[440] in *E.coli* σ^{70}, are exposed, facing the major groove of the DNA near -12 (the only double–stranded portion of -10 hexamer) and could interact with either template strand A or non–template strand T [Murakami et al. 2002b].

3.1.5. *Regions 2.3 and 2.2.*

A large body of data has implicated the highly conserved aromatic residues Tyr[425], Tyr[430], Trp[433], and Trp[434] in *E.coli* σ^{70} region 2.3 as potentially involved in promoter melting, at least partly via sequence–specific recognition of the non–template strand [reviewed in Fenton et al. 2002]. Mutation analyses identified that Tyr[430] and Trp[433] as being particularly important for the initiation of DNA opening, suggesting that these residues interact with the bases at the -11 and/or -10 positions, whereas Tyr[425] and Trp[434] are critical for duplex DNA binding [Fenton et al. 2002]. Analysis of the effects of serine substitution for Tyr[425] and Trp[434] confirmed, that Tyr[425] affects recognition of DNA duplex downstream of the -12 nucleotide, whereas Trp[434] reflects interaction at -12 and further upstream. Moreover, it has been reported that universally conserved basic residues Lys[414] and Lys[418] in *E.coli* σ^{70} regions 2.2 and 2.3 are important for promoter binding. The role of the two positively charged residues would be to hold the promoter DNA in the proper orientation and allow the aromatic amino acid residues Tyr[430] and Trp[433] to nucleate the strand separation process, likely by flipping the highly conserved non–template strand A at -11 out of the helix by a mechanism that is not yet fully understood. However, it was proposed that Trp[433] may participate in "forcing" the flipped base out of the DNA duplex, whereas Tyr[430] would interact with the flipped out base subsequent to the action of Trp[433] on duplex DNA [Tomsic et al. 2001]. Together, these data indicate that σ region 2.3 is involved in promoter melting and it has also a role in closed complex formation along with regions 2.4 and 3.0.

In the structure of the *Taq* holoenzyme– fork-junction DNA complex, amino acid residues Phe[248], Tyr[253], and Trp[256], corresponding to Tyr[425], Tyr[430], and Trp[433] in *E.coli* σ^{70}, appear ideally positioned to interact with unpaired bases of the single–stranded tail of the non–template strand DNA. Phe[248] is closest to bases at the -8/-9 positions, whereas Tyr[253] is

closest to bases at the -9/-10. Trp^{256} is positioned to stack on the exposed face of the -12 base pair and may also be able to interact with the exposed base at the -11 position. Universally conserved basic residues Arg^{237} and Lys^{241} in *Taq* σ^A, corresponding to Lys^{414} and Lys^{418} in *E.coli* σ^{70}, are positioned to interact with the negatively charged DNA backbone of the non–template strand at the -13/-14 positions (Arg^{237}) or at -15 (Lys^{241}) [Murakami et al. 2002b].

On the basis of genetic analysis, the most highly conserved σ region 2.2 is considered to be an important determinant of core RNAP binding [Joo et al. 1997, Sharp et al. 1999]. σ^{70} region 2.2 has been shown to interact with the coiled–coil within β' pincer, and residues Leu^{402}, Asp^{403}, Gln^{406}, Glu^{407}, Asn^{409}, and Met^{413} within region 2.2 and residues 275, 295, and 302 within the β' coiled–coil have been established to be involved in this interaction [Sharp et al. 1999, reviewed in Mekler et al. 2002].

3.1.6. Region 2.1.

Several lines of evidence indicate that the conserved σ region 2.1 is involved in core RNAP binding [reviewed in Sharp et al. 1999]. It has been observed that derivatives of *E.coli* σ^{70} and σ^{32} lacking region 2.1 are unable to bind to core RNAP. Mutations in region 2.1 of the σ^{70} family member *Bacillus subtilis* σ^E factor have been also shown to cause defects in binding to both *B.subtilis* and *E.coli* core RNAP. In addition, the σ^{54} factor, which is unrelated in sequence and mechanism to the σ^{70} protein family, has the only short stretch of amino acid residues that bears resemblance to σ^{70} residues 381-385 within region 2.1. This portion of σ^{54} is implicated in core RNAP binding as well.

3.1.7. Region 1.1.

The structural model of the σ subunit lacks the disordered N–terminal domain, which includes the poorly conserved region 1.1. This is a self–inhibitory domain, which is known to mask the DNA binding determinants of the σ factor in the absence of the core RNAP. In fact, in free σ^{70}, regions 2.4 and 4.2 are incorrectly positioned to interact with -10 and -35 hexamers of the promoter DNA. Binding of the σ^{70} factor to core RNAP results in the repositioning of regions 1.1, 2.4, and 4.2, allowing promoter binding [reviewed in Vuthoori et al. 2001]. Moreover, it has been shown that region 1.1 can accelerate open complex formation

at some promoters [reviewed in Dombroski 1997, Vuthoori et al. 2001]. Fluorescence resonance energy transfer (FRET) measurements of the distances between different specific sites on σ^{70} factor and core subunits provided direct evidence that in the *E.coli* RNAP holoenzyme, region 1.1 is located deep within the RNAP active–site channel. Upon formation of the promoter open complex, however, region 1.1 is displaced outside the channel and is positioned to interact with the tip of downstream lobe of the β pincer [Mekler et al. 2002], which explains how region 1.1 can affect the kinetics of open complex formation. It was proposed that the positioning of region 1.1 in the RNAP active–site channel may widen the channel to facilitate the entry of double–stranded DNA, but the precise role of region 1.1 in transcription initiation is not understood.

In summary, these data clearly indicate the central importance of σ factor in transcription initiation. After binding to RNAP core enzyme, σ factor directs the process of transcription initiation by first locating the promoter through sequence–specific recognition of -35 and -10 hexamers. Then the σ factor plays a key role in promoter melting, as well as in promoter clearance. Furthermore, σ factor is a target for transcription activators that bind to promoter regions overlapping the -35 hexamer.

3.2.   Open complex structure.

Based on the crystal structures of the *Taq* holoenzyme– fork–junction DNA complex and the *Taq*σ^A_4 domain––35 element DNA complex, as well as the known structure of the B form of DNA, and based on the numerous data of footprinting and crosslinking studies, Murakami and co–workers proposed the structural model of the open complex, that includes both strand of DNA from -60 to +25 [Murakami et al. 2002b].

In this model, the interactions of the upstream portion of double–stranded DNA, from -60 to -17, with αCTDs, σ region 4.2 and β'ZBD result in DNA wrapping around the RNAP (Figure 8 (a)). The conformation of upstream DNA is characterized by the three bends: at around -45, in the -35 region at about -35 (36°) and in the spacer region at -25 (8°). Moreover, unlike the model proposed for the closed complex, in the open complex, at -16 the DNA makes another sharp bend (37°C) toward the holoenzyme. The two DNA strands then separate at position -11, and take drastically different paths downstream for ~ 15 nucleotides, until they reanneal at position +3, thus creating the "transcription bubble".
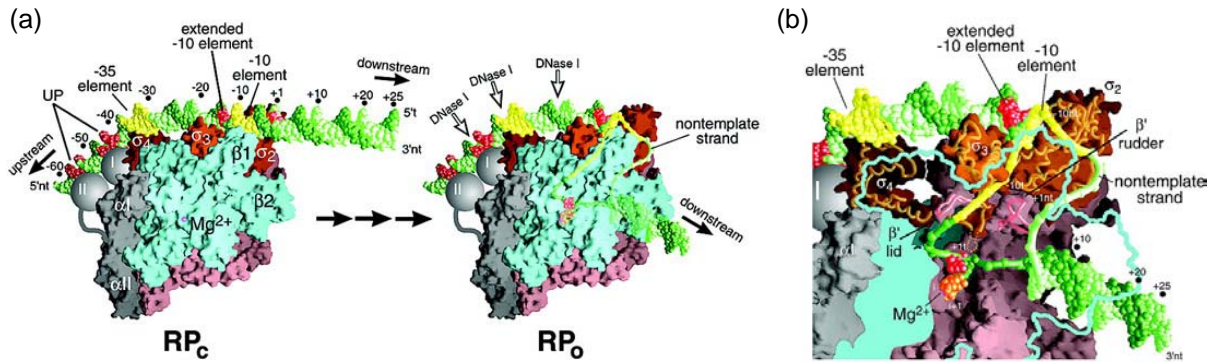
Figure 8. Structural models of the closed (RP_c) and open (RP_o) complexes [Murakami et al. 2002b]. The RNAP is shown as a molecular surface. Subunits are color−coded as follows: α^I, α^{II}, gray; β, cyan; β', pink; σ, orange. The DNA is shown as phosphate backbone worms with only the phosphate atoms visible. The template strand is green, the non−template strand is light green, except for the -35 and -10 elements, which are yellow; and the UP element, extended -10 element, and transcription start site on the template strand (+1) are red. The upstream and downstream directions on the DNA are indicated and labeled. The possible disposition of the αCTDs (drawn as gray spheres, labeled "I" and "II") on the UP element is shown. (**a**) Models of RP_c (left) and the final RP_o (right). The arrows in between denote that several intermediate steps exist along the pathway between these two states. The β subunit is rendered partially transparent to reveal the RNAP active site Mg^{+2} (magenta sphere) inside the main channel and the transcription bubble and downstream DNA enclosed inside the primary channel in RPo. In RPo, RNA occupying the i and i+1 sites is shown as orange atoms. (**b**) Magnified view of RP_o, showing the details of the core promoter interactions, transcription bubble, and downstream DNA. Obscuring portions of the β subunit in front have been removed (the outline of β is shown as a cyan line) to reveal the structural elements inside the primary RNAP channel. The template strand DNA within the transcription bubble is directed through a protein tunnel framed by σ_2 and the LD of σ factor underneath, an α−helix of σ_3 and the β' lid on one side, σ_2 and the β' rudder on the other side, and a β1 domain in front.

The single−stranded template DNA must enter the active site of the protein in order to base pair with initiating rNTPs. To reach the active site, the template strand passes through a tunnel that is completely enclosed by σ_2 and σ_3 domains, the β' lid and rudder, and the β1 (Figure 8 (b)). More specifically, the entrance to this tunnel is lined with highly conserved basic amino acid residues of σ regions 2.4 and 3.0, which presumably play a key role in directing the negatively charged single−stranded template DNA into a tunnel. The DNA then moves between the active site wall and σ LD hairpin loop, juxtaposing DNA +1 position to the catalytic center.

The single−stranded non−template DNA, at first, crosses the σ_2 domain making interactions primarily with highly conserved aromatic residues of σ regions 2.3, and then continues its path in a groove formed between two lobes of β pincer, β1 and β2.

The downstream double−stranded DNA from +5 to about +12 is enclosed in downstream DNA channel of RNAP.

4. Contribution of discrete promoter regions for optimal promoter activity.

Promoters direct not only the site of transcription initiation but also its rate. The strength of promoter in initiating productive transcription is primarily determined by the balance of promoter binding and activation (isomerisation from closed to open promoter complex), and RNA chain initiation and promoter escape resulting in transition of transcript initiation to elongation. The binding affinity of promoter to RNAP and the rate of activation correlate to a large extent with the degree of similarity of the -35 and -10 hexamers to their consensus sequence and with the length of the spacer between them (usually 17 bps). For instance, mutations in -35 sequence appear to affect both the RNAP binding and the subsequent isomerisation step resulting in the open complex formation [Shin and Gussin 1983, Hawley and McClure 1982]. It has also been shown that changes in the length of the 17 bp spacer separating the -35 and -10 hexamers of the λP_R promoter primarily result in a decrease of efficiency of conversion from closed to open complex [McKane and Gussin 2000]. More specifically the -10 sequence plays a critical role at all steps in the pathway leading to formation of final open complex [Fenton et al. 2001, McKane et al. 2001, Heyduk et al. 2006].

4.1. Function of the bacterial -10 hexamer.

The first promoter element to be discovered (first named as "Pribnow box") and the most conserved is the -10 hexamer. On the non−template strand the consensus sequence of -10 hexamer is <u>TA</u>TAA<u>T</u>, from -12 to -7, where the underlined nucleotides are 80% conserved and the others are 60% conserved [Lisser et al. 1993]. The -10 hexamer is recognized predominantly by the σ_2 domain of σ^{70} and is involved both in initial promoter binding and in subsequent promoter melting leading to formation of the open complex.

Competition binding studies demonstrated that the upstream half of the -10 hexamer (TAT---) is dominant for DNA duplex recognition and binding by partial polypeptide (lacking σ_1 domain) of the σ factor alone [Dombroski 1997]. The electrophoretic mobility–shift assays (EMSA) of the RNAP binding to promoter fragments characterized by different substitutions in the -10 sequence confirmed that the first two highly conserved positions of the -10 hexamer (TA----) are most important for general duplex binding by RNAP with all other positions

making an accessory contribution [Fenton et al. 2001]. For example, it has been established that mutation of -10 T:A to G:C in the *lacUV5* promoter fragment gives a 3–fold reduction in the extent of promoter binding [Fenton et al. 2001]. The same mutation substantially decreases the occupancy of the λP_R promoter [McKane et al. 2001]. Additionally, analysis of the effects of A:T, C:G, and G:C substitutions for -10 T:A in the λP_R promoter results in finding that all three mutations at position -10 primarily affect isomerisation step in open complex formation, which precedes DNA strand separation [McKane et al. 2001], and is thought to involve both a conformational change in RNAP, and DNA untwisting [reviewed in McKane et al. 2001]. However, it has been reported that mutation from -10 T:A to G:C may also inhibit promoter melting at 37°C [McKane et al. 2001].

Numerous studies indicated that promoter melting is a stepwise process that can be divided into at least two steps: nucleation of melting involving a very small subset of promoter region, which eventually becomes single–stranded in the open complex, and subsequent expansion of the melted region roughly to position +3 [reviewed in Heyduk et al. 2006]. It was determined that the base–specific interactions of the polymerase with consensus non−template strand adenine at position -11 are directly involved in facilitating DNA strand separation by stimulating initial melting nucleation at the upstream edge of the -10 hexamer [Heyduk et al. 2006]. The exact mechanism by which these interactions could facilitate promoter melting nucleation is not yet understood. However, it was suggested that initiation of DNA strand separation by RNAP could involve a base–flipping event. RNAP could either actively promote the flipping of -11A out of the DNA helix or passively take advantage of spontaneous dynamics of this base and use the -11A–specific interactions to stabilize extrahelical conformation of the base [Tomsic et al. 2001, Heyduk et al. 2006].

The EMSA experiments on DNA fork probes, in which the -11 to -7 sequences are present in single–stranded form, revealed that specific nucleotide sequences within the non−template strand are critical for the conversion of the RNAP−DNA complex to a form that resist heparin challenge [Guo et al. 1998, Fenton et al. 2001]. The strongest effects on ability of RNAP to form a heparin–resistant complex were observed for certain substitutions for -12T and -11A. Furthermore, any mutation of non−template strand T at position -7 in *lacUV5* promoter fragment was found to lead to a 10–fold reduction in the level of heparin–resistant complex [Fenton et al. 2001]. However, the same substitutions in the $\lambda P_{R'}$ promoter fragment have almost no effect on open complex formation [Roberts et al. 1996] and on the mode of

fork–junction probe binding [Guo et al. 1998]. Additionally, studies of fork–junction templates showed that, when the melted DNA encompasses both the -10 hexamer and the start site, even the substitutions in the most important positions within the consensus -10 hexamer have very little effect on complex resistance to heparin [Fenton et al. 2001].

Taken together, these data indicate that the -10 hexamer has its primary effects on the binding of DNA duplex by RNAP, on the stabilization of the closed complex and subsequent isomerisation events *via* interactions in both its double–stranded and single–stranded form as DNA melting takes place, rather than on the stability of the final functional complex.

4.2.   UP element, interaction with α subunit.

It has been assumed that optimal transcription activity could be achieved by combinations of promoter elements, including not only the -35 and -10 hexamers, but also the sequences outside the core promoter region. In agreement with this assumption, it has been shown in footprints that RNAP protects regions both upstream and downstream of the -35 and -10 hexamers [Schickor et al. 1990, Ozoline et al. 1995, Craig et al. 1995] and that, in some promoters in *E.coli* and in other bacterial species, sequences upstream of the -35 hexamer increase transcription in the absence of additional factors [Rao et al. 1994, reviewed in Nikerson et al. 1995, Estrem et al. 1998, Ross et al. 1998]. These upstream sequences are generally A+T–rich, and some contain multiple A–tracts in phase with the DNA helical repeat (phased A–tracts).

4.2.1. *A–tract sequences and α subunit recognition.*

Phased A–tracts inserted upstream of the -35 region in various promoter constructs were found to increase transcription from the promoter, which are rate–limited in complex formation, by stimulation of RNAP binding to promoter DNA [Ellinger et al. 1994, Aiyar et al. 1998]. It has been reported that an A–tract placed upstream of the *E.coli lac* promoter accelerates transcription 5– to 20–fold *in vivo*, depending on the position of A–tract (A–tract functions best when positioned close to the -35 hexamer rather than one helical turn further upstream) [Aiyar et al. 1998], and that *lac* promoter activity is increased progressively by insertion of one, two or three A–tracts upstream of -35 hexamer [reviewed in Ellinger et al. 1994]. However, a single A–tract placed upstream of P_{SI} promoter is as effective as three

[Ellinger et al. 1994]. These findings indicate that response to multiple A–tracts can differ between promoters.

It has been also observed that A–tracts fail to stimulate expression when promoters are transcribed with RNAPs lacking the DNA–binding domain of α subunit, and protection of the A–tract sequences in footprints requires the αCTD [Aiyar et al. 1998]. These data, along with other studies, suggested that direct effects of A–tracts on transcription result from DNA–α subunit interactions, rather than from the macroscopic DNA bending associated with the multiple in phase A–tracts. However, it was proposed that the unusual structural features of A–tract DNA (for example, narrow minor groove width; high degree of propeller twisting of bases) might facilitate α subunit binding.

4.2.2. rrnB P1 UP element.

In the *E.coli* rRNA promoter *rrnB P1*, an A+T–rich sequence functions as a promoter recognition element, the UP element, which increases transcription 30– to 70–fold generally by stimulation of initial closed complex formation, although it might also affect process after DNA binding by RNAP [Rao et al. 1994, Aiyar et al. 1998, Estrem et al. 1998]. The *rrnB P1* UP element also has an approximately 2– to 10–fold effect on the isomerisation rate constant of the λP_{PM} promoter in the chimeric constructs [Tang et al. 1996, Strainic et al. 1998]. The extent to which the presence of UP element accelerates open complex formation was found to be temperature–sensitive and depend on the sequence of the core promoter. A more likely possibility is that the mechanism whereby the UP element stimulates open complex formation is promoter dependent.

4.2.3. Full UP element and subsite consensus sequences.

The optimal (consensus) UP element sequence was identified by *in vitro* selection for upstream sequences that promote RNAP binding to the *rrnB P1* promoter, followed by *in vivo* screening for high transcription activity using promoter–*lacZ* fusion [Estrem et al. 1998]. The consensus full UP element sequence contains an alternating A– and T–tracts (-59 5'–nnAAA(A/T)(A/T)T(A/T)TTTTnnAAAAnnn–3' -38), and increases promoter activity about 330–fold *in vivo*, 5– to 10–fold more than the natural *rrnB P1* UP element.

The results of these studies together with other data [reviewed in Estrem et al. 1999] suggested that UP element contains two parts: an 11−bp distal subsite, centered at about position -52, and a 4−bp proximal subsite, centered at about position -42. Mutational analyses indicated that specific positions within the consensus sequence (-53 to -51 and -43 to -41) are most critical to function and that each UP element subsite can stimulate transcription alone, with the proximal subsite conferring larger effects on the *rrnB P1*core promoter (>100−fold *in vivo*) than the distal subsite (~15−fold) [Estrem et al. 1999]. Consensus sequences for the distal and proximal subsites were then estimated individually by *in vitro* selection and *in vivo* screen. The sequences of the consensus distal and proximal subsites are both purine–rich but are significantly different (-57 5'−A(A/T)(A/T)(A/T)(A/T)(A/T)TTTTT−3' -47 versus -46 5'−AAAAAA(A/G)n(A/G) −3' -38). Furthermore, the sequence of the consensus proximal subsite differs from the sequence of the corresponding segment of the consensus full UP element, and proximal subsite substitution (at most critical positions -43 to -41) have larger effects on transcription when the consensus distal subsite is absent [Estrem et al. 1998, Estrem et al. 1999]. These findings suggested that the proximal subsite plays a different role alone, than in a full UP element.

To estimate the frequency of potential UP elements in naturally occurring promoters, the *E.coli* genome sequence was screened for matches to the consensus full UP element and to the consensus proximal or distal subsite [Estrem et al. 1999]. Several conclusions were drawn from this analysis. First, numerous *E.coli* promoters contain single near−consensus subsites. Second, promoters with a close match to consensus in only one subsite are more common then promoters with near−consensus full UP element. Third, stable RNA (rRNA and tRNA) promoters are significantly enriched for UP elements.

4.2.4. *UP elements of different strengths.*

UP elements have been identified in many bacterial and phage promoters and can function with holoenzymes containing different σ factors [Fredrick et al. 1995, Ross et al. 1998, reviewed in Estrem et al. 1998]. It has been reported that the effect of the UP element on a promoter activity correlates generally with its degree of similarity to the UP element consensus sequence. The *rrnD P1* and *rrnB P1* UP elements both greatly stimulate transcription (90−fold and >30−fold, respectively) and contain relatively good matches to the consensus [Ross et al. 1998]. UP elements in certain other promoters exhibit poorer matches

to the consensus and increase transcription only 2– to 13–fold (*rrnB P2*, *PNA II* and *merT* [Ross et al. 1998], λP_{L2}, phage *Mu Pe* [reviewed in Estrem et al. 1998]).

In addition, the positioning of an UP element with respect to the core promoter affects the promoter function. For example, RNAP forms a heparin–resistant nonproductive initiation complex at the *malT* promoter which has an A+T–rich sequence that begins 9–bp upstream of the -35 hexamer. The deletion of 5–bp between the A+T–rich sequence and -35 hexamer increases the promoter activity by stimulation of productive complex formation [Tagami and Aiba 1999]. The same effect of the location of the *rrnB P1* UP element on *malT* core promoter was observed.

Together, these results support the model that bacterial promoters consist of at least three RNAP recognition modules, not just -35 and -10 hexamers. In this general view, promoter activity correlates positively with the number of promoter elements present and positioned correctly, with the extent of similarity of each element to the consensus, and with the relative importance of individual matching positions within each module. In addition, there may be negative contributions from nucleotides least favored at specific positions. In this context, the effectiveness of a particular UP element will be determined not only by its similarity to the UP element consensus, but also by the strength and kinetic characteristics of the core promoter. In extreme cases, an increased match to consensus may decrease transcription by reducing promoter clearance [Ellinger et al. 1994, Strainic et al. 1998].

4.2.5. *Sequence–specific αCTD – UP element interaction.*

Detailed information about the α subunit–UP element interaction is based primarily on footprinting studies. Footprints of RNAP on *rrnB P1* and other promoters extend about 60–bp upstream of transcription start site, and protection upstream of ~ -40 is attributable to interactions with αCTD [Ross et al. 1998, Burns et al. 1999]. The αCTD binds to the *rrnB P1* UP element as a purified peptide (although with lower affinity than the intact α subunit and with much lower affinity than RNAP holoenzyme), confirming its identity as an independent domain responsible for UP element binding [Blatter et al. 1994].

Mutational analyses identified seven amino acid residues in the αCTD critical for DNA binding [reviewed in Ross et al. 2001]. These residues reside in two helix–hairpin–helix (HhH) motifs that interact with UP element DNA in and across the minor groove [Ross et al.

2001]. A high resolution X−ray structure of αCTD bound to DNA confirmed the roles of the two HhH motifs of αCTD in DNA recognition, and of five of the seven crucial amino acid residues (Arg[265], Asn[268], Gly[296], Lys[298] and Ser[299]) in direct or water−mediated DNA contacts [Benoff et al. 2002]. It has been shown that mutations in the αCTD that prevent DNA binding eliminate UP element function [Ross et al. 1998, Estrem et al. 1998, Estrem et al. 1999].

4.2.6. *Sequence−independent αCTD – upstream DNA interaction.*

In addition to sequence−specific interaction with UP element, αCTD also interacts nonspecifically with the upstream DNA in promoters that lack UP elements. These include the well characterized *lacUV5* and *λP$_R$* promoters, in which upstream sequences do not closely match the UP element consensus and do not function in a sequence−specific manner [Ross et al. 1998]. Replacement of these upstream sequences with other non–UP–element sequences has negligible (<2−fold) if any effect on promoter activities *in vitro* and *in vivo* [Ross et al. 1998, Cellai et al. 2007]. Nevertheless, αCTD interactions with a promoter upstream region were indicated in the protein–DNA photocrosslinking experiments with *lacUV5* [Naryshkin et al. 2000] and αCTD–dependent protection of DNA sequences upstream of ∼ -40 was observed in RNAP footprints of *lacUV5* and *λP$_R$* [Ross et al. 1998, Ross et al. 2005]. Furthermore, removal of αCTD or upstream DNA diminishes the overall association rate for RNAP with *lacUV5* promoter ≈ 10−fold. A major part of this effect is attributable to a decrease in the rate–limiting conformational change (isomerisation) in open complex formation [Ross et al. 2005].

Kinetic studies with the *λP$_R$* promoter and *E.coli* RNAP also showed that the presence of DNA upstream of base pair -47 greatly increases the rate of open complex formation [Davis et al. 2005]. In addition, DNase I footprints revealed that the presence of upstream DNA permits downstream DNA to fully enter the active site cleft of RNAP early in the process. It was proposed that upstream DNA interactions with αCTD at *λP$_R$* promoter directly influence movements in the transcriptional machinery that place the start site in the jaws of RNAP.

At two other promoters that lack UP elements, *λP$_{RM}$* and *galP1–SUB(-93 to -35)*, higher levels of open (KMnO$_4$–reactive) complexes were observed with wild type RNAP relative to RNAP lacking αCTD [Tang et al. 1996, Burns et al. 1999]. These data confirm that

the αCTD has a functional role in promoting open complex formation which is independent of its ability to form a specific protein–DNA complex.

Together, these findings indicate that αCTD–upstream DNA interactions are likely to play a significant role in RNAP association at many or all promoters, not merely promoters that have UP elements or that use transcription factors.

4.2.7. *Arrangement of α subunits on upstream region of DNA.*

RNAP contains two α subunits and therefore two αCTDs. Although αCTD purifies as a dimer in solution [Blatter et al. 1994], there is no evidence that the two αCTDs interact when bound to DNA. In fact, the binding and dimerization interfaces appear to overlap, and therefore DNA binding and dimerization may be mutually exclusive events [reviewed in Gourse et al. 2000].

In the *rrnB P1* UP element, DNA backbone regions centered at ~ -52 (distal subsite) and ~ -42 (proximal subsite) are protected against hydroxyl radical cleavage by α subunits [Estrem et al. 1998]. Several conclusions could be drawn from footprinting and transcription experiments with the RNAP mutants lacking either α^ICTD or α^{II}CTD at *rrnB P1* hybrid promoters. First, each UP element subsite constitutes a site for interaction with one copy of αCTD and as a result, an UP element containing good matches to consensus in both subsites requires both αCTDs for maximal stimulation of transcription. Second, the two αCTDs function interchangeably with respect to UP element subsite recognition. Third, only one αCTD is required for function of an UP element with only consensus proximal subsite, but promoter containing only a consensus distal subsite requires both αCTDs for efficient transcription, perhaps because of sequence–nonspecific interactions between the proximal subsite region and the second αCTD or because the second αCTD in some other way affects the overall stability of the complex [Estrem et al. 1999].

Although these studies provided some information about αCTD–UP element interaction, a full understanding of the mode of α subunit binding was hindered by complicating features of its interactions with different promoter upstream regions. Not only can α subunit bind to DNA in either a sequence–specific or sequence–nonspecific manner, but αCTD–dependent protection signals were also observed upstream of -60, and sometimes more that two protected regions appear in footprints upstream of a single promoter [Aiyar et

al. 1998, Ross et al. 2005]. A more likely possibility is that, in these cases, multiple potential binding sites for αCTD are present. Experiments with RNAP derivatives containing chemical nuclease covalently attached to αCTDs (FeBABE–αCTDs) [Lee et al. 2003] and protein–DNA photocrosslinking studies [Naryshkin et al. 2000] further suggest that one, or even two αCTDs may oscillate among different binding sites, or alternatively may occupy different sites in different molecules in the population of DNA fragments. The remarkable degree of flexibility in the positioning of αCTDs with respect to the rest of the RNAP–promoter complex likely results from the long unstructured linker present between the two domains of the α subunit [Blatter et al. 1994], as well as from intrinsic DNA distortions [Ross et al. 1998, Estrem et al. 1999] and protein–induced DNA bending in initiation complexes containing activator [Naryshkin et al. 2000].

4.2.8. *Potential interaction between α and σ subunits.*

The location of the proximal UP element subsite, where αCTD binds centered at ~ -42, suggested that αCTD might interact with the region of σ^{70} bound to the -35 hexamer (region 4.2) [Estrem et al. 1999, Ross et al. 2001]. Mutational analyses identified two acidic amino acid residues in the αCTD (Asp^{259} and Gly^{261}, which are outside of the DNA–binding determinant) and the basic amino acid residue in region 4.2 of σ^{70} (Arg^{603}) to be most critical for function of promoters containing some, but not all, UP elements [Ross et al. 2003]. It was shown that alanine substitutions at these positions lead to specific defects in transcription stimulation by the proximal UP element subsite *in vivo* and *in vitro* resulting from a decrease in the initial equilibrium constant for RNAP binding, but have little if any effect on full UP element function. The open complexes formed by wild type and mutant RNAPs (α E261A RNAP or σ R603A RNAP) were also reported to differ in DNase I sensitivity at the junction of the binding sites for αCTD and σ region 4.2 (position -38 on the template strand), but only at promoters where the mutant RNAPs affect UP element–dependent transcription. Furthermore, a model of the DNA–αCTD–σ region 4.2 ternary complex, constructed from the previously determined X–ray structures of the *Taq* σ^A region 4.2–DNA complex [Campbell et al. 2002] and the *E.coli* αCTD bound to a DNA fragment containing an A–tract [Benoff et al. 2002], indicates that the residues identified by mutation in αCTD and in σ region 4.2 are in very close proximity [Ross et al. 2003]. Experiments with RNAP derivatives containing

FeBABE–αCTDs also revealed that the orientation of the αCTD bound near position -42 is such that the 261 determinant is directed toward the promoter -35 hexamer [Lee et al. 2003].

Together, these data strongly suggest that αCTD, when bound to a proximal UP element subsite, contacts the σ^{70} region 4.2, increasing transcription primarily by stabilization of initial RNAP binding.

4.2.9. *DNA wrapping around RNAP.*

Several lines of evidence suggested that upstream DNA may wrap around *E.coli* RNAP during initiation forming a nucleosome–like structure. For example, hydroxyl radical footprinting studies of the open complex formed at several promoters (λP_R, *lacUV5*, *T7A1*) revealed a pattern of periodic protection of upstream DNA extending to approximately -70 [Schickor et al. 1990, Craig et al. 1995, Ross et al. 2005]. To reconcile the length (~240–320 Å) and periodicity of the open complex footprint with the dimensions of *E.coli* RNAP (160 Å x 95 Å x 90 Å) determined by electron microscopy [reviewed in Rivetti et al. 1999], upstream DNA was proposed to wrap in a surface groove of RNAP [Craig et al. 1995]. Evidence for DNA wrapping was also emerged from protein–DNA crosslinking and from atomic force microscopy analysis [Rippe et al. 1997, Rivetti et al. 1999].

Atomic force microscopy (AFM) allows straightforward detection and quantification of DNA compaction by DNA–binding proteins. DNA compaction is observed as a reduction of the DNA contour length in the presence of the DNA–binding protein of interest with respect to free DNA [Rivetti et al. 1999]. Using AFM, it has been shown that *E.coli* RNAP results in massive (~30 nm) apparent DNA compaction upon formation of a catalytically competent RNAP–promoter open complex at the λP_R promoter. This result is indicative of DNA wrapping by nearly 300° around the surface of the RNAP in the open complex. However, it has been recently found that stable DNA wrapping is not a general feature of open complex at all promoters, but rather depends on the promoter sequence and, in particular, on sequence determinants in the upstream region of the promoter (~30 nm DNA compaction by RNAP at λP_R promoter *versus* ~4–6 nm DNA compaction by RNAP at *lacUV5* and λP_R–*SUB(-463 to -36)* promoters) [Cellai et al. 2007]. In addition, it has been estimated that full UP element (not a single proximal UP element subsite) represents the sequence determinants for large DNA compaction in open complex (~21 nm DNA

compaction by RNAP at *lacUV5(UP full)* promoter *versus* ~2–4 nm DNA compaction by RNAP at *lacUV5* and *lacUV5(UP prox)* promoters) and that the presence of both αCTDs and an intact α–linker is required to maintain stable DNA wrapping.

In summary, all these data clearly indicate that the sequence of promoter can affect the recognition process and the isomerisation events in the pathway of promoter binding by RNAP as well as the conformation of the final open complex.

5. Footprinting technique and its application for the study of DNA−protein interactions.

Footprinting is a widely used method for the study of DNA−protein interactions. Nuclease protection or footprinting typically refers to assays in which the cleavage of the phosphodiester backbone of a nucleic acid polymer by an enzymatic or chemical nuclease is inhibited by the binding of a ligand to specific sequences of bases or by the conformation of the nucleic acid.

Under ideal footprinting experiments conditions each nucleic acid molecule is cleaved on average only once (single−hit kinetics regime). If a nucleic acid polymer is labeled at one end with either ^{32}P or a fluorophore, then the nuclease cleavage products, the DNA or RNA fragments of varying length, can be identified uniquely by electrophoretic or chromatographic separation (Figure 9).

Although the term 'footprint' refers to a decrease in the intensity of bands corresponding to nucleotides within a protected region, enhancements of reactivity can also be highly informative about nucleic acid structure.

The original implementation of footprinting used the endonuclease DNase I as a probe of specific sequences of DNA bound by proteins [Galas et al., 1978]. Since that time, additional footprinting assays have been developed using a wide range of enzymatic and chemical nucleases [Sigman et al., 1993; Armitage, 1998]. Among these, the hydroxyl radical (•OH) is one of the most sensitive probes of nucleic acid−ligand interactions and nucleic acid structure. Indeed, because of the high reactivity, diffusible nature and small size of the •OH, the nucleic acid cleavage by this probe is mostly dependent upon the solvent accessibility of the phosphodiester backbone and is relatively insensitive to base sequence and whether the

nucleic acid is single or double stranded [reviewed in Sclavi et al., 1997]. Thus, using •OH as a cleavage reagent, footprints can be obtained with resolution as fine as a single nucleotide.

The •OHs can be generated by chemical means (Fe−EDTA catalyzed Fenton chemistry [Sigman et al., 1993], peroxonitrous acid decomposition at neutral pH [reviewed in Sclavi et al., 1997]) but the time resolution of these chemical methods is on the seconds or minutes timescales, not sufficient for examining the short−lived intermediates in the formation of nucleoprotein complexes. The radiolysis of water by various forms of radiation, including γ−rays, β particles, and fast neutrons is also an effective mean of producing •OHs for footprinting [reviewed in Sclavi et al., 1997]. An advantage of radiolysis is that it does not require the addition of high concentrations of nucleases or nucleic acid modifying reagents to the sample, thus minimizing perturbation of the equilibrium and transition being studied. A limiting feature of method is that the irradiation of aqueous solutions from high energy sources with a low total flux requires exposures of tens of minutes to hours in order to conduct footprinting experiments.
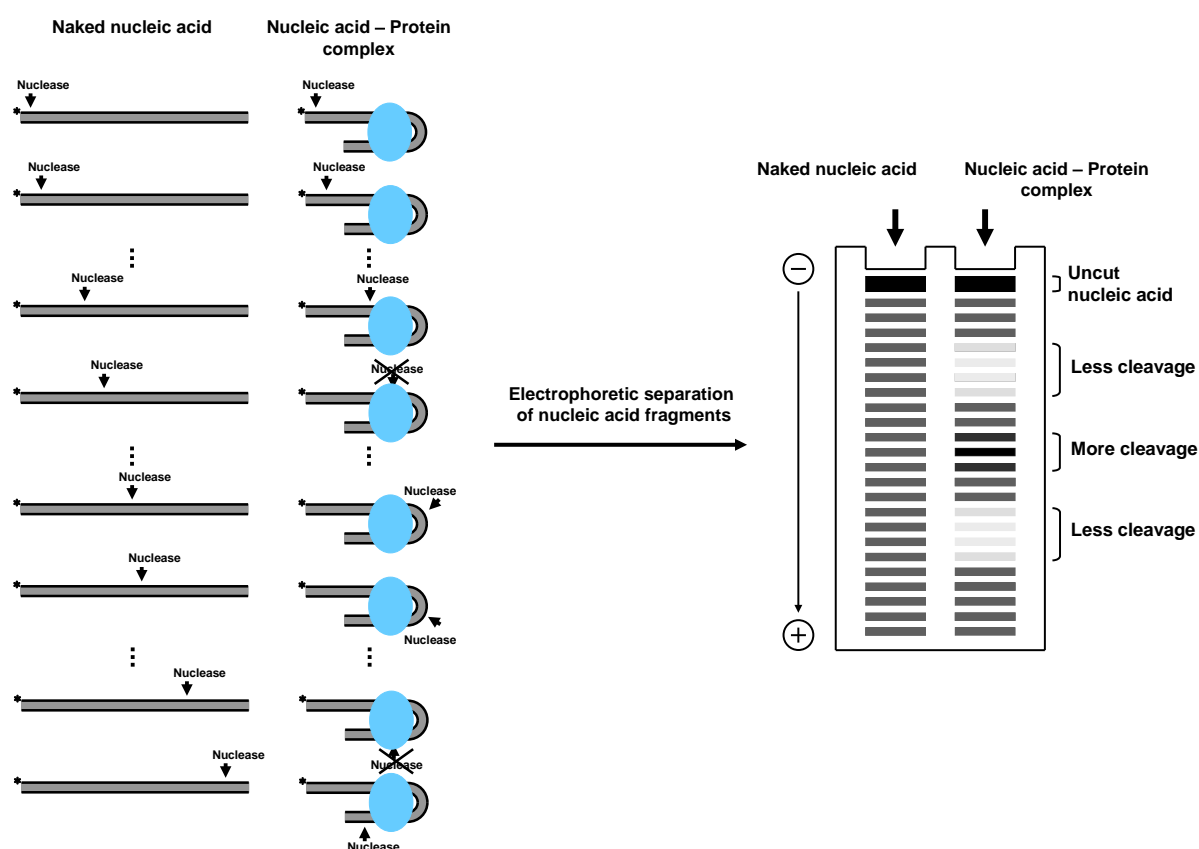
Figure 9. A schematic illustration of nuclease footprinting of nucleoprotein complex.

In contrast, by using a high flux "white-beam" of a synchrotron enough $^{\bullet}$OHs for the footprinting experiments can be produced in tens of milliseconds, thus considerably increasing the time resolution of this technique [Sclavi et al., 1997]. The radiolysis of water by X−rays generates $^{\bullet}$OH according to the overall reaction illustrated in equation [1]:

$$H_2O \xrightarrow{h\nu} H_2O^+ + e_{dry}^- \xrightarrow{H_2O} H_3O^+ + {}^{\bullet}OH + e_{aq}^- \qquad\qquad [1]$$
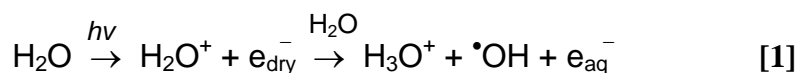
Upon the initial interactions of ionizing radiation with water, the absorption of the high−energy photons promotes electrons to an unbound state. These electrons and the water ions are responsible for all subsequent radiation chemistry in dilute aqueous solutions. The electrons are thermalized and deposit their energy in discrete ionizations of other water molecules. The ionized water molecules react with other water molecules to yield the hydroxyl radicals.

Because the reactant concentrations are extremely low compared to the concentration of water (~ 55 M), all effects of radiation chemistry (for ionizing radiation) in dilute aqueous solutions involve direct interactions only with water whereas the radiation effects on the solutes (nucleic acid polymers or the proteins) are entirely indirect. For nucleic acid polymers, the hydroxyl radicals produced by radiolysis of water initially abstract the hydrogen atom predominantly from the C5' or C4' carbon of the ribose moiety on the phosphodiester backbone, leading to a single strand nick [Balasubramanian et al., 1998].

Several factors affect the amount of cleavage of nucleic acid polymer and must be controlled and calibrated. For example, the synchrotron ring energy and ring current affect the amount of X−ray radiation incident on the sample and thus the concentration of $^{\bullet}$OH radicals generated [Sclavi et al., 1998]. In addition, the nucleic acid cutting by $^{\bullet}$OHs depends on the solution conditions such as temperature, pH, and the chemical composition of the buffer. Sodium cacodylate is the standard buffer for $^{\bullet}$OH footprinting experiment since it does not interfere with radiolysis chemistry.

In order to obtain quantitative data about the interaction of nucleic acid polymer and protein, the time of sample exposure to the X−ray beam has to be adjusted in a such a way that not more than 10−30 % of the DNA molecules are cut (single−hit kinetics regime). In order to achieve this level of cleavage, an exposure time calibration (dose−response experiment) should be completed under the reaction condition of experiment.

6.   Results.

Our laboratory has been studying the interaction of *E.coli* σ^{70} RNAP with promoter DNA for several years using footprinting techniques applied to binary complexes trapped at different temperatures and to ternary complexes stalled at different steps of the transcription cycle. While the static description of the complexes formed between RNAP and the DNA has given important insight into the structure of these complexes, this kind of approach does not give information on the mechanism by which these complexes are formed and the structure of the intermediates in their real time pathway of formation.

Recent technical developments make it possible to obtain DNA footprinting at faster timescales. In fact, in a short time (a few microseconds) the synchrotron X−ray radiolysis of water produces high concentration of hydroxyl radicals. Coupled with a rapid mixing technique, it enables to follow the macromolecular transitions with a millisecond time resolution. Time−resolved hydroxyl radical footprinting based on synchrotron X−rays has been used for the past 10 years at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratories in the United States to monitor the process of RNA folding and of DNA−protein interactions [Sclavi et al, 1998; Dhavan et al., 2002]. For the first time, we applied this technique to study in real time the structural events taking place upon promoter binding by RNAP. In order to find the adequate X−ray source, we tested a number of beamlines: the Deutsches Elektronen−Synchrotron (DESY) in Hamburg, the Berliner Electronenspeicherring−Gesellschaft für Synchrotronstrahlung (BESSY) in Berlin, and three beamlines (ID9, BM5 and ID10) at the European Synchrotron Radiation Facility (ESRF) in Grenoble. The result was that only ESRF beamlines ID9 and ID10 have enough brilliance to allow DNA footprinting at a millisecond timescale. Finally we set up an X−ray footprinting technique at the beamline ID10A at ESRF (Grenoble, France).

6.1.   Improvements in the technique.

In the course of establishing the X−ray generated hydroxyl radicals footprinting method we made several improvements. The first was to use DNA labeled with a fluorescent tag instead of a radioactive one. This allows us to work with fewer safety restrictions.

The second improvement was to use the gel−electrophoretic instrument with fluorescence detectors (ALF Express II DNA analyzer, Pharmacia). Its advantages are:

increased resolution of DNA fragments; and faster and more accurate quantitative analysis of each fragment's intensity. This can be determined while the electrophoresis is running instead of having to wait for the radioactive gel to be exposed onto X−ray film or imager screen, thus saving a significant amount of time and allowing us to determine whether an experiment worked, just a few hours after it was completed.

The important instrumental part of the method is the stopped−flow machine. The first experiments were performed using a home−made apparatus. This machine was successfully used to elaborate the X−ray footprinting technique and to obtain the first structural kinetic data on RNAP binding to promoter [Sclavi et al. 2005]. However, the time resolution of this machine was not high enough to measure the fast conformational changes taking place in the formation of the DNA−protein complex. Therefore we decided to use a commercially available stopped−flow machine. The independently controlled four−syringe stopped−flow machine (SFM−400), configured to our needs, was purchased from BioLogic (Claix, France). We then incorporated in it our own mixing device, exposure chamber and sample collector. The use of this machine resulted in improved data quality and reproducibility because of an increased time resolution and a more stable and uniform temperature control. Besides being faster, this machine is equipped with a sample collector that allows us to mix and expose several samples in a sequence. This results in faster sample processing, omitting the necessity to enter the hutch to retrieve each sample and reload the machine. In ten minutes we can now mix and expose the same amount of samples that we used to do in more than half an hour. The detailed description of the experimental set−up is presented in the section "Materials and Methods".

6.2. Time−resolved X−ray generated hydroxyl radical footprinting of the binary complex.

In this study we used time−resolved hydroxyl radical footprinting to directly characterize the structural intermediates present in the pathway leading to final open complex formation on the wild type *T7A1* promoter at 37°C and 20°C and on a promoter variant containing a consensus -10 hexamer.

The *A1* promoter is an early promoter in the lifetime of phage *T7*. Its high level of activity allows it to successfully compete for the host's pool of RNAP. The *T7A1* promoter's wild type sequence is characterized by a nearly consensus -35 hexamer (TTGAC**T** instead of

Wild type *T7A1* promoter

```
 -90       -80       -70       -60       -50       -40       -30       -20       -10       +1       +10
  .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
CGAGGCCAACTTAAAGAGACTTAAAAGATTAATTTAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGGA.........
GCTCCGGTTGAATTTCTCTGAATTTTCTAATTAAATTTTAAATAGTTTTTCTCATAACTGAATTTCAGATTGGATATCCTATGAATGTCGGTAGCTCTCCCT.........
                                                      „-35"                       „-10"
```

Mutant-1.   Consensus -10 region

```
 -90       -80       -70       -60       -50       -40       -30       -20       -10       +1       +10
  .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
CGAGGCCAACTTAAAGAGACTTAAAAGATTAATTTAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGTATAATTACAGCCATCGAGAGGGA.........
GCTCCGGTTGAATTTCTCTGAATTTTCTAATTAAATTTTAAATAGTTTTTCTCATAACTGAATTTCAGATTGGATATCATATTAATGTCGGTAGCTCTCCCT.........
```

Figure 10. Sequences of the wild type *T7A1* promoter and the promoter variant containing consensus -10 hexamer.

TTGACA) and a non−consensus -10 hexamer (**G**ATA**C**T instead of TATAAT) separated by a 17nt−long spacer. Moreover, this promoter contains an A+T−rich UP element extending from -42 to -71 (Figure 10).

6.2.1. *Experimental setup, raw data generation and quantitative analysis.*

To study the process of formation of the binary complex between *E.coli* RNAP and *A1* promoter of phage *T7* we rapidly mixed the protein solution with a solution containing an end−labeled DNA promoter fragment inside the BioLogic stopped−flow machine. After incubation of the reagents for a variable time (in the range from 50 ms to 5 min), the sample was exposed to the X−ray beam by pushing the solution through a quartz capillary at a specific speed, thus determining the exposure time (for details see section "Materials and Methods"). The hydroxyl radicals produced by the radiolysis of water cleave the DNA backbone at any solvent accessible nucleotide. The cleavage products were resolved on 8 % denaturing polyacrylamide gel using an ALF Express II DNA analyzer. A fluorescence detector at the bottom of the gel produces a profile of the mobility of DNA fragments containing a fluorescently labeled end and provides a quantitative measurement of each fragment's amount.

Hydroxyl radicals not only cleave the DNA but also result in modification/oxidation of the side chains of the protein. However, this does not interfere with the interpretation of the footprinting data, since the DNA is cut more efficiently than the protein, due to its high sensitivity towards the hydroxyl radical.

An example of the gel profiles obtained with the wild type *T7A1* promoter DNA labeled at the 5'−terminus of either the non−template or template strand with the fluorophore Alexa 647 is shown in Figure 11.

35

Non−template strand		Template strand	
	0 s		0 s
	0.1 s		0.1 s
	0.5 s		0.5 s
	2 s		2 s
	10 s		10 s
	30 s		30 s
	120 s		120 s

-80 -70 -60 -50 -40 -30 -20 -10 +1 +10 +20 +30 +40   -80 -70 -60 -50 -40 -30 -20 -10 +1 +10 +20 +30 +40
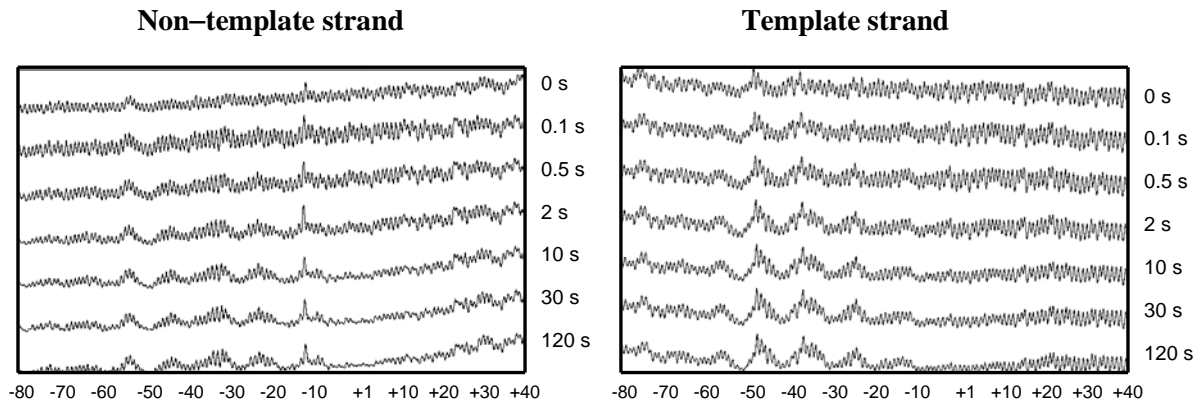
Figure 11. Hydroxyl radical footprinting profiles of the wild type *T7A1* promoter fragments labeled with fluorophore Alexa 647 at the 5'−end of either non−template strand (NTS, left panel) or template strand (TS, right panel) as a function of incubation time with RNAP. Each line graph corresponds to one lane on the gel. The incubation time for each lane is shown next to the line graph.

The profile in the first row from the top (0 s) represents the cleavage pattern of the DNA by hydroxyl radicals in the absence of RNAP. Each peak in the row represents a DNA fragment of a specific size. The position of the peak corresponds to the number of bases in the fragment. Shorter fragments migrate on the gel faster than longer ones. The difference between one fragment and the next is only one base. The peak intensity correlates with the cleavage efficiency at a given base. The important consequence of the lack of base selectivity of the hydroxyl radicals is that any particular backbone position will be cleaved at almost the same frequency, therefore a plot of peaks of nearly equal intensity, one for each base position, is expected to be seen on the gel profile. Indeed, a nearly uniform cleavage pattern is observed, implying that the free DNA in solution is cut by hydroxyl radicals rather evenly along its length. Only the intensity of the peaks corresponding to the nucleotides within the stretches of adenine and thymine residues in the far upstream region of the promoter (positions from m71 to m42; the nucleotide positions will be designated with "m" (minus) for nucleotides upstream and "p" (plus) for those downstream of the site of the start of transcription (designated as p1)) is lower than the average intensity of other peaks. This reduction in cleavage arises from a narrowing of the minor groove associated with a change in DNA structure at this particular sequence [Tullius, 1987].

The interaction of RNAP with the *T7A1* promoter was analyzed in real time by monitoring changes in the hydroxyl radical reactivity of the nucleotides upon nucleoprotein complex formation. One can see that upon DNA incubation with RNAP, the intensity of the peaks in specific regions decreases (Figure 11). These peaks correspond to nucleotides which enter in close contact with the enzyme and are therefore less accessible to the hydroxyl

radicals. In the presence of the protein, the enhancement of hydroxyl radical reactivity of nucleotides at several sites is also observed. These sites are the regions of the DNA where the double helix has been deformed upon RNAP binding. Indeed, if the DNA is bent, some of the sugars in the backbone will become more solvent accessible compared to those in a regular double helix and will be cleaved more easily by the radicals.

Figure 12 represents the profiles of the peak area measured by peak fitting (for details see "Data Analysis") relative to that in naked DNA (Φ). One can observe that the regions of the promoter to become protected at the earliest time points (50 ms – 200 ms) are situated far upstream from the transcription start site (p1). During the later steps the rest of the promoter comes into contact with the different domains of the RNAP. The final footprint consists of several blocs of modulated protection from m75 to m11, typical of a case where the protein is sitting on one face of the DNA double helix, and of the region of continuous protection covering the stretch between positions m9 and p20 resulting from the protein completely enveloping the DNA.
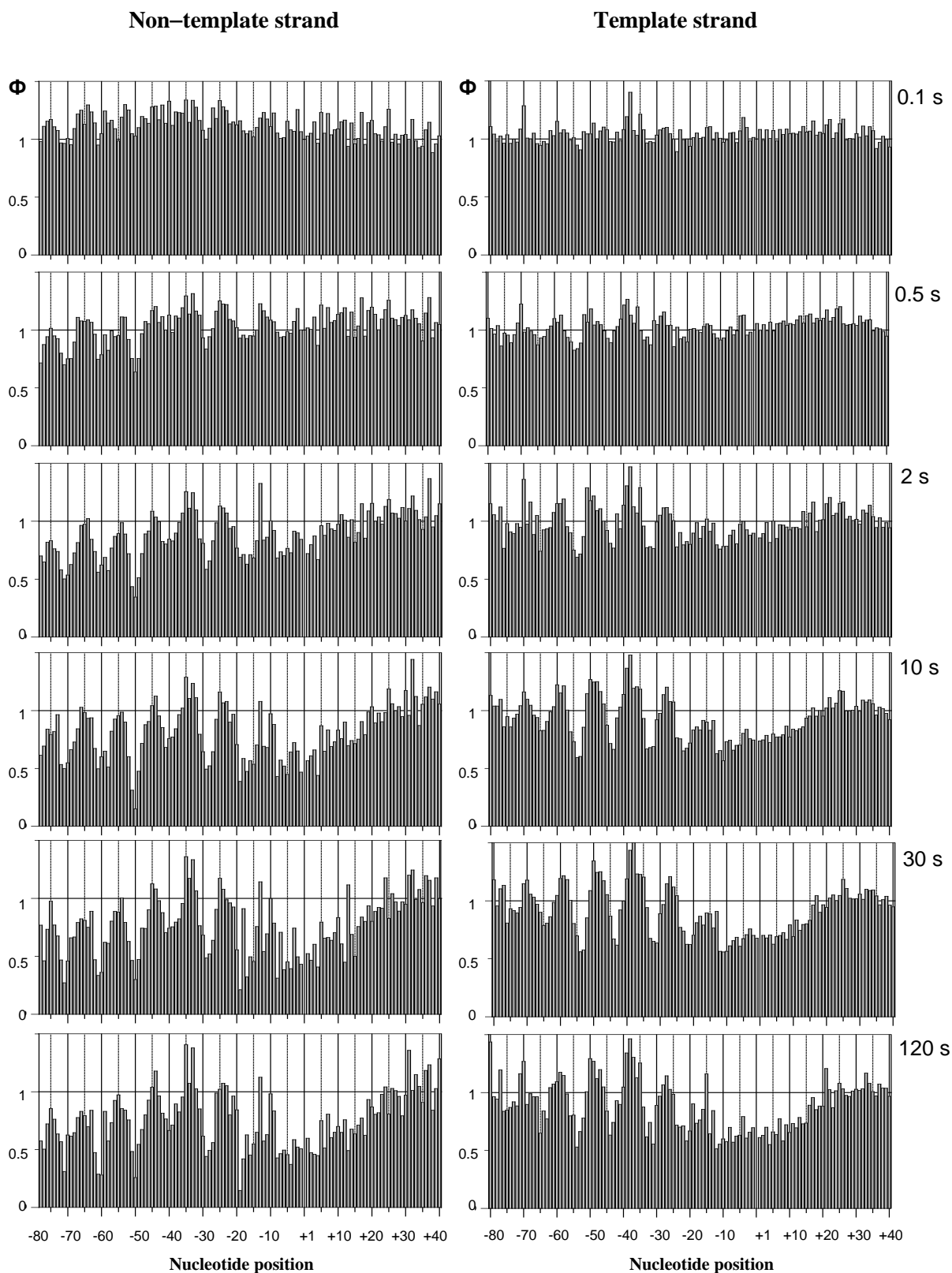
Figure 12. The bar plots obtained in the case of RNAP binding to the wild type *T7A1* promoter at 37°C. The bar plots show Φ, the ratio of the area of each peak relative to that in the naked DNA (lane at 0 s in Figure 11), for the two strands (non−template in the left panel and template in the right panel), in the course of the binding reaction. The bar at nucleotide position 0 is missing because in the conventional numbering of the nucleotides in a promoter 0 is not used.

6.2.2. Determination of kinetic of protection appearance at different promoter regions.

As shown in Figure 12, at the earliest time points, 50 ms – 200 ms, a weak pattern of the modulated protection is present in the upstream area between positions m73 to m48 on the NTS and from m76 to m53 on the TS. As the incubation time between RNAP and DNA is increased, the extent of protection at these sites increases and additional protections appear, finally resulting in an extension of the protected region down to p20. The kinetic of the subsequent increase of protection can be described by either a single or double exponential equation (**[2]** and **[3]**, see "Data Analysis"). Analysis of the data revealed that at 37°C the kinetic of the appearance of protection in different promoter regions is best described by a double exponential expression. For instance, Figure 13 represents the comparison of fits to a single (red curve) and double (blue curve) exponential equation for the data sets m73−m70 NTS and m23−m20 TS, obtained on the wild type *T7A1* promoter at 37°C.

The panel below each plot shows the residuals for the fit (for details see "Data Analysis"). One can observe that the residuals for the single exponential fit show a periodic pattern of distribution, whereas the residuals for the double exponential fit are almost evenly distributed on either side of the fit curve indicating that a double exponential expression better fits data. In addition, an F test comparison of the fits revealed that the double exponential fit is statistically better than the single exponential fit (see "Data Analysis"). A visual analysis of the residuals and an F test were carried out for each data set (see Figure 1S and Table 1S in "Supporting Materials").
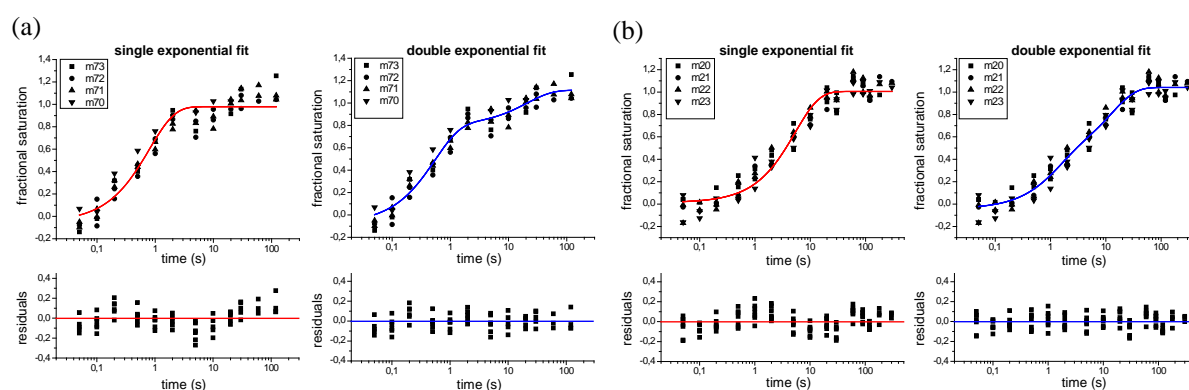
Figure 13. Comparison of the fits of kinetics of the protection appearance at m73−m70 NTS (**a**) and m23−m20 TS (**b**) upon RNAP binding to the wild type *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. Bottom panels show the residuals for the fits.

(a) wild type *T7A1* promoter, 37°C

(b) *T7A1* promoter variant with consensus -10 region, 37°C

(c) wild type *T7A1* promoter, 20°C
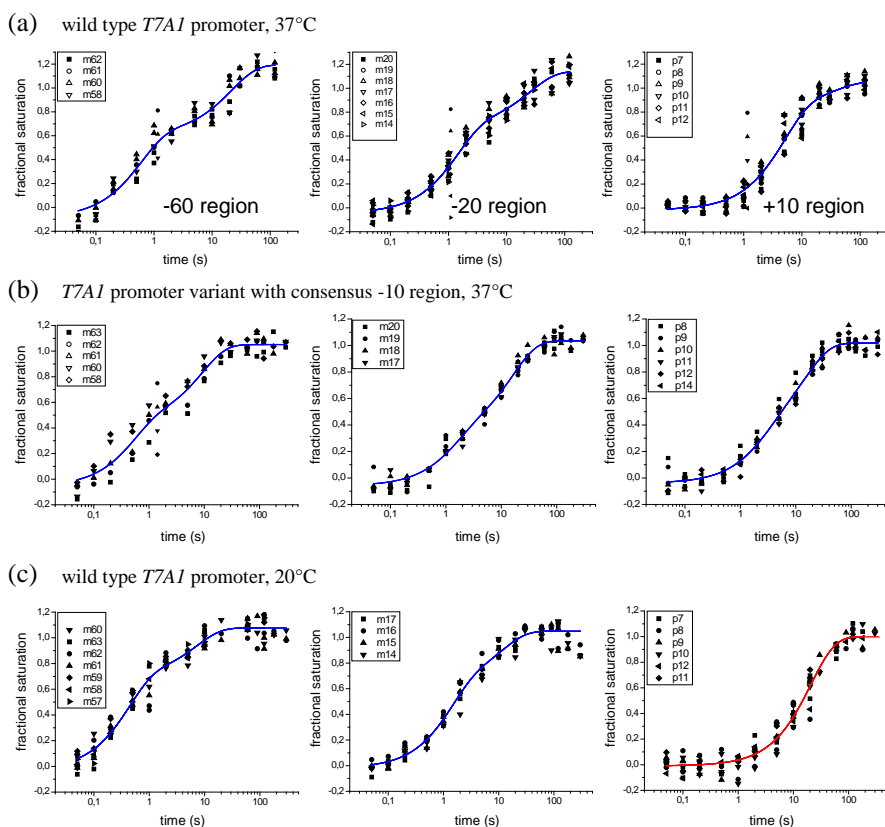
Figure 14.
Plots of the kinetics of protection appearance at three representative regions on the NTS upon the binary complex formation on the wild type *T7A1* promoter at 37°C (**a**), the promoter mutant characterized by the consensus -10 region at 37°C (**b**), the wild type *T7A1* promoter at 20°C (**c**).

Figure 14 (a) shows a representative sample of the plots of the change in the fraction of protection as a function of time in different regions of the wild type *T7A1* promoter at 37°C. The rates and amplitudes of both phases change depending on the position of the base on the promoter. Since the extent of protection of the DNA backbone from hydroxyl radical cleavage at each base position correlates with the solvent accessibility of the base, the differences in the rates and amplitudes of the protection appearance allowed us to assign the protection at each site of the promoter to a specific intermediate in the pathway to the final complex. Moreover, based on the published data from crystallographic studies, as well as numerous genetic and biochemical studies, describing the contacts of the RNAP subunits with different regions of the promoter, we were able to assign at each step of promoter recognition and binding the specific domains of distinct subunits of the enzyme interacting with DNA.

The approach described in the previous paragraphs was applied to generate and evaluate also the data in the cases of *E.coli* RNAP binding to wild type *A1* promoter of phage *T7* at 20°C, and to the mutant promoter with a consensus -10 region at 37°C. Comparison of the results obtained on two different promoters (for example, Figure 14 (a) *versus* (b)) and at two different temperatures (for example, Figure 14 (a) *versus* (c)) allowed us to determine

how the DNA sequence in the -10 hexamer and temperature may affect the structure of intermediates in the process of RNAP–DNA complex formation.

6.3. Real–time identification and structural characterization of the intermediates formed upon *E.coli* RNAP binding to the wild type *T7A1* promoter at 37°C.

6.3.1. Detection of the specific intermediate RNAP–DNA complexes on the basis of kinetic data, obtained by X–ray hydroxyl radical footprinting.

On the wild type *T7A1* promoter at 37°C the kinetic of appearance of protection at all promoter regions follows a biphasic behavior. The apparent rate constant for the first fast phase (k_A) is a macroscopic rate constant that describes the overall kinetics of promoter binding and dissociation. We can define the fast phase as the formation of unstable, reversible protein–DNA contacts present in the early intermediates and assign the process associated with the slow second phase to the formation of a more stable complex.

Values for the apparent rate constants are listed with their confidence limits in Table 1. Figure 15 represents the summary plots of the apparent rate constant and amplitude of the fast and slow phases *versus* the position of the nucleotides on both strands of wild type *T7A1* promoter. One can see that the value of the apparent rate constant for the fast phase displays a stepwise decrease as a function of nucleotide position (Figure 15 (a1) and (a2)). Indeed, the nucleotides at the upstream promoter region become protected at a faster rate than those downstream. The difference in the rates with which nucleotides become protected as a function of their position on the promoter allowed us to define several steps in the pathway to formation of the final binary complex. Moreover, differences in the amplitude of each phase (Figure 15 (c1) and (c2)) gave an additional criterion for identification of intermediate structures.

Wild type T7A1 promoter, 37°C, Template strand				
Nucleotide position	**Apparent rates (s⁻¹)**		**% of total amplitude**	
	k_A	k_B	**A**	**B**
m55–m53	1.15 ± 0.15	0.032 ± 0.006	57.1	42.9
m45–m42	0.98 ± 0.29	0.067 ± 0.018	46.5	53.5
m34–m31	0.99 ± 0.23	0.069 ± 0.014	48.6	51.4
m23–m20	0.82 ± 0.22	0.079 ± 0.014	41.4	58.6
m19–m13	0.31 ± 0.06	0.016 ± 0.006	57.2	42.8
m12–m9	0.67 ± 0.12	0.062 ± 0.010	47.5	52.5
m8–m6	0.36 ± 0.09	0.049 ± 0.013	51.5	48.5
m5–m3	0.17 ± 0.03	0.018 ± 0.012	74.4	25.6
m2–p2	0.34 ± 0.09	0.051 ± 0.011	46.1	53.9
p3–p6	0.16 ± 0.02	0.015 ± 0.005	67.3	32.7
p7–p9	0.16 ± 0.06	0.037 ± 0.023	57.9	42.1
p10–p13	0.18 ± 0.03	0.028 ± 0.015	71.0	29.0
p14–p20	0.20 ± 0.06	0.021 ± 0.006	45.3	54.7

Wild type T7A1 promoter, 37°C, Non–template strand				
Nucleotide position	**Apparent rates (s⁻¹)**		**% of total amplitude**	
	k_A	k_B	**A**	**B**
m73–m70	1.86 ± 0.25	0.048 ± 0.020	73.5	26.5
m62–m58	1.70 ± 0.29	0.046 ± 0.010	54.6	45.4
m52–m48	1.73 ± 0.24	0.025 ± 0.014	74.2	25.8
m41–m38	0.86 ± 0.10	0.018 ± 0.013	72.7	27.3
m31–m27	0.91 ± 0.22	0.150 ± 0.055	60.4	39.6
m20–m14	0.71 ± 0.07	0.039 ± 0.008	61.3	38.7
m12–m11, m9–m5	0.54 ± 0.06	0.064 ± 0.016	66.6	33.4
m4–p2	0.33 ± 0.05	0.034 ± 0.012	67.0	33.0
p3–p6	0.36 ± 0.04	0.026 ± 0.012	72.4	27.6
p7–p12	0.20 ± 0.03	0.027 ± 0.028	80.4	19.6
p15–p20	0.22 ± 0.04	0.041 ± 0.015	55.5	44.5

Table 1. Values for the apparent rate constants and the amplitudes of both phases as a function of nucleotide position on the template (top panel) and non–template strand (bottom panel) for the case of *E.coli* RNAP holoenzyme–wild type *T7A1* complex formation at 37°C.

Template strand

Non−template strand

■ − fast phase   ▨ − slow phase
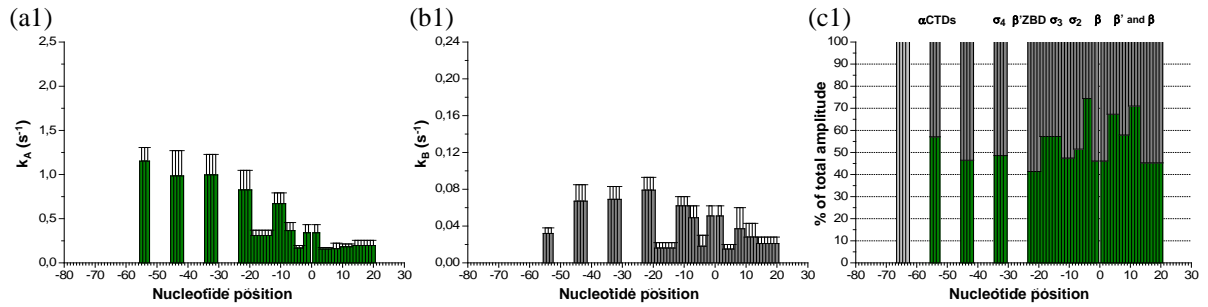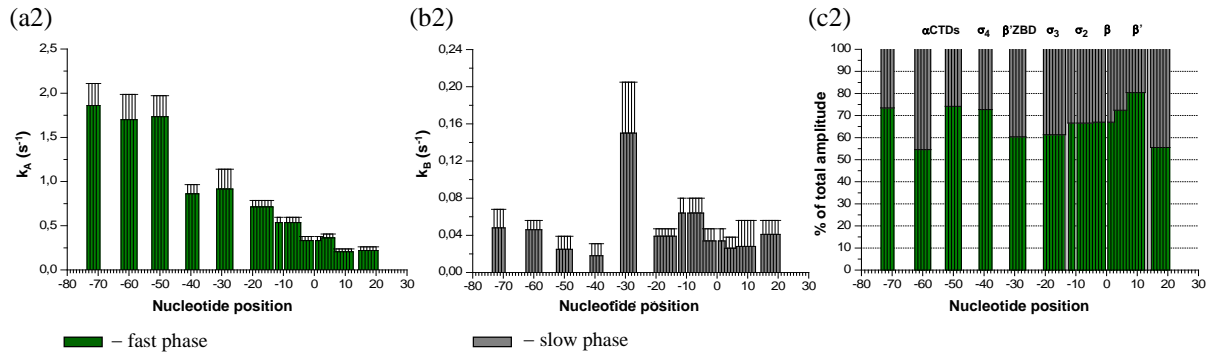
Figure 15. Summary plots of the apparent rate constants and the amplitudes of appearance of the protection on both strands of wild type *T7A1* promoter upon RNAP binding at 37°C.

(**a1**) and (**a2**) Apparent rate constants for the fast phase *versus* the position of the nucleotides on the TS and NTS, respectively. (**b1**) and (**b2**) Apparent rate constants for the slow phase *versus* the position of the nucleotides on the TS and NTS, respectively. (**c1**) and (**c2**) Amplitude of each phase as the percentage of total amplitude *versus* the position of the nucleotides on the TS and NTS, respectively. The RNAP domains known to be responsible for a given protection are shown above their corresponding signals [Naryshkin et al. 2000, Murakami et al. 2002b].

Thus, based on the data summarized in Figure 15 and the results from the literature describing specific RNAP−promoter interactions, we conclude that the earliest intermediates observed upon RNAP binding to the wild type *T7A1* promoter at 37°C are characterized by the protection of the distal UP element subsites (m55−m53 TS; m73−m48 NTS) by the αCTDs. In the proposed kinetic model, Figure 16, these intermediates are designated with the letter A. On the NTS the protection extends over three sites within a 25 base region, however one can see that the amplitude of the fast phase for the protection at the m62−m58 is lower compared to the ones on either side. Further, Figure 12 shows that the "early" protection is also observed near m80 NTS, but the low resolution of the fragments upstream of m75 NTS did not allow a precise measurement of the kinetic. In addition, the weaker protection on the template strand at the sites upstream of m55 results in a low signal to noise ratio in the quantification and indicates that the two strands are protected by the αCTDs in an asymmetric fashion.

43

Figure 15 shows that the protection of the nucleotides from m45 to m31 on the TS and from m41 to m27 on the NTS by the αCTDs, the σ region 4.2 and the β'ZBD appears at a similar slower rate. However, one can see that the amplitude of the first, fast phase for the protection at the m41−m38 NTS differs from those at the m31−m27 NTS (Figure 15 (c2)), a signature of the presence of several intermediates in rapid equilibrium. Thus, to account for the different amplitudes, we propose that a first isomerisation step leads to two complexes with footprints of different size, one with additional protection from m45 to m31 on the TS and from m41 to m38 on the NTS, and a second one with protection extending to position m27 on the NTS (Figure 16, B and B' intermediates, respectively). The difference in the apparent rates for the fast phase between the A and B intermediates is more evident in the results obtained using the DNA labeled at the NTS. In this set of experiments the rate of protection appearance at the distal UP element subsites is significantly faster, probably due to a higher concentration of active protein in these samples. The rates of the subsequent slower steps however do not change significantly between the datasets obtained on the two strands, suggesting that they result from an isomerisation of the intermediates characterized by interactions of RNAP with the distal UP element subsites.

The next step results in the appearance of protection at positions m23−m20 and m12−m9 on the TS and m20−m5 on the NTS (C intermediate). The protection from m23 to m21 TS is due to contact with the β'ZBD [Naryshkin et al. 2000, Murakami et al. 2002b]. The continuous protection of nucleotides at m20−m14 NTS involves the σ_3 domain of the *E.coli* σ^{70} factor [Barne et al. 1997, Bown et al. 1999], whereas the protection of -10 sequence is due to σ_2 domain protruding from the surface of the protein [Murakami et al. 2002b]. According to the structural model of the RNAP−DNA complex proposed by Murakami and co−workers on the basis of the crystal structure of *Taq* RNAP holoenzyme in complex with fork−junction promoter DNA, the protection can extend beyond m9 on the non−template strand, only if the DNA is curved towards the surface of the protein [Murakami et al. 2002b].

In the following step the protected region extends to the transcription initiation site (m19−m13, m8−m6, m2−p2 TS; m4−p6 NTS) (D intermediate in Figure 16). The protection on both strands may result from further bending of the DNA towards the β and β' pincers, facilitated by a partial melting of the DNA double helix (see below). The significant difference in the rates of formation of the C and D intermediates suggests that they are separated by the presence of a significant kinetic barrier.
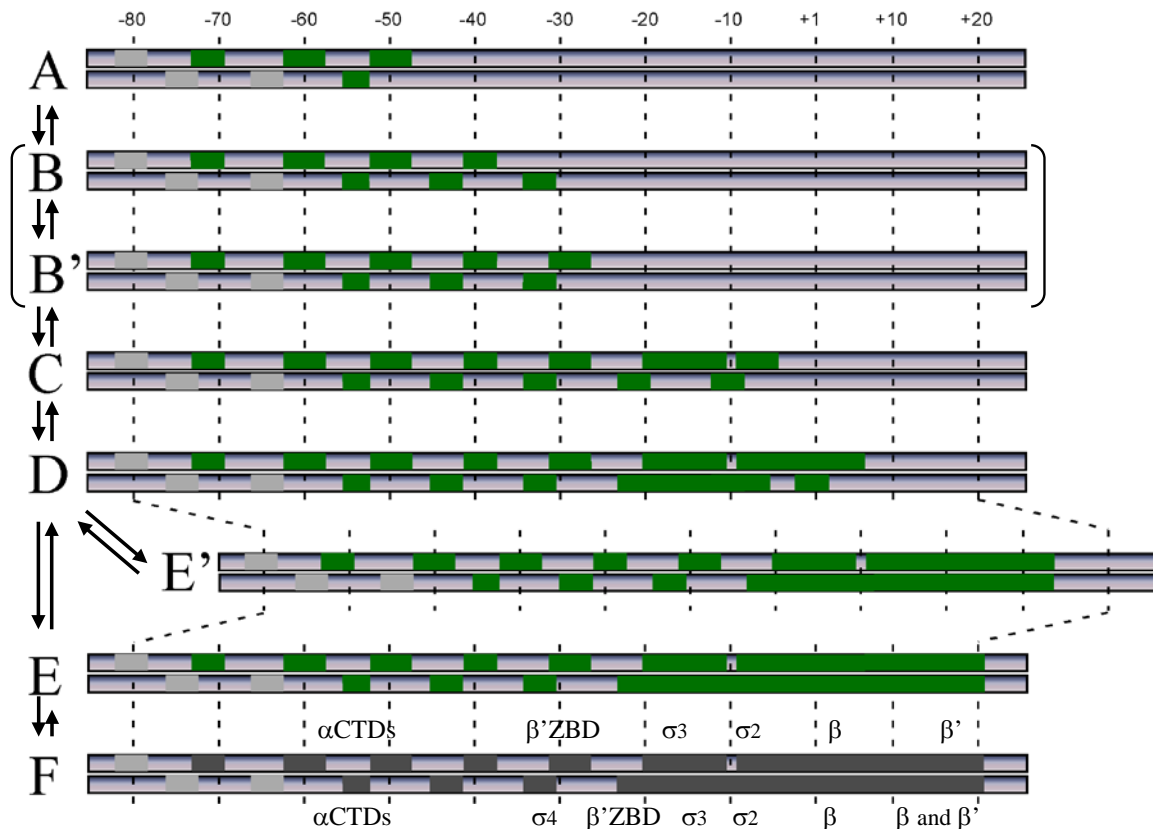
Figure 16. Proposed kinetic model for the binding of *E.coli* RNAP holoenzyme to the wild type *T7A1* promoter at 37°C. The DNA strands are shown as rectangles. The non−template strand (NTS) is top rectangle. The template strand (TS) is bottom rectangle. The regions that become protected at each step of RNAP binding to promoter are shown as colored boxes. Light gray boxes correspond to the regions with weak protection; green boxes correspond to the regions which become protected during the first, fast phase; dark gray boxes correspond to the regions which extent of protection increases in the second phase of the kinetics. The RNAP domains known to be responsible for a given protection are shown for the final complex above (NTS) or below (TS) their corresponding signals [Naryshkin et al. 2000, Murakami et al. 2002b].

The protection at positions m5−m3 and p3−p13 TS and p7−p13 NTS appears at the same rate as the protection at positions p14−p20 on both strands, but is characterized by the larger amplitude for the first phase (Figure 15 (a1) and (a2), (c1) and (c2)). We propose that this larger amplitude (higher extent of protection) results from the presence of two complexes, in each of which a DNA is protected at these nucleotides (intermediates E and E' in Figure 16). In the first complex, E, the contacts with these nucleotides are formed at the same time as the contacts with the downstream DNA (p14−p20 TS and NTS); in the second complex, E', the same protection pattern appears in the absence of the downstream contacts. These results indicate that there is a branching of the pathway where a fraction of the intermediates, E', is in a conformation that cannot efficiently isomerise to form contacts between RNAP domains and

the double−stranded DNA downstream of position p13. This can occur only following a conversion of E' back to the D intermediate.

The rates of the second exponential phase correspond to the isomerisation steps leading to the stabilization of the RNAP−DNA complex (F complex in the kinetic model). Two sets of rates of the second phase can be distinguished depending on the position of the nucleotides on the wild type *T7A1* promoter: a faster rate for the formation of stable contacts with the core promoter sequence, and a slower rate for the sites which become protected only in the E and E' complexes. These differences in the rate of the second phase are more evident in the dataset for the template strand, where the protection of the far upstream region (m55−m53) also becomes stabilized at a slower rate rather than at the core promoter (Figure 15 (b1)). The stabilization of the contacts at positions m31−m27 on the non−template strand happens faster than at the other sites (Figure 15 (b2)). This protection is probably caused by an interaction of the non−template strand with the RNAP β'ZBD. Thus, a change in conformation of the DNA in the spacer region precedes the final isomerisation step. The much less pronounced differences in the rate of the second phase between the equivalent site on the template strand (m23−m20 TS) and the surrounding core promoter sites may reflect the asymmetry resulting from a change in the twist associated with the last steps of open complex formation, as the single−stranded template strand is led deep into the core of the enzyme towards the active site.

In summary, our data indicate that the relatively slow rates for the isomerisation steps leading to the final functional complex at the wild type *T7A1* promoter seem to be due to the limited amount of the intermediates in the correct conformation able to isomerise to the next step.

6.3.2. Determination of kinetic of DNA melting by RNAP on the wild type T7A1 promoter at 37°C, using time−resolved permanganate footprinting.

A critical step in transcription initiation is the formation of the "open" binary complex between RNAP and promoter, in which the stands of DNA duplex, approximately from the -10 region to the transcription start site, become separated, forming a so called "transcription bubble". The final complex formed by RNAP on the wild type *A1* promoter of bacteriophage *T7* at 37°C in the presence of Mg^{2+} has a bubble extending from position m12 to position p2 [Zaychikov et al. 1997]. In order to determine the timing of DNA melting with respect to the

appearance of protections from hydroxyl radical cleavage we carried out time−resolved potassium permanganate footprinting experiments.

The use of potassium permanganate ($KMnO_4$) as a probe to map the melted region is based on the finding that this chemical reagent is able to oxidize the C5−C6 double bond in thymines in single−stranded DNA regions at a higher rate than those within the DNA duplex [reviewed in Łoziński and Wierzchowski 2003] making the neighbouring phosphodiester bond labile to alkali (piperidine) cleavage.

To study the kinetic of strand separation upon binary complex formation between *E.coli* RNAP and the wild type *T7A1* promoter at 37°C we rapidly mixed the protein solution with a solution containing promoter fragment inside the stopped−flow machine of our own construction. The DNA fragment was radioactively labeled at the 5'−end of the template strand (it was found that fluorescent labels do not withstand $KMnO_4$ treatment). The reagents were incubated for a specific amount of time (in the range from 1 s to 4 min) following which the sample was exposed to the $KMnO_4$ for 0.2 s. Subsequently the DNA backbone was cleaved at the thymine, modified by $KMnO_4$ oxidation, by reaction with piperidine. The cleavage products were analyzed on a 7 % denaturing polyacrylamide gel. The details are described in the section "Materials and Methods". An example of the time−resolved permanganate footprinting patterns obtained at the template strand of DNA is shown in Figure 17. Process of promoter melting was analyzed by monitoring changes in reactivity of thymines toward $KMnO_4$ upon nucleoprotein complex formation.
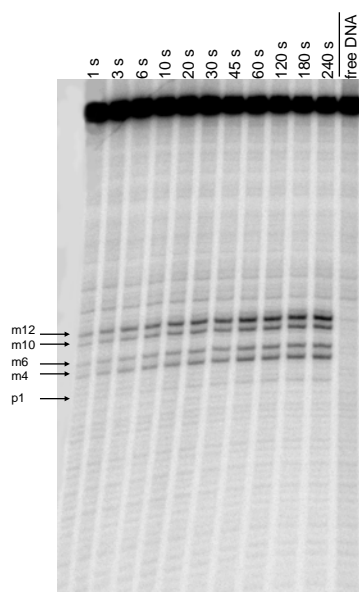
Figure 17. Analysis of the promoter melting upon binary complex formation using single strand−specific reagent, $KMnO_4$. The panel shows the electrophoretic patterns of 5'−[32]P labeled *T7A1* promoter fragment cleaved with piperidine at the thymine modified by $KMnO_4$ as a function of time of its incubation with RNAP at 37°C. Arrows indicate the positions of bands representing cleavage at modified thymines.

The kinetic curves describing the increase of thymine reactivity *versus* time are presented in Figure 18. The kinetic data are described by double rather than single exponential as follows from the analysis of residuals and F test statistical analysis (see Figure 2S and Table 2S in "Supporting Materials", respectively). Values for the apparent rate constants and amplitudes of both phases are summarized in Table 2.

Comparison of the footprinting data shows that the permanganate signals on the template strand at thymines m12, m10 and m6 appear with rates similar to those measured for the appearance of protection of DNA backbone in the -10 region (C intermediate), while the modification of thymine at m4 appears at the rate similar to the rate of the protection extension down to p2 on the TS and p6 on the NTS (D intermediate). The reactivity of the thymine at position p1 toward $KMnO_4$ was too weak to obtain a precise value of the kinetic constant. Since the thymine at position m4 becomes accessible to modification at a slower rate than the thymines in the -10 hexamer, we conclude that, upon RNAP binding to the wild type *T7A1* promoter at 37 °C, DNA melting takes place in at least two steps. Moreover, our results indicate that this process occurs in the absence of the extended protection to p20 and it is not the rate−limiting step in the pathway leading to final functional complex formation.
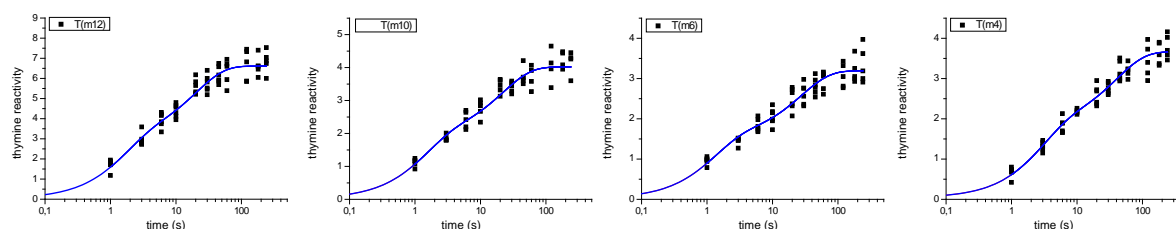
Figure 18. Time dependence of the thymines accessibility to the $KMnO_4$ modification in the template strand upon the binary complex formation on the wild type *A1* promoter of phage *T7* at 37°C.

Wild type T7A1 promoter, 37°C, Template strand				
	Apparent rates (s^{-1})		% of total amplitude	
Thymine position	k_A	k_B	A	B
m12	0.59 ± 0.13	0.042 ± 0.008	50.7	49.3
m10	0.57 ± 0.14	0.034 ± 0.007	52.2	47.8
m6	0.54 ± 0.14	0.022 ± 0.005	53.4	46.6
m4	0.35 ± 0.07	0.025 ± 0.004	49.0	51.0

Table 2. Kinetic study of transcription bubble formation upon RNAP binding to wild type *T7A1* promoter at 37°C. Values for the apparent rate constants and the amplitudes of both phases as a function of thymine position on template strand.

6.4. Real−time study of a dynamic of RNAP−DNA interactions upon binary complex formation on the *T7A1* promoter variant with a consensus -10 hexamer at 37°C.

6.4.1. Kinetic characterization of the intermediates formed upon RNAP binding to the mutant T7A1 promoter, using X−ray hydroxyl radical footprinting.

As detailed in literature reviews, the identity of the bases within the -10 region plays a critical role at different steps during the process of open complex formation [Guo et al. 1998, Fenton et al. 2001, McKane et al. 2001]. The wild type *T7A1* promoter contains a non−consensus -10 hexamer (**G**ATA**C**T instead of TATAAT). It was shown that the guanosine at position m12 (position m13 on *T7A1* promoter) is especially detrimental to the stability of the intermediates leading to the open complex [Fenton et al. 2001]. In order to determine how the lack of optimal contacts with this region may influence the structure and stability of the intermediates in this process, we carried out time−resolved X−ray footprinting experiments on a mutant *T7A1* promoter with a consensus -10 sequence.

The changes in the hydroxyl radical reactivity of nucleotides upon RNAP binding to this *T7A1* promoter variant at 37°C were quantified and the kinetics of appearance of protection at discrete promoter regions were obtained in the same way as described above for the wild type promoter (for more details see "Data analysis" and Figures 3S, 4S and 5S as well as Table 3S in "Supporting Materials"). The results of the data evaluation show that the kinetic of appearance of protection at all regions is characterized by a biphasic behavior. The apparent rate constants and the amplitudes of both phases for different promoter regions are summarized in Table 3 and Figure 19.

One of the first steps in the data interpretation was to assign the protection appearing at each promoter region to a specific intermediate RNAP−DNA complex on the basis of kinetic rate constants. This was difficult for the mutant promoter, since in this case the formation of nucleoprotein complex is characterized by a less pronounced difference in the rates with which nucleotides become protected, depending on their position on the promoter (Compare kinetic constants in the Tables 1 and 3 and Figures 15 and 19). While a higher rate of protection appearance at the upstream end of the mutant promoter is still observed, the differences between the other sites are more gradual (Figure 19 (a1) and (a2)). The simplest interpretation of this result is that the consensus -10 hexamer lowers the energetic barriers between the corresponding intermediates. Therefore, the isomerisation can take place at higher rates along the whole pathway.

Mutant T7A1 promoter (consensus -10 region), 37°C, Template strand				
Nucleotide position	**Apparent rates (s⁻¹)**		**% of total amplitude**	
	k_A	k_B	A	B
m55–m53	2.46 ± 0.62	0.183 ± 0.075	60.6	39.4
m45–m42	2.01 ± 0.43	0.260 ± 0.072	56.4	43.6
m33–m31	2.02 ± 0.78	0.301 ± 0.099	45.9	54.1
m23–m20	1.68 ± 0.53	0.239 ± 0.054	43.6	56.4
m18–m14	1.32 ± 0.71	0.239 ± 0.094	43.6	56.4
m13–m10	1.07 ± 0.23	0.151 ± 0.042	56.9	43.1
m9–m6	1.18 ± 0.32	0.196 ± 0.043	46.7	53.3
m5–m3	0.99 ± 0.19	0.138 ± 0.030	52.5	47.5
m2–p1	1.06 ± 0.26	0.165 ± 0.035	47.5	52.5
p2–p4	0.94 ± 0.35	0.179 ± 0.040	39.4	60.6
p5–p9	0.92 ± 0.19	0.163 ± 0.028	47.2	52.8
p10–p13	0.65 ± 0.16	0.166 ± 0.037	48.2	51.8
p14–p20	0.69 ± 0.24	0.187 ± 0.069	51.7	48.3

Mutant T7A1 promoter (consensus -10 region), 37°C, Non–template strand				
Nucleotide position	**Apparent rates (s⁻¹)**		**% of total amplitude**	
	k_A	k_B	A	B
m72–m68	2.07 ± 0.29	0.107 ± 0.038	74.0	26.0
m63–m58	1.94 ± 0.55	0.109 ± 0.023	45.8	54.2
m53–m48	1.41 ± 0.39	0.104 ± 0.013	34.1	65.9
m42–m38	1.13 ± 0.38	0.099 ± 0.013	30.9	69.1
m31–m26	0.78 ± 0.24	0.085 ± 0.010	28.5	71.5
m20–m17	0.72 ± 0.14	0.062 ± 0.006	35.2	64.8
m16–m15	0.90 ± 0.33	0.097 ± 0.017	34.1	65.9
m14–m13	0.64 ± 0.31	0.078 ± 0.017	31.7	68.3
m12–m11	0.71 ± 0.19	0.070 ± 0.009	34.4	65.6
m9–m4	0.63 ± 0.13	0.073 ± 0.007	33.2	66.8
m3–p2	0.44 ± 0.08	0.059 ± 0.006	37.5	62.5
p3–p7	0.44 ± 0.09	0.058 ± 0.007	39.4	60.6
p8–p13	0.42 ± 0.11	0.065 ± 0.009	34.8	65.2
p14–p20	0.19 ± 0.04	0.032 ± 0.007	49.0	51.0

Table 3. Values for the apparent rate constants and the amplitudes of both phases as a function of nucleotide position on template (top panel) and non–template strand (bottom panel) for the case of *E.coli* RNAP holoenzyme binding to the *T7A1* promoter mutant with consensus -10 sequence at 37°C.

Template strand

(a1) (b1) (c1)

Non−template strand

(a2) (b2) (c2)

█ − fast phase   ▬ − slow phase
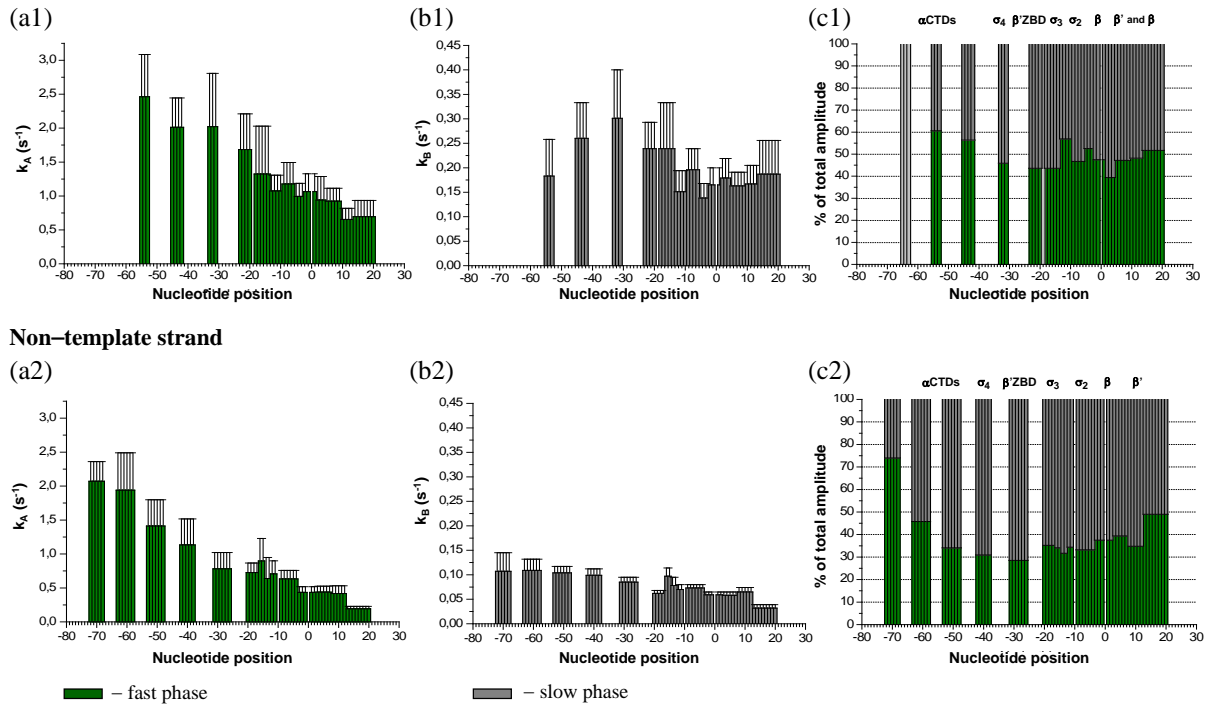
Figure 19. Summary plots of the apparent rate constants and the amplitudes of appearance of the protection on the both strands of *T7A1* promoter mutant with consensus -10 region upon RNAP binding at 37°C.

(**a1**) and (**a2**) Apparent rate constants for the fast phase *versus* the position of the nucleotides on the TS and NTS, respectively. (**b1**) and (**b2**) Apparent rate constants for the slow phase *versus* the position of the nucleotides on the TS and NTS, respectively. (**c1**) and (**c2**) Amplitude of each phase as the percentage of total amplitude *versus* the position of the nucleotides on the TS and NTS, respectively. The RNAP domains known to be responsible for a given protection are shown above their corresponding signals [Naryshkin et al. 2000, Murakami et al. 2002b].

Despite the small differences in the kinetic rate constants of appearance of protection at the different regions of the mutant promoter, we grouped the regions as described previously for the wild type promoter. We attributed the observed "early" protection of the distal UP element subsites (m55−m53 TS and m72−m48 NTS) by the αCTDs to the intermediates A, the first intermediates on the pathway to final binary complex formation (see Figure 20). Protection of the proximal UP element subsite (m45−m42 on the TS and m42−m38 on the NTS) by the αCTDs and of the -35 region (m33−m31 on the TS) by σ region 4.2 appears with a slower rate and corresponds to intermediate B.

Figure 19 shows that on a *T7A1* promoter variant containing a consensus -10 hexamer protection of nucleotides at positions m31−m26 NTS by β'ZBD appears at the same rate as protection downstream, at m20−m4 NTS, reflecting more cooperative conformational changes, probably favored by an improved interaction of σ factor with the -10 region

(intermediate C in the proposed kinetic model, Figure 20). We assigned the extension of protection to position m14 on the TS to the C intermediate.

In the following step the nucleotides from m3 to p13 on the NTS and from m13 to p9 on the TS become protected from hydroxyl radical cleavage as DNA enters the cleft defined by the β and β' subunits of RNAP (D intermediate). Subsequently protection extends down to p20 (E intermediate).

The amplitude of the first exponential phase on both strands of mutant promoter shows a different pattern than the one detected on the wild type promoter (Figure 19 (c1) and (c2) *versus* Figure 15 (c1) and (c2)). In these datasets there is no strongly pronounced difference in the values for the amplitudes between the nucleotide positions m5−m3, p3−p13 TS, p7−p13 NTS and those at the surrounding protected sites, as it was observed on the wild type promoter. This indicates that in this sequence context the formation of the "off−pathway" intermediate is less favoured, probably due to the formation of more stable interactions in the early intermediates that may pull the equilibrium towards the main pathway (see discussion).
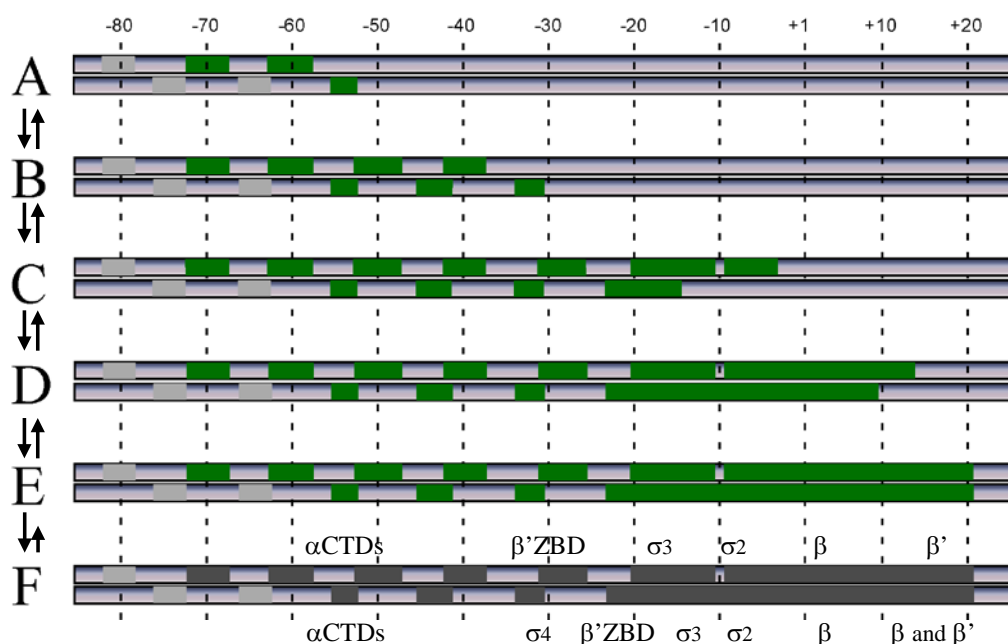
Figure 20. Proposed kinetic model for the binding of *E.coli* RNAP holoenzyme to the *T7A1* promoter mutant with a consensus -10 sequence at 37°C. The DNA strands are shown as rectangles. The non−template strand (NTS) is top rectangle. The template strand (TS) is bottom rectangle. The regions that become protected at each step of RNAP binding to promoter are shown as colored boxes. Light gray boxes correspond to the regions with weak protection; green boxes correspond to the regions which become protected during the first, fast phase; dark gray boxes correspond to the regions which extent of protection increases in the second phase of the kinetics. The RNAP domains known to be responsible for a given protection are shown above (NTS) or below (TS) their corresponding signals [Naryshkin et al. 2000, Murakami et al. 2002b].

Finally, in agreement with the absence of formation of the incorrect complexes, the values for the rates in the second phase are also more uniform in the case of the promoter with consensus -10 hexamer (Figure 19 (b1) and (b2)). The larger values of the apparent rate constants for the second phase compared to values detected with the wild type promoter, more evident in the template strand dataset, may reflect the presence of lower energy barriers for the conformational changes leading to stabilization of the RNAP−DNA complex (F complex).

6.4.2. Kinetic of DNA opening upon binary complex formation on the "-10" consensus promoter, obtained by time−resolved permanganate footprinting experiments.

Literature data, along with our footprinting data, indicate that the formation of an open complex is dependent on the sequence of the -10 hexamer. Two reasons are discussed: the importance of the presence of the consensus base pattern for facilitating proper contact formation between protein and promoter; and a lower energy requirement for the melting of A+T−rich -10 region, which facilitates DNA strand opening.

Therefore, we analyzed the process of DNA melting upon binary complex formation on the *T7A1* promoter variant containing a consensus -10 hexamer by using time−resolved permanganate footprinting to monitor changes in the thymine's accessibility, as has been done for the wild type promoter. The data show that the permanganate signals on the template strand appear with a double exponential (see Figure 6S and Table 4S), with rates indicating that DNA strand separation occurs at the states preceding the entrance of downstream duplex into the ββ' cleft. However, in contrast to what is observed on wild type promoter, the signals at the m12 and m4 thymines appear at the same rate (Table 4). Thus, on the promoter characterized by consensus -10 sequence, DNA melting nucleation and bubble propagation appear to take place in a more cooperative way.

Mutant T7A1 promoter (consensus -10 region), 37°C, Template strand				
	Apparent rates (s^{-1})		% of total amplitude	
Thymine position	k_A	k_B	A	B
m12	0.81 ± 0.29	0.025 ± 0.006	37.2	62.8
m4	0.72 ± 0.25	0.018 ± 0.006	47.6	52.4

Table 4. Kinetic study of transcription bubble formation upon RNAP binding to the mutant *T7A1* promoter at 37°C. Values for the apparent rate constants and the amplitudes of both phases as a function of thymine position on template strand.

6.5. Real–time description of a process of binary complex formation on the wild type *T7A1* promoter at 20°C.

Changing the -10 sequence to its consensus form increases the efficiency of binary complex formation, probably by favouring specific promoter–protein interactions, as shown above. A decrease in the temperature should have an opposite effect on the efficiency of formation of those intermediates in the pathway that are dependent on the presence of single–stranded DNA due to disfavouring melting of the DNA double helix.

We have compared the time–resolved hydroxyl radical footprints obtained at 20°C with those formed at 37°C and described above, with the aim of identifying conformational changes responsible for the strong temperature effects previously observed at *T7A1* promoter (see "Discussion") [Zaychikov et al. 1997, Johnson and Chester 1998].

The time–dependent change in the hydroxyl radical footprinting pattern was determined in the same way as described above. For more methodological details see "Data analysis" and Figures 7S, 8S and 9S as well as Table 5S in "Supporting Materials". The results of the data evaluation are summarized in Table 5 and Figure 21.

The rates of the first phase in the formation of the intermediates at 20°C are very similar to those measured at 37°C. However, the amplitude of the first phase shows a pattern which is very different from those observed at 37°C. While the differences in the rates of the first phase on the non–template strand allow the identification of different intermediate structures, on the template strands all the rates are more or less the same within the experimental error. On the non–template strand there is a strong protection in the distal UP element subsites that appears at the first step of nucleoprotein complex formation (A intermediates in the proposed kinetic model, Figure 22), while on the template strand the weaker protection in this region does not allow a precise quantification of the kinetics.

Wild type T7A1 promoter, 20°C, Template strand				
Nucleotide position	Apparent rates (s⁻¹)		% of total amplitude	
	k_A	k_B	A	B
m46–m42	0.88 ± 0.12	0.050 ± 0.024	77.7	22.3
m34–m29	1.06 ± 0.18	0.106 ± 0.036	65.0	35.0
m23–m19	0.96 ± 0.11	0.076 ± 0.014	62.1	37.9
m18–m16	0.70 ± 0.18	0.046 ± 0.015	53.7	46.3
m15–m8	0.86 ± 0.09	0.089 ± 0.013	56.1	43.9
m7–m6	0.81 ± 0.19	0.041 ± 0.007	41.0	59.0
m5–m3	0.66 ± 0.18	0.046 ± 0.006	33.3	66.7
m2–p3	1.18 ± 0.30	0.078 ± 0.010	36.7	63.3
p4–p9	0.59 ± 0.11	0.032 ± 0.003	31.2	68.8
p10–p17	0.72 ± 0.17	0.029 ± 0.003	31.0	69.0

Wild type T7A1 promoter, 37°C, Non–template strand				
Nucleotide position	Apparent rates (s⁻¹)		% of total amplitude	
	k_A	k_B	A	B
m72–m70	2.52 ± 0.39	0.142 ± 0.100	81.9	18.1
m63–m57	2.36 ± 0.31	0.137 ± 0.033	65.8	34.2
m53,m52,m48	1.38 ± 0.23	0.072 ± 0.030	71.4	28.6
m51–m49	1.75 ± 0.22	0.130 ± 0.038	71.3	28.7
m43,m42,m38	1.13 ± 0.30	0.160 ± 0.104	70.8	29.2
m41–m39	1.07 ± 0.14	0.080 ± 0.033	72.6	27.4
m32–m30,m26	0.84 ± 0.14	0.084 ± 0.027	65.3	34.7
m29–m27	0.85 ± 0.12	0.084 ± 0.030	70.7	29.3
m20–m18	0.84 ± 0.17	0.102 ± 0.040	64.5	35.5
m17–m14	0.75 ± 0.11	0.087 ± 0.021	61.6	38.4
m9–m7	0.76 ± 0.24	0.127 ± 0.057	58.2	41.8
m6–m3	0.68 ± 0.16	0.058 ± 0.011	42.1	57.9
m2–p2	0.39 ± 0.13	0.047 ± 0.008	32.3	67.7
p3–p6	0.41 ± 0.11	0.049 ± 0.008	36.0	64.0
p7–p12		0.048 ± 0.002		
p13–p20		0.056 ± 0.003		

Table 5. Values for the apparent rate constants and the amplitudes of both phases as a function of nucleotide position on template (top panel) and non–template strand (bottom panel) for the case of *E.coli* RNAP holoenzyme–wild type *T7A1* complex formation at 20°C.

Template strand

(a1) (b1) (c1)

Non−template strand

(a2) (b2) (c2)

■ − fast phase   ■ − slow phase
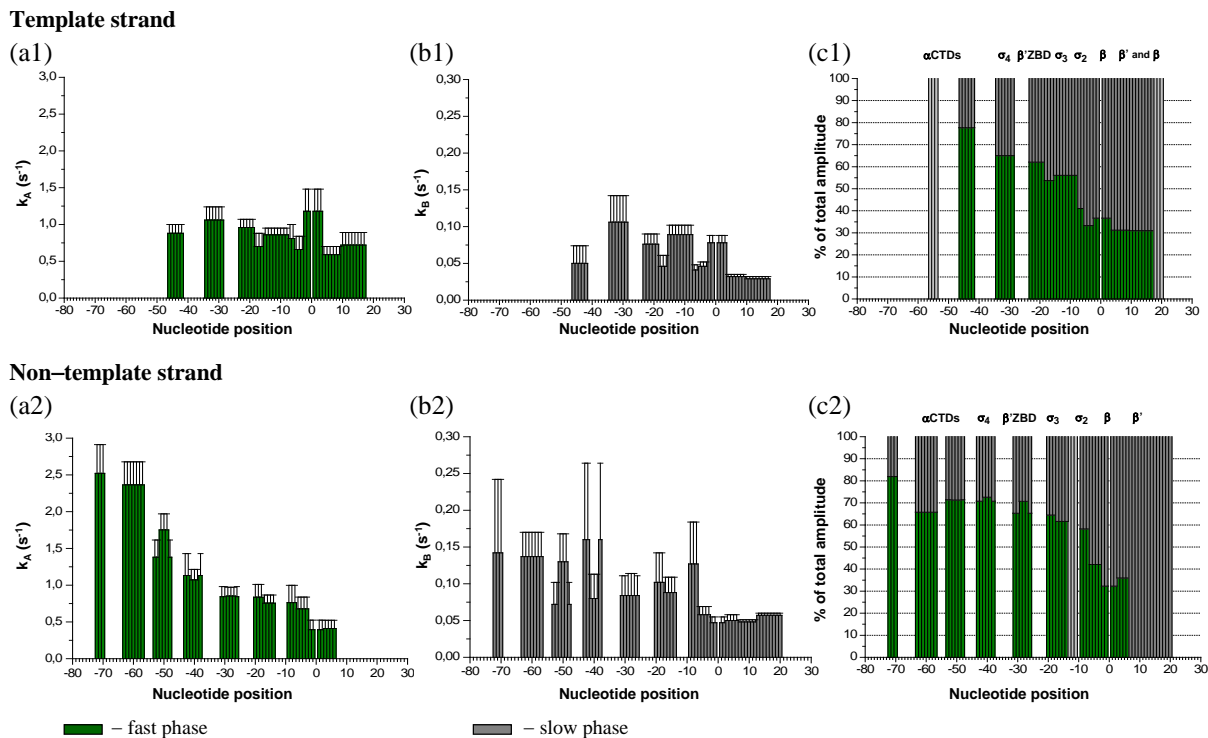
Figure 21. Summary plots of the apparent rate constants and the amplitudes of appearance of the protection to hydroxyl radical cleavage on the both strands of wild type *T7A1* promoter upon RNAP binding at 20°C.

(**a1**) and (**a2**) Apparent rate constants for the fast phase *versus* the position of the nucleotides on the TS and NTS, respectively. (**b1**) and (**b2**) Apparent rate constants for the slow phase *versus* the position of the nucleotides on the TS and NTS, respectively. (**c1**) and (**c2**) Amplitude of each phase as the percentage of total amplitude *versus* the position of the nucleotides on the TS and NTS, respectively. The RNAP domains known to be responsible for a given protection are shown above their corresponding signals [Naryshkin et al. 2000, Murakami et al. 2002b].

The data of Figure 21 indicate that the next step results in protection of the proximal UP element subsite (m43−m38 NTS and m46−m42 TS) and of the -35 region (m34−m29 TS) (B intermediate in the proposed kinetic model). Furthermore, as observed for the consensus -10 promoter at 37°C, on the wild type *T7A1* promoter at 20°C the protection of the nucleotides at positions m31−m26 on the NTS appears at the same rate as the protection downstream, at m20−m3 NTS, suggesting that there is a decrease in a kinetic barrier during the formation of contacts with the -10 region compared to what observed at 37°C. This could be due to the specific recognition of the still double−stranded DNA by the σ subunit, as observed by Fenton and co−workers [Fenton et al. 2001], and to a decrease in the flexibility of double−stranded DNA which reduces the number of possible orientations of the double helix. The difference in amplitudes of the first, fast phase for the protection at the m20−m7 NTS and m6−m3 NTS (Figure 21 (c2)) allows us to propose the existence of two complexes in rapid equilibrium, one whose protection extends to m7 NTS and a second one with

56

protection to m3 NTS (C and C' intermediates in Figure 22) We also propose that, on the template strand, the protection to m8 corresponds to the C intermediate, and the additional protection at positions from m7 to m6 and from m2 to p3 corresponds to the C' intermediate.

In the subsequent step the protected region extends to p6 NTS (D intermediate in the proposed kinetic model). The kinetic of protection appearance at the nucleotides from m2 to p6 follows a biphasic behavior with the fast phase accounting for ~ 35% of the total amplitude. We assigned the protection of the template strand to position p17 to the D intermediate.

The notable difference, compared to the other two data sets, is that at 20°C the protection downstream of p6 on the non−template strand is monophasic and appears at the rate which corresponds to the rate of the second, slow phase present at the other sites (E complex in the proposed kinetic model). Thus, at low temperature the conformational change resulting in the formation of protein contacts with downstream double−stranded DNA becomes rate−limiting.
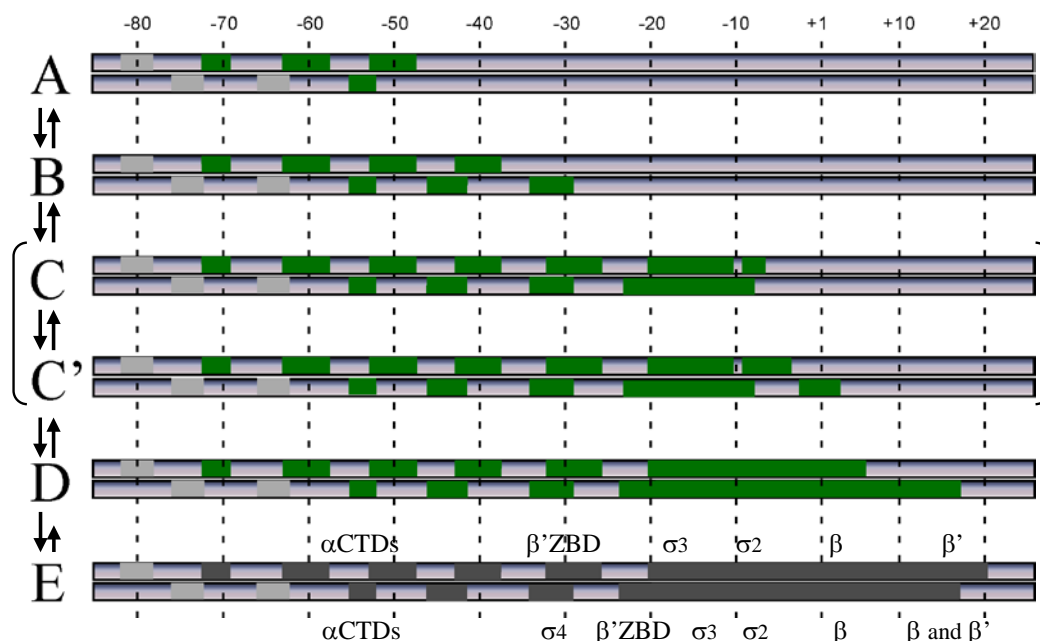
Figure 22. Proposed kinetic model for the binding of *E.coli* RNAP holoenzyme to the wild type *T7A1* promoter at 20°C. The DNA strands are shown as rectangles. The non−template strand (NTS) is top rectangle. The template strand (TS) is bottom rectangle. The regions that become protected at each step of RNAP binding to promoter are shown as colored boxes. Light gray boxes correspond to the regions with weak protection; green boxes correspond to the regions which become protected during the first, fast phase; dark gray boxes correspond to the regions which extent of protection increases in the second phase of the kinetics. The RNAP domains known to be responsible for a given protection are shown above (NTS) or below (TS) their corresponding signals [Naryshkin et al. 2000, Murakami et al. 2002b].

The absence of the larger amplitude for the first phase of the protection appearance at positions m5−m3 and p3−p13 TS and p7−p13 NTS compared to the surrounding protected sites, in contrast to what detected at 37°C, indicates that under these conditions there is no significant accumulation of the "off−pathway" intermediate.

Figures 21 (b1) and (b2) show that the rates for the second phase of these datasets can also be divided into two groups, the slower rates corresponding to the stabilization of contacts with downstream DNA (m6−p20 NTS and p4−p20 TS).

6.6. Biochemical characterization of the final open complexes formed with the *T7A1* promoters having mutations in different regions.

The results of time−resolved footprinting studies described in previous paragraphs show that the pathway to the final binary complex depends on the sequence of -10 hexamer. To check whether the changes in the promoter sequence affect the conformation of the transcriptionally active complex and the efficiency of productive RNA synthesis, we performed a set of additional tests, which will be discussed below.

6.6.1. Stability of the final open complex.

The open complex has two possible fates: it can convert into a stable elongation complex when ribonucleoside triphosphate substrates (rNTPs) are added or it can convert back to closed complex and eventually dissociate into free DNA and RNAP. The amount of dissociated RNAP can be determined by trapping with the polyanion heparin, which acts as a DNA competitor. This so called heparin challenge test can be used to follow the dissociation of the RNAP−promoter complex.

Wild type *T7A1* promoter is one of the few which form an open binary complex with RNAP, equilibrating reversibly with the free components, as measured by heparin challenge [Susa et al. 2002 and reverences therein]. The data of our time−resolved footprinting studies indicate that mutation of the -10 region towards the consensus sequence leads to a more efficient formation of the final open complex. In order to elucidate whether the consensus sequence in the -10 region also results in a higher stability of the final binary complex, we subjected the open complex, formed on the wild type *T7A1* promoter, as well as on the promoter variant with a consensus -10 region, to heparin challenge.
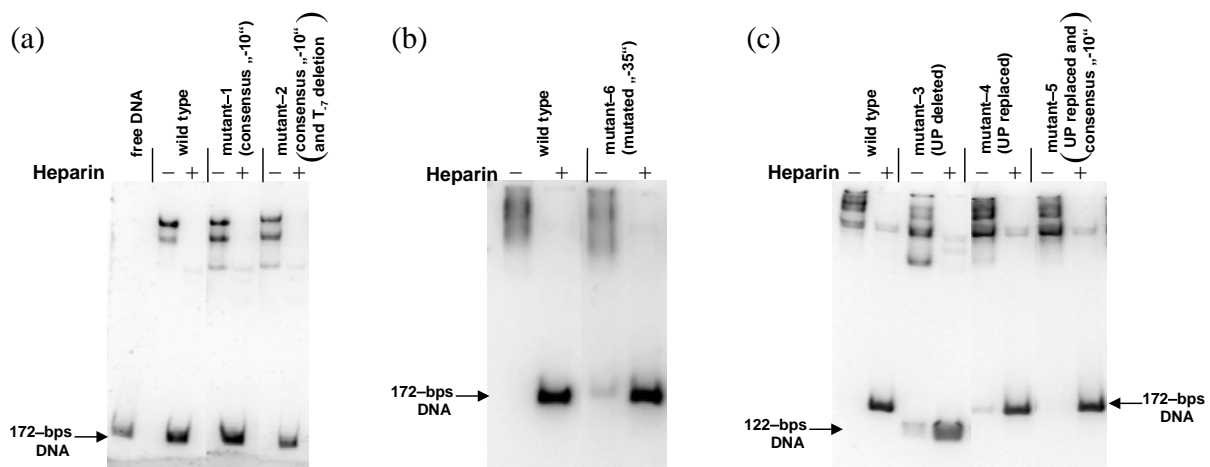
Figure 23. EMSA of *E.coli* σ70 RNAP with wild type *T7A1* promoter and different promoter mutants. In each case, heparin was added after formation of final open complex.

Electrophoretic mobility−shift assay (EMSA) data show that in both cases addition of heparin to a solution containing RNAP bound to a promoter results in the displacement of the promoter (Figure 23 (a)). Supporting the results of studies made by Fenton and his colleagues [Fenton et al. 2001] on different fork−junction DNA fragments, these data indicate that changing the base identity in the -10 region has little effect on the resistance of the final open complex to heparin.

Numerous studies revealed that not only -10 region but also the other core promoter regions, as well as upstream element, influence the efficiency of transcription initiation which is primarily determined by the balance of promoter binding and activation, and RNA chain initiation and promoter escape. We therefore tested which of these steps could be affected by mutations within the -35 hexamer and the upstream region of *T7A1* promoter.

The band−shift assay performed with the *T7A1* promoter mutant having a non−consensus -35 region (mutant−6), as well as with mutants having the UP element deleted (mutant−3) or replaced for non−A+T−rich sequence (mutant−4) (the sequences of these mutants are represented in Figure 24), shows less effective binding to RNAP (Figures 23 (b) and (c)). The substitution of the wild type -10 sequence for the consensus one in the UP−replaced mutant (mutant−5) however results in restoring of binding efficiency. Each of these mutants forms with RNAP a final complex that is sensitive to heparin challenge.

Mutant–2. **Consensus -10 region and T(-7) deletion** (→ consensus distance to +1)

```
 -90       -80       -70       -60       -50       -40       -30       -20       -10       +1        +10
  .         .         .         .         .         .         .         .         .         .         .
CGAGGCCAACTTAAAGAGACTTAAAAGATTAATTTAAAATTTATCAAAAGAGTATTGACTTAAAGTCTAACCTATAGTATAAT-ACAGCCATCGAGAGGGA.........
GCTCCGGTTGAATTTCTCTGAATTTTCTAATTAAATTTTAAATAGTTTTTCTCATAACTGAATTTCAGATTGGATATCATATTA-TGTCGGTAGCTCTCCCT.........
```

Mutant–3. **UP element deleted** (→ 122nt–long DNA fragment)

```
                                                       -40       -30       -20       -10       +1        +10
                                                        .         .         .         .         .         .
                                                     GAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGGA.........
                                                     CTCATAACTGAATTTCAGATTGGATATCCTATGAATGTCGGTAGCTCTCCCT.........
```

Mutant–4. **UP element replaced**

```
 -90       -80       -70       -60       -50       -40       -30       -20       -10       +1        +10
  .         .         .         .         .         .         .         .         .         .         .
CGAGGCCAACTTAAAGAGACCCCTTTCGTCTTCAAGAATTCCTCGAGGAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGGA.........
GCTCCGGTTGAATTTCTCTGGGGAAAGCAGAAGTTCTTAAGGAGCTCCTTCTCATAACTGAATTTCAGATTGGATATCCTATGAATGTCGGTAGCTCTCCCT.........
```

Mutant–5. **UP element replaced and consensus -10 region**

```
 -90       -80       -70       -60       -50       -40       -30       -20       -10       +1        +10
  .         .         .         .         .         .         .         .         .         .         .
CGAGGCCAACTTAAAGAGACCCCTTTCGTCTTCAAGAATTCCTCGAGGAAGAGTATTGACTTAAAGTCTAACCTATAGTATAATTACAGCCATCGAGAGGGA.........
GCTCCGGTTGAATTTCTCTGGGGAAAGCAGAAGTTCTTAAGGAGCTCCTTCTCATAACTGAATTTCAGATTGGATATCATATTAATGTCGGTAGCTCTCCCT.........
```

Mutant–6. **Mutant -35 region**

```
 -90       -80       -70       -60       -50       -40       -30       -20       -10       +1        +10
  .         .         .         .         .         .         .         .         .         .         .
CGAGGCCAACTTAAAGAGACTTAAAAGATTAATTTAAAATTTATCAAAAGAGTAGAATTCTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGGA.........
GCTCCGGTTGAATTTCTCTGAATTTTCTAATTAAATTTTAAATAGTTTTTCTCATCTTAAGATTTCAGATTGGATATCCTATGAATGTCGGTAGCTCTCCCT.........
```

Figure 24. Sequences of the *T7A1* promoter mutants.

6.6.2. *The efficiency of promoter escape.*

We also analyzed the efficiency of transition of the open complex to an elongation complex (promoter escape) for each sequence context. A large body of published data indicates that after addition of rNTPs, the transition from initiation to the elongation phase of transcription is typically accompanied by the production of an abundant level of short RNA transcripts ranging from 2 to 15 nucleotides in length (abortive transcripts) due to repetitive release of the nascent RNA by the initial transcribing complex followed by the reinitiation of RNA synthesis [Hsu et al. 2003 and references therein]. Abortive initiation ceases when RNAP succeeds to continue the elongation of the RNA product and moves away from the promoter. Abortive initiation is thought to reflect the strong promoter–holoenzyme interactions that prevent escape of RNAP into the elongation mode and instead result in the dissociation of the incipient transcripts [Vo et al. 2003].

Figures 25 (a) and (b) illustrate the time courses of synthesis of 11nt–long and 20nt–long RNA products respectively from the wild type *T7A1* promoter and the promoter mutants. One can see that "improvement" of the -10 sequence leads to both more abortive and run–through products in an assay that results in 11nt–long RNA transcription (Figure 25 (a)

mutants 1, 2, 5). In an assay that results in 20nt−long RNA transcription, increased amounts of pre−terminated RNAs (12–15nt in length) are also formed in contrast to what observed on the wild type promoter (Figure 25 (b) mutant−1). We infer therefore that the presence of a consensus -10 sequence decreases the efficiency of transcription probably by impairing the ability of the enzyme to release the promoter, in agreement with the proposal of Vo and co−workers [Vo et al. 2003]. This is also confirmed by the finding that the deletion (mutant−3) or replacement of UP element for non−A+T−rich sequence (mutant−4), resulting in weaker interaction of the RNAP α subunits with upstream promoter region, leads to enhanced productive synthesis.

Moreover, Figures 25 (a) and (b) show that a mutation in the -35 region (mutant−6) drastically decreases the rate of transcription.

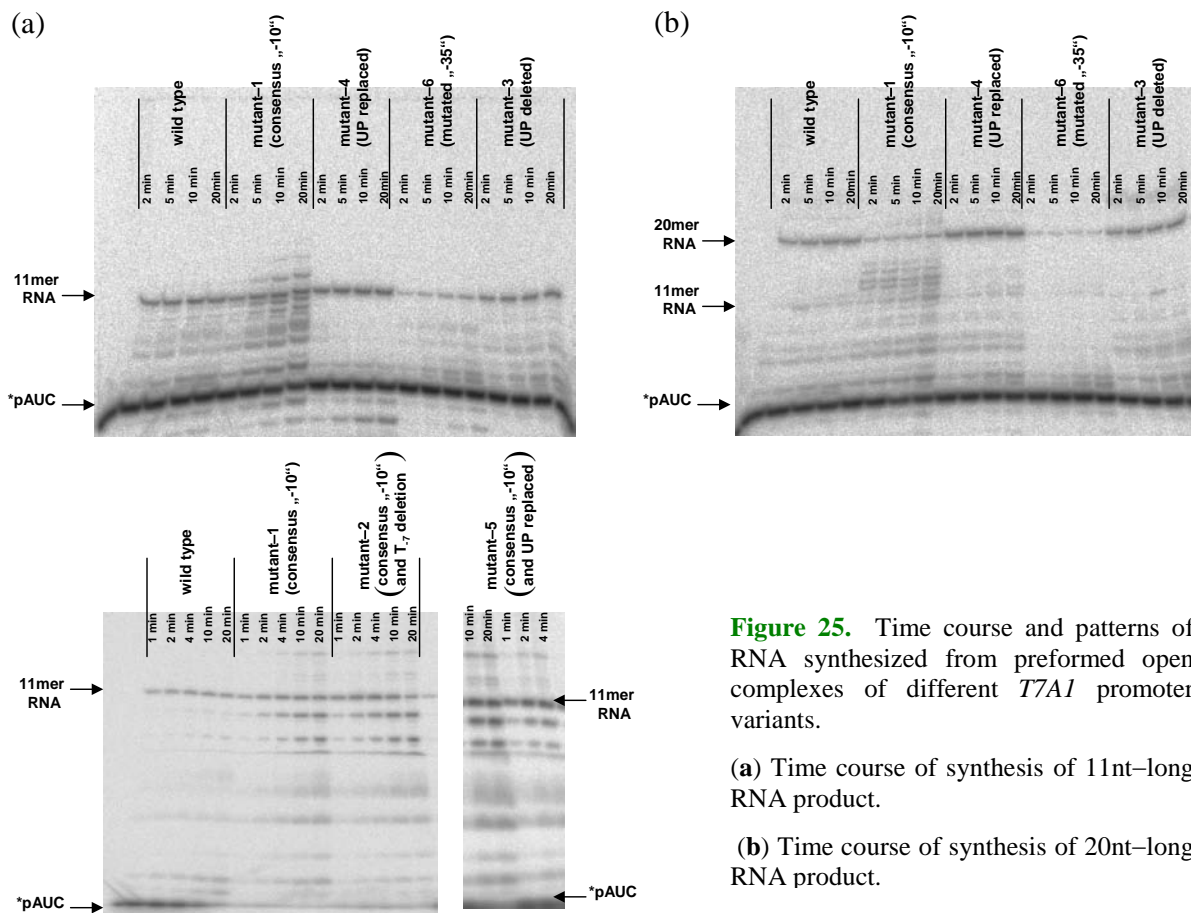**Figure 25.** Time course and patterns of RNA synthesized from preformed open complexes of different *T7A1* promoter variants.

(**a**) Time course of synthesis of 11nt−long RNA product.

(**b**) Time course of synthesis of 20nt−long RNA product.

6.6.3. *Mapping size and position of transcription bubble.*

In order to determine if the differences in transcription properties are associated with the differences in the structure of the transcription bubble, we performed potassium permanganate probing of the final open complexes formed with the *T7A1* promoter mutants 1−6. No changes in the bubble position and size were revealed in the complexes formed with mutants 1−5 compared to those observed with the wild type *T7A1* promoter (data not shown). However the transcription bubble in the binary complex formed with mutant−6 having a different -35 hexamer sequence shows considerable particularities (Figure 26). While the binding of RNAP to the wild type promoter and other mutants results in open complexes having a bubble from m12 to p2, in the case of the promoter derivative with the non−consensus -35 region, the open complex is characterized by the transcription bubble extending from position m15 to at least m4. Moreover, in this sequence context, the non−template strand thymine at m8 is accessible to permanganate oxidation confirming that the interactions between RNAP and the melted DNA strands are not the same as those formed on the wild type promoter. This reflects the observed differences in the ability of RNAP initiate RNA synthesis in this sequence context.
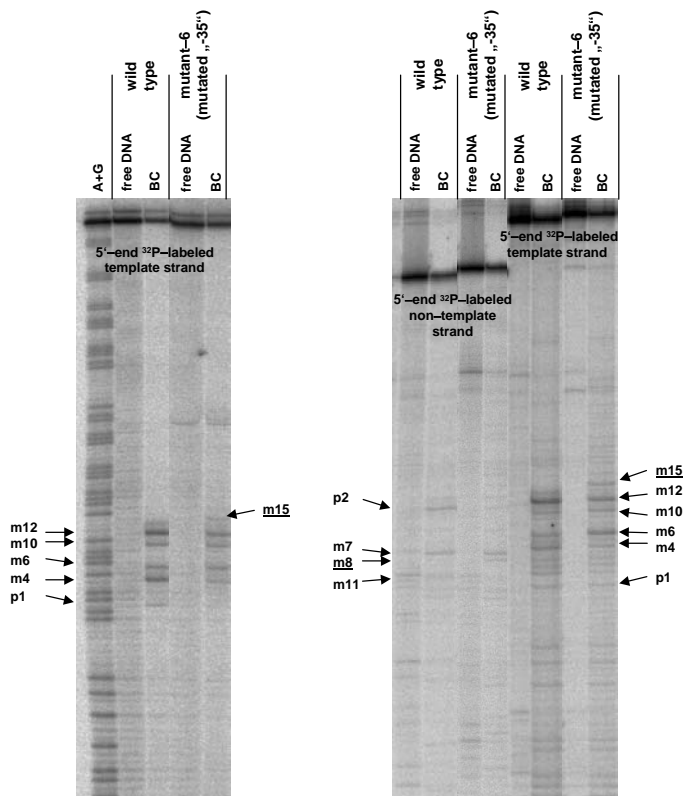

Figure 26. Comparative probing of the binary complexes formed on wild type and mutant−6 *T7A1* promoters. Complexes were formed using promoter fragments radioactively labeled at either the template or non−template strand and probed with potassium permanganate (for details see "Materials and Methods").

7.   Discussion.

The binding of RNAP to promoter DNA is a key step in the regulation of gene expression. Thus, in order to understand how regulatory mechanisms control RNAP activity at the promoter, it is necessary to define the cascade of events that occur during the process of transcription initiation. This process has been described as a multi−step pathway that begins with the promoter recognition and the formation of a closed complex, followed by a series of large−scale conformational changes in both the protein and the DNA resulting in the formation of a transcriptionally active open complex [Roe et al. 1985, Buc and McClure 1985, Craig et al. 1998, Kontur et al. 2006, Davis et al. 2007]. However, the precise description of structural intermediates in the pathway leading to the formation of an active complex remains a challenge mainly because these intermediates are short−lived complexes.

In this work, we used both time−resolved X−ray hydroxyl radical and potassium permanganate footprinting in order to directly identify and characterize the intermediates in the pathway of *T7A1* promoter recognition and open complex formation by *E. coli* σ^{70} RNA polymerase. The comparison of the kinetic and structural properties of these intermediates on two different promoter sequences and at two different temperatures allows us to propose specific conformational changes which take place along the pathway and to associate specific structures to the rate−limiting steps. In addition, we applied classical methods, such as heparin challenge test and promoter escape assay, to describe the properties of final binary complexes formed at wild type *T7A1* promoter as well as its mutant variants.

7.1.   Characterization of the kinetically determined intermediates on the basis of structural information.

As noted in the section "Footprinting technique and its application for the study of DNA−protein interactions", hydroxyl radicals can cut the DNA backbone at all solvent accessible sites. Thus, each pattern of DNA protection observed after incubation of the promoter with RNAP for a specific time corresponds to a specific structure of the binary complex. Based on the results of the experiments described in this work and the available crystallographic and biochemical data on RNAP−promoter interactions at equilibrium, we will discuss the mechanism of the promoter recognition and open complex formation in terms of the conformational rearrangements of both protein and DNA.

7.1.1. *A complexes.*

The results of our studies reveal that the earliest complexes in the pathway of the *T7A1* promoter binding by RNAP are characterized by the modulated protection of the extensive upstream promoter region containing several A+T−stretches (UP element) (Figure 16, page 45). Based on photocrosslinking [Naryshkin et al. 2000] and numerous footprinting studies [Aiyar et al. 1998, Ross et al. 1998, Ross et al. 2005, Davis et al. 2005], this protection (at base positions m72−m68, m63−m58, m53−m48 on the NTS and m76−m73, m66−m63, m55−m53 on the TS) is attributed to an interaction with the two αCTDs of RNAP (Figure 27, page 65). The αCTDs may occupy different sites in different molecules in the population of DNA fragments, or may oscillate between different binding sites. Supporting the idea arising from the our results that the binding of αCTDs at the distal upstream promoter region plays a significant role in the process of promoter recognition by RNAP, data of several other studies showed that the truncation of αCTDs or DNA upstream of position m45 greatly reduces the overall rate of RNAP association with promoter [Ross et al. 2005, Davis et al. 2005]. Hence, the most likely functional role of αCTDs is directing the DNA double helix and the rest of the protein into the correct orientation, facilitating the next isomerisation steps.

7.1.2. *B complexes.*

On the wild type promoter at 37°C we show that the first isomerisation step (A to B, Figure 16, page 45 and Figure 27, page 65) results in the appearance of protection at the proximal UP element subsite (m45−m42 TS and m41−m38 NTS) recognized by the αCTDs, and at the -35 sequence (m34−m31 TS) recognized by σ region 4.2 [Siegele et al. 1989, Moyle et al. 1989, Campbell et al. 2002]. An attractive possibility, based on the close proximity of αCTD bound to a proximal UP element subsite to σ region 4.2 bound to -35 region, is that these RNAP domains may specifically interact with each other and this may further assist the recruitment of RNAP to the promoter. Such possibility was confirmed by the data reported by several groups [Ross et al. 2003, Lee et al. 2003].

Furthermore, on the *T7A1* promoter at 37°C the nucleotides at positions m31−m27 on the NTS become protected in the same timescale, however show a lower amplitude of the fast phase than the amplitude measured for the protected region just upstream. We therefore attribute the protection extending down to m27 to an intermediate B' that is in rapid equilibrium with B.

DNA   RNAP

σ4 σ3 σ2
αCTD-I   β1
αCTD-II   β2   βDR1
           β' jaw   β'GNCD

+

A   A'   A''

B   B'
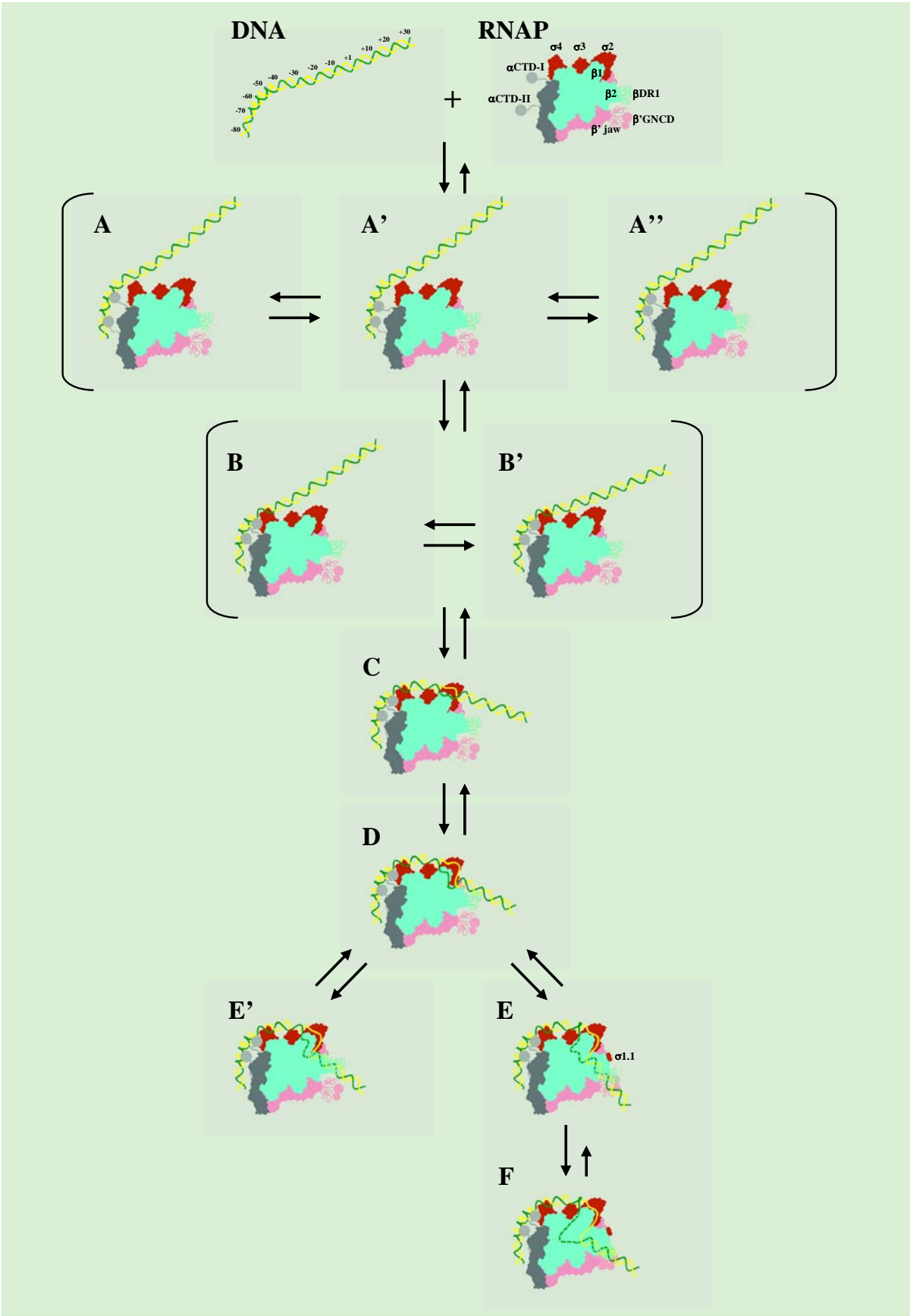
C

D

E'   E   σ1.1

F

Figure 27. Proposed structures for the key intermediates in the pathway leading to the formation of a transcriptionally active complex on the wild type *T7A1* promoter at 37°C. The model of the closed complex and the model of the open complex proposed by Murakami and co−workers from the crystal structure of the *Taq* RNAP holoenzyme in complex with fork−junction promoter DNA were used as the starting points to create these images [Murakami et al. 2002b]. The RNAP subunits are color coded as follows: αCTDs, light gray; αNTDs, dark gray; β, green−cyan; *Ec*βDR1, light green−cyan; β', pink; *Ec*β'GNCD, light pink; σ, dark orange. The template strand is colored green. The non−template strand is yellow. The shaded parts of the DNA are those that have entered the active site channel and the ββ' cleft and are therefore placed behind the β subunit. Following the formation of early complexes stabilized by the interactions of the αCTDs with the UP element and σ$_4$ domain with the -35 region of the promoter (A and B intermediates) the DNA is bent towards σ$_3$ and σ$_2$ domains where contacts are made with the spacer and the upstream end of the -10 region (C intermediate). At this stage, the separation of the DNA strands from m12 to m6 takes place. In the subsequent step, further bending of the DNA towards the protein surface is associated with the propagation of DNA melting (D intermediate). DNA then enters the cleft defined by β and β' subunits and becomes protected downstream to p20 (E intermediate). Finally, while the pattern of DNA protection does not changes, the extent of protection increases as the RNAP−DNA complex isomerises into a transcriptionally active complex where the template strand is properly placed at the active site (F complex). Protection of the DNA from p14 to p20 is likely due to interactions with *Ec*βDR1 and *Ec*β'GNCD and their subsequent folding stabilizing the transcriptionally active complex [Davis et al. 2007]. The E' intermediate corresponds to the off−pathway complex.

In the crystal structure of *Taq*σ$^A_4$ domain in complex with -35 element DNA, a 36° bend was observed in the -35 region [Campbell et al. 2002]. Results of our X−ray footprinting experiments show increased sensitivity to hydroxyl radicals at nucleotides m39−m37 TS and m35−m33 NTS (see Figure 12, page 38). These data suggest that the isomerisation step leads to the bending of DNA double helix in the -35 region. Moreover, in line with the structural model of the binary complex proposed by Vassylyev and co−workers [Vassylyev et al. 2002], the enhancement of hydroxyl radical reactivity of nucleotides at about m27−m25 TS suggests that the DNA double helix seems to be locally bent or kinked also in the spacer region. This deformation of the DNA might be caused by an interaction with β'ZBD that intervenes between the σ$_4$ and σ$_2$ domains and is necessary to bring the -10 promoter element and σ$_2$ domain into contact [Vassylyev et al. 2002, Murakami et al. 2002b].

7.1.3. C complex.

Our hydroxyl radical footprinting data show that in the subsequent step the protections appear at positions m23−m20 and m12−m9 on the TS and m20−m5 on the NTS (C intermediate, Figure 16, page 45 and Figure 27, page 65). Crystallographic studies revealed that most of these protections are due to σ regions 3.0, 2.4, and 2.3 protruding above the surface of the polymerase [Murakami et al. 2002b].

Based on the crystal structure of the *Taq* RNAP holoenzyme in complex with fork−junction promoter DNA, Murakami suggested the model of the closed complex

[Murakami et al. 2002b, Murakami and Darst 2003]. In this model the DNA double helix continues on a straight path from σ_2 in such a way, that the nucleotides downstream of m9 NTS are not near to any of the polymerase's subunits. Taking into consideration this model and the observation that nucleotides from m9 to m5 on the NTS are protected in the C complex, we conclude that, upon isomerisation from intermediate B to intermediate C, DNA has been further curved towards the β and β' pincers. This DNA bending can be facilitated by the nucleation of promoter melting that sufficiently increases the DNA flexibility, as already proposed by several other groups [Vassylyev et al. 2002, Saecker et al. 2002].

As outlined in the literature review, the non−template strand adenine at position m11 and thymine at position m10, corresponding to position m12 and m11 on the *T7A1* promoter, respectively, are the key nucleotides for the nucleation of DNA opening. It has been reported that the disruption of base−pairing at the upstream edge of the -10 hexamer followed by specific recognition of this region by highly conserved aromatic residues of σ region 2.3 leads to the separation of the DNA strands down to position p3 [Heyduk et al. 2006 and references therein]. But it is still unknown at which steps in the pathway to final complex formation these events take place.

The use of both time−resolved X−ray hydroxyl radical and potassium permanganate footprinting has allowed us to directly show that the thymines m12, m10 and m6 on the template strand of DNA become accessible to $KMnO_4$ modification at a rate that is similar to the rate of appearance of protection of the non−template strand in the -10 region (C complex), while the modification of thymine at m4 is associated with the next, slower step during which protection extends to include the transcription start site (D complex).

Our results thus indicate that on the *T7A1* promoter at 37°C DNA melting takes place within the early intermediates, before RNAP forms full contacts with downstream DNA, and that propagation of the bubble is associated with a slow isomerisation from the C to the D intermediate. This observation contrasts with the previously proposed model for the formation of a transcriptionally active complex at *T7A1* promoter that placed DNA melting within the last step of the pathway [Schickor et al. 1990]. This discrepancy can be explained by the fact that the low time resolution of the experimental approach available in the previous study did not allow direct analysis of the dynamics of RNAP−promoter interactions. This study was based on the postulation that the basic sequence of events leading to a trancsriptionally active complex is not affected by temperature and therefore the RNAP−promoter complexes trapped

at low temperatures are equivalent to the intermediates formed in the kinetic pathway at 37°C. Our results indicate that such postulation is not correct.

7.1.4. *D complex.*

We suggest that the partial melting of the double helix in the C intermediate increases the flexibility of the DNA at the bubble, allowing further bending or kinking of the downstream duplex towards the β and β' pincers. This DNA bending results in the appearance of an additional protection at m19−m13, m8−m6, m2−p2 on the TS and m4−p6 on the NTS (D intermediate, Figure 16, page 45 and Figure 27, page 65). In the D intermediate, the protein already forms interactions with the double−stranded DNA downstream of the opened bubble but the single−stranded DNA is probably not yet in the correct position, as suggested by the finding that the nucleotides at m5−m3 TS remain exposed to the solvent.

The data of numerous studies on the structure of RNAP−promoter complexes allow us to suggest that in the D intermediate contacts with nucleotides at m8−m6, m2−p2 on the TS and m4−p6 on the NTS are made by σ regions 1.2, 2.2 and 2.3 and the β subunit [Naryshkin et al. 2000, Mekler et al. 2002, Murakami et al. 2002b, Haugen et al. 2006], in agreement with the proposed role of these interactions in the melting process [Naryshkin et al. 2000, Nechaev et al. 2000, Brodolin et al. 2005].

Johnson and Chester measured the rate at which the intrinsic fluorescence of the protein changes as RNAP binds to the *T7A1* promoter [Johnson and Chester 1998]. They observed biphasic kinetics and interpreted these results by assigning the first, fast phase to the formation of an early, closed complex and the second phase to a subsequent isomerisation step resulting in DNA melting. The rate of this isomerisation step is similar to the rate of the formation of the D complex reported here, suggesting that this step in the pathway (C to D), characterized by the propagation of DNA strand separation, is also associated with the large conformational changes in the protein. These conformational changes seem to be less favourable on the wild type *T7A1* promoter, apparently due to non−optimal contacts of σ_2 domain with non−conserved guanosine at m13 and cytosine at m9 on the NTS, as suggested by a significant difference in the rates of formation of the C and D intermediates, in contrast to what is observed on promoter variant with a consensus -10 hexamer.

In summary, the results of our potassium permanganate and hydroxyl radical X−ray footprinting studies indicate that the D intermediate is an open binary complex characterized

by a transcription bubble extending at least to position m4 and by a downstream protection extending only to position p2 on the TS and to position p6 on the NTS. Such protection pattern reflects the fact that the DNA double helix is not yet positioned in the cleft defined by β and β' pincers. This complex is not yet functionally active, since the rate of formation of an active complex on *T7A1* promoter, measured by abortive initiation assay, is about 10 times slower [Johnson and Chester 1998].

7.1.5. E complex.

In the following step in the pathway leading to the formation of a transcriptionally active complex on *T7A1* promoter (D to E), protection appears at m5−m3, p3−p13 and p14−p20 on the TS and p7−p13, p14−p20 on the NTS (Figure 16, page 45 and Figure 27, page 65). Based on the structural models proposed for the final open complex, we conclude that this protection results from the entry of downstream DNA within the ββ' cleft and interactions of the DNA with the β2 domain (also named downstream lobe) of the β pincer on one side and the downstream half of the β' pincer (named β' downstream jaw) on the other [Naryshkin et al. 2000, Murakami et al. 2002b, Mekler et al. 2002]. More specifically, the protection from p14 to p20 is dependent on an interaction with the *E.coli*−specific (*Ec*) β'GNCD (β' G Non−Conserved Domain) and possibly the *Ec*βDR1 (β Dispensable Region 1) positioned near the top of the β2 domain [Saecker et al. 2002, Chlenov et al. 2005, Kontur et al. 2006]. It has been shown that both *Ec*βDR1 and in particular *Ec*β'GNCD of *E.coli* RNAP contribute to the stability of the final open complex formed at the *T7A1* promoter [Artsimovitch et al. 2003] as well as at the λP_R promoter [Kontur et al. 2006].

Studies of the process of binary complex formation at the λP_R promoter suggested that the entry of the downstream DNA into the cleft defined by the pincers is accompanied by large conformational changes in the β subunit, such as repositioning of the downstream lobe of β and folding of *Ec*βDR1, followed by the rate−limiting step in which folding of the conserved G and G' regions of the β' subunit upon interaction with downstream DNA and repositioning of *Ec*β'GNCD leading to stabilization of the RNAP−DNA complex take place [Saecker et al. 2002, Kontur et al. 2006].

7.1.6. F complex.

We observe that on the *T7A1* promoter at 37°C the appearance of protection at all sites covering DNA to p20 follows a biphasic behaviour. The rate of the slow, second kinetic phase characterizes the last isomerisation step (E to F, Figure 16, page 45 and Figure 27, page 65). The rate of this step, ~ 0.03 ± 0.01 s^{-1}, is similar to the rate of active complex formation measured by abortive initiation assay [reviewed in Johnson and Chester 1998]. This isomerisation step, leading to the formation of a more stable, transcriptionally active complex (F complex), is neither accompanied by a significant change in the pattern of DNA protection from hydroxyl radical cleavage as well as in the pattern of permanganate reactivity of thymines (our data), nor in a change in protein fluorescence [Johnson and Chester 1998]. The rate of this step is however influenced by the concentration of the initiating nucleotide [reviewed in Johnson and Chester 1998]. This suggests that even though the DNA is melted and fully protected within the first, fast phase of the kinetics (E complex), the template strand is not yet properly positioned at the active site in the core of the enzyme. We propose that unstable contacts, formed by RNAP with the DNA downstream from transcription start site in E complex, induce additional conformational changes taking place within the active site and the downstream channel of the enzyme leading to a reorientation of the DNA at the active site. We further suggest that this internal rearrangement of the DNA within the active site channel could be coupled with the closing of the clamp and the folding of the conserved β' G and G' regions and of the non−conserved *Ec*βDR1 and *Ec*β'GNCD regions upon binding to downstream DNA, like it has been proposed for *λP*$_R$ promoter [Craig et al. 1998, Saecker et al. 2002, Kontur et al. 2006]. This suggestion is consistent with a lack of change in the protein's fluorescence in this step, since only one of the nine tryptophan residues of the β' subunit is found within these β' regions.

An open complex can form in the absence of these downstream interactions, as shown above and for some RNAP mutants [Severinov and Darst 1997, Young et al. 2004], however their presence is necessary for subsequent steps in the transcription process. The closing of the downstream clamp and the interaction of non−conserved elements of the *E. coli* RNAP with the DNA must not be too stable, or sequence specific, in order to allow for a rotation of the enzyme along the groove during RNA synthesis [Sakata-Sogawa and Shimamoto 2004], however they are necessary to constrain downstream DNA in the correct orientation resulting in propagation of the transcription bubble [Nechaev et al. 2000, Young et al. 2004] and for

increased stability of both transcription initiation and elongation complexes [Korzheva et al. 2000, Artsimovitch et al. 2003, Chlenov et al. 2005, Kontur et al. 2006].

Even though the last isomerisation step (E to F) results in a much tighter complex than the previous intermediates, due to the large RNAP and DNA conformational changes and the amount of interactions created between RNAP and DNA, the final open complex formed between RNAP and *T7A1* promoter is not absolutely irreversible, as shown by its sensitivity to heparin challenge.

7.1.7. *The off−pathway intermediate (E' complex).*

On the wild type *T7A1* promoter at 37°C, the protection at positions m5−m3 and p3−p13 on the TS and p7−p13 on the NTS appears at the same rate as the protection further downstream (p14−p20 on both strands), but is characterized by a larger amplitude for the first phase in the kinetic (Figure 15 (c1) and (c2), page 43). This larger amplitude indicates that the nucleotides at m5−m3, p3−p13 TS and p7−p13 NTS are less accessible to the hydroxyl radicals than the nucleotides at p14−p20. We propose that this difference is due to the presence of two complexes formed at the same time, but characterized by different protection patterns (E and E' complexes in Figure 16, page 45 and Figure 27, page 65). In the first one, E, the DNA backbone is additionally protected down to position p20 on both strands, while in the second one, E', the protection extends only to position p13. We propose that either the orientation of the DNA, or the structure of the protein in the E' intermediate does not allow for the downstream double−stranded DNA to properly enter the cleft so as to be protected from cleavage from p14 to p20. In order to achieve the formation of all downstream contacts, the E' intermediate has to be converted back to the D intermediate. The absence of the off−pathway E' intermediate, in the case of -10 consensus promoter, allows us to suggest that in the case of the wild type *T7A1* promoter this complex arises from the lack of proper contacts between the σ_2 domain and the -10 region. These poor contacts could result in increased flexibility of the DNA after melting occurred allowing the downstream DNA to interact with the pincers at an incorrect orientation.

The negatively charged σ region 1.1 may play an important role in the formation of the binary complex. This was originally suggested by the fact that substitutions and deletions in this region impair binary complex formation at some promoters [reviewed in Dombroski 1997, Vuthoori et al. 2001]. Fluorescence resonance energy transfer studies revealed that the

σ region 1.1 in *E.coli* RNAP holoenzyme is located at the same place as the DNA downstream double helix in the final open complex [Mekler et al. 2002]. It was suggested that σ region 1.1 in the holoenzyme serves as molecular placeholder for the downstream double helix. Upon binary complex formation, σ region 1.1 must exchange places with the DNA to sit outside the channel. Thus, the accumulation of the off−pathway E' intermediate could result from the competition of the DNA with σ region 1.1 for binding within the channel.

The theory that the mechanism of transcription initiation, at least at some promoter, follows a branched pathway is not new. For example, such pathway was established for the initiation at the $\lambda P_R AL$ [reviewed in Susa et al. 2002] and the *E.coli malT* promoter [Tagami and Aiba 1999]. Moreover, work by Shimamoto and co−workers provided evidence for the presence of a branched pathway in transcription initiation by *E.coli* RNAP within different promoter contexts, showing that the stability of the off−pathway intermediates leading to transcriptionally "moribund" complexes depends on promoter sequence and may be affected by environmental parameters such as pH and salt concentration [Susa et al. 2002]. In addition, the possible role of the presence of these off−pathway structures in the regulation of transcription initiation was underlined by the effect of the GreA and GreB factors on their stability and the probability of accumulation of stalled initiation complexes [Susa et al. 2006]. Here we directly identify a step in the recognition of *T7A1* promoter that may lead to the formation of such a branched pathway.

7.2. The role of the -10 consensus sequence in the process of transcription initiation.

Changes in the consensus sequence of the -10 region can affect all steps in the transcription initiation pathway [Fenton et al. 2001, McKane et al. 2001, Heyduk et al. 2006]. The rate of closed complex formation depends on the interaction of the protein with the still double−stranded DNA in this region; the rate of isomerisation steps is affected by both, the formation of specific interactions with the fork−junction structure at the upstream end of the -10 region, as well as by the formation of specific interactions between the single−stranded DNA regions, especially at the non−template strand, and aromatic amino acids on the σ_2 domain [Fenton et al. 2002, Tomsic et al. 2001].

The absence of a perfect consensus -10 sequence in the wild type *T7A1* promoter (**G**ATACT instead of TATAAT) results in a non−optimal interaction with the upstream

fork−junction and with the base at m9. In particular, the guanosine at m12 (position m13 on the *T7A1* promoter, Figure 10, page 35) was shown to have a detrimental effect on transcription activity at some promoters [Roberts and Roberts 1996].

The results obtained on the -10 consensus mutant of the *T7A1* promoter in fact show that nearly every step in the pathway is affected by this change in the DNA sequence. For example, in this case protection at m20−m4 NTS appears at the same rate as protection of nucleotides at positions m31−m26 NTS, reflecting a more cooperative conformational change, involving contacts with both the spacer and the -10 region (Figure 16, page 45 *versus* Figure 20, page 52). In addition, in the C intermediate the template strand of the -10 consensus promoter is protected continuously from m23 to m14 by interactions with β'ZBD and the σ_3 domain, while in the case of the wild type promoter the protection due to σ_3 is lacking. We propose that this could be the effect of the interaction of RNAP with non−consensus guanosine at m13 and cytosine at m9 on the NTS of the wild type promoter on the position of the DNA within the complex.

Furthermore, in the case of the promoter with a consensus -10 region, the difference in the rates of formation of the C and D intermediates is not as large as for the wild type promoter, in agreement with the more cooperative DNA melting process. Moreover, in this case the D intermediate is characterized by a protection pattern which differs significantly from that on the wild type *T7A1* promoter. In the D complex formed on the mutant promoter, the DNA backbone is additionally protected from cleavage at nucleotide positions m5−m3 TS as well as at p3−p9 TS and at p7−p13 NTS. These data could be explained by proposing that the presence of optimal contacts between the σ_2 domain and the consensus -10 region formed in C complex may be transmitted via the β' coiled−coil domain to the other domains of the enzyme driving a possible conformational change necessary for the subsequent isomerisation. This would result in the entry of single−stranded template DNA into the active site and the burial of downstream DNA duplex within the ββ' cleft (D complex).

The process of RNAP binding to a promoter with a consensus -10 region is characterized by the lack of the signatures of the "off−pathway" complex. The possible explanation for the absence of this intermediate is that the specific contacts formed between the protein and the consensus bases in the -10 region favours the formation of the complex where σ region 1.1 has been displaced out of the channel. This then allows the successful

interactions of the protein with the downstream DNA, the folding of the downstream polymerase domains and the isomerisation to a final complex.

The rate of the last isomerisation step (slow, second phase in the kinetics) in fact is more uniform in the datasets of the promoter with a consensus -10 region. This could be due to a direct effect caused by the increased stabilization of the DNA−protein interactions, as well as to an indirect effect, resulting from the absence of the off−pathway complex. The lower rate of the last isomerisation step in the case of wild type promoter could be due to the necessity to break the interactions formed in the off−pathway complex before the final structure can be adopted.

Once more, the use of time−resolved potassium permanganate footprinting allowed us to show that on the wild type *T7A1* promoter DNA melting from m12 to m4 takes place in two steps preceding the formation of a transcriptionally active complex. Furthermore, we established that the substitution of the wild type -10 sequence (**G**ATA**C**T) by the consensus sequence (TATAAT) in the context of *T7A1* promoter does not have a large effect on the process of DNA melting nucleation at 37°C, while, by favouring the subsequent isomerisation step, it accelerates propagation of the transcription bubble downstream. The view, that DNA strand separation occurs after promoter recognition and initial binding of RNAP but before the formation of a stable binary complex, is supported by data recently obtained by deHaseth and colleagues from measurements of the fluorescence of 2−aminopurine (2−AP) substituted at different positions of the -10 region of a consensus promoter [Schroeder et al. 2009]. Moreover, this study revealed that upon binary complex formation the changes in fluorescence of 2−AP at m8 and m4 appear at the same rate as that of 2−AP at m11. These data indicate that DNA melting from m12 to m4 occurs in one step on the consensus promoter. This result is in agreement with what we observed here on the *T7A1* promoter with a consensus -10 hexamer by measurements of the change in permanganate reactivity. In contrast, in the case of another naturally strong promoter that has a near consensus -10 region (TAT**G**AT), the wild type *tyrT* promoter, a two−step DNA opening process was detected by potassium permanganate footprinting [Auner et al. 2003]. An explanation for this difference would be that *tyrT* promoter also has G+C−rich discriminator region upstream of the transcription start site, providing a thermodynamic barrier to propagation of DNA melting. In summary, all these data clearly show that the process of DNA opening is not the same for all

promoters. Instead it depends on the sequence of -10 region as well as on the sequence of other promoter elements.

The results of our studies revealed that "improvement" of the -10 sequence of *T7A1* promoter leads to more effective formation of final binary complex. However, as in the case of the wild type *T7A1* promoter, this complex is not resistant to heparin challenge. Furthermore, although the final complex formed by *E.coli* σ^{70} RNAP on the promoter variant with a consensus -10 region is functionally competent, it is characterized by an increased level of abortive products accumulating during initiation of RNA synthesis compared to wild type promoter. This observation confirms the idea that promoter sequences conferring stronger RNAP−promoter interactions as a rule decrease transcription by reducing promoter clearance [Hsu et al. 2003, Vo et al. 2003].

It seems thus that often during evolution promoters with sequences that result in a less efficient DNA melting and less efficient isomerisation steps were selected rather than promoters with fully consensus sequences. Such strategy could have some advantages. For example, at the strong ribosomal and tRNA promoters the presence of G+C−rich discriminator region destabilizes the complex and renders it sensitive to regulation by DNA supercoiling and changing ppGpp and NTP concentrations [Travers et al. 2005, Paul et al. 2004], which allows cells to adapt faster to new environmental conditions. In the case of *T7A1* promoter, as noted above, less efficient formation of final binary complex is however compensated by a more efficient initiation of productive RNA synthesis.

7.3. The effect of low temperature on the mechanism of promoter binding and activation.

The transcriptional activity of the binary complex formed by RNAP at the *T7A1* promoter shows a steep temperature dependence between 10°C and 37°C [Rosenberg et al. 1982, Zaychikov et al. 1997]. At 20°C only 10 % of the complexes are transcriptionally active. Furthermore, the mapping studies on the size of the melted region (the bubble) in the final binary complexes formed at different temperatures revealed that at 20°C the DNA is melted from m12 to m7. The thymines at m6 and m4 (TS), at p2 (NTS) and at p1 (TS) are only 60 %, 40 % and 10% accessible to $KMnO_4$ modification respectively compared to what is observed at 37°C. The accessibility of these bases toward $KMnO_4$ increases with increasing

temperature and the increase of accessibility of thymine at p1 (TS) correlates with the appearance of transcription activity [Zaychikov et al. 1997].

In this study we used time−resolved hydroxyl radical footprinting to define in details the pathway of transcription initiation at 37°C and 20°C. By comparing the protection pattern observed at each step of *E.coli* polymerase binding to *T7A1* promoter at 20°C to that at 37°C, we concluded that the process of formation of the final binary complex at low temperature is characterized by the accumulation of structural intermediates significantly different from those at normal physiological conditions (37°C). For example, at 20°C we were able to distinguish two intermediate complexes, C and C', which appear at the same time, but are characterized by different pattern of protection in the discriminator region upstream of the transcription start site. Compared to the C intermediate, in the C' intermediate DNA is additionally protected at positions m6−m3 on the NTS and at m7−m6, m2−p3 on the TS (Figure 22, page 57). These data may indicate that at low temperature the formation of the sharp bend of the DNA at the upstream end of the -10 region becomes impaired.

Moreover, we observed that at 20°C only the last isomerisation step results in the appearance of protection of non−template strand from nucleotide position p7 to p20 (Figures 21, page 56 and 22, page 57, E complex). We propose that the structure of the intermediates formed prior to this rate−limiting step reflects the inability of the RNAP to build full contacts with downstream DNA until the temperature−dependent protein conformational changes have taken place. In addition, we assume that the short footprint observed at low temperature on the non−template strand before the DNA−protein complex isomerises to the final complex is due to β' not being correctly folded on the DNA. This assumption is based on the data of crosslinking studies performed by Naryshkin and co−workers which showed that the protection of the template strand of the DNA duplex downstream of p7 is likely due to interactions with β', *Ec*βDR1 and the downstream lobe of β, while the non−template strand down to p14 only crosslinks to the β' subunit [Naryshkin et al. 2000].

It is necessary to point out one more time, that our experimental approach gives us information about a process of binary complex formation only at the level of DNA. Several other groups studied the process of transcription initiation using techniques providing a probe for protein conformational changes [Johnson and Chester 1998, Kontur et al. 2006]. The technique used by Johnson and Chester is based on monitoring alterations in the intrinsic fluorescence of RNAP upon it's binding to DNA. They measured the rate of these alterations

during binary complex formation at the *T7A1* promoter over a wide temperature range [Johnson and Chester 1998]. They observed that the rate of change of the protein's intrinsic fluorescence exhibits a non−linear temperature dependence with a break point at 28°C, after which the isomerisation rate decreases more steeply as a function of temperature. As described above, the comparison of results of our DNA footprinting studies with results of studies performed by Johnson and Chester allows us to suggest that at 37°C these protein conformational changes take place within the C to D step on the pathway, before RNAP makes full contacts with the downstream promoter region, and are associated with the propagation of DNA melting. The extrapolation of the data from Johnson and Chester to 20°C (our experimental temperature) would allow us to estimate the rate of this protein isomerisation to about 0.03 s^{-1}, which is similar to the rate we measured for the appearance of the downstream protection on the non−template strand at 20°C (E complex, Figure 22, page 57), namely, 0.048 ± 0.002 s^{-1} for p7−p12 and 0.056 ± 0.003 s^{-1} for p13−p20 (Table 5, page 55 and Figure 21, page 56). It is thus possible that at low temperature (20°C) the less favourable DNA melting nucleation occurs later in the pathway, only after the entry of the downstream DNA within the ββ' cleft distorts the double helix at the -10 region enough to destabilize it. If such an assumption is correct, than the decreased flexibility of double−stranded DNA, compared to a melted bubble, restricting the number of possible orientations of the double helix, and the different sequence requirement for the formation of stable contacts between the σ subunit and the still double−stranded DNA [Fenton et al. 2001] may explain the absence of an off−pathway complex under this condition.. We also propose that the interactions of RNAP with the partly melted DNA induce changes in the protein conformation resulting in a decrease of fluorescence. Direct prove of this speculation would however require time−resolved potassium permanganate footprinting experiments at 20°C.

As described above, it seems that at 20°C the β' subunit of RNAP is able to contact with the non−template strand at positions p7−p20 only after the temperature−dependent protein conformational changes have occurred. Although our experimental methods don't allow us to directly define the nature of these protein rearrangements, we assume however that they are defined by the β subunit. This assumption is based on the data available from the literature. It is evident that the β and β' subunits of RNAP cooperate with the σ factor in the process of DNA melting [Nechaev et al. 2000, Brodolin et al. 2000, Brodolin et al. 2005]. Moreover, it has been reported that a large deletion in β subunit that removes substantial part

of the downstream lobe, including *Ec*βDR1, results in an RNAP [β(Δ186–433)] that, unlike wild type RNAP, can form an open promoter complex even at 0°C, albeit this complex is shortened in the downstream direction (there was no protection beyond position p5) and has a transcriptional bubble extended down only to position m7 in the absence of NTPs [Severinov and Darst 1997, Nechaev et al. 2000]. This supports the idea that, as the temperature is decreased, the protein conformational changes, dependent on the presence of this domain of the β subunit, become rate–limiting, and eventually inhibiting, for DNA melting. Severinov and colleagues proposed that these conformational changes are dependent on a "two stroke" coupling of the upstream and the downstream β lobes [Nechaev et al. 2000]. This "two stroke" model suggests that upstream β lobe cooperates with the σ_2 region and maintains the initial strand separation at the upstream edge of the transcriptional bubble. The locking of the upstream β lobe into the interactions with the -10 hexamer and discriminator region then engages, via an allosteric mechanism, the downstream lobe into the contacts with the downstream DNA. This, in turn, may help to direct downstream DNA towards β' jaw [Brodolin et al. 2005].

In summary, all these data clearly indicate that the structure of intermediates in the pathway of promoter binding by RNAP, the process of DNA melting as well as the conformation of the final binary complex differ at low temperatures compared with normal physiological conditions, 37°C. The presence of multiple possible pathways within the same promoter context endows robustness and specificity on the mechanism of formation of a transcriptionally active complex.

8.   Summary.

Transcription initiation is the first step in gene expression and a major point of cellular regulation. In bacteria, transcription initiation is carried out by the RNAP holoenzyme which recognizes and binds a specific DNA sequence at the beginning of a gene, eventually resulting in the formation of the transcription–competent binary complex. Numerous kinetic and footprinting studies provide evidences that this process involves a series of structural intermediates where protein conformational changes are coupled with DNA wrapping and melting [Roe et al. 1985, Buc and McClure 1985, Craig et al. 1998, Johnson and Chester 1998, Kontur et al. 2006, Davis et al. 2007]. However, real–time identification and detailed structural characterization of these short–lived complexes are still lacking.

For the first time, we applied X–ray hydroxyl radical footprinting to study the dynamics of RNAP–DNA interactions. The milliseconds time resolution as well as the single nucleotide structural resolution of this method allowed us to follow RNAP binding to *T7A1* promoter on–line and analysed how it is affected by the substitution of the wild type -10 sequence (**G**ATA**C**T) by the consensus sequence (TATAAT) and by the decrease of temperature from 37°C to 20°C. In addition, for the first time, we used time–resolved potassium permanganate footprinting to determine kinetic of DNA strand separation in order to directly define at which step in the pathway to the final complex DNA melting takes place.

Our experiments revealed that at least five groups of kinetically distinguishable intermediates are present before formation of the transcriptionally active complex at the wild type *T7A1* promoter at 37°C (A, B, C, D and E groups in Figure 27, page 65). Moreover, our results allowed us to postulate that some of these groups (A, B and E groups) contain more than one structural intermediate. Based on these data, as well as crystallographic and biochemical data on RNAP–promoter interactions available from literature, we described the mechanism of promoter binding and activation in terms of conformational rearrangements of both protein and DNA.

A feature of our model is a key role attributed to the RNAP interaction with UP element in transcription initiation. Two αCTDs of RNAP contact the UP element subsites of the promoter upon its first encounter (A, A', A'' complexes). These contacts are maintained as these early complexes progress to the final complex through a series of isomerisation events. Following the first isomerisation step (A, A', A'' complexes to B, B' complexes)

leading to binding of -35 region by σ_4 domain, the DNA is bent toward σ_3 and σ_2 domains where contacts are made with the spacer and the upstream end of the -10 region (C intermediate). At this stage, the separation of the DNA strands from m12 to m6 takes place. In the subsequent step, further bending of the DNA towards the protein surface is associated with the propagation of DNA melting (D intermediate). The DNA then enters the cleft defined by β and β' subunits and becomes protected downstream to p20 (E intermediate). Finally, the RNAP−DNA complex isomerises into a transcriptionally active complex, where the template strand is properly placed at the active site (F complex). This internal rearrangement of the DNA within the active site channel could be coupled with closing of the clamp and folding of the β' G and G' regions as well as $Ec\beta$DR1 and $Ec\beta'$GNCD. In addition our data showed that on the wild type *T7A1* promoter at 37°C formation of non−optimal contacts between the σ_2 domain and the melted non−consensus -10 region results in the accumulation of an off−pathway E' complex in which downstream contacts are not properly formed.

It is important to point out, that results of time−resolved potassium permanganate footprinting experiments performed in this work provided new insights into the process of transcriptional bubble formation. In fact, contrary to the generally accepted model that places DNA melting within the last step on the pathway, the results of our studies demonstrated that, at *T7A1* promoter at 37°C, DNA opening occurs before RNAP makes full contacts with the downstream promoter region. Moreover, we suggested that the propagation of DNA strand separation is associated with significant protein conformational changes. Although these changes seem to be less favoured at the wild type *T7A1* promoter than at the promoter with consensus -10 region, they precede the rate−determining step in the formation of an active complex. We reconciled our results with the previous model by showing that at 20°C a different pathway becomes dominant. In this case the entry of the still double−stranded DNA into the $\beta\beta'$ cleft may be necessary in order to allow the less favourable DNA melting nucleation to take place. Furthermore, we suggested that temperature−dependent protein conformational changes become a rate−limiting factor for the propagation of transcriptional bubble. Thus, the results of our experiments revealed that the pathway to the formation of a final binary complex is not always characterized by the same sequence of events, but that it is highly dependent on the environmental parameters, rendering this process more flexible and adapt to external perturbations.

In addition here we showed that the "improvement" of the -10 region to obtain a consensus sequence results in a more efficient formation of final open complex, probably by favouring specific promoter–protein interactions, but in a less efficient initiation of productive RNA synthesis, probably by impairing the ability of RNAP to release the promoter. Together with the data available from literature, these data indicate that the efficiency of gene expression is determined by the balance of promoter binding and activation, and RNA chain initiation and promoter escape.

9.   Materials and Methods.

Chemicals used in this work were of highest possible quality. They were purchased from Sigma, Roth, Serva, Fluka, Merck (Germany). Water for all buffers was purified in Millipore system.

9.1.   Preparation of T7A1 promoter fragments.

9.1.1. Primers.

Unmodified and modified with amino group at the 5'−end *T7A1*−up(-91) (5'−CGAGGCCAACTTAAAGAGAC−3') and *T7A1*−down(+81) (5'−TCGTGCGACTTATCAGGCTG−3') primers were obtained from IBA. (Göttingen, Germany)

9.1.2. Fluorescence labeling of primers.

The *T7A1*−up(-91) primer or *T7A1*−down(+81) primer containing the amino group at the 5'−end was incubated with Alexa Fluor 647 fluorescent dye (Invitrogen) (at a primer to dye molar ratio equals 1 to 3) in 50 mM Li−Borate buffer, pH 9.0, at 37°C for 4−5 hours. After completion the reaction, the oligonucleotide material was precipitated with cold ethanol, dissolved in small volume (~50 μl) of 10 mM Li−Hepes, pH 7.5, and then the 5'−Alexa Fluor 647−labeled primer was separated from unreacted primer by HPLC chromatography on a monoQ column using a MiLiChrom chromatograph (EcoNova, Novosibirsk, Russia). This was followed by precipitation with ethanol. The precipitated primers were resuspended in 10 mM Li−Hepes, pH 7.5 to final concentration of 0.1 mM. The yield of 5'−Alexa Fluor 647−labeled primer usually equals to 30 %.

9.1.3   Radioactive labeling of primers.

The unmodified *T7A1*−up(-91) primer or *T7A1*−down(+81) primer was incubated with [γ–^{32}P]ATP (811 Ci/mmole) (Hartmann Analytic) (at a primer to [γ–^{32}P]ATP molar ratio equals 1000 to 1) in the presence of *T4* polynucleotide kinase in 1xKinase buffer (50 mM Tris-HCl, pH 7.6 at 25°C; 10 mM $MgCl_2$; 5 mM DTT; 0.1 mM spermidine; 0.1 mM EDTA) at 37°C for 20 min following which the cold ATP at 8−fold molar excess relative to a primer was added to the reaction mixture and the reaction of primer phosphorylation was continued for another 10 min. Unincorporated [γ–^{32}P]ATP and ATP were removed by HPLC on a

monoQ column using a MiLiChrom chromatograph (EcoNova, Novosibirsk, Russia). The radioactively labeled at the 5'−end primer was precipitated with ethanol and then resuspended in 10 mM Li−Hepes, pH 7.5.

9.1.4. *Isolation of plasmid pDS1−A1$_{220}$ containing wild type T7A1 promoter.*

Bacteria *Escherichia coli* containing plasmid *pDS1−A1$_{220}$* were grown in LB medium (1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl) containing 75 µg/ml ampicillin at 37°C. When the culture reached a cell density of ~ $1x10^9$ cells per ml (OD$_{600}$ = 1.0), bacterial cells were harvested by centrifugation for 20 min at 7000 rpm at 4°C. Plasmid was isolated by using QIAGEN column (QIAGEN−tip 100) according to the manufacturer's protocol.

9.1.5 *Synthesis of the labeled DNA fragment containing wild type T7A1 promoter.*

Radioactively or fluorescently labeled at the 5'−end 172bp−long DNA fragment containing wild type *T7A1* promoter was prepared by PCR using *pDS1−A1$_{220}$* plasmid DNA as a template and primer pairs *T7A1−*up(-91) for the non−template strand (NTS) and *T7A1−*down(+81) for the template strand (TS). One of these primers was labeled with ^{32}P or Alexa Fluor 647 fluorescent dye. Reaction mixtures (100 µl) contained: 20 mM Tris−HCl, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 0.1 % TritonX−100, 0.01 mg BSA, 2.5 mM $MgCl_2$, 0.4 mM dNTPs, 2 µM each primer, 3 ng *pDS1−A1$_{220}$* and 5 U *Pfu* DNA polymerase (Novosibirsk, Russia). After denaturation at 94°C for 2 min, 40 cycles of 94°C for 40 sec, 60°C for 40 sec and 74°C for 1 min were used, followed by a final incubation at 74°C for 5 min. PCR products were precipitated with cold ethanol, rinsed with 80 % ethanol, dried, dissolved in small volume (~ 200 µl) of 10 mM Li−Hepes, pH 7.5, and then the 172bp−long DNA fragment was purified by HPLC on a monoQ column using a MiLiChrom chromatograph (EcoNova, Novosibirsk, Russia).

9.1.6 *Synthesis of the labeled mutants of T7A1 promoter.*

T7A1 promoter mutants were received from Evgeny Zaychikov and then amplified by PCR as described above.

9.2. RNAP preparation.

9.2.1. Cell growing.

Escherichia coli strain RL916 was used as the source of His−tag RNAP. Bacterial cells were grown at 37°C with vigorous shaking in LB medium, harvested at the late logarithmic phase by centrifugation for 20 min at 7000 rpm at 4°C, frozen in liquid nitrogen and then stored at -80°C until use.

9.2.2. RNAP purification.

All purification procedures were carried out at 4°C.

9.2.2.1. Disruption of cells.

Frozen *E. coli* RL916 cells, 70 g, were suspended in 200 ml of Grinding buffer (50 mM TrisHCl, pH 8.0; 2 mM EDTA; 5% (v/v) glycerol; 0.1 mM DTT; 1 mM β−mercaptoethanol; 20 µg/ml PMSF), blended until the mixture was homogeneous and then disrupted by three passages through a "French press" at 150 MPa. 270 ml of TEDG buffer (10 mM TrisHCl, pH 8.0; 0.1 mM EDTA; 5% (v/v) glycerol; 0.1 mM DTT; 20 µg/ml PMSF) were added to the mixture, after that the debris was removed by centrifugation for 30 min at 12000 rpm in Beckman Coulter J2−HS centrifuge.

9.2.2.2. Polymin−P fractionation.

Under gentle stirring 18 ml of 10% polyethylenimine, pH 8.0, were added slowly to the crude supernatant (ca. 520 ml) (final concentration of polyethylenimine ~ 0.35 %). After another 5 min of stirring, the mixture was centrifuged for 20 min at 12000 rpm. The pellet was suspended in 400 ml of TEDG buffer containing 0.5 M NaCl and the mixture was centrifuged as above. This washing step was repeated 2 times more, subsequently the RNAP was eluted from the polyethylenimine pellet with TEDG buffer containing 1 M NaCl (Figure 28). After centrifugation of the mixture as above, RNAP was
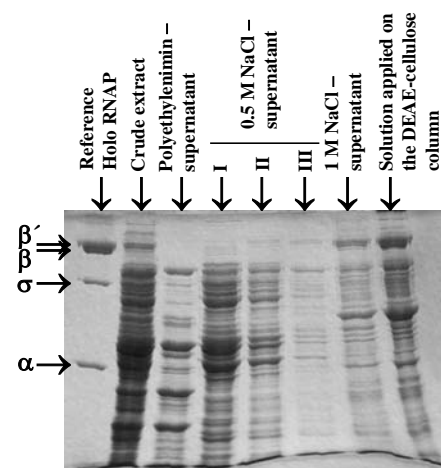

Figure 28. 0.1% SDS−10% PAGE analysis of the steps of the RNAP purification. Crude extract preparation and polyethyleninime fractionation.

precipitated from resulting supernatant (330 ml) by adding of 115 g of ammonium sulfate (0.35 g/ml). After dissolving of the salt, the mixture was stirred for further 20 min and centrifuged for 30 min at 12000 rpm. The precipitate was dissolved in 20 ml of TEDG buffer (25 ml total volume) and dialyzed against two changes of 400 ml of the TDEG buffer containing 0.1 M NaCl overnight.

9.2.2.3. DEAE−cellulose chromatography.

The sample (about 35 ml) was applied to the 70 ml DEAE−cellulose DE32 column (Sigma) preequilibrated with TEDG buffer containing 0.1 M NaCl and washed with the same buffer until no more proteins eluted from the column, as monitored by UV absorbance at 280 nm. RNAP was then eluted from the column with a linearly increasing gradient (200 ml total volume) from 0.1 M to 0.5 M NaCl in TEDG buffer at 110 ml/h. Eleven−milliliter fractions were collected and an aliquot of each fraction was analyzed by SDS−10 % polyacrylamide gel electrophoresis (PAGE) (see below). The fractions containing RNAP (fractions 8−23, Figure 29) were pooled (190 ml total volume), and the RNAP was precipitated by adding of ammonium sulphate as described above. The precipitate was dissolved in 12 ml of TEDG buffer (15 ml total volume) and dialyzed against two changes of 300 ml of the TDEG buffer containing 0.1 M NaCl overnight.
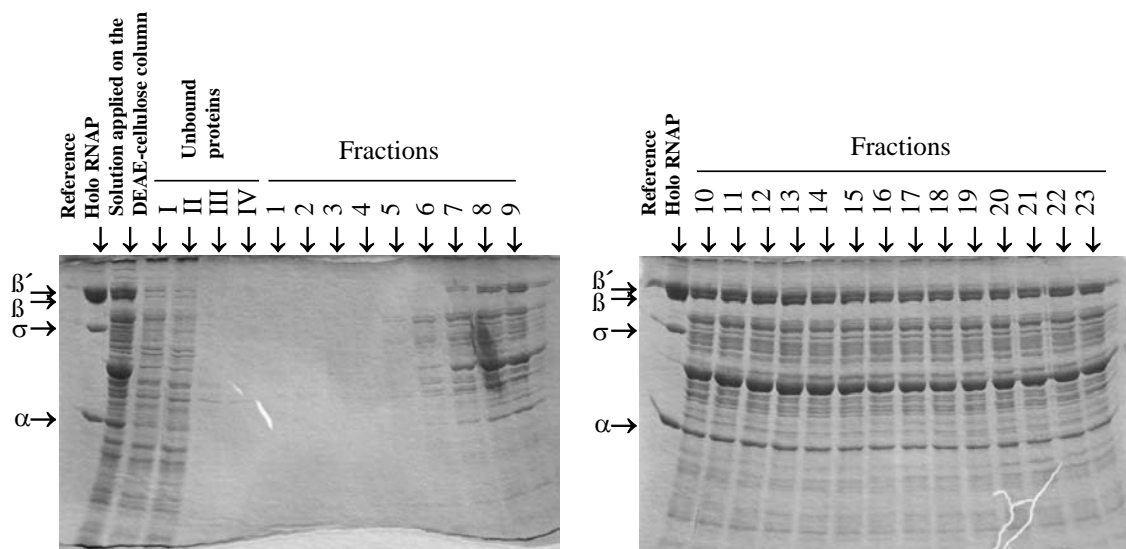

Figure 29. 0.1% SDS−10% PAGE analysis of the DEAE−cellulose column fractions.

9.2.2.4. Heparin−superose chromatography.

The dialyzed sample (about 21 ml) was applied to a 10 ml heparin−superose column (Pharmacia) preequilibrated with TEDG buffer containing 0.1 M NaCl, consecutively washed with the same buffer until the absorbance at 280 nm returned to baseline and eluted with a linearly increasing gradient (200 ml total volume) from 0.1 M to 1 M NaCl in TEDG buffer at 80 ml/h. Eleven−milliliter fractions were collected and an aliquot of each fraction was analyzed by SDS−10 % PAGE. The fractions enriched in RNAP holoenzyme (fractions 9−10, Figure 30) were combined (22 ml total volume), and ammonium sulphate (8.8 g, final concentration of 0.4 g/ml) was added to precipitate the protein. Core RNAP contained in fraction 11 was precipitated by adding of 4.4 g of ammonium sulphate.
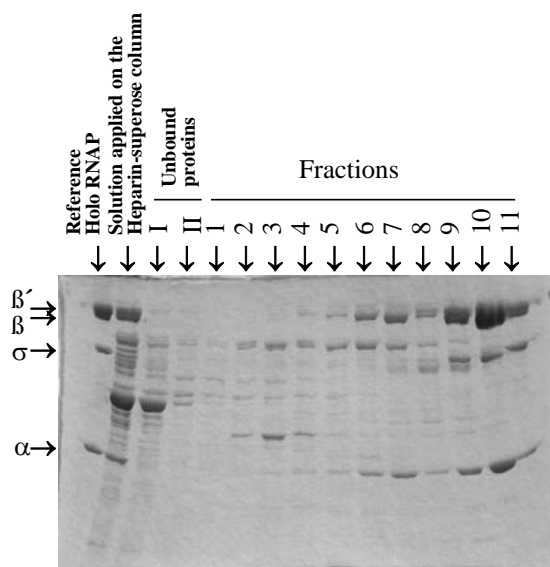

Figure 30. 0.1% SDS−10% PAGE analysis of the Heparin−superose column fractions.

9.2.2.5. MonoQ chromatography.

The precipitate containing RNAP holoenzyme was collected by centrifugation for 20 min at 18000 rpm, dissolved in 3 ml of TEDG buffer, dialyzed against two changes of 250 ml of the TDEG buffer containing 0.1 M NaCl and loaded onto 8 ml MonoQ column (Pharmacia) preequilibrated with TEDG buffer containing 0.1 M NaCl. After the column was washed with 10 ml of the same buffer, the bound proteins were eluted using FPLC system (Sykam) with a gradient of TEDG buffer containing 1 M NaCl (0−20 % in 0.8 ml, then 20−40 % in 16 ml and 40−100 % in 2.4 ml). Portions of the fractions were analyzed by SDS−10 % PAGE. Trace amounts of the core polymerase eluted first (fraction 5), followed by the holoenzyme (fractions 6−7) (Figure 31 (a)). Core RNAP was precipitated by adding of ammonium sulphate. The fractions containing pure RNAP holoenzyme were pooled (4 ml

total volume, D_{280}=0.77 AU/ml $\Rightarrow$ m(holo RNAP)= 4.8 mg), concentrated–desalted with TEM buffer (10 mM TrisHCl pH 8.0; 1 mM EDTA; 1 mM ß-mercaptoethanol; 0.05 M NaCl) by using Microsep–100K centrifugal device (Pall Filtron) and then supplied with an equal volume of glycerol. The pure RNAP holoenzyme was stored at the concentration of 10 mg/ml at -20°C or in liquid nitrogen in small aliquots.

9.2.2.6. BioRex chromatography.

The precipitate containing core RNAP (fraction 11 of heparin–superose chromatography and fraction 5 of MonoQ chromatography) was collected by centrifugation for 20 min at 18000 rpm, dissolved in 3 ml of TEDG buffer, dialyzed against two changes of 250 ml of the TDEG buffer containing 0.1 M NaCl and loaded onto 35 ml BioRex column (Pharmacia) preequilibrated with TEDG buffer containing 0.1 M NaCl. After the column was washed with 30 ml of the same buffer, the RNAP was eluted from the column with a linearly increasing gradient (200 ml total volume) from 0.1 M to 1.0 M NaCl in TEDG buffer at 83 ml/h. Twelve–milliliter fractions were collected and an aliquot of each fraction was analyzed by SDS–10 % PAGE. The fractions containing pure RNAP core enzyme (fractions 7–8, Figure 31 (b)) were pooled (total volume 24 ml, m(core RNAP)= 6.6 mg), and the protein was precipitated by ammonium sulfate (10 g). After centrifugation for 20 min at 18000 rpm, the precipitate was dissolved in 2 ml of TEM buffer. RNAP solution was concentrated–desalted on Microsep–100K and supplied with an equal volume of glycerol. The RNAP core enzyme was stored at the concentration of 10 mg/ml at -20°C or in liquid nitrogen in small aliquots.
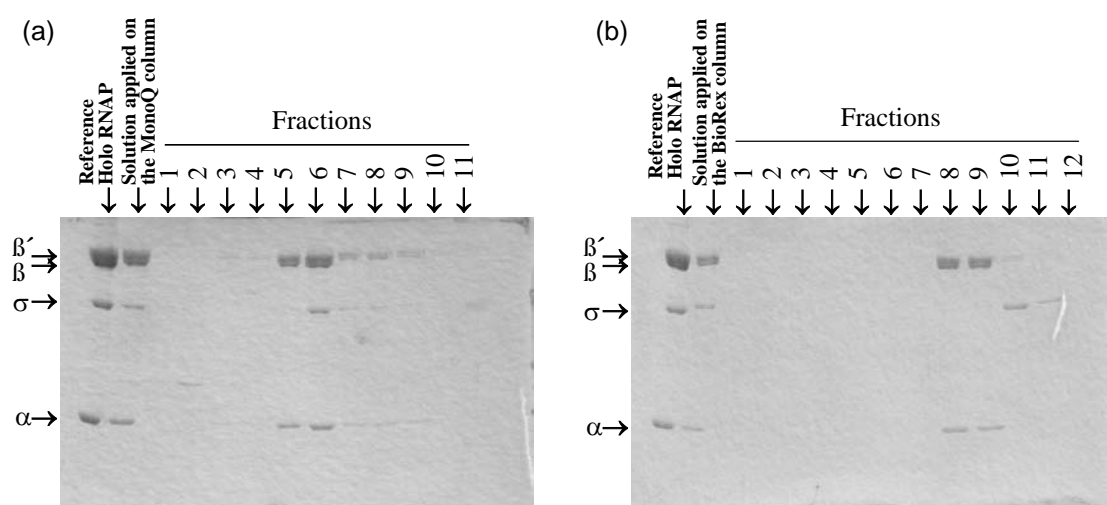

Figure 31. 0.1% SDS–10% PAGE analysis of the steps of RNAP purification. (**a**) MonoQ column fractions. (**b**) BioRex column fractions.

9.2.2.7. Gel electrophoresis.

An aliquot of each chromatographic fraction (0.001 volume of fraction) was supplemented with 0.3 volume of 5 x sample loading buffer (50 mM TrisHCl, pH 6.8; 5 % SDS; 50 % glycerol; 5 % ß-mercaptoethanol; 0.1 % bromphenol blue) and heated at 80°C for 3 min. Samples were then applied on 9 x 8 x 0.1−cm slab of 10 % resolving gel (10 % polyacrylamide (acrylamide : bisacrylamide = 30 : 1); 0.375 M TrisHCl, pH 8.8; 0.1 % SDS) overlaid with 9 x 1.5 x 0.1−cm slab of 4 % stacking gel (4 % polyacrylamide; 0.125 M TrisHCl, pH 6.8; 0.1 % SDS). Gel was run in Tris−glycin buffer (25 mM TrisOH; 250 mM glycin; 0.1 % SDS) at 150 V until the bromphenol blue dye front reached the bottom edge of the gel. The location of proteins in the gel was visualized by staining with Coomassi brilliant blue R−250.

9.2.3. Characterization of holoenzyme.

Protein sample was analysed by performing SDS−10 % polyacrylamide gel electrophoresis. The quantification of the bands on the Coomassi blue stained gel by using ImageQuant indicated that the holoenzyme was 100 % saturated with σ factor. In addition, holoenzyme was almost 100 % active in promoter binding as measured by electrophoretic mobility−shift assay (EMSA) (see below Figure 32).

9.2.3.1. EMSA.

Electrophoretic mobility−shift assay (EMSA) was done as follows. RNAP holoenzyme was mixed with 172bp−long *T7A1* promoter fragment at three molar ratios (RNAP to DNA molar ratio equals 1 to 1 or 1.5 to 1 or 2 to 1) in BBcac20 buffer (50 mM sodium cacodylate, pH 7.5; 20 mM NaCl; 6 mM MgCl$_2$). After a 5−min incubation at 37°C, samples were supplemented with equal volume of 2 x loading buffer (20 mM TrisHCl, pH 7.5; 20 % glycerol; 0.04 % bromphenol blue) and directly applied on 4 % / 20 % non−denaturing polyacrylamide gel (9 x 5−cm slab of 4 % gel joined to the 9 x 4−cm slab of 20 % gel) with 1 x TBE buffer. After electrophoresis, in order to visualize
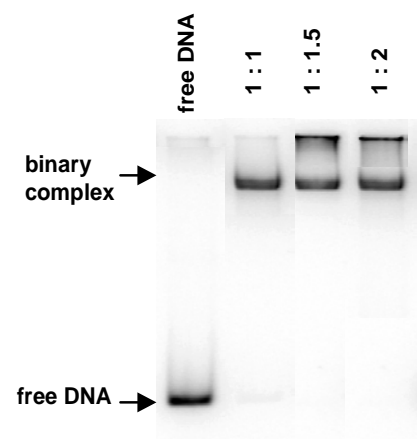

Figure 32. EMSA analysis of binding of *E.coli* RNAP holoenzyme to the promoter fragment at diverse enzyme to DNA molar ratios.

the DNA, gel was incubated for 15 min in water containing 0.2 µg/ml fluorescent reagent ethidiumbromide (EtdBr), washed three times with water (total washing time: 30 min) and scanned with BioRad Gel Doc 1000 instrument (Figure 32).

9.3. Rapid mixing X−ray footprinting experiments.

Time−resolved hydroxyl radical footprinting experiments were performed at the ID10A beamline at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France) using the modified BioLogic stopped−flow machine (SFM−400) (Claix, France).

9.3.1. Beamline characteristics.

The ID10A beamline at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France) is a multi−purpose, high−brilliance undulator beamline. The insertion devices serving the beamline consist of three undulators producing a white beam with enough flux (flux at sample: ~ 5 x 10^{13} photons/s/mm^2 at 8 keV) to carry out time−resolved footprinting experiments in 16 bunch mode at ring currents from 88 to 55 mA. The beam size at the sample was 1 mm V x 1 mm H.

9.3.2. BioLogic stopped−flow machine characteristics.

Figure 33 (a) represents a scheme of experimental set−up used at the ESRF. Briefly, the BioLogic stopped−flow module (SFM−400) (Claix, France) consists of a mechanical subsystem and a motor power supply. Each of four syringes of the SFM−400 is driven by independent stepping−motor whose function and speed are precisely controlled by computer (BioKine32 software). The independent control of each syringe allows a high versatility in the injection sequence. It is possible, for example, to make an injection of one syringe only, unequal filling of syringes, variable mixing ratios. The outlets of the syringes are connected to a mixing device. The mixing device of apparatus comprises the cross mixer of our own construction. Check valves are inserted to eliminate a diffusion of reactants into the mixing area ("reaction loop") and effects due to backpressure buildup. The additional modification of the mixing device aimed to X−ray footprinting experiments is the insertion of an X−ray transparent "exposure chamber" following the reaction loop in the solution flow path. The exposure chamber, a schematic drawing of which is shown in Figure 33 (b), comprises a quartz capillary within a plexiglass holder. All SFM−400 drive syringes, valves and flow lines

are immersed into a water jacket to provide an accurate temperature control. The design of the stopped−flow apparatus and the sample collector at the end of the machine allow the sequential mixing and exposing of up to 13 samples.
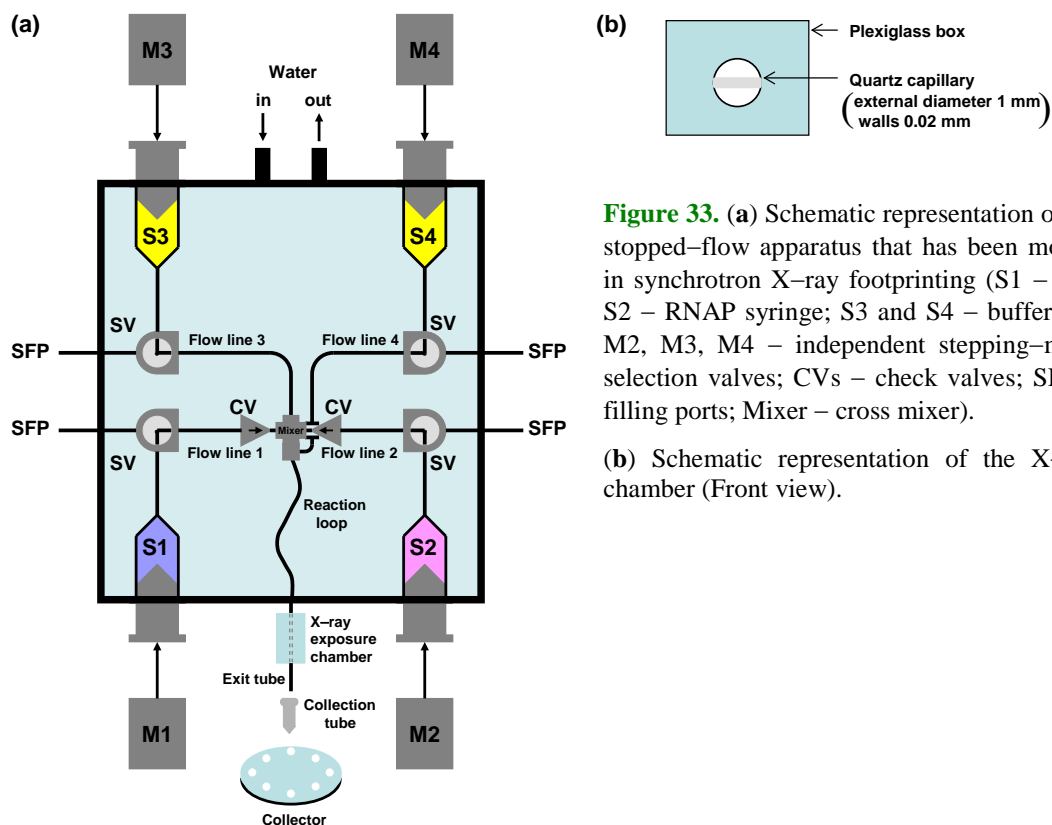

Figure 33. (**a**) Schematic representation of the BioLogic stopped−flow apparatus that has been modified for use in synchrotron X−ray footprinting (S1 − DNA syringe; S2 − RNAP syringe; S3 and S4 − buffer syringes; M1, M2, M3, M4 − independent stepping−motors; SVs − selection valves; CVs − check valves; SFPs − syringe filling ports; Mixer − cross mixer).

(**b**) Schematic representation of the X−ray exposure chamber (Front view).

9.3.3. Time−resolved hydroxyl radical footprinting experiments.

At the beginning of a typical experiment, the circulating water bath was connected to the stopped−flow apparatus, switched on, and allowed to equilibrate at the desired temperature of the experiment. Meanwhile, all syringes and flow lines of the SFM−400 were filled (no air bubbles!) with appropriate solutions. In case of binding experiments, syringe 1 and flow line 1 were loaded with 50 nM fluorescently labeled DNA in BBcac20 buffer (50 mM sodium cacodylate, pH 7.5; 20 mM NaCl; 6 mM $MgCl_2$), syringe 2 and flow line 2 were loaded with 200 nM RNAP in BBcac20 buffer, syringes 3 and 4 as well as flow lines 3 and 4 were loaded with BBcac20 buffer. The selector valves were then turned so that the driving syringes were opened toward the flow lines. The Eppendorf tubes for exposed samples were placed into the collector.

At this point the investigator exited the beamline hutch and initiated the safety procedures required for enabling the beamline. The time sequences (time to mix and incubate the reactants and subsequently expose them to the beam) for the samples were entered into the drive motor controller. Each injection consisted of a "push motor 3 (first push) – move the collector – simultaneously push motors 1 and 2 (second push) – pause – push motor 4 (third push) – move the collector – push motor 3 (fourth push) – move the collector". The first push washed the reaction loop, exposure chamber and exit line with buffer. The second push rapidly mixed 13 µl of DNA solution with 13 µl of RNAP solution in the cross mixer and brought the reactants into the reaction loop that has a sufficient volume to hold the combined volumes of the reactants (26 µl). The pause was the time the reactants were allowed to react. This was the only timing parameter that was changed during the experiment. The third push moved the sample through the X–ray exposure chamber at a specific speed, which determines the exposure time and then expelled the exposed sample into the Eppendorf collection tube. There was no need to quench the cleavage reaction, since the lifetime of the hydroxyl radicals is at most a few microseconds. The fourth push washed the reaction loop, exposure chamber and exit line with buffer.

At this point the safety shutter of the beamline hutch was enabled, allowing entry of the X–ray beam. The time sequences were initiated and the irradiated samples were collected. The safety shutter was closed immediately after finishing the experiment (turning off the X–ray beam), and the operator entered the hutch to collect the exposed samples and to reload the machine for the next experiment.

After completion of the experiment, each irradiated sample (final volume 100 µl) was mixed with 10 µl of 3 M sodium acetate (NaAc, pH 5.0) containing carrier DNA (0.1 mg/ml) and with four volumes (400 µl) of cold ethanol to precipitate the DNA. After 2 hours at -70°C and centrifugation the precipitate was dissolved in 16 µl of 80 % formamide and heated for 2 min at 90°C. DNA fragments were resolved on 8 % denaturing polyacrylamide gel using an ALF Express II DNA analyzer (Amersham Pharmacia Biotech).

Important:
1. Before starting the experiment, the exposure capillary in the experimental set–up needs to be aligned with the X–ray beam.

2. Proper operating of the stopped−flow system requires that no air bubbles are present in the driving syringes. Should this occur, the solutions flow through the SFM−400 flow lines will not be correctly controlled by the plunger movement and artifacts may be observed. Therefore, the buffer was always degassed before filling the machine and any bubbles in each SFM−400 syringe were eliminated by driving the plunger of SFM−400 syringe up and down several times while it was connected to the reservoir syringe.

3. In order to obtain quantitative evaluation of the footprinting patterns, the hydroxyl radical cleavage of DNA has to fall within single−hit kinetics regime. It means that each DNA molecule that is cleaved by hydroxyl radical is cleaved only once, and since each backbone position on the DNA has the same probability to be cleaved, the equally accessible nucleotides are represented with the equimolar amounts of the corresponding cleavage fragments. Theoretical calculations show that cleavage on average of 10 % of the DNA molecules fulfills single−hit kinetics regime for more than 99 % molecules which are cut by hydroxyl radicals. In order to find the experimental parameters fulfilling this criterion, an exposure time calibration (dose−response experiment) was performed under the reaction condition of experiment. The exposure time was varied by changing the speed at which the sample was pushed through the X−ray exposure chamber (the speed of the movement of plunger of syringe 4, third push) and DNA cleavage yield, along with cleavage fragments distribution, was analyzed by electrophoresis using ALF machine. The dose−response experiments revealed that at ring currents from 88 to 68 mA, the optimal speeds of the movement of plunger of syringe 4 are between 4.4 ml/s in the case of naked promoter DNA and 3 ml/s in the case of protein binding. These speeds result in ~ 0.18 and 0.26 millisecond exposure times, respectively. Since the small quartz capillary volume and 1 mm x 1 mm beam size allow exposure of about one thirtieth of the sample, the resulting dead times for exposure were between ~ 6 and 8.7 milliseconds.

4. Glycerol, present in many biological reagent storage buffers, is a potent radical scavenger; as little as 0.1 % (v/v) will inhibit nucleic acid cleavage by 50 % [Sclavi et al., 1998]. Therefore, RNAP was always dialyzed against BBcac20 buffer on a

floating membrane VSWP02500 (Millipore) to remove glycerol immediately before the experiment.

5. The X–ray beam causes a rise in the temperature of the irradiated material. Therefore, for samples with reaction times longer than 20 sec, the timing sequence was initiated prior to enabling the X–ray beam in order to minimize the time that the X–rays are entering the experimental hutch and thus preventing the breakdown of the quartz capillary which could result from overheating. The controller software was programmed with a countdown timer that allows the experimenter to know when the sample is poised to enter the exposure chamber.

9.4. Rapid mixing permanganate footprinting experiments (single–strand probing).

Time–resolved permanganate footprinting experiments were performed in two steps. First, using the stopped–flow machine of our own construction, the RNAP and DNA promoter fragment were rapidly mixed and incubated for a given amount of time following which the sample was exposed to the single–strand specific reagent, potassium permanganate ($KMnO_4$). Subsequently the DNA was cleaved at the modified thymines by applying piperidine.

9.4.1. Characteristics of stopped–flow machine of our own construction.

Figure 34 shows a scheme of stopped–flow machine used in time–resolved permanganate footprinting studies. Basically, the instrument consists of four Hamilton syringes driven by three independent stepping–motors. The stepping–motors are controlled by computer–based electronics that generate the necessary pulses with the appropriate timing. The outlets of the syringes are connected to a mixing device which is enclosed in a circulating water bath providing the temperature control. The mixing device of machine comprises a system of reagent loops, mixing chambers and reaction loops which is set up according to the desired kinetic experiments. The first mixing chamber is a cross mixer designed to mix two reagents, RNAP and DNA promoter fragment, while the second mixing chamber is a T–mixer designed to mix the reacted solution with the potassium permanganate. The end of the reaction loop 2 is connected to a short piece of capillary tubing (exit tube) that feeds into an Eppendorf tube containing quenching reagent. The design of the stopped–flow machine and

the sample collector at the end of the machine allow the sequential mixing of up to 11 samples.
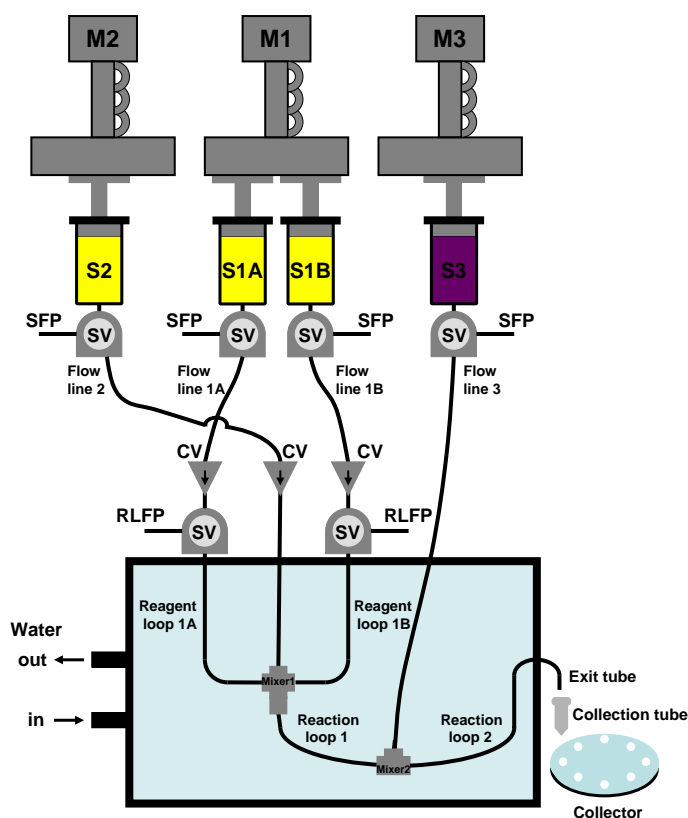

Figure 34. Schematic representation of the stopped−flow machine of our own construction (S1A, S1B, S2, S3 − Hamilton syringes; M1, M2, M3 − independent spindle−driven stepping−motors; SVs − selection valves; CVs − check valves; SFPs − syringe filling ports; RLFPs − reagent loop filling ports; Mixer 1 − cross mixer; Mixer 2 − T−mixer).

9.4.2. *Modifications of thymines using potassium permanganate.*

At the beginning of a typical experiment, the circulating water bath of the stopped−flow machine was switched on, and allowed to equilibrate at the desired temperature of the experiment. Meanwhile, syringes 1A, 1B and 2 and flow lines connected to them were filled (no air bubbles!) with BBcac20 buffer (50 mM sodium cacodylate, pH 7.5; 20 mM NaCl; 6 mM MgCl$_2$), while syringe 3 and flow line 3 were filled with 60 mM KMnO$_4$ solution. The two reagents were then loaded in the two reagent loops, 200 nM RNAP in BBcac20 buffer on one side and 50 nM radioactively labeled DNA in BBcac20 buffer on the other. Eppendorf tubes, each containing 100 µl of 2 x stop−KMnO$_4$−solution (10% β−mercaptoethanol; 0.6 M NaAc, pH5.0; 20 µg/ml carrier DNA), were placed into the collector. The selector valves were then turned so that the driving syringes were opened toward the flow lines.

The time sequences (time to mix and incubate the reagents and subsequently expose them to the $KMnO_4$) for the samples were entered into the drive motor controller. Each injection consisted of a "push motor 1 (first push) – pause 1 – simultaneously push motors 2 and 3 (second push) – pause 2 – push motor 2 (third push) – move the collector – push motor 2 (fourth push) – move the collector". The first push rapidly mixed 10 µl of DNA solution with 10 µl of RNAP solution in the cross mixer and brought the reactants into the reaction loop that has a sufficient volume to hold the combined volumes of the reactants (20 µl). The pause 1 was the time the reactants (RNAP and promoter DNA) were allowed to react. This was the only timing parameter that was changed during the experiment. The second push rapidly mixed the sample (20 µl) with 20 µl of $KMnO_4$ solution. The next pause determined the time of sample exposure to the single–strand specific reagent (0.2 s). The third push expelled the exposed sample into the Eppendorf collection tube where the oxidation reaction was rapidly quenched. The fourth push washed the mixing chambers, reaction loops and exit line with buffer.

The time sequences were then initiated and the samples for data points were collected. After the completion of experiment, each sample (final volume ~ 200 µl) was mixed with four volumes (800 µl) of cold ethanol to precipitate the DNA. After precipitation with ethanol, the DNA was subjected to piperidine treatment.

9.4.3. Piperidine treatment.

The pellet was dissolved in 90 µl of freshly prepared 10 % piperidine and incubated at 90°C for 20 min. Ten µl of 5 M LiCl was then added and DNA fragments were precipitated with 400 µl of ethanol, rinsed with 80 % ethanol, dried, dissolved in 10 µl of 80 % formamide containing 0.02 % of bromphenol blue and resolved by gel electrophoresis.

9.4.4. Gel electrophoresis.

The samples were heated at 90°C for 2 min, chilled on ice, and then 3 µl aliquots of each sample were applied on 22 x 55-cm (spacer thickness from 0.2 to 0.4-mm) slab of 7 % polyacrylamide gel containing 8 M urea and TBE buffer (50 mM Tris, 50 mM boric acid, 1 mM EDTA). Gels were run at 50 W for ~1.5 h using heating plates (50°C). Gels were then incubated for 20 min in 10 % acetic acid, washed twice with water (total washing time: 1 h), dried and exposed to X–ray film. For quantification, gels were exposed to Fuji Imager plate

BAS IIIS, which was scanned with Bas−1000 PhosphorImager instrument. Scans were processed with MacBas software.

9.5. Characterization of open complexes formed on different T7A1 promoter variants.

9.5.1. *Band shift experiments.*

The band shift experiments were done to assess how well each of *T7A1* promoter variants bound to RNAP holoenzyme. In brief, *E.coli* RNAP holoenzyme was mixed with the test promoter fragment in BBcac20 buffer (at an enzyme to promoter ratio equals 1.5 to 1) and then incubated for 10 min at 37°C. For heparin challenge experiments, heparin at 5−fold molar excess relative to a RNAP was added to the reaction mixture for an additional 2 min. Each sample was then supplemented with an equal volume of 2 x loading buffer and run on 4% / 20% non−denaturing polyacrylamide gel. After electrophoresis, DNA was visualized with ethidiumbromide (for details see section *9.2.3.1.*). All experiments were repeated at least two times.

9.5.2. *In vitro transcription.*

The *E.coli* RNAP holoenzyme (7.5 pmoles; → 750 nM) was incubated with the unlabeled test promoter fragment (5 pmoles; → 500 nM) in 9 µl of BBcac20 buffer at 37°C for 10 min to form open complex. The synthesis of 11nt−long or 20nt−long RNA products was initiated by addition of 1 µl of 10xStart−RNA11 (50 µM [γ−^{32}P]ApUpC primer; 500 µM rATP+rGTP) or 10xStart−RNA20 (50 µM [γ−^{32}P]ApUpC primer; 500 µM rATP+rGTP+rCTP) respectively. The reaction mixtures were further incubated at 37°C. Aliquots (1.5 µl) were withdrawn at the desired times (in the range from 1 min to 20 min), mixed with 5 µl of 100 % formamide containing 0.02 % of bromphenol blue, and then analyzed by denaturing gel electrophoresis (20 x 20-cm slab of 20 % denaturing polyacrylamide gel). In order to visualize the RNA transcripts, after electrophoresis the gels were exposed directly to a phosphorimager screen and evaluated with AIDA Image Analyser software.

9.5.3. *Probing of transcription bubble using potassium permanganate.*

Typically, the *E.coli* RNAP holoenzyme was incubated with the test promoter fragment radioactively labeled at 5'−end of either template or non−template strand (at an enzyme to promoter ratio equals 1.5 to 1) in BBcac20 buffer at 37°C for 10−15 min to form open binary complex (BC). After incubation period, each sample was mixed with an equal volume of 2 mM freshly prepared potassium permanganate, incubated at 37°C for 10 sec, rapidly supplemented with an equal volume of 2 x stop−KMnO$_4$−solution, and processed for piperidine treatment and gel analysis as described above (see sections *9.4.3.* and *9.4.4.*, respectively).

10.   Data analysis.

10.1.   Analysis of hydroxyl radical footprinting data.

10.1.1. Quantification and normalization of time−resolved footprints.

The output file of an electrophoresis run on the ALF Express DNA analyzer II (fluorescence intensity versus time for each lane on the gel, see Figure 11 in "Results") was converted into ASCII format. By using the software OriginPro (OriginLab, Northampton, MA) a linear baseline was subtracted from each dataset and a smaller version of the file was created containing the time region of the run to be analyzed. The Peak Fitting Module of the software OriginPro was used to fit the peaks in each lane to a Lorenzian curve. Approximately 150 peaks were fit simultaneously, covering the whole region interacting with the polymerase. The values for peak areas within each lane were divided by the average value of the area of several peaks within the same lane that were not protected by the polymerase in the course of the binding reaction. Normalized area values were then divided by the values for the corresponding peaks in the lane of DNA cleaved in the absence of protein, the resulting values were then relative change compared with naked DNA, Φ (Figure 12 in "Results").

10.1.2. Fit of the kinetic data to single and double exponential equations.

The progression curves (Φ *versus* time) of appearance of protection from hydroxyl radical cleavage at each nucleotide position were fit individually to single or double exponential expressions as follows:

$$y = L + (U\text{-}L)\, e^{-kt}, \tag{2}$$

and

$$y = L + A\, e^{-k_A t} + B\, e^{-k_B t} \qquad U = L + A + B, \tag{3}$$

where A and B are the signal amplitudes and k, k_A and k_B are the observed, apparent rate constants. The upper (U) and lower (L) limits from these fits were used to normalize the data from 0 to 1, where 0 corresponds to the value in the absence of polymerase and 1 corresponds to the value for the area of the peak at its minimum, resulting in the value of the fractional saturation at that site. We observed that often the difference in the value of the rate and amplitude measured for different nucleotides within a certain region was smaller than the error derived from the fit. The data points for these nucleotides, taken from at least two independent experiments, were combined and then fit again to reduce the error. In order to

determine whether a single or a double exponential expression better describes the results in each dataset, a visual analysis of the residuals was carried out (for example, see Figure 1S in "Supporting materials") as well as an F test (for example, see Table 1S in "Supporting materials").

10.1.3. Residuals from nonlinear regression.

Nonlinear regression finds parameters that make a model fit the data as closely as possible (given some assumptions). It does not automatically ask whether another model might work better. The graphical residuals analysis is one of the key tools for model validation.

Residuals are the differences between the observed and predicted responses. In other words, residuals are the vertical distances of each corresponding data point from the fit curve. Residuals are positive when the points are above the curve, and are negative when the points are below the curve. Carefully looking at residuals can tell whether the made assumptions are reasonable and the choice of model is appropriate. The general assumption applied to the group of residuals is that one expects them to be roughly normal and (approximately) independently distributed around the fit curve [Motulsky et al., 2003].

10.1.4. Extra sum−of−squares F test.

As outlined above, the goal of nonlinear regression is to find parameter values that make the curve come near the data points or, in other words, to minimize the sum of squares of the vertical distances of the data points from the curve. So it seems that the model which fit data with the smallest sum−of−squares is the best. In fact, that approach is too simple. The problem is that a more complicated model (more parameters) can fit the data better (can come close to the points) just because it can have more inflection points. For instance, two−phase model almost always fits data better than one−phase model. So any method to compare a simple model with more complicated model has to balance the decrease in sum−of−squares with the increase in the number of parameters.

The F test (extra sum−of−squares) is one of the key statistical approaches used to compare related models (two models are related when one is a simpler case of the other) [Motulsky et al., 2003]. It is based on the difference between the sum−of−squares of the two

models and also takes into account the number of data points and the number of parameters of each model.

If the simpler model (fewer parameters) is correct, then the relative increase in the sum−of−squares (going from complicated to simple model) would be approximately equal the relative increase in degrees of freedom (DF), which equal the number of data points minus the number of parameters. If the more complicated model is correct, then relative increase in the sum−of−squares would be greater than the relative increase in degrees of freedom.

The F ratio equals the relative difference in the sum−of−squares divided by the relative difference in decrees of freedom:

$$F = \frac{(SS1-SS2)/SS2}{(DF1-DF2)/DF2} \qquad \textbf{[4]}$$

That equation is more commonly shown in a following equivalent form:

$$F = \frac{(SS1-SS2)/(DF1-DF2)}{SS2/DF2} \qquad \textbf{[5]}$$

where the number 1 and 2 refer to the simple (for example, single exponential) and more complex (for example, double exponential) models, respectively. (DF1-DF2) is the degrees of freedom for the numerator (DFn), and DF2 is the degrees of freedom for the denominator (DFd).

If the simpler model is correct, then an F ratio is expected to be near 1.0. If the F ratio is much greater than 1.0, then there are two possibilities:

• The more complicated model is correct.

• The simpler model is correct, but random scatter in the data led the more complicated model to fit better.

The P value, which can be calculated from the F ratio and the two DF values, tells how frequently the second possibility would happen. P is the probability (ranging from 0 to 1) that the results observed in study could have occurred by chance. If the P value is low (0.05 or below), then one can conclude that more complicated model is significantly better than the simple model.

10.2. Analysis of potassium permanganate footprinting data.

In order to obtain the image of potassium permanganate footprints of DNA bound by RNAP, the radioactive (^{32}P) gel was exposed to Fuji Imager plate BAS IIIS and scanned using Bas−1000 PhosphorImager instrument. Band intensity profiles along each gel lane were determined using the MacBas software. For integration of the area of the peaks in each lane, the lowest intensity points in the lane were used as the horizontal baseline and peaks were fit to a Lorenzian curve. The values for the peaks' area were then normalized on the total amount of DNA material applied on corresponding gel lane. Normalized area values of the peaks in the lane of DNA treated with $KMnO_4$ in the absence RNAP (local background intensity) were then subtracted from the normalized area values of corresponding peaks within each lane. The resulting values were then change compared with naked DNA, Φ.

The plots of the increase of $KMnO_4$−accessibility at each thymine position (Φ *versus* time) were fit to single (**[2]**) or double (**[3]**) exponential equations and processed in the same way as in case of hydroxyl radicals footprints (see section *10.1.2.*). Analysis of residuals as well as statistical F−test have shown that in all cases the plots fit better to double than to single exponential equation (see Figure 2S and Table 2S in "Supporting materials", respectively).
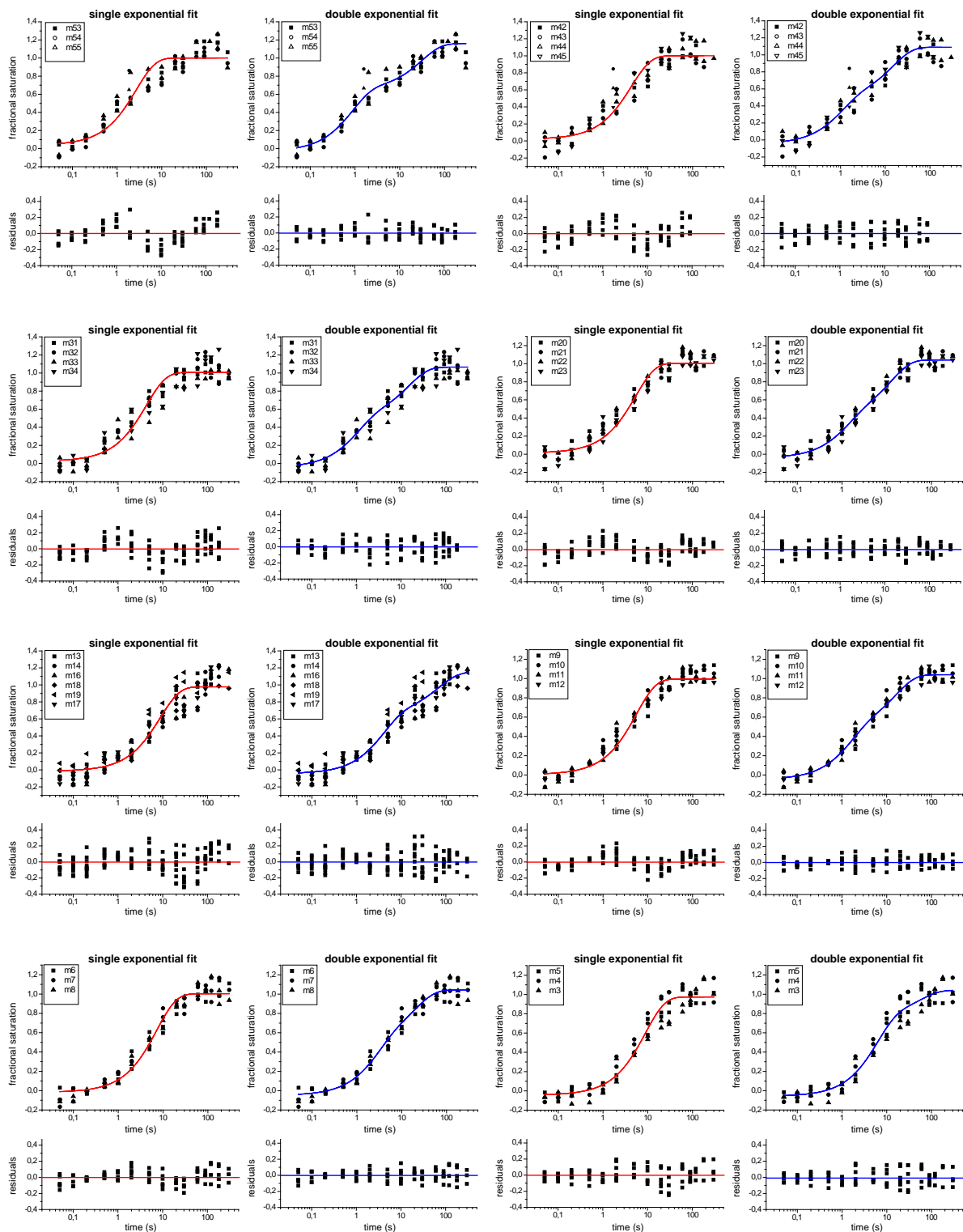
11.   Supporting materials.

Figure 1S (a). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the wild type *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. At the bottom of each panel shown are the residuals (the difference between the data points and the fit curve, see "Data analysis"). The residuals for the single exponential fit show a periodic pattern of distribution, whereas the residuals for the double exponential fit are almost evenly distributed on either side of the fit curve indicating that a double exponential expression better fits data
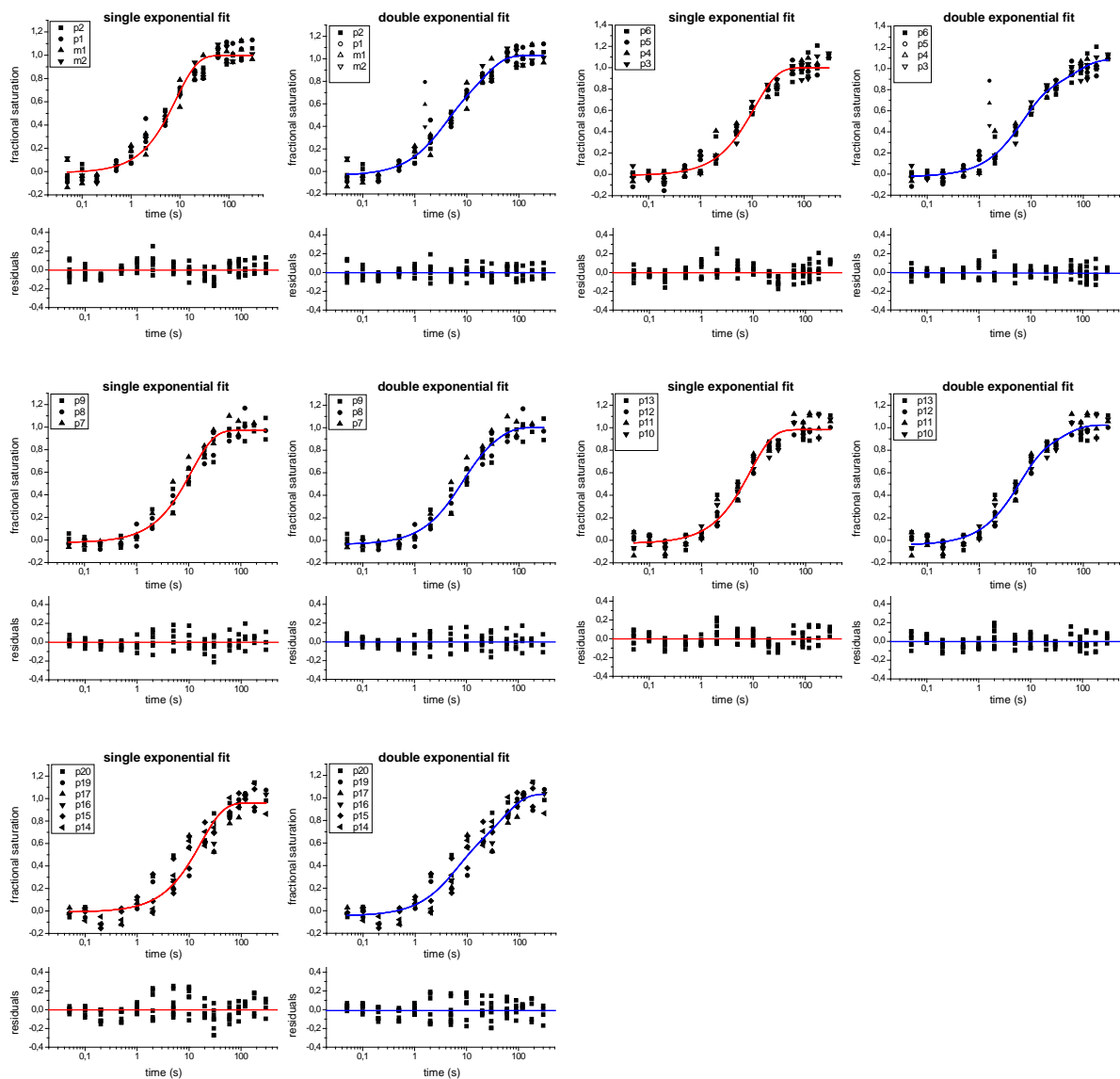
Figure 1S (a) (continuation). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the wild type *A1* promoter of phage *T7* at 37°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel.
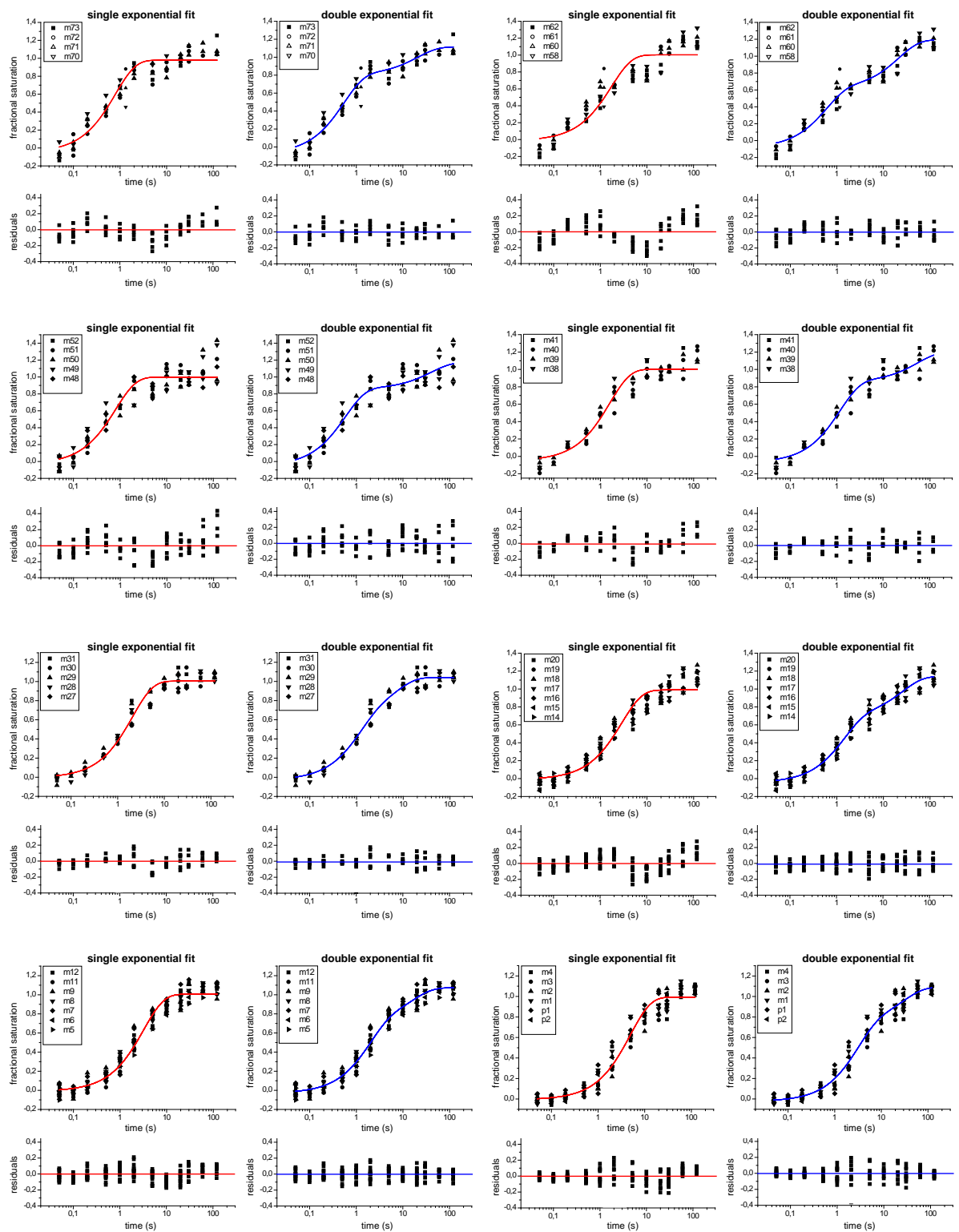
Figure 1S (b). Comparison of the fits of the kinetics of the protection appearance at specific sites on the non−template strand upon *E.coli* RNAP binding to the wild type *A1* promoter of phage *T7* at 37°C to a single (red curve) and double (blue curve) exponential equation. Shown are the residuals for the two fits.

104

Figure 1S (b) (continuation). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the wild type *A1* promoter of phage *T7* at 37°C to a single (red curve) and double (blue curve) exponential equation. Shown are the residuals for the two fits.
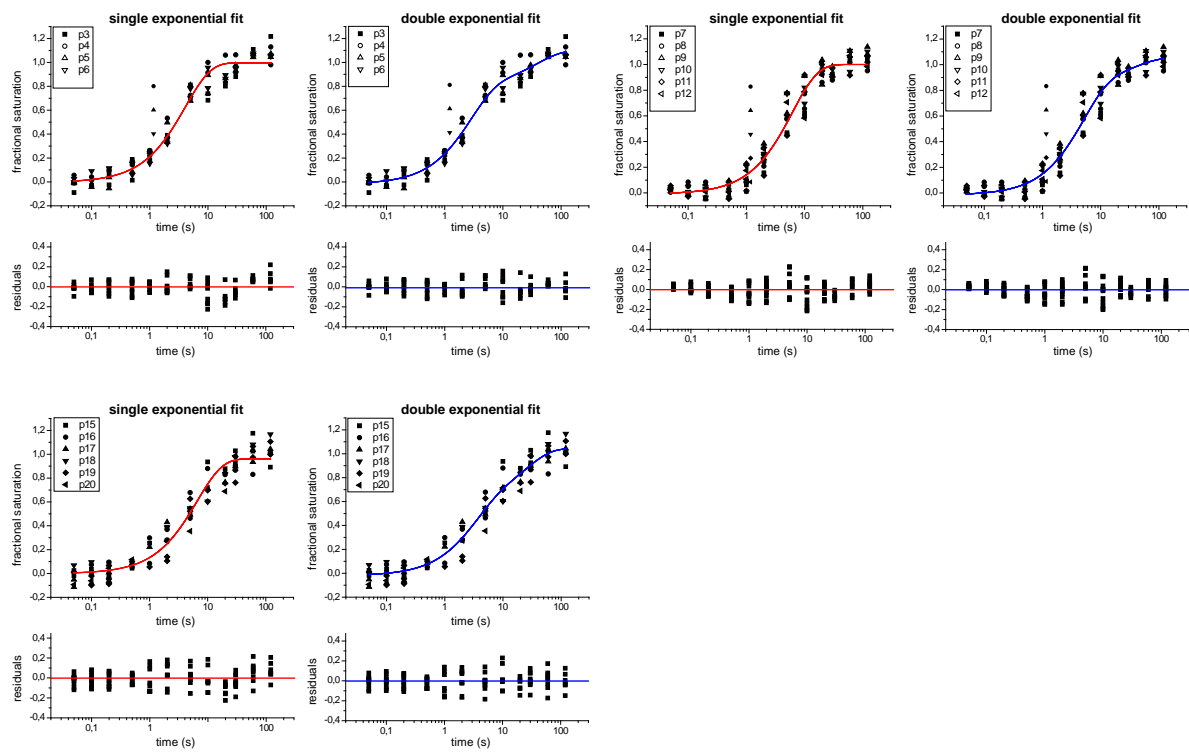
Wild type T7A1 promoter, 37°C, Template strand					
Nucleotide position	DFn	DFd	*F* value	*F* value for P=0.05 (95%)	*F* value for P=0.01 (99%)
m55-m53	2	74	75.893	3.120	4.904
m45-m42	2	68	19.259	3.132	4.932
m34-m31	2	99	34.043	3.088	4.826
m23-m20	2	97	22.707	3.090	4.831
m19-m13	2	146	23.097	3.058	4.754
m12-m9	2	98	42.788	3.089	4.829
m8-m6	2	83	17.634	3.107	4.870
m5-m3	2	80	5,687	3.111	4.881
m2-p2	2	103	18.434	3.085	4.817
p3-p6	2	118	27.051	3.073	4.790
p7-p9	2	76	4.490	3.117	4.896
p10-p13	2	107	10.109	3.081	4.809
p14-p20	2	94	15.370	3.093	4.838

Wild type T7A1 promoter, 37°C, Non–template strand					
Nucleotide position	DFn	DFd	*F* value	*F* value for P=0.05 (95%)	*F* value for P=0.01 (99%)
m73-m70	2	64	17.884	3.140	4.953
m62-m58	2	70	60.452	3.128	4.922
m52-m48	2	88	15.403	3.100	4.855
m41-m38	2	64	14.317	3.140	4.953
m31-m27	2	73	12.761	3.122	4.908
m20-m14	2	137	64.407	3.062	4,763
m12-m11, m9-m5	2	154	32.927	3.055	4.746
m4-p2	2	121	22.419	3.071	4.785
p3-p6	2	89	21.235	3.099	4.852
p7-p12	2	126	4.824	3.068	4.778
p15-p20	2	99	13.600	3.088	4.826

Table 1S. *F* values for the data sets that were fit to the double exponential equation.

The *F* test comparison of fits was performed to determine whether the greater number of variables in the double exponential equation was responsible for the better fit compared with the single exponential equation (for details see section "Data analysis"). The *F* values corresponding to the 95 % and 99 % confidence levels are shown. If the *F* value for a given data set is greater than its 95 % confidence level, for example, it means that there is at most a 5% probability that the scatter in the data could result in a better fit to the simpler model (a single exponential); therefore, we can say with a 95 % confidence level that a double exponential results in a better fit of the data. This analysis was performed by using the software PRISM (GraphPad, San Diego).
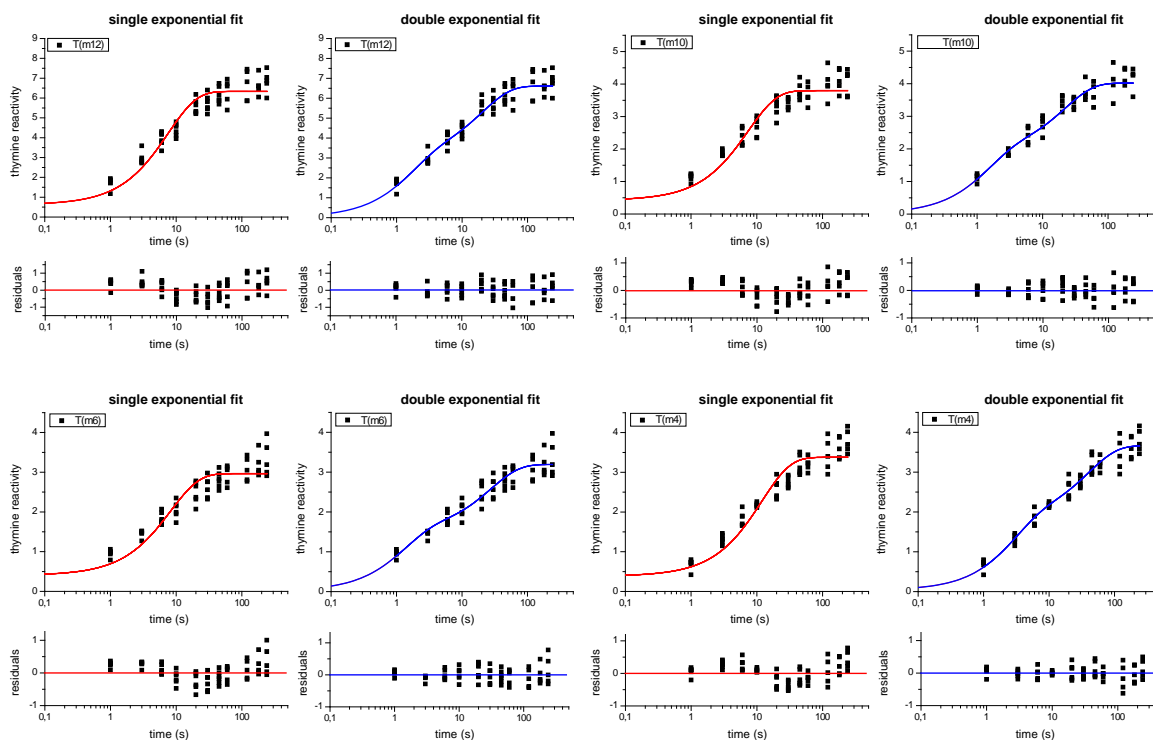
Figure 2S. Comparison of the fits of the kinetics of increase of template strand thymines reactivity to $KMnO_4$ upon open complex formation on the wild type *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel (for details see section "Data analysis").

Wild type T7A1 promoter, 37°C, Template strand					
Tymine position	DFn	DFd	*F* value	*F* value for P=0.05 (95%)	*F* value for P=0.01 (99%)
m12	2	63	21.029	3.143	4.959
m10	2	63	28.168	3.143	4.959
m6	2	63	33.012	3.143	4.959
m4	2	62	45.060	3.145	4.965

Table 2S. *F* test statistical analysis confirming that the increase of $KMnO_4$–accessibility of thymines in the template strand upon binary complex formation between *E.coli* RNAP and wild type *T7A1* promoter at 37°C is better described by double rather than single exponential equation. Shown are the *F* values for each data set.
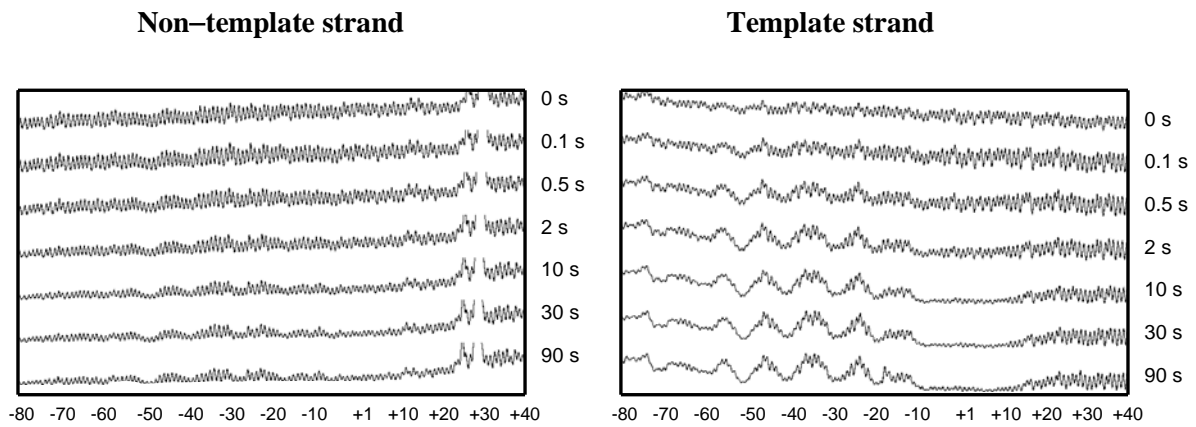
Non−template strand   **Template strand**

Figure 3S. Hydroxyl radical footprinting profiles of the mutant−1 *T7A1* promoter fragments after different incubation time with RNAP. The DNA was labeled with fluorophore Alexa 647 at the 5'−end of either non−template strand (NTS, left panel) or template strand (TS, right panel). The incubation time for each lane is shown next to the line graph.
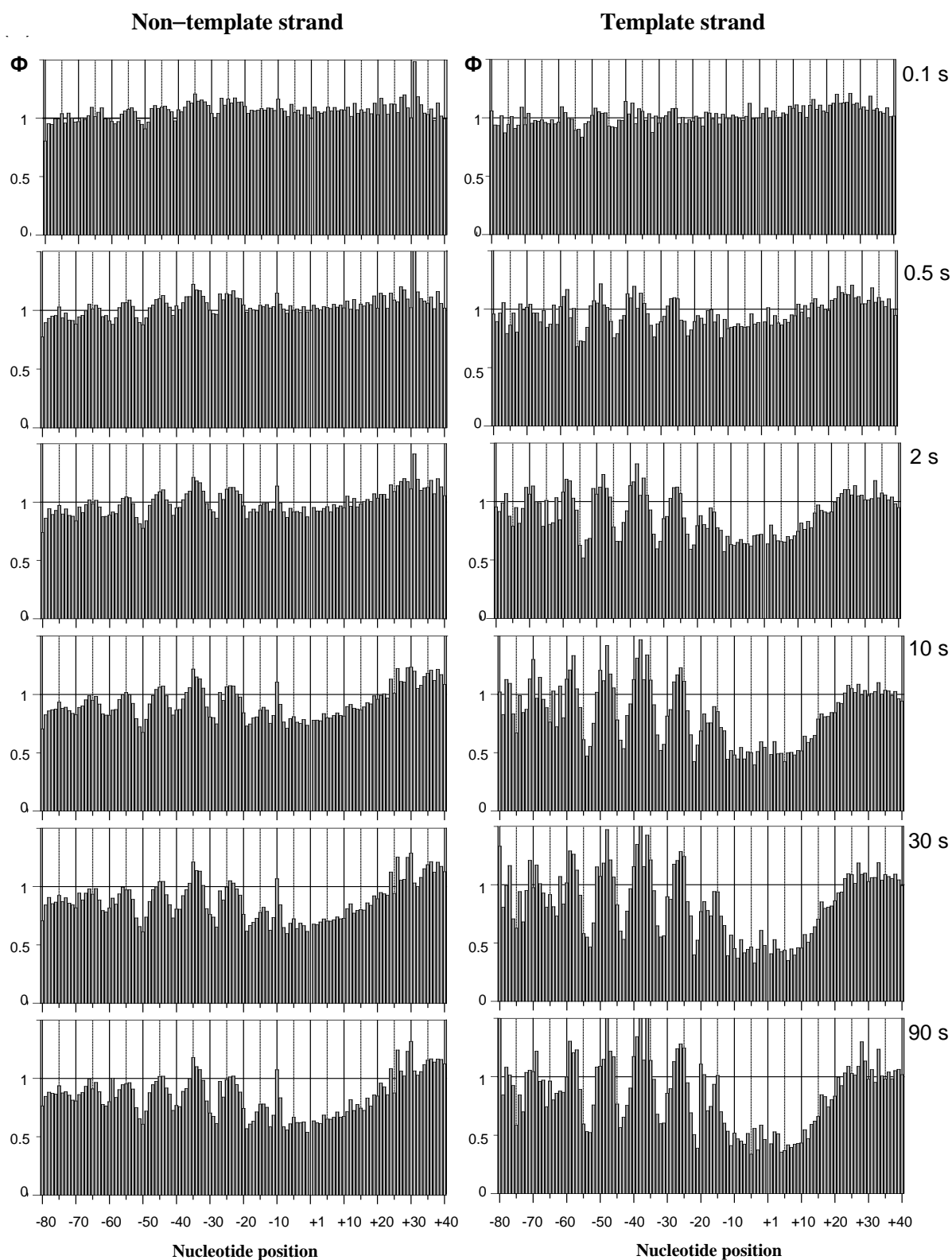
Figure 4S. Representative bar plots obtained in the case of RNAP binding to the mutant-1 *T7A1* promoter at 37°C. The bars show Φ, the ratio of the area of each peak relative to that in the naked DNA (lane at 0 s), in the course of the binding reaction. The bar at nucleotide position 0 is missing because in the conventional numbering of the nucleotides in a promoter 0 is not used. Incubation times are shown at the right edge of each raw for the two strands (non−template at the left panel and template at the right panel).
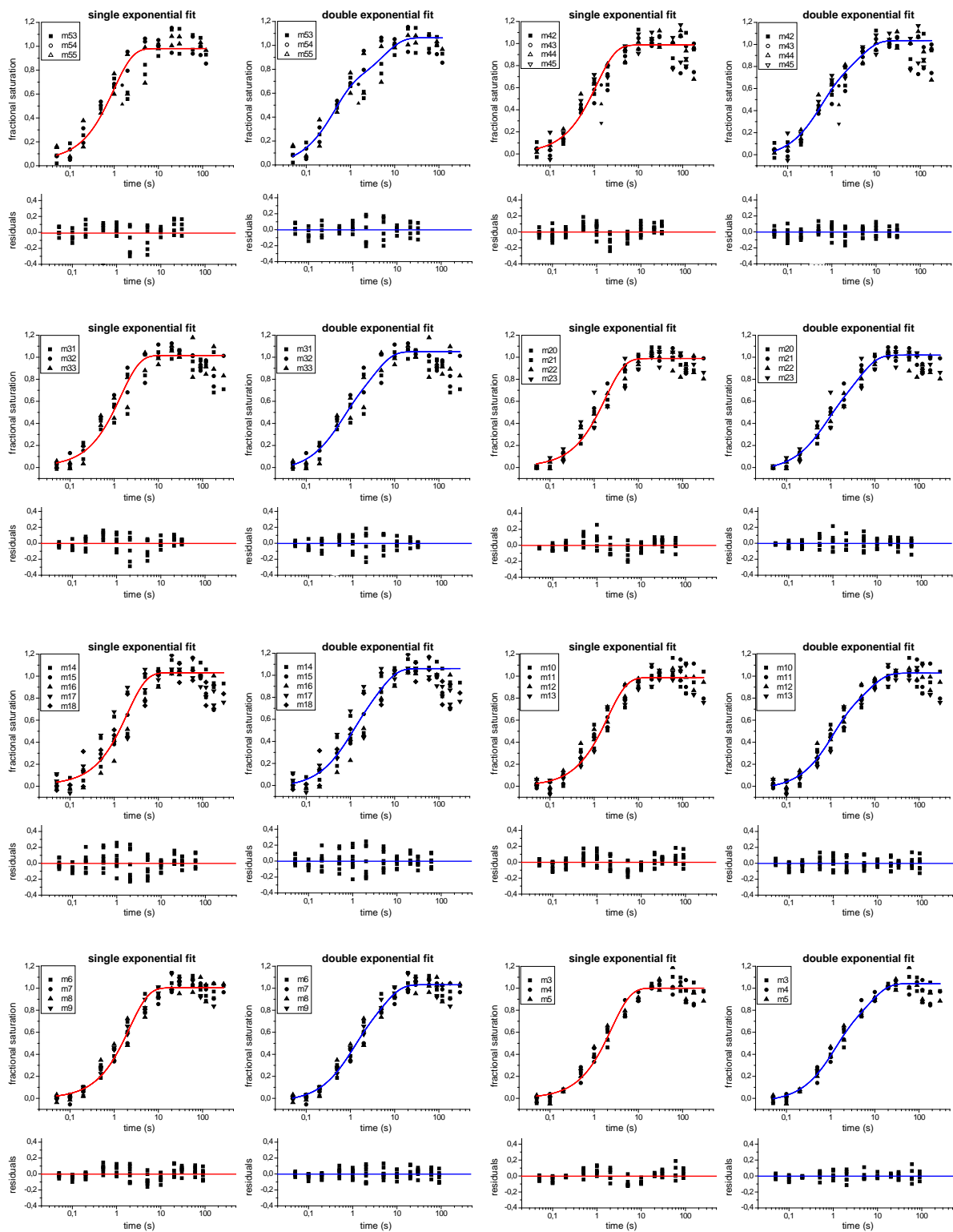
109

Figure 5S (a). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the mutant−1 *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. At the bottom of each panel shown are the residuals (for details see section "Data analysis"). In the case then a decrease in fractional saturation was observed at the longer time points, the fits were carried out in the absence of the points in the decreasing portion of the curve.

Figure 5S (a) (continuation). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the mutant−1 *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. At the bottom of each panel shown are the residuals (for details see section "Data analysis"). In the case then a decrease in fractional saturation was observed at the longer time points, the fits were carried out in the absence of the points in the decreasing portion of the curve.
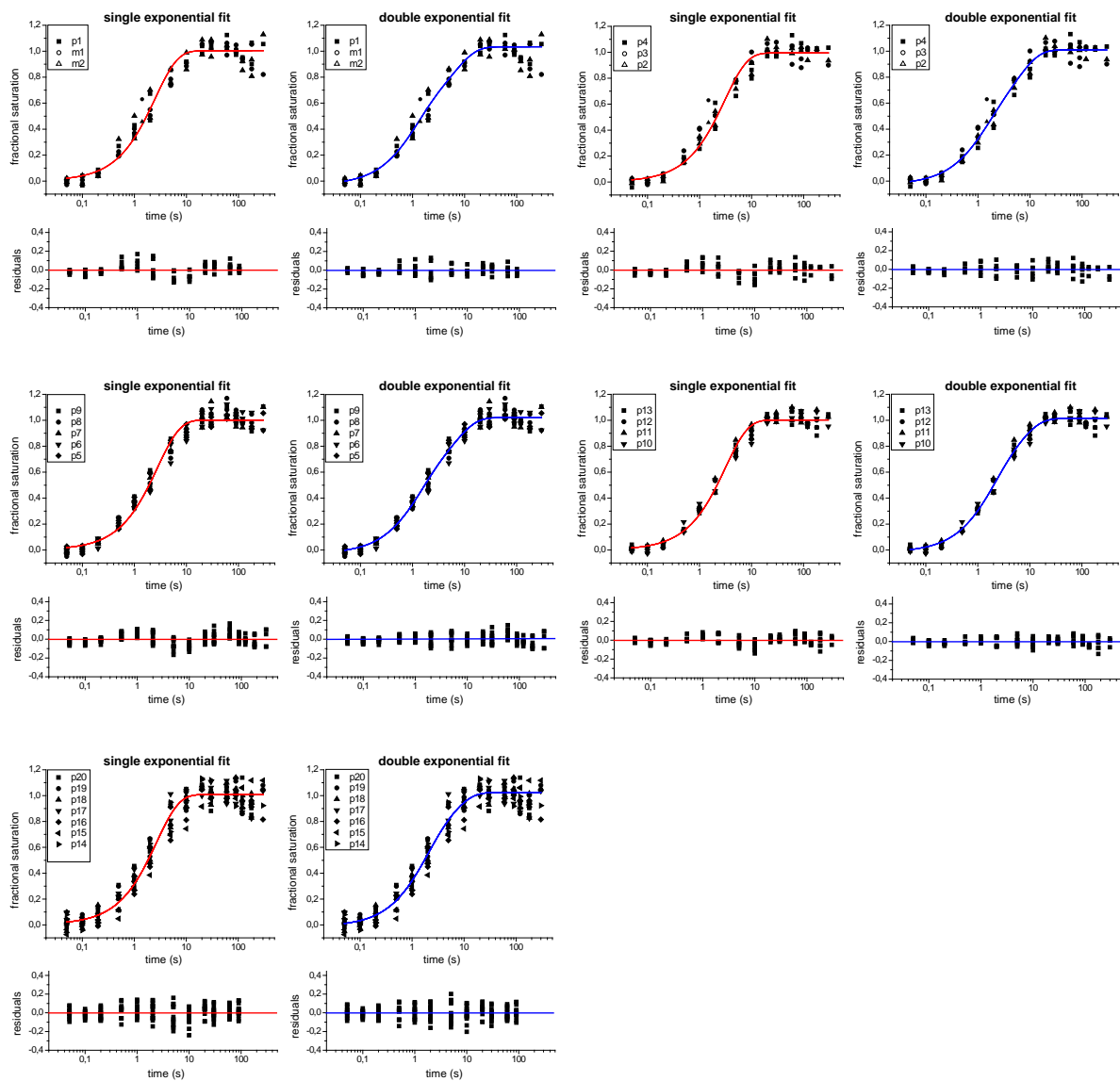
Figure 5S (b). Comparison of the fits of the kinetics of the protection appearance at specific sites on the non−template strand upon *E.coli* RNAP binding to the mutant−1 *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. At the bottom of each panel shown are the residuals (for details see section "Data analysis").
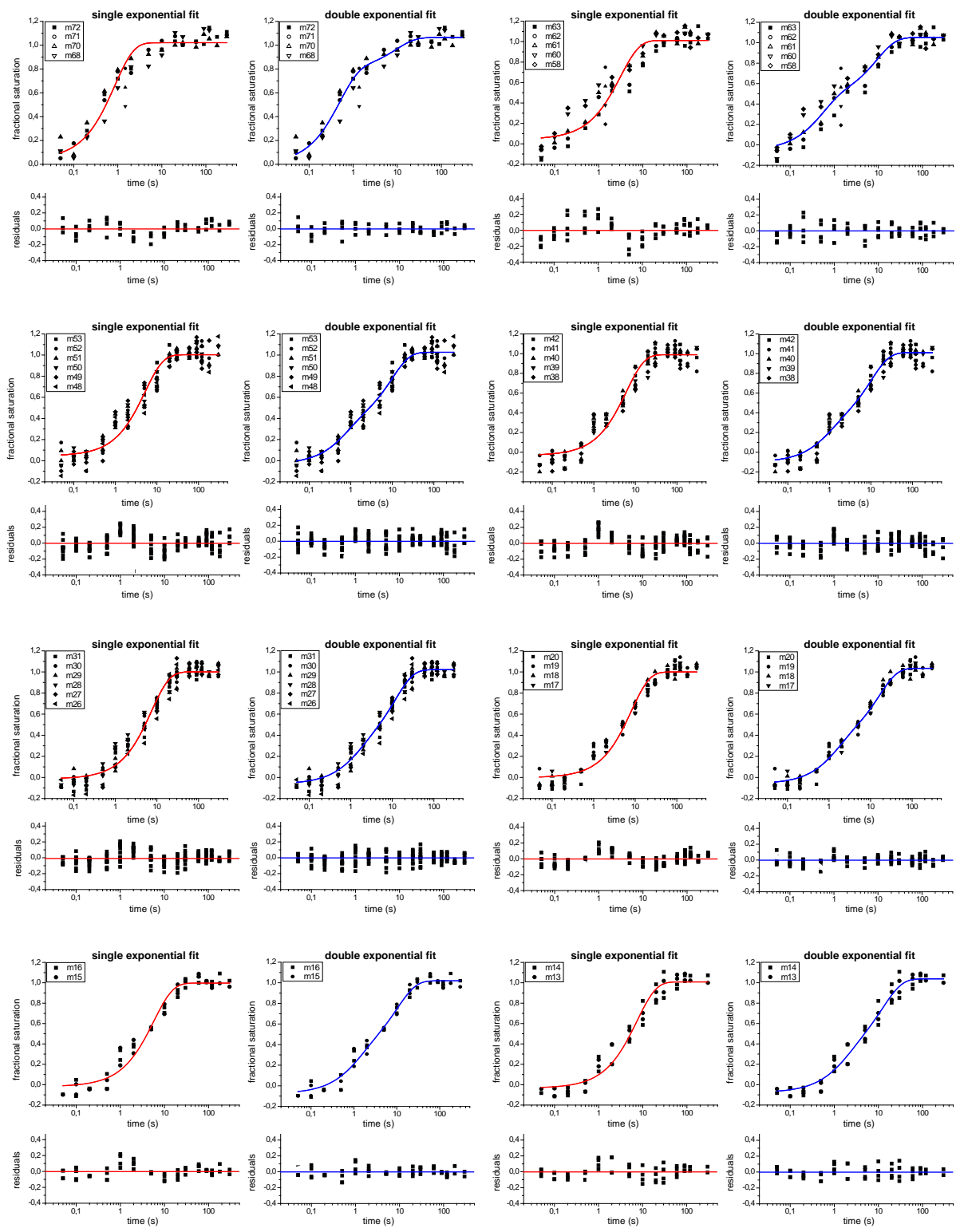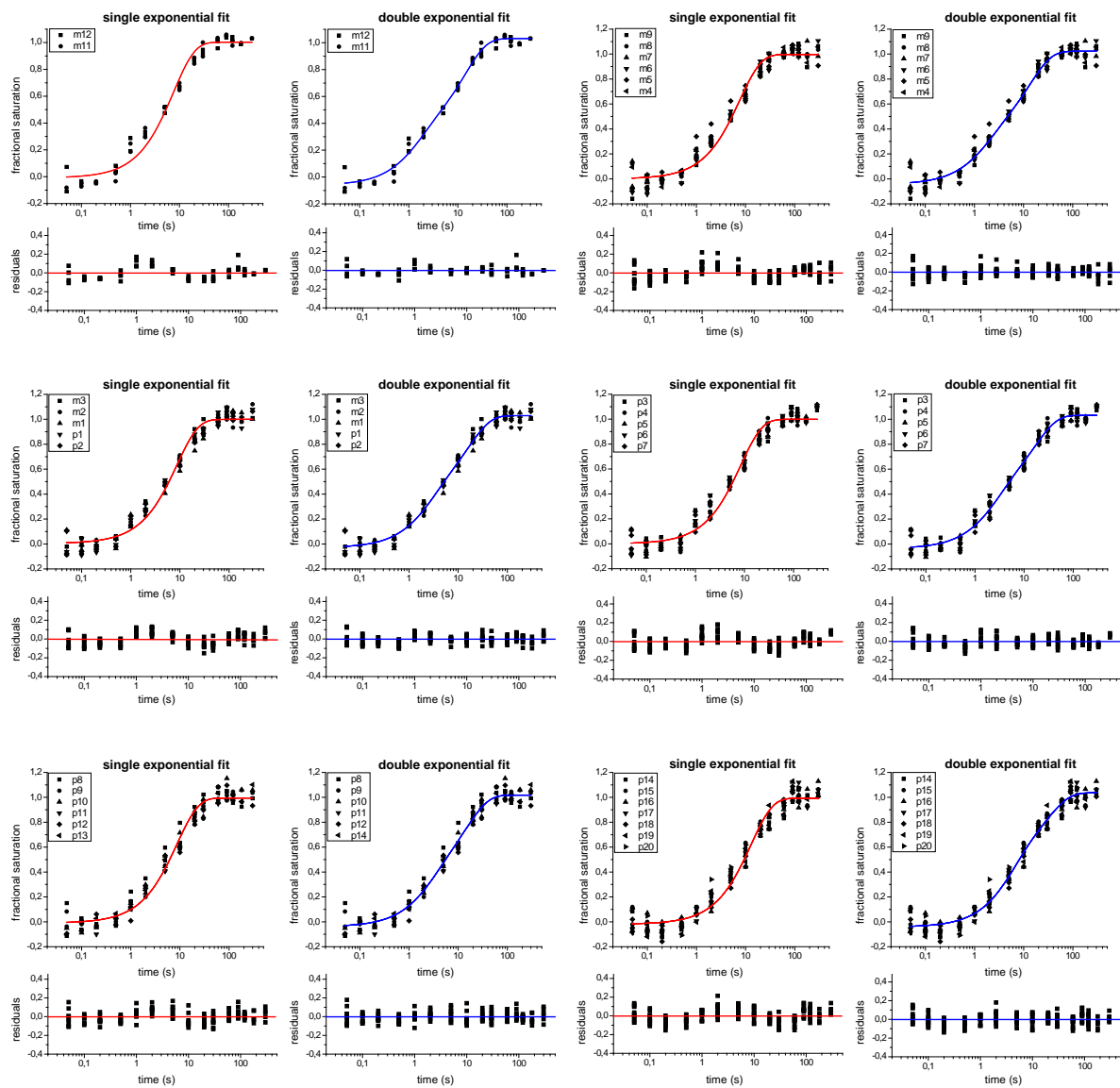
112

Figure 5S (b) (continuation). Comparison of the fits of the kinetics of the protection appearance at specific sites on the non−template strand upon *E.coli* RNAP binding to the mutant−1 *T7A1* promoter at 37°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel.

113

Mutant T7A1 promoter (consensus -10 region), 37°C, Template strand					
Nucleotide position	DFn	DFd	*F* value	*F* value for P=0.05 (95%)	*F* value for P=0.01 (99%)
m55-m53	2	54	12.555	3.186	5.021
m45-m42	2	74	17.846	3.120	4.904
m33-m31	2	59	7.353	3.153	4.984
m23-m20	2	76	15.305	3.117	4.896
m18-m14	2	84	4.665	3.105	4.867
m13-m10	2	93	20.368	3.094	4.841
m9-m6	2	95	19.986	3.092	4.836
m5-m3	2	71	31.351	3.126	4.917
m2-p1	2	73	24.333	3.122	4.908
p2-p4	2	82	13.959	3.108	4.874
p5-p9	2	144	34.030	3.059	4.756
p10-p13	2	116	18.635	3.074	4.793
p14-p20	2	175	8.188	3.048	4.729

Mutant T7A1 promoter (consensus -10 region), 37°C, Non–template strand					
Nucleotide position	DFn	DFd	*F* value	*F* value for P=0.05 (95%)	*F* value for P=0.01 (99%)
m72-m68	2	51	19.976	3.179	5.047
m63-m58	2	68	29.175	3.132	4.932
m53-m48	2	149	39.591	3.057	4.750
m42-m38	2	137	25.188	3.062	4.763
m31-m26	2	145	24.896	3.058	4.755
m20-m17	2	108	56.127	3.080	4.807
m16-m15	2	53	17.047	3.172	5.030
m14-m13	2	49	8.068	3.187	5.066
m12-m11	2	54	27.334	3.168	5.021
m9-m4	2	157	48.806	3.054	4.743
m3-p2	2	142	46.723	3.060	4.758
p3-p7	2	144	41.355	3.059	4.756
p8-p13	2	127	23.687	3.068	4.776
p14-p20	2	172	29.996	3.049	4.731

Table 3S. *F* test statistical analysis confirming that the double exponential equation results in a better fit of the data.
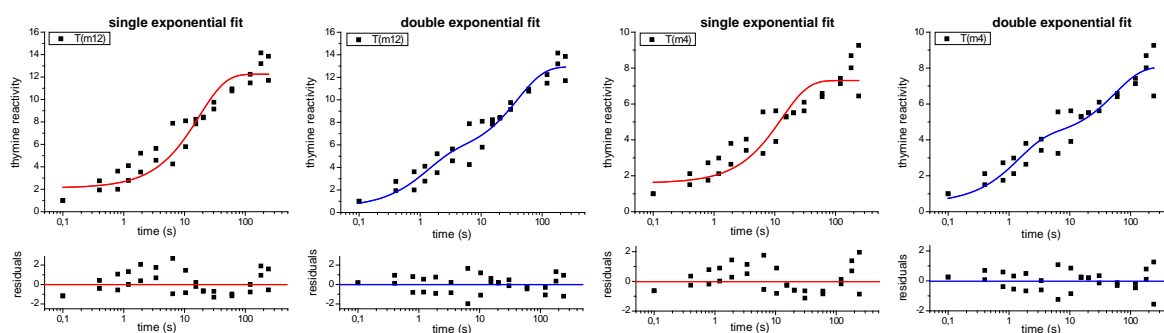
Figure 6S. Comparison of the fits of the kinetics of increase of template strand thymines reactivity to $KMnO_4$ upon open complex formation on the *T7A1* promoter with consensus -10 region at 37°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel (for details see section "Data analysis").

Mutant T7A1 promoter (consensus -10 region), 37°C, Template strand					
Tymine position	DFn	DFd	*F* value	*F* value for P=0.05 (95%)	*F* value for P=0.01 (99%)
m12	2	27	16.244	3.354	5.488
m4	2	27	15.411	3.354	5.488

Table 4S. *F* test statistical analysis confirming that the increase of $KMnO_4$−accessibility of thymines in the template strand upon binary complex formation between *E.coli* RNAP and mutant−1 *T7A1* promoter at 37°C is better described by double rather than single exponential equation. Shown are the *F* values for each data set.
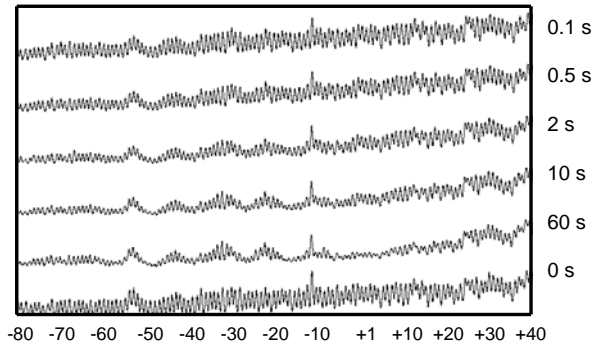
Figure 7S. Representative footprinting profiles of the wild type *T7A1* promoter fragments labeled with fluorescent tag at the 5'−end of non−template strand as a function of incubation time with RNAP at 20°C. Each line graph corresponds to one lane on the gel. The incubation time for each lane is shown next to the line graph.
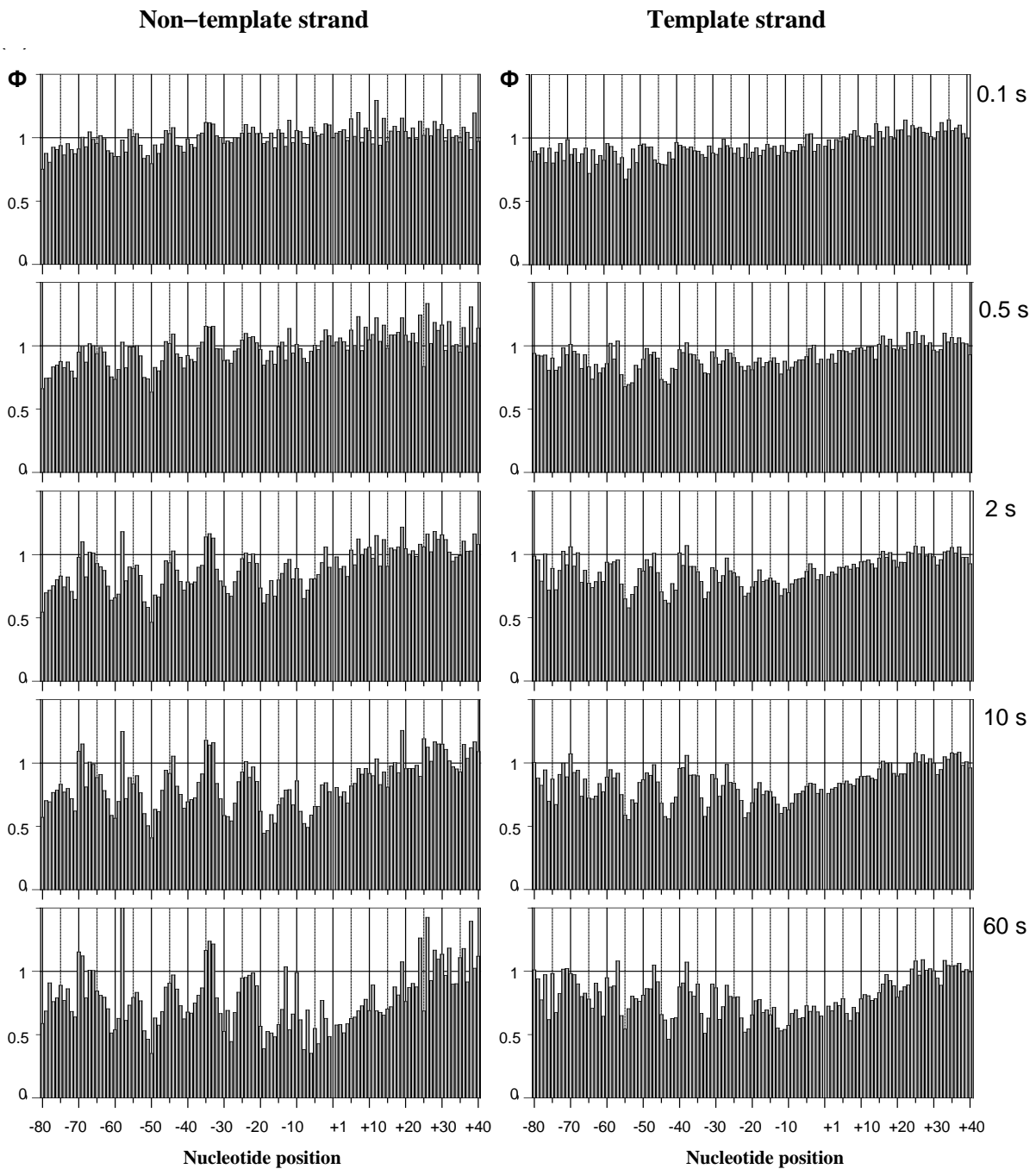

Figure 8S. The bar plots obtained in the case of RNAP binding to the wild type *T7A1* promoter at 20°C.

Figure 9S (a). Comparison of the fits of the kinetics of the protection appearance at specific sites on the non−template strand upon *E.coli* RNAP binding to the wild type *T7A1* promoter at 20°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel (for details see section "Data analysis"). In the case then a decrease in fractional saturation was observed at the longer time points, the fits were carried out in the absence of the points in the decreasing portion of the curve.

117

Figure 9S (a) (continuation). Comparison of the fits of the kinetics of the protection appearance at specific sites on the non−template strand upon *E.coli* RNAP binding to the wild type *T7A1* promoter at 20°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel.
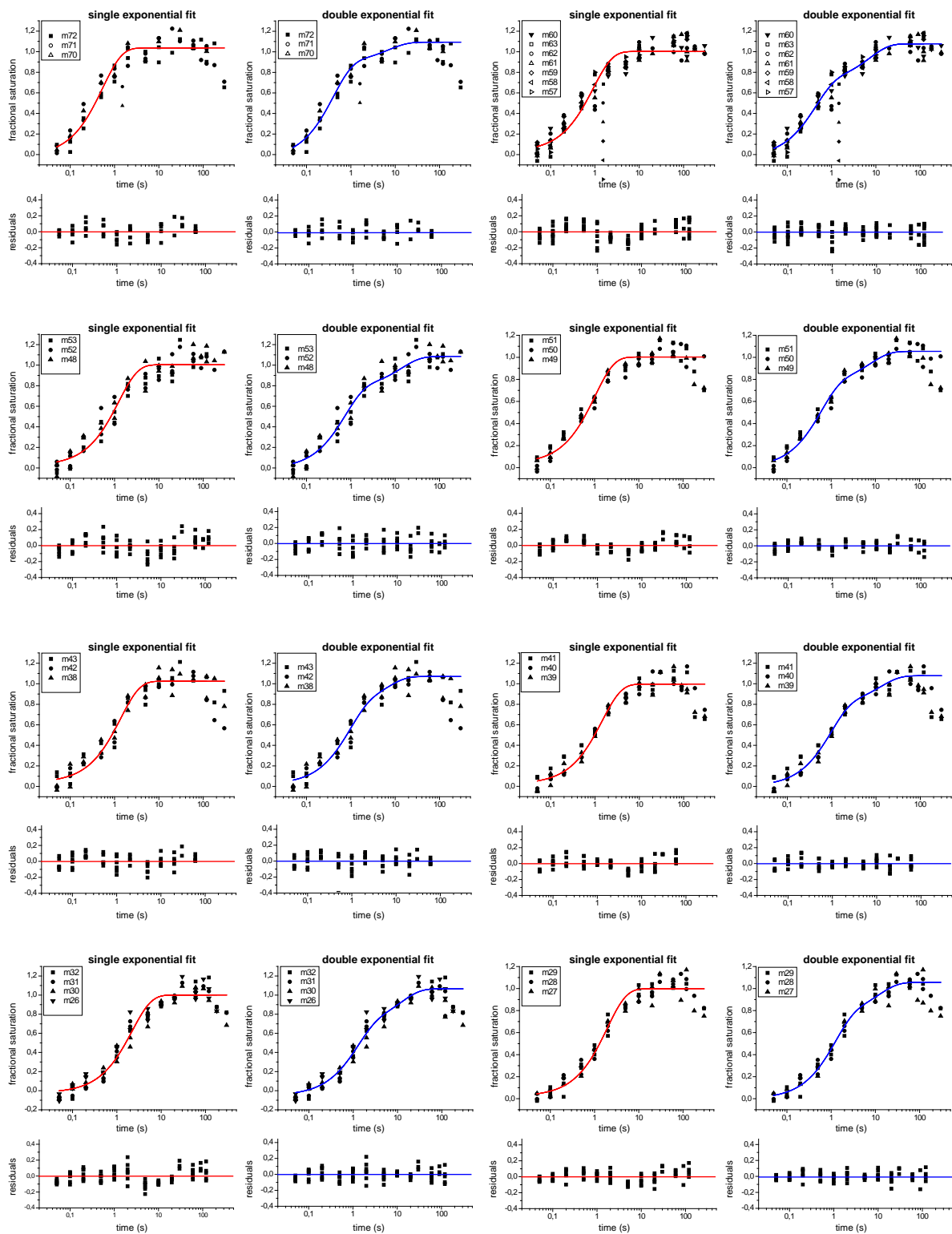
118

Figure 9S (b). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the wild type *T7A1* promoter at 20°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel. In the case then a decrease in fractional saturation was observed at the longer time points, the fits were carried out in the absence of the points in the decreasing portion of the curve.
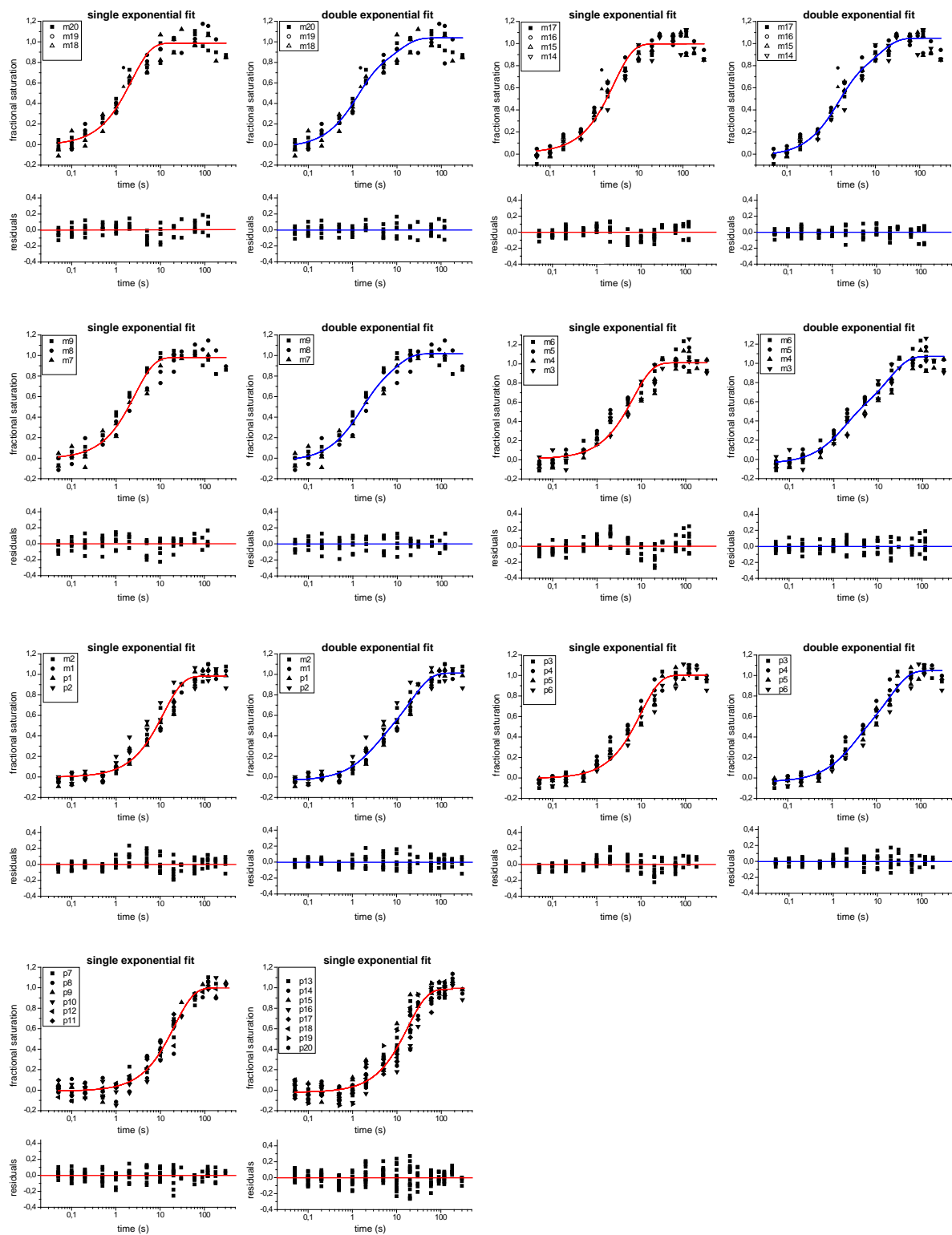
Figure 9S (b) (continuation). Comparison of the fits of the kinetics of the protection appearance at specific sites on the template strand upon *E.coli* RNAP binding to the wild type *T7A1* promoter at 20°C to a single (red curve) and double (blue curve) exponential equation. The residuals are shown at the bottom of each panel.
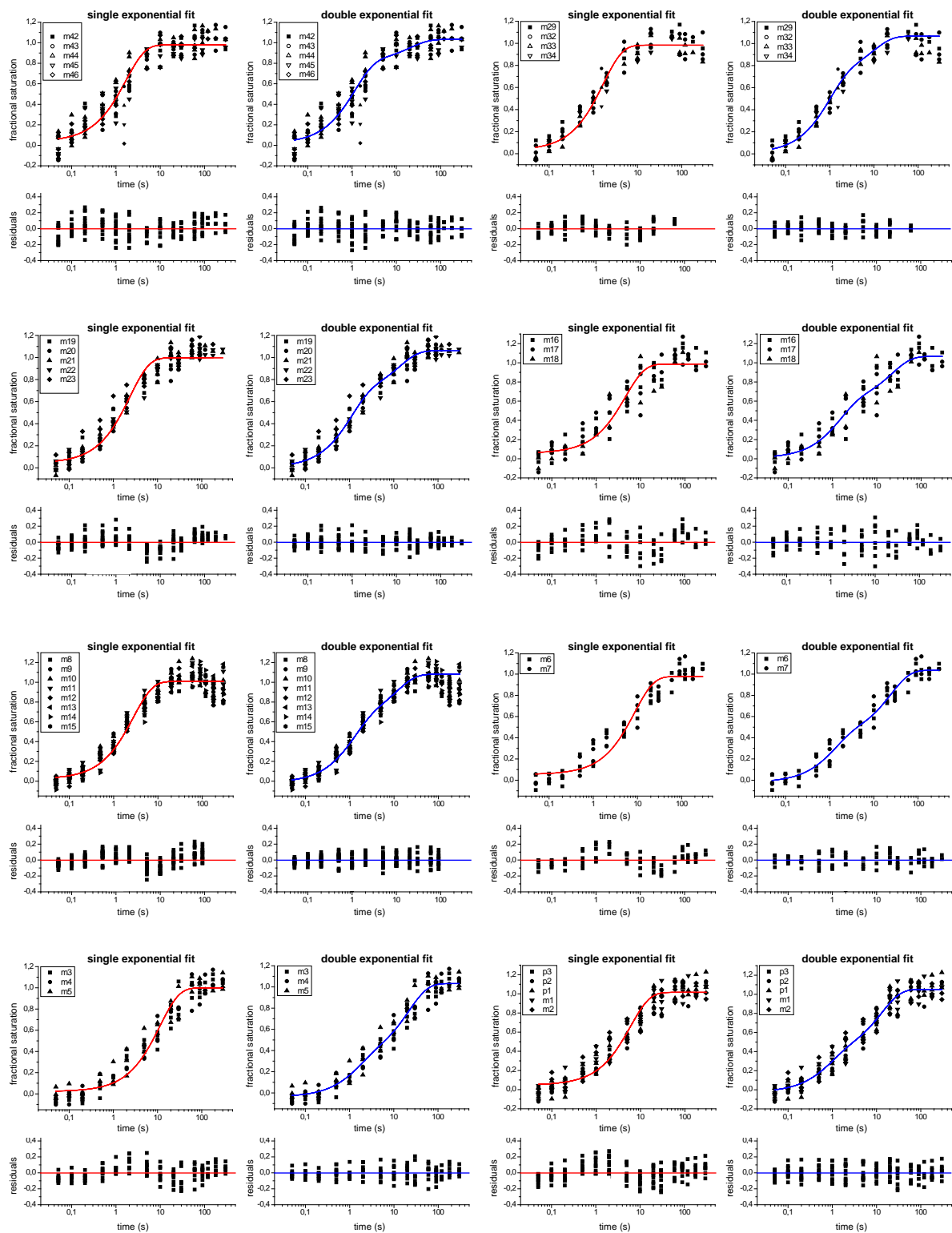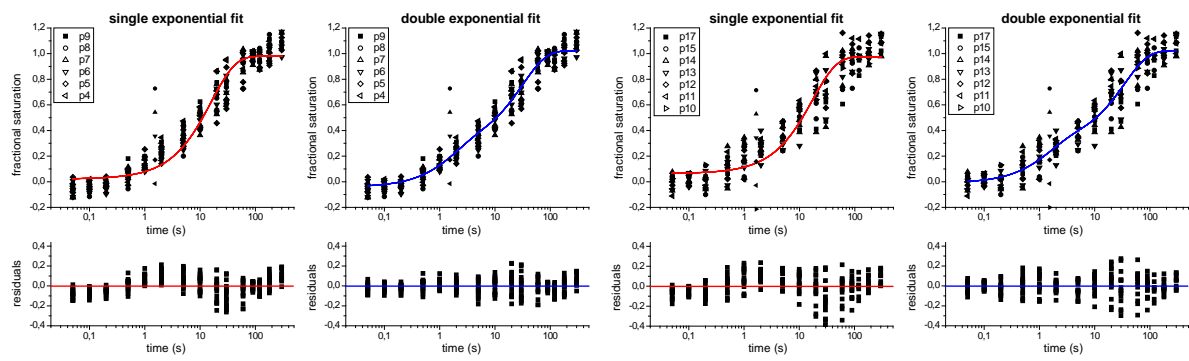
Wild type T7A1 promoter, 20°C, Template strand					
Nucleotide position	**DFn**	**DFd**	**_F_ value**	**_F_ value for P=0.05 (95%)**	**_F_ value for P=0.01 (99%)**
m46-m42	2	160	11.132	3.053	4.740
m34-m29	2	77	15.907	3.115	4.892
m23-m19	2	159	56.896	3.053	4.741
m18-m16	2	101	18.817	3.086	4.822
m15-m8	2	219	85.754	3.037	4.703
m7-m6	2	76	39.588	3.117	4.896
m5-m3	2	130	32.854	3.066	4.772
m2-p3	2	163	40.505	3.051	4.738
p4-p9	2	253	73.947	3.073	4.790
p10-p17	2	228	45.596	3.035	4.699

Wild type T7A1 promoter, 20°C, Non–template strand					
Nucleotide position	**DFn**	**DFd**	**_F_ value**	**_F_ value for P=0.05 (95%)**	**_F_ value for P=0.01 (99%)**
m72-m70	2	50	5.626	3.183	5.057
m63-m57	2	98	37.377	3.089	4.829
m53,m52,m48	2	70	15.116	3.128	4.922
m51-m49	2	70	24.653	3.128	4.922
m43,m42,m38	2	59	5.186	3.153	4.984
m41-m39	2	63	16.808	3.143	4.959
m32-m30,m26	2	81	23.493	3.109	4.877
m29-m27	2	68	18.414	3.132	4.932
m20-m18	2	68	13.554	3.132	4.932
m17-m14	2	95	31.069	3.092	4.836
m9-m7	2	69	8.069	3.130	4.927
m6-m3	2	98	32.205	3.089	4.829
m2-p2	2	105	18.951	3.083	4.813
p3-p6	2	98	24.810	3.089	4.829

Table 5S. _F_ test statistical analysis confirming that the double exponential equation results in a better fit of the data. Shown are the _F_ values for the data sets that were fit to the double exponential equation.

12. References.

Aiyar, S.E., Gourse, R.L. and Ross, W. (1998) Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase α subunit. PNAS. **95**: 14652-14657.

Armitage, B. (1998) Photocleavage of nucleic acids. Chem. Rev. **98**: 1171-1200.

Artsimovitch, I., Svetlov, V., Murakami, K.S. and Landick, R. (2003) Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. J. Biol. Chem. **278**: 12344-12355.

Auner, H., Buckle, M., Deufel, A., Kutateladze, T., Lazarus, L., Mavathur, R., Muskhelishvili, G., Pemberton, I., Schneider, R. and Travers, A. (2003) Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. J. Mol. Biol. **331**: 331-344.

Balasubramanian, B., Pogozelski, W.K. and Tullius, T.D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. PNAS. **95**: 9738-9743.

Barne, K.A., Bown, J.A., Busby, S.J.W. and Minchin, S.D. (1997) Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended -10' motif at promoters. EMBO J. **16**: 4034-4040.

Benoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Bermann, H.M. and Ebright, R.H. (2002) Structural basis of transcription anitiation: the CAP – αCTD – DNA complex. Science. **297**: 1562-1566.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L. and Ebright, R.H. (1994) Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute an independently folded domain capable of dimerization and DNA binding. Cell. **78**: 889-896.

Bown, J.A., Owens, J.T., Meares, C.F., Fujita, N., Ishihama, A., Busby, S.J.w. and Minchin, S.D. (1999) Organization of open complexes at *Escherichia coli* promoters. Location of promoter DNA sites close to region 2.5 of the σ^{70} subunit of RNA polymerase. J. Biol.Chem. **274**: 2263-2270.

Brodolin, K., Mustaev, A., Severinov, K. and Nikiforov, V. (2000) Identification of RNA polymerase β' subunit segment contacting the melted region of the *lacUV5* promoter. J. Biol. Chem. **275**: 3661-3666.

Brodolin, K., Zenkin, N. and Severinov, K. (2005) Remodeling of the σ^{70} subunit non-template DNA strand contacts during the final step of transcription initiation. J. Mol. Biol. **350**: 930-937.

Browning, D.F. and Busby, S. J. W. (2004) The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. **2**: 1-8.

Buc, H. and McClure, W.R. (1985) Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the *lac* UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry **24**: 2712-2723.

Buckle, M., Pemberton, I.K., Jacquet, M-A. and Buc, H. (1999) The kinetics of sigma subunit directed promoter recognition by *E.coli* RNA polymerase. J. Mol. Biol. **285**: 955-964.

Burns, H.D., Ishihama, A. and Minchin, S.D. (1999) Open complex formation during transcription initiation at the *Escherichia coli galP1* promoter: the role of the RNA polymerase α subunit at promoters lacking an UP-element. Nucleic Acids Res. **27**: 2051-2056.

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

Cellai, S., Mangiarotti, L., Vannini, N., Naryshkin, N., Kortkhonjia, E., Ebright, R.H. and Rivetti, C. (2007) Upstream promoter sequences and αCTD mediate stable DNA wrapping within the RNA polymerase – promoter open complex. EMBO reports. **8**: 271-278.

Chlenov, M., Masuda, S., Murakami, K.S., Nikiforov, V., Darst, S.A. and Mustaev, A. (2005) Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase β' subunit. J. Mol. Biol. **353**: 138-154.

Craig, M.L, Tsodikov, O.V., McQuade, K.L, Schlax, P.E., Jr., Capp, M.W., Saecker, R.M. and Record, M.T., Jr. (1998) DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase–promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. J. Mol. Biol. **283**: 741-756.

Craig, M.L., Won-Chul Suh, Record, M.T., Jr. (1995) HO$^\bullet$ and DNase I probing of Eσ^{70} RNA polymerase - λ P$_R$ promoter open complex: Mg^{2+} binding and its structural consequences at the transcription start site. Biochem. **34**: 15624-15632.

Darst, S.A., Opalka, N., Chacon, P., Polykov, A., Richter, C., Zhang, G. and Wriggers, W. (2002) Conformational flexibility of bacterial RNA polymerase. PNAS. **99**: 4296-4301.

Darst, S.A., Polykov, A., Richter, C. and Zhang, G. (1998) Insights into *Escherichia coli* RNA polymerase structure from a combination of X–ray and electron crystallography. J. Struct. Biol. **124**: 115-122.

Davis, C.A., Bingman, C.A., Landick, R., Record, T., Jr. and Saecker, R.M. (2007) Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. PNAS. **104**: 7833-7838.

Davis, C.A., Capp, M.W., Record, M.T., Jr. and Saecker, R.M. (2005) The effect of upstream DNA on open complex formation by *Escherichia coli* RNA polymerase. PNAS. **102**: 285-290.

Dhavan, G.M., Crothers, D.M., Chance, M.R and Brenowitz, M. (2002) Concerted binding and bending of DNA by *Escherichia coli* integration host factor. J. Mol. Biol. **315**: 1027-1037.

Dombroski, A.J. (1997) Recognition of the -10 promoter sequence by a partial polypeptide of σ^{70} *in vitro*. J. Biol. Chem. **272**: 3487-3494.

Ellinger, T., Behnke, D., Knaus, R., Bujard, H. and Gralla, J.D. (1994) Stimulation or inhibition of *Escherichia coli* promoter function. J. Mol. Biol. **239**: 466-475.

Estrem, S.T., Gaal, T., Ross, W. and Gourse, R.L. (1998) Identification of an UP element consensus sequence for bacterial promoters. PNAS. **95**: 9761-9766.

Estrem, S.T., Ross, W., Gaal, T., Chen, Z.W.S., Niu, W., Ebright, R.H. and Gourse, R.L. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. Genes & Development. **13**: 2134-2147.

Fenton, M.S. and Gralla, J.D. (2001) Function of the bacterial TATAAT -10 element as single-stranded DNA during RNA polymerase isomerization. PNAS. **98**: 9020-9025.

Fenton, M.S., Lee, S.J. and Gralla, J.D. (2002) *Escherichia coli* promoter opening and -10 recognition: mutational analysis of σ70. EMBO J. **19**: 1130-1137.

Fredrick, K., Caramori, T., Chen, Ya-Fen, Galizzi, A. and Helmann, J.D. (1995) Promoter architecture in the flagellar regulon of *Bacillus subtilis*: high-level expression of flagellin by the σ^{D} RNA polymerase requires an upstream promoter element. PNAS. **92:** 2582-2586.

Galas, D. and Schmitz, A. (1978) DNase footprinting: a simple method for detecting protein DNA binding specificity. Nucleic Acids Res. **5**: 3157-3170.

Gourse, R.L., Ross, W. and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Mol. Microbiol. **37**: 687-695.

Guo, Y. and Gralla, J.D. (1998) Promoter opening via a DNA fork junction binding activity. PNAS. **95**: 11655-11660.

Hampsey, M. (2001) Omega meets its match. Trends. Genet. **17**: 190-191.

Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C. and Gourse, R.L. (2006) rRNA promoter regulation by nonoptimal binding of σ region 1.2: an additional recognition element for RNA polymerase. Cell. **125**: 1069-1082.

Hawley, D.K. and McClure, W.R. (1982) Mechanism of activation of transcription initiation from the λP_{RM} promoter. J. Mol. Biol. **157**: 493-525.

Heyduk, E., Kuznedelov, K., Severinov, K. and Heyduk, T. (2006) A consensus adenine at position -11 of the nontemplate strand of bacterial promoter is important for nucleation of promoter melting. J. Biol. Chem. **281**: 12362-12369.

Hsu, L.M, Vo, N.V., Kane, C.M. and Chamberlin, M.J. (2003) *In vitro* studies of transcription initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. Biochem. **42**: 3777-3786.

Ishihama, A. (2000) Functional modulation of *Escherichia coli* RNA polymerase. Annu. Rev. Microbiol. **54**: 499-518.

Ishihama, A. (1988) Promoter selectivity of prokaryotic RNA polymerase. TIG. **4**: 282-286.

Johnson, R.S. and Chester, R.E. (1998) Stopped-flow kinetic analysis of the interaction of *Escherichia coli* RNA polymerase with the bacteriophage T7 A1 promoter. J. Mol. Biol. **283**: 353-370.

Joo, D.M., NG, N. and Calendar, R. (1997) A σ^{32} mutant with a single amino acid change in the highly conserved region 2.2 exhibits reduced core RNA polymerase affinity. PNAS. **94**: 4907-4912.

Kontur, W.S., Saecker, R.M., Davis, C.A., Capp, M.W. and Record, M.T.Jr. (2006) Solute probes of conformational changes in open complex (RP_o) formation by *Escherichia coli* RNA polymerase at the λP_R promoter: evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to RP_o. Biochem. **45**: 2161-2177.

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

Lee, D.J., Busby, S.J.W. and Lloyd, G.S. (2003) Exploitation of a chemical nuclease to investigate the location of the *Escherichia coli* RNA polymerase α subunit C-terminal domains at simple promoters that are activated by cyclic AMP receptor protein. J. Biol. Chem. **278**: 52944-52952.

Li, X-Y. and McClure, W.R. (1998) Characterization of the closed complex intermediate formed during transcription initiation by *Escherichia coli* RNA polymerase. J. Biol. Chem. **273**: 23549-23557.

Lisser, S. and Margalit, H. (1993) Compilation of *E.coli* mRNA promoter sequences. Nucleic Acids Research. **21**: 1507-1516.

Liu, M., Tolstorukov, M., Zhurkin, V., Garges, S. and Adhya, S. (2004) A mutant spacer sequence between -35 and -10 elements makes the P_{lac} promoter hyperactive and cAMP receptor protein-independent. PNAS. **101**: 6911-6916.

Łoziński, T. and Wierzchowski, K.L. (2003) Inactivation and destruction by $KMnO_4$ of *Escherichia coli* RNA polymerase open transcription complex: recommendations for footprinting experiments. Anal. Biochem. **320**: 239-251.

Maleknia, S.D., Ralston, C.Y., Brenowitz, M.D., Downard, K.M. and Chance, M.R. (2001) Determination of macromolecular folding and structure by synchrotron X-ray radiolysis techniques. Anal. Biochem. **289**: 103-115.

Malhotra, A. and Severinova, E. (1996) Crystal structure of a σ^{70} subunit fragment from *E.coli* RNA polymerase. Cell. **87**: 127-136.

McKane, M. and Gussin, G.N. (2000) Changes in the 17 bp spacer in the P_R promoter of bacteriophage λ affect steps in open complex formation that precede DNA strand separation. J. Mol. Biol. **299**: 337-349.

McKane, M., Malone, C. and Gussin, G.N. (2001) Mutations at position -10 in the λ P_R promoter primarily affect conversion of the initial closed complex (RP_C) to a stable, closed intermediate (RP_i). Biochemistry. **40**: 2023-2031.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Renyakin, A., Kapanidis, A.N., Niu, W., Edright, Y.W., Levy, R. and Ebright, R.H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and polymerase-promoter open complex. Cell. **108**: 599-614.

Mooney, R.A., Darst, S.A. and Landick, R. (2005) Sigma and RNA polymerase: an on-again, off-again relationship? Mol. Cell **20**: 335-345.

Motulsky, H. J. and Christopoulos, A. (2003) Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting. GraphPad Software Inc., San Diego CA.

Moyle, H. and Susskind, M.M. (1989) A mutant *Escherichia coli* σ^{70} subunit of RNA polymerase with altered promoter specificity. J. Mol. Biol. **206**: 579-590.

Murakami, K.S. and Darst, S.A. (2003) Bacterial RNA polymerase: the wholo story. Current Opinion in Structural Biology. **13**: 31-39.

Murakami, K.S., Masuda, S. and Darst, S.A. (2002a) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science. **296**: 1280-1284.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. (2002b) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science. **296**: 1285-1290.

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.

Nechaev, S., Chlenov, M. and Severinov, K. (2000) Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. J. Biol. Chem. **275**: 25516-25522.

Nickerson, C.A. and Achberger, E.C. (1995) Role of curved DNA in binding of *Escherichia coli* RNA polymerase to promoters. J. Bacter. **177**: 5756-5761.

Ozoline, O.N., Tsyganov, M.A. (1995) Structure of open promoter complexes with *Escherichia coli* RNA polymerase as revealed by the DNase I footprinting technique: complication analysis. Nucleic Acids Res. **23**: 4533-4541.

Paul, B.J., Ross, W., Gaal, T. and Gourse, R.L. (2004) rRNA transcription in *Escherichia coli*. Annu. Rev. Genet. **38**: 749-770.

Rao, L., Ross, W., Appleman, A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., Jr. and Gourse, R.L. (1994) An "extended" promoter with an upstream element that dramatically increases promoter strength. J. Mol. Biol. **235**: 1421-1435.

Rippe, K., Guthold, M., von Hippel, P.H. and Bustamante, C. (1997) Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of *E.coli* RNA polymerase · σ^{54} holoenzyme by scanning force microscopy. J. Mol. Biol. **270**: 125-138.

Rivetti, C., Guthold, M. and Bustamante, C. (1999) Wrapping of DNA around the *E.coli* RNA polymerase open promoter complex. EMBO J. **18**: 4464-4475.

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.

Roe, J.H., Burgess, R.R. and Record, M., Jr. (1985) Temperature dependence of the rate constants of the *Escherichia coli* RNA polymerase-lambda PR promoter interaction. Assignment of the kinetic steps corresponding to protein conformational change and DNA opening. J. Mol. Biol. **184**: 441-453.

Rosenberg, S., Kadesch, T.R. and Chamberlin, M.J. (1982) Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of the rate of open complex formation at T7 promoter A1. J. Mol. Biol. **155**: 31-51.

Ross, W., Aiyar, S.E., Salomon, J. and Gourse, R.L. (1998) *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. J. Bacter. **180**: 5375-5383.

Ross, W., Ernst, A. and Gourse, R.L. (2001) Fine structure of *Escherichia coli* RNA polymerase – promoter interactions: α subunit binding to the UP element minor groove. Genes & Development. **15**: 491-506.

Ross, W. and Gourse, R.L. (2005) Sequence-independent upstream DNA – αCTD interactions strongly stimulate *Escherichia coli* RNA polymerase – lacUV5 promoter association. PNAS. **102**: 291-296.

Ross, W., Schneider, D.A., Paul, B.J., Mertens, A. and Gourse, R.L. (2003) An intersubunit contact stimulating transcription initiation by *E.coli* RNA polymerase: interaction of the α C-terminal domain and σ region 4. Genes & Development. **17**: 1293-1307.

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 modules of the association of *E.coli*

σ^{70} RNA polymerase with the λP_R promoter: large scale conformational changes in forming the kinetically significant intermediates. J. Mol. Biol. **319**: 649-671.

Sakata-Sogawa, K. and Shimamoto, N. (2004) RNA polymerase can track a DNA groove during promoter search. PNAS. **101**: 14731-14735.

Schickor, P., Metzger, W., Werel, W., Lederer, H. and Heumann, H. (1990) Topography of intermediates in transcription of *E.coli*. EMBO J. **9**: 2215-2220.

Schroeder, L.A., Gries, T.J., Saecker, R.M., Record, M.T., Jr, Harris, M.E. and deHaseth, P.L. (2009) Evidence for tyrosine-adenine stacking interaction and for a short-lived open intermediate subsequent to initial binding of *Escherichia coli* RNA polymerase to promote DNA. J. Mol. Biol. **385**: 339-349.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R. and Brenowitz, M. (1997) Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. J. Mol. Biol. **266**: 144-159.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M. and Brenowitz, M. (1998) Following the folding of RNA with time-resolved synchrotron X-ray footprinting. Methods Enzimology. **295**: 379-402.

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. PNAS. **102**: 4706-4711.

Severinov, K. and Darst, S.A. (1997) A mutant RNA polymerase that forms unusual open promoter complex. PNAS. **94**: 13481-13486.

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

Shin, M.-C. and Gussin, G.N. (1983) Mutations affecting two different steps in transcription initiation at the phage λP_{RM} promoter. PNAS. **80**: 496-500.

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

Sigman, D.S., Mazumder, A. and Perrin, D.M. (1993) Chemical nucleases. Chem. Rev. **93**: 2295-2316.

Strainic, M.G., Jr., Sullivan, J.J., Velevis, A. and deHaseth, P.L. (1998) Promoter recognition by *Escherichia coli* RNA polymerase: effects of UP element on open complex formation and promoter clearance. Biochem. **37**: 18074-18080.

Susa, M., Kurobi, T. and Shimamoto, N. (2006) A pathway branching in transcription initiation in *Escherichia coli*. Mol. Microbiol. **59**: 1807-1817.

Susa, M., Sen, R. and Shimamoto, N. (2002) Generality of the branched pathway in transcription initiation by *Escherichia coli* RNA polymerase. J. Biol. Chem. **277**: 15407-15412.

Tagami, H. and Aiba, H. (1999) An inactive open complex mediated by an UP element at *Escherichia coli* promoters. PNAS. **96**: 7202-7207.

Tang, Y., Murakami, K., Ishihama, A. and deHaseth, P.L. (1996) Upstream interactions at the lambda P_{RM} promoter are sequence nonspecific and activate the promoter to a lesser extent than an introduced UP element of an rRNA promoter. J. Bacter. **178**: 6945-6951.

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* σ^{70} in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complex. J. Biol. Chem. **276**: 31891-31896.

Travers, A. and Muskhelishvili, G. (2005) DNA supercoiling – a global transcriptional regulator for enterobacterial growth? Nat. Rev. Microbiol. **3**: 157-169.

Tullius, T.D. (1987) Chemical 'snapshots' of DNA: using the hydroxyl radical to study the structure of DNA and DNA-protein complexes. TIBS. **12**: 297-300.

Vassylyev, D.G., Seklne, S.-I., Laptenko, O., Lee, J., Vassylyeva, 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.

Vo, N.V., Hsu, L.M, Kane, C.M. and Chamberlin, M.J. (2003) *In vitro* studies of transcription initiation by *Escherichia coli* RNA polymerase. 3. Influence of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. Biochem. **42**: 3798-3811.

Vuthoori, S., Bowers, C.W., McCracken, A., Dombroski, A.J. and Hinton, D.M. (2001) Domain 1.1 of the σ^{70} subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. J. Mol. Biol. **309**: 561-572.

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.

Zaychikov, E., Denissova, L., Meier, T., Götte, M. and Heumann, H. (1997) Influence of Mg^{2+} and temperature on formation of the transcription bubble. J. Biol. Chem. **272**: 2259-2267.

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

13. Abbreviations.

AFM	Atomic force microscopy
ATP	Adenosine 5'–triphosphate
cryo–EM	Cryo–electron microscopy
αCTD	α Carboxy–terminal domain
CTP	Cytosine 5'–triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
***Ec*βDR**	*E.coli* β dispensable region
***Ec*β'GNCD**	*E.coli* β' G non–conserved domain
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility–shift assay
ESRF	European Synchrotron Radiation Facility
EtdBr	Ethidiumbromide
FeBABE	Iron (S)–1–(p–bromoacetamidobenzyl)–EDTA
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
GTP	Gyanosine 5'–triphosphate
HhH	Helix–hairpin–helix
HPLC	High–performance liquid chromatography
LD	Linker domain
mRNA	Messenger ribonucleic acid
αNTD	α Amino–terminal domain
NTS	Non–template strand
OD	Optical density
•OH	Hydroxyl radical
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonylfluoride
RNA	Ribonucleic acid
RNAP	RNA polymerase
rNTPs	Ribonucleoside 5'–triphosphates
rRNA	Ribosomal RNA
SBHM	Sandwich–barrel hybrid motif
SDS	Sodium dodecyl sulfate
SFM	Stopped–flow module
Taq	*Thermus aquaticus*
TEC	Ternary elongation complex
Tris	Tris[hydroxymethyl]aminomethan
tRNA	Transfer RNA
TS	Template strand
Tth	*Thermus thermophillus*
β'ZBD	β' Zinc–binding domain

Acknowledgments.

CURRICULUM VITAE

Anastasia Rogozina

Max-Planck Institute of Biochemistry   Phone: (49) (89) 8578-2217
Am Klopferspitz 18A   FAX: (49) (89) 8578-2822
D-82152 Martinsried, Germany   E-mail: rogozina_nastya@mail.ru

PERSONAL INFORMATION

Date and Place of Birth	July 11, 1978, Svirsk, Russia
Nationality	Russian
Citizenship	Russian
Status in Germany	Unlimited stay permission
Family status	Single

EDUCATION BACKGROUND

1995-2001	Novosibirsk State University
	Faculty of Natural Sciences
	Department of Chemistry
	Specialization in Biochemistry
	Novosibirsk, Russia
1999	Bachelor degree
2001	Master degree
1994–1995	Specialized Educational Scientific Centre of
	Novosibirsk State University
	Novosibirsk, Russia
1985–1994	Elementary and High School
	Svirsk, Russia

RESEARCH EXPERIENCE

2001-present	Laboratory of Dr. Hermann Heumann
	Max-Planck-Institute of Biochemistry
	Martinsried, Germany
1999-2001	Laboratory of Bioorganic Chemistry of Enzymes
	Institute of Bioorganic Chemistry
	Russian Academy of Sciences
	Novosibirsk, Russia
1998-1999	Laboratory of Immobilized Enzymes
	Institute of Bioorganic Chemistry
	Russian Academy of Sciences
	Novosibirsk, Russia

PUBLICATION

Bianca Sclavi, Evgeny Zaychikov, <u>Anastasia Rogozina</u>, Ferdinand Walther, Malcolm Buckle, and Hermann Heumann. (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 USA.* **102**; 4706-4711.

Rashid O. Anarbaev, <u>Anastasia L. Rogozina</u>, Olga. I. Lavrik. (2009) DNA polymerase β reveals enhanced activity and processivity in reverse micelles. *Biophysical Chemistry.* **141**; 11-20.

Rashid O. Anarbaev, <u>Anastasia L. Rogozina</u>, Olga. I. Lavrik. (2009) DNA polymerase β reveals enhanced stability in reverse microemulsions. *Journal of Molecular Catalysis B: Enzymatic*. **59**; 64-69.

<u>Anastasia Rogozina</u>, Evgeny Zaychikov, Malcolm Buckle, Hermann Heumann, Bianca Sclavi. (2009) DNA melting by RNA polymerase at the T7A1 promoter precedes the rate limiting step at 37°C and results in the accumulation of an off-pathway intermediate. *Nucleic Acids Research*. **37**; 5390-5404.

München, am 17.07.2009

Anastasia Rogozina