Sequence Determinants for the Recognition of the Fork Junction DNA Containing the −10 Region of Promoter DNA by *E. coli* RNA Polymerase†

Dennis L. Matlock and Tomasz Heyduk*

E. A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, 1402 South Grand Boulevard, St. Louis, Missouri 63128

Received June 22, 2000; Revised Manuscript Received August 9, 2000

ABSTRACT: It has been recently suggested that *E. coli* RNA polymerase can specifically recognize a fork junction DNA structure, suggesting a possible role for such interaction in promoter DNA melting [Guo, Y., and Gralla, J. D. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11655−11660]. We have determined here quantitatively, using a site-specific binding assay, the effects of base substitutions within the conserved −10 hexamer in the context of a short fork junction DNA on binding to RNA polymerase. Adenine at position −11 and thymine at position −7 were found to be critical for sequence-specific recognition of the DNA. The identities of bases at positions −9 and −8 were found to be not important for the binding whereas replacement of bases at positions −12 and −10 had a mild negative effect on the binding affinity. It was found that for the binding of fork DNA to RNA polymerase, specific sequence recognition was more important than specific recognition of fork junction DNA structure. The pattern of relative importance of bases in the −10 region for binding RNA polymerase was generally consistent with the sequence conservation pattern observed in nature where positions −11 and −7 are the most conserved. Binding experiments with a series of adenine analogues at position −11 revealed that the N1 nitrogen of adenine was a critical determinant for the preference of the adenine at this position, suggesting a mechanism for the nucleation of promoter DNA melting initiation in which RNA polymerase destabilizes duplex DNA by directly competing with the thymine of the A-T base pair for hydrogen bonding to the N1 position of the −11 nontemplate strand adenine.

Transcription initiation in *E. coli* is performed by the multisubunit RNA polymerase holoenzyme (RNAP)[1] with a subunit composition of $\alpha_2\beta\beta'\sigma$ (*1*). Under various growth conditions, RNAP is able to utilize different σ subunits. However, under normal growth conditions, σ^{70} is the predominant σ subunit found (*2*). Previous genetic and biochemical data have revealed a consensus structure for the σ^{70}-RNAP core promoter consisting of two conserved hexamers centered around positions −10 and −35 from the transcription start position and separated by a 17 base pair (bp) linker (*3*). Naturally occurring promoters which have been characterized have been found to have varying degrees of departure from the consensus sequence and exhibit a broad range of promoter strength (*4*, *5*). Therefore, it is reasonable to suggest that sequence-dependent variability of the interactions between RNA polymerase and promoter DNA plays an important role as a determinant of the promoter strength.

The σ^{70} subunit within RNA polymerase holoenzyme is responsible for sequence-specific recognition of promoter DNA (*6*). This recognition involves two main steps: initial recognition of −35 and −10 promoter elements in the ds DNA form and a unimolecular isomerization reaction to form an open complex in which the duplex DNA is partially melted to expose the template strand. The role of the σ^{70} subunit in initial recognition of promoter DNA has been generally accepted since the discovery of the protein (*7*). Recent studies suggested that this subunit actively participates in the promoter melting reaction as well (*8*). A likely mechanism for the role of σ^{70} subunit in melting promoter DNA could be to provide favorable interaction with promoter DNA in the conformation found in the open complex to facilitate isomerization between the closed and open complex. The characteristic feature of promoter DNA in the open complex is a fork junction between ds and ss DNA. Several lines of evidence suggest that the σ^{70} subunit is involved in specific recognition and interaction with the fork junction structure in the open complex. First, this subunit binds single-stranded oligonucleotides corresponding to the nontemplate strand ss portion of the transcription bubble (*9*−*13*). The binding was shown to be sequence-specific and required binding σ to the core polymerase (*14*). These oligonucleotides could be cross-linked to the σ subunit in a sequence-specific manner (*12*, *13*). Second, the site-specific cross-linking between promoter DNA and polymerase revealed that σ subunit was cross-linked to the promoter DNA region corresponding to the fork DNA junction of the open complex (*15*). Additional cross-linking in this region was found to occur to the β' subunit, suggesting that the σ and β' subunits collaborate in interaction with this region of promoter DNA

† This work was supported by NIH Grant GM50514.

* Corresponding author. E-mail: heydukt@slu.edu. Phone: (314) 577-8152. Fax: (314) 577−8156.

[1] Abbreviations: RNAP, DNA-directed RNA polymerase holoenzyme; bp, base pair; ds, double stranded; ss, single stranded; A, adenine; T, thymine; C, cytosine; G, guanine; DTPA−AMCA, diethylenetriaminepentaacetic acid−7-amino-4-methylcoumarin; S9, spacer 9 phosphoramidite; wt, wild-type.

10.1021/bi001433h CCC: $19.00
Published on Web 09/15/2000

(*15*, *16*). Finally, Guo and Gralla have shown recently that RNA polymerase can form a heparin-resistant complex with short DNA molecules containing fork junction DNA (*17*). This interaction required σ subunit, and mutagenesis of several residues in this subunit affected the interaction (*8*). Thus, it appears that σ^{70} subunit may provide critical interactions required for the initiation of sequence-specific melting of promoter DNA.

We present here a quantitative analysis of DNA sequence determinants for the interaction of RNA polymerase with fork junction DNA containing the −10 promoter element. We show that adenine at position −11 and thymine at position −7 of the nontemplate strand are essential for the binding of fork junction DNA. We further show that the N1 position of the −11 adenine of the nontemplate strand is critical for the recognition of adenine at this position.

MATERIALS AND METHODS

Materials. Cy5 monosuccinimidyl ester (*18*) was purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Diethylenetriaminepentaacetic acid−7-amino-4-methylcoumarin−maleimide (DTPA−AMCA−maleimide, the luminescent donor used in LRET experiments) was prepared in our laboratory as previously described (*19*). The reagents (to include phosphoramidites for normal DNA bases and the various adenine analogues) for oligonucleotide synthesis were purchased from Glen Research (Sterling, VA). Core RNAP was purified from *E. coli* K12 cells (University of Alabama fermentation facility) using the method of Burgess and Jendrisak (*20*) as modified by Polyakov et al. (*21*) and Hager et al. (*22*)

Oligonucleotides. The oligonucleotides used for this study were synthesized on an Applied Biosystems (Foster City, CA) model 394 DNA synthesizer using standard phosphoramidite chemistry. The sequence of the consensus oligonucleotide was 5′-TAACCGCCAGAGGTAA(S9)-TTACCTCTGGCGGTTATAATGGTP-3′ where (S9) represents the position at which a spacer 9 phosphoramidite (Glen Research) was incorporated during synthesis. This spacer was used to increase the stability of a short DNA duplex by facilitating hairpin formation (see Figure 1). Computer modeling (not shown) indicated that this spacer would allow hairpin formation without perturbation of the DNA duplex. The underlined sequence is the nontemplate −10 region (−12 to −7 positions), and the consensus sequence is shown. A total of 41 such oligonucleotides were synthesized which included the 1 consensus sequence, the 18 possible single-base substitutions, 10 selected double-base substitutions in the −10 region, and 12 sequences which contained adenine analogues at the −11 nontemplate position. The 3′-phosphate group was incorporated during synthesis to facilitate labeling with an acceptor fluorophore (described below). When base substitutions at −12 or −11 of the nontemplate strand were made, the corresponding bases in the template strand were also mutated to preserve Watson−Crick base-pairing and to prevent end-fraying. Thus, with the exception of some experiments with adenine analogues, the experiments involved testing −12 and −11 positions in the context of duplex DNA whereas −10, −9, −8, and −7 were tested in the context of ss DNA. Following synthesis and purification, the oligonucleotides were heated to 95 °C and allowed to cool to room temperature (in order to facilitate duplex formation). The final fork junction DNA product was a duplex from the −26 to −11 positions and single-stranded (nontemplate strand) from the −10 to −4 positions (Figure 1). The correct structure of the fork junction DNA after annealing was confirmed by gel electrophoresis and by S1 nuclease digestion experiments (not shown). Also, DNA melting experiments were performed in which the hairpins used in this study exhibited a single cooperative transition with the melting temperature which was much higher compared to the melting temperature of a corresponding DNA duplex (without spacer 9 linker) and which was independent of DNA concentration. These melting studies thus further confirmed the DNA hairpin structure. Additionally, the observation of a single cooperative melting transition indicated that the incorporation of the spacer 9 linker did not affect the ability to form a correct duplex in the upstream portion of the hairpin DNA.

For the binding experiments, the hairpin/fork oligonucleotide which contained a consensus −10 hexamer sequence (TATAAT) was labeled at the −4 position of the single-stranded (nontemplate) overhang with the acceptor fluorophore, Cy5. Labeling the oligonucleotide with the acceptor fluorophore was performed by converting the 3′-phosphate group to an ethylenediamine derivative followed by a reaction with Cy5 monosuccinimidyl ester as previously described (*23*, *24*). For the stopped-flow kinetic experiments, the DNA fork and the corresponding nontemplate ss DNA were labeled with 7-diethylaminocoumarin-3-carboxylic acid succinimidyl ester (Molecular Probes, Eugene, OR) in the same manner as with Cy5.

RNAP Holoenzyme Labeled with (Eu^{3+})DTPA−AMCA−Maleimide. A single cysteine mutant of σ^{70} ([A59C]σ^{70}) was expressed and purified as previously described with the exception that a 10 mL Q Sepharose column was used instead of a Resource Q column (*14*, *19*). [A59C]σ^{70} was used to prepare RNAP holoenzyme labeled with DTPA−AMCA−maleimide at a specific site. Briefly, purified [A59C]σ^{70} was denatured using 6 M guanidine hydrochloride and incubated with a ∼2−3-fold excess of DTPA−AMCA−maleimide. Excess label was removed by the use of a G-50 spin column. [A59C]σ^{70} was subsequently refolded by dialysis. The entire labeling procedure has been described elsewhere (*25*). $EuCl_3$ was added to DTPA−AMCA-[A59C]σ^{70} at a 1:1 ratio. Holoenzyme was reconstituted by mixing (Eu^{3+})DTPA−AMCA-[A59C]σ^{70} and core polymerase at a final ratio of 1.5:1, respectively, and incubating at 4 °C for 30 min followed by purification on a Superdex 200 HR sizing column (Pharmacia) equilibrated in 50 mM HEPES (pH 7.9), 100 mM KCl, and 10 μM EDTA.

Luminescence Resonance Energy Transfer (LRET) Competition Binding Measurements. Fork junction DNA binding experiments were performed at 25 °C in 50 mM HEPES (pH 7.9), 10 mM $MgCl_2$, 10 μM EDTA, 0.05 μg/mL BSA, 1 mM DTT, 250 mM KCl, 2.8% poly(ethylene glycol) (molecular weight 8000), and 6.0 ng/μL double-stranded poly(dI·dC) (Amersham). The BSA, DTT, and poly(dI·dC) were added fresh prior to each experiment. The relative binding affinity of fork junction DNA molecules with base substitutions in the −10 region was determined by their ability to compete with an equimolar amount of Cy5-labeled fork junction DNA containing the consensus −10 sequence

(TATAAT). The final concentrations of Cy5-labeled and unlabeled competitor fork DNA were 15 nM each. RNAP holoenzyme containing (Eu^{3+})DTPA−AMCA-[A59C]σ^{70} was added to the reaction mixture last at a final concentration of 5 nM. After a 5 min incubation period, the sensitized acceptor emission at 670 nm was determined using pulsed nitrogen laser excitation at 337 nm. Sensitized acceptor emission was measured on a laboratory-built two-channel instrument described previously (*26*). The time-resolved sensitized acceptor signal resulting from 1000 excitation pulses was accumulated, and the total sensitized acceptor signal was calculated by integration of the time-resolved signal from 60 to 1000 μs. The total sensitized acceptor signal was normalized by dividing by the total initial donor signal (measured simultaneously with the sensitized acceptor signal) to correct for the variability due to the fluctuation of laser pulse power or due to a nonspecific quenching of the donor signal. The total initial donor signal was calculated as a sum of donor decay amplitudes obtained by the fitting of donor decay curves to a double exponential decay function. The data presented represent the results of at least three independent experiments.

The binding of the consensus fork junction DNA under our experimental conditions appeared to be tight (Figure 2), preventing accurate determination of absolute values of the equilibrium binding constants. However, the relative affinities of fork DNA molecules containing mutations in the −10 region could be determined from the measurement of the relative degree of saturation of RNA polymerase with labeled consensus fork and unlabeled competitor when both were present in the binding reaction. Under these conditions:

$$\nu_{cons} = K_{cons}[DNA_{cons}]/(1 + K_{cons}[DNA_{cons}] + K_{mut}[DNA_{mut}]) \quad (1)$$

$$\nu_{mut} = K_{mut}[DNA_{mut}]/(1 + K_{cons}[DNA_{cons}] + K_{mut}[DNA_{mut}]) \quad (2)$$

$$\nu_{mut}/\nu_{cons} = K_{mut}[DNA_{mut}]/K_{cons}[DNA_{cons}] \quad (3)$$

where ν_{cons} is the degree of saturation of RNAP with consensus DNA in the presence of mutant DNA, ν_{mut} is the degree of saturation of RNAP with mutant DNA in the presence of consensus DNA, [DNA_{cons}] and [DNA_{mut}] are the free concentrations of consensus and mutant DNA, respectively, and K_{cons} and K_{mut} are the equilibrium association constants for consensus and mutant DNA, respectively. Since [DNA_{cons}] = [DNA_{cons}]$_{tot}$ − ν_{cons}[RNAP]$_{tot}$ and [DNA_{mut}] = [DNA_{mut}]$_{tot}$ − ν_{mut}[RNAP]$_{tot}$, eq 3 can be rewritten:

$$\{\nu_{mut}([DNA_{cons}]_{tot} - \nu_{cons}[RNAP]_{tot})\}/\{\nu_{con}([DNA_{mut}]_{tot} - \nu_{mut}[RNAP]_{tot})\} = K_{mut}/K_{cons} \quad (4)$$

where [DNA_{cons}]$_{tot}$ and [DNA_{mut}]$_{tot}$ are the total concentrations of consensus and mutant DNA, respectively, and [RNAP]$_{tot}$ is the total concentration of RNA polymerase. Thus, measurement of ν_{cons} and ν_{mut} allows determination of the ratio of equilibrium binding constants for the consensus and mutant DNA from which the free energy loss or gain due to a mutation can be calculated:

$$\Delta\Delta G_{cons-mut} = \Delta G_{cons} - \Delta G_{mut} = -RT \ln K_{cons} + RT \ln K_{mut} = RT \ln (K_{mut}/K_{cons}) \quad (5)$$

Under our experimental conditions, ν_{cons} and ν_{mut} could be easily determined because when the Cy5-labeled consensus fork DNA was present at a final concentration of 15 nM, RNA polymerase (5 nM, final) was 100% saturated (Figure 2). Thus, the LRET signal in the absence of unlabeled competitor (F_o) corresponds to 100% saturation with Cy5-labeled consensus fork (i.e., corresponds to $\nu_{cons} = 1$). In the presence of unlabeled competitor:

$$\nu_{cons} = F_{mut}/F_o \text{ and } \nu_{mut} = 1 - \nu_{cons} \quad (6)$$

where F_{mut} is the signal observed in the presence of unlabeled competitor. To eliminate possible effects of the fluorescence probe on the affinity of Cy5-labeled consensus fork DNA, the competition with unlabeled consensus fork DNA was included in the experiments, and all $\Delta\Delta G_{cons-mut}$ values were calculated using unlabeled consensus fork DNA as the reference.

Stopped-Flow Measurements. These measurements were performed on an Aminco Bowman Series 2 Luminescence Spectrometer equipped with a stopped-flow accessory (Spectronic Instruments, Rochester, NY). The binding of 7-diethylaminocoumarin-3-carboxylic acid-labeled fork junction DNA or nontemplate ss DNA to RNA polymerase was initiated by rapid mixing of equal volumes of DNA (5 nM, final) and RNA polymerase (15−100 nM, final). A change of fluorescence emission at 472 nm (excitation at 432 nm) was monitored as a function of time at 25 °C in the same buffer as used in DNA fork binding experiments. For both the fork junction and single-stranded DNA experiments, the curves from at least five independent experiments were averaged and analyzed using SCIENTIST (Micromath, Salt Lake City, UT) according to the following kinetic equation:

$$F = A^*[1 - \exp(-k^*t)] + C \quad (7)$$

where F is the observed fluorescence of 7-diethylaminocoumarin-3-carboxylic acid-labeled DNA at time t, A and k are the amplitude and observed rate constant, respectively, and C is the fluorescence intensity at $t = 0$.

RESULTS

Site-Specific Assay for Fork DNA Binding. Our main goal in this study was to quantitatively determine the relative contribution of each position within the −10 region of promoter DNA on the binding of a simple model fork junction DNA. The model fork junction DNA we used for the experiments presented herein is shown in Figure 1. The sequence is based on the λP_R promoter and extends from the −26 to −4 positions. The fork DNA is duplex from −26 to −11 with single-stranded nontemplate strand from the −10 to the −4 position. This DNA did not contain a −35 region. Thus, the effects of substitutions in the −10 region were studied here in the absence of possible correlations of such mutations with the recognition of the −35 element. This DNA contained only the upstream fork junction and only the nontemplate strand ss DNA extension since these were shown to be the most important determinants for the binding to polymerase (*17*). The junction between duplex and ss

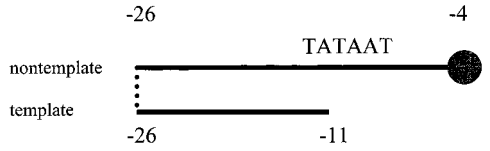

FIGURE 1: Schematic representation of fork junction DNA which was used for this study. The consensus sequence for the nontemplate strand is depicted. The last base pair is at the −11 position which results in a single-stranded overhang that extends from the −10 to −4 positions on the nontemplate strand. The hairpin was formed by incorporation of a Spacer 9 phosphoramidite (Glen Research) during synthesis immediately following the −26 position. The position of the fluorescent probe is indicated (filled circle).

DNA was between the −11 and −10 positions since it was shown by chemical and enzymatic probing that the transcription bubble in the open complex extends up to position −11 or −10 (*27*, *28*). Furthermore, it was shown that specific fork DNA binding was the strongest with a junction at either −12 or −11 (*17*).

A difficulty in binding studies of model short DNA molecules to RNA polymerase is that in addition to the σ^{70} subunit the other subunits are also capable of binding different DNA molecules. Since we wanted to determine the effect of the −10 hexamer sequence on binding the σ^{70} subunit within RNA polymerase, it was necessary to develop a binding assay that would report DNA bound specifically to the σ^{70} subunit. The σ^{70} subunit-specific binding assay was based on luminescence resonance energy transfer (LRET) (*29, 30*) between donor-labeled σ^{70} subunit reconstituted to the RNAP holoenzyme and acceptor-labeled fork DNA (Figure 2A). In this assay energy transfer can occur only upon formation of a complex between Cy5-labeled fork junction DNA and the (Eu^{3+})DTPA−AMCA-[A59C]σ^{70} subunit (when present within RNAP holoenzyme). The limited range of resonance energy transfer *(29, 30)* implies that this was effectively a σ^{70} subunit-specific binding assay and that possible nonspecific complexes formed with other subunits of the polymerase should not contribute to the signal. Based upon quantitative measurement of energy transfer (not shown), we estimate that the distance between residue 59 of σ^{70} and the polymerase-bound fork junction DNA was ∼43 Å, which was well within the distance allowed by the dimensions of σ^{70}. Energy transfer can be conveniently measured by sensitized acceptor emission (i.e., emission of the acceptor upon excitation of the donor). The use of the (Eu^{3+})DTPA−AMCA−Cy5 donor−acceptor pair allows the measurement of sensitized acceptor emission with very little background using pulsed excitation and time-gated sensitized acceptor intensity measurement (*23*). As a control experiment, we incubated the (Eu^{3+})DTPA−AMCA-[A59C]σ^{70} subunit with Cy5-labeled fork DNA (−10 region, TATAAT) in the absence of the other subunits of RNA polymerase (data not shown). Upon excitation of the donor, the observed emission of the acceptor was essentially that of the background signal. This indicated that the σ^{70} subunit, in the absence of the other subunits, was unable to bind fork junction DNA. This is in agreement with previous observations that specific recognition of ss nontemplate strand oligonucleotides required core−σ^{70} interaction (*14*).

Figure 2B shows a titration of (Eu^{3+})DTPA−AMCA-[A59C]σ^{70} holoenzyme with Cy5-labeled fork junction DNA containing the consensus −10 element. The sensitized acceptor signal increased with the increase of DNA concentration, and the signal reached essentially a plateau at ∼3-fold excess of DNA. This suggests that under the experimental conditions used in this experiment the binding was tight (almost stoichiometric).

Effects of Single-Base Substitutions on Fork Junction DNA Binding. To measure the effects of a base substitution in the −10 region, competition binding experiments were performed in which the ability of fork junction DNA's with base substitutions in the −10 region to compete with the binding of Cy5-labeled consensus fork junction DNA was measured. Under the experimental conditions used (saturation of the polymerase holoenzyme with the Cy5-labeled consensus fork junction DNA), the relative binding affinity of fork junction DNA containing a mutant −10 sequence could be determined from the measurement of the relative degree of saturation of polymerase with both DNA molecules (see Materials and Methods). Figure 3 summarizes the results of such competition experiments with all 18 possible variants of single-base-substituted fork junction DNA molecules. The first three bars in the upper panel illustrate the methodology. In the absence of labeled consensus fork junction DNA, only a very small background signal was observed (bar "1"), while in the presence of labeled consensus fork junction DNA efficient sensitized acceptor was detected (bar "2"), illustrating an excellent signal-to-noise ratio of the methodology employed. As expected, in the presence of an equimolar amount of unlabeled consensus fork junction DNA ∼50% decrease of sensitized acceptor signal was observed (bar "TATAAT"), suggesting ∼50% decrease in saturation of polymerase with labeled DNA. The equimolar concentrations of a corresponding full duplex DNA (i.e., DNA identical to the one shown in Figure 1 except that the bases in template strand between positions −10 and −4 were added) did not compete at all (not shown). Also, the "anticonsensus" DNA fork (in which all A's and T's in the −10 region were mutated to G' and C's, respectively) was a very poor competitor (not shown). We estimate that the free energy difference for the binding of consensus vs anticonsensus fork was on the order of ∼4.0 kcal/mol.

The competition experiments using fork junction DNA with base substitutions at every position within the conserved −10 hexamer revealed that single-base substitutions at the −11 or −7 position were the most deleterious to binding by RNAP holoenzyme. Base substitutions at positions −12 and −10 had a small negative effect on the affinity whereas base substitutions at positions −9 and −8 had essentially an insignificant effect on the binding. The sensitized acceptor emission intensities shown in the upper panel were used to calculate the free energy loss (or gain) due to the mutation (see Materials and Methods), and the calculated $\Delta\Delta G°$ values are shown in the lower panel. The above discussed pattern (i.e., the importance of positions −11 and −7) is more clearly seen in this representation of the data. The most deleterious single-base substitutions at the −11 and −7 positions resulted in a free energy difference of about 0.8 and 1 kcal/mol, respectively. This value is roughly comparable to the free energy difference due to a loss of a single protein−DNA hydrogen bond (*31*). There seem to be no simple rules regarding the relationship between the identity of the base replacing the consensus base and the effect on the binding (for example, there is no preference for replacing purine with

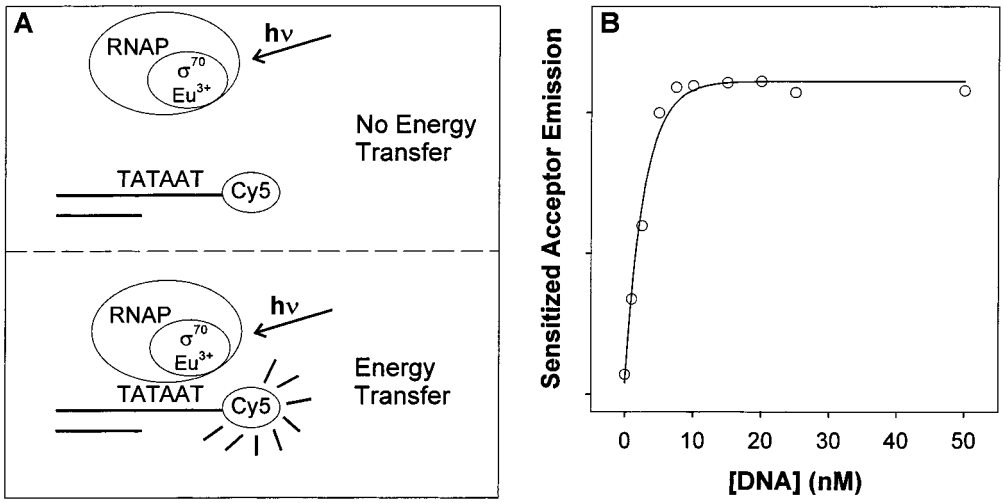

FIGURE 2: (A) Design of luminescence resonance energy transfer (LRET) competition binding assay. (B) Saturation of donor-labeled RNA polymerase holoenzyme (RNAP) with acceptor-labeled fork junction DNA that contains a consensus −10 sequence (TATAAT). The concentration of RNAP was 5 nM. Based on these results, 15 nM DNA was used for all subsequent competition experiments.

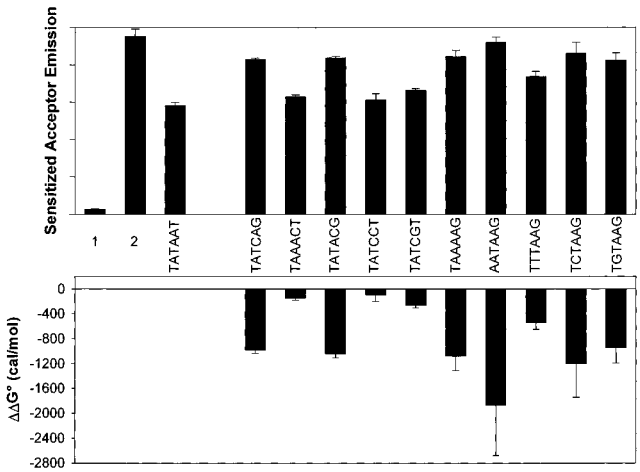

FIGURE 3: Effect of a single-base substitution in the −10 hexamer region of promoter DNA on binding RNAP. Upper panel: Normalized acceptor intensity for each single-base substitution. For each experiment, the concentrations of each component were as follows: donor-labeled RNAP, 5 nM; acceptor-labeled fork junction DNA, 15 nM; unlabeled competitor fork junction DNA (where indicated), 15 nM. Donor-labeled RNAP was added last to the mixture and incubated for 5 min at the indicated temperature prior to measuring the acceptor intensity. The control experiments at the far left were as follows: bar "1", donor-labeled RNAP only (5 nM, final); bar "2", donor-labeled RNAP (5 nM, final) + acceptor-labeled DNA (15 nM, final); bar "TATAAT", donor-labeled RNAP (5 nM, final) + acceptor-labeled consensus DNA (15 nM, final) + unlabeled consensus DNA (15 nM, final). Lower panel: Free energy calculations from the results obtained in (A). Calculations were performed as described in the text. Positive values indicate a more preferred −10 sequence (over the consensus TATAAT).

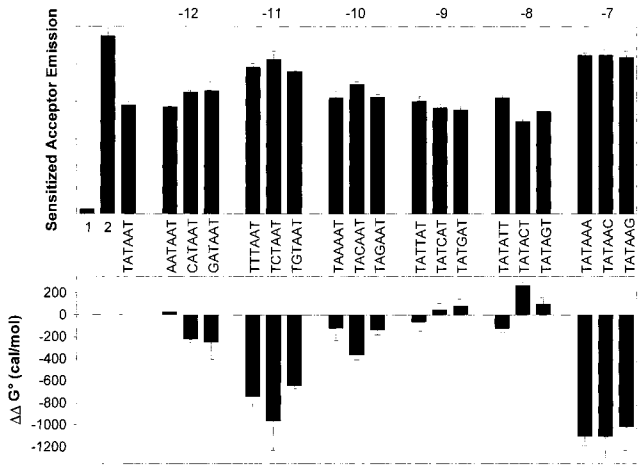

FIGURE 4: Results of selected double-base substitutions in the −10 hexamer region on binding RNAP. The upper and lower panels are as in Figure 3.

purine etc.). Based on the data shown in Figure 3, we synthesized a DNA fork (the −10 sequence: GCCTTC) which should exhibit the weakest affinity for binding of the polymerase. The free energy difference for the binding of this fork DNA in comparison to consensus DNA was ∼4.5 kcal/mol, which is consistent with the theoretical predictions (such as, for example, that this DNA should be a worse binder compared to the "anticonsensus" DNA).

To test the additivity of the effects of the individual base substitutions, we next synthesized a limited set of fork junction oligonucleotides which contained double-base substitutions in the −10 hexamer region, and we determined their effect on binding to RNAP via the same competition assay. The observed acceptor fluorescence intensity values when these double substitution mutants were used as competitors and their resulting free energy differences are shown in Figure 4. Several double mutants exhibited simple additivity of the effects observed in corresponding single-base substitutions. For example, TATCAG mutant consisting of base substitution at position −9 (no effect on binding) and at position −7 (1.0 kcal/mol loss of free energy of binding) exhibited ∼1.0 kcal/mol loss of free energy, consistent with summation of the effects of single-base substitutions. However, some of the double-base substitutions showed lack of such simple additivity. The most apparent example is the AATAAG mutant which contains a base substitution at position −12 (minimum negative effect on binding) and a base substitution at position −7 (1.0 kcal/mol loss of free energy of binding). The combined effect of both substitutions was ∼1.9 kcal/mol, which is much more than expected from the summation of the effects of both single-base substitutions. Interestingly, it was observed previously that some promoter mutations can be context-dependent. For example, mutations at position −12 and −8

were found to have a grossly nonadditive effect on the transcription from P22 promoter *(32)*.

Comparison between Fork Junction DNA and Single-Stranded (Nontemplate Strand) DNA Binding. To determine the relative contribution of sequence-specific recognition of fork DNA and the general preference of polymerase for binding DNA molecules containing a junction between single-stranded and double-stranded DNA, we quantitatively compared the affinities for binding consensus fork DNA and the corresponding nontemplate single-stranded DNA (i.e., ss oligonucleotide corresponding to the top strand of the fork DNA shown in Figure 1). Also, the affinities of fork DNA and corresponding ss DNA in which all A's and T's in the −10 region were mutated to G' and C's, respectively ("anticonsensus" DNA), were also compared. The competition binding experiments (not shown) revealed that in both of the above cases the DNA fork molecules displayed a slight preference over the corresponding ss oligonucleotides. The $\Delta\Delta G(ss_{cons}/fork_{cons})$ was ∼0.1 kcal/mol whereas $\Delta\Delta G(ss_{anti}/fork_{anti})$ was ∼0.15 kcal/mol.

To further probe possible mechanistic differences in the association of RNA polymerase with fork versus single-stranded DNA, the kinetics of binding of consensus fork and consensus ss DNA were examined. For these experiments, we labeled each oligonucleotide with 7-diethylaminocoumarin-3-carboxylic acid succinimidyl ester at the −4 position of the nontemplate strand. This probe was placed at exactly the same position as the Cy5 probe used in the LRET competition binding experiments. This fluorescent probe (i.e., 7-diethylaminocoumarin-3-carboxylic acid), when covalently attached to promoter DNA in the vicinity of the −10 region, has previously been shown to be able to report the initial binding of promoter DNA by RNAP (*33*). Upon addition of wt RNAP to 7-diethylaminocoumarin-3-carboxylic acid-labeled fork DNA or ss DNA, a ∼50% increase in fluorescence intensity was observed (not shown). The time course of this fluorescence change upon rapid mixing of RNA polymerase with the 7-diethylaminocoumarin-3-carboxylic acid-labeled fork DNA is shown in Figure 5 (inset). The observed change in fluorescence intensity was described well by a single-exponential kinetic equation. The time-dependent fluorescence changes observed with the ss DNA (not shown) were very similar and could also be described by a simple first-order kinetic equation. Plotting the observed rate constants as a function of polymerase concentration revealed a simple linear dependence between these two parameters, suggesting a simple kinetic scheme:

$$k_{obs} = k_1[\mathrm{RNAP}] + k_{-1} \quad (8)$$

where k_{obs} is the observed rate constant, k_1 is the rate constant for the binding of DNA to polymerase, and k_{-1} is the rate constant for the dissociation of protein−DNA complex. The value of k_1 was determined from the slopes of plots shown in Figure 5. The calculated bimolecular association rate constants for fork and ss (nontemplate) DNA were (the values in parentheses correspond to 95% confidence intervals) $k_{1(fork)} = 5.00 \times 10^6\ M^{-1}\ s^{-1}$ (4.98×10^6; 5.02×10^6) and $k_{1(ss)} = 4.21 \times 10^6\ M^{-1}\ s^{-1}$ (4.19×10^6; 4.23×10^6), respectively. The values of k_{-1} (intercept) could not be determined accurately since they were much smaller than the rate of the forward reaction. However, the rate of dissociation for the fork and ss DNAs seem to be very similar. Taken together, these results suggest that RNA polymerase exhibits a very slight preference for binding the fork DNA over ss DNA and this preference is the result of a faster rate of association of the fork DNA compared to the ss DNA.

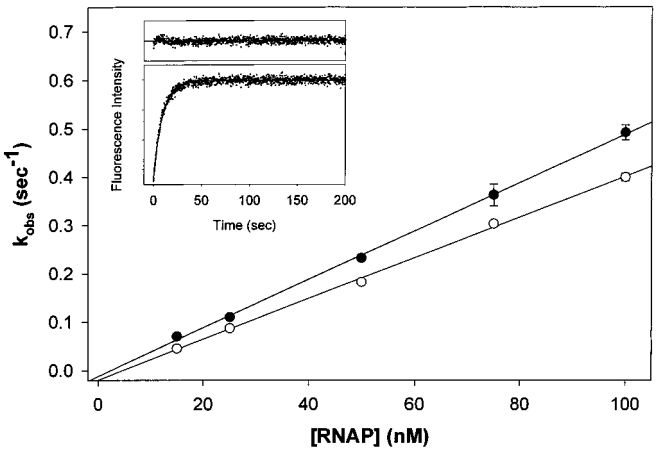

FIGURE 5: Comparison of the binding of fork junction DNA and single-stranded DNA. Both DNA species contained the consensus −10 hexamer region (TATAAT). The only difference is the structural context in which it is presented. Filled circles, fork junction DNA; open circles, single-stranded DNA. See text for kinetic parameters. Each data point (for fork and single-stranded DNA experiments) is the result of five individual experiments. The error bar for each point is plotted; however, in most instances it is smaller than the size of the plotted symbol. The inset depicts a representative fit to a single-exponential kinetic equation in which fork DNA was used (25 nM RNAP, 5 nM DNA).

Determinants of the Recognition of Adenine at Position −11. Having determined that a base substitution at the −11 or −7 position was deleterious to binding by RNAP, we were interested in the nature of the selectivity for particular bases at these positions. We were particularly interested in the interactions with the adenine at the −11 position as this position is widely accepted to be at or about the first position which undergoes strand separation leading to an open complex from which productive transcription initiation can proceed (*34*).

The approach taken to address this question was to synthesize a series of DNA forks with the consensus sequence (TATAAT) and with a number of adenine analogues incorporated into the −11 nontemplate position. The use of base analogues to address which functional groups are important in the T7 RNA polymerase promoter has been previously reported (*35*). Figure 6 shows the analogues which were used in these studies. Since some of these analogues would perturb or even eliminate base-paring of the base analogue at position −11 in the nontemplate strand with thymine in the template strand, the binding studies with adenine analogues were conducted using fork DNA with the ds/ss DNA boundary between positions −12 and −11. It has been shown previously that a fork with the last base pair at position −11 or at position −12 could bind polymerase with a similar affinity (*17*). Figure 7A shows a comparison of the competition binding experiments with these two fork DNA's. The affinity of both DNA's for binding RNA polymerase was indeed very similar. As an additional control, we incorporated cytosine at the −11 position in the nontemplate strand of a fork DNA that had its last base pair at

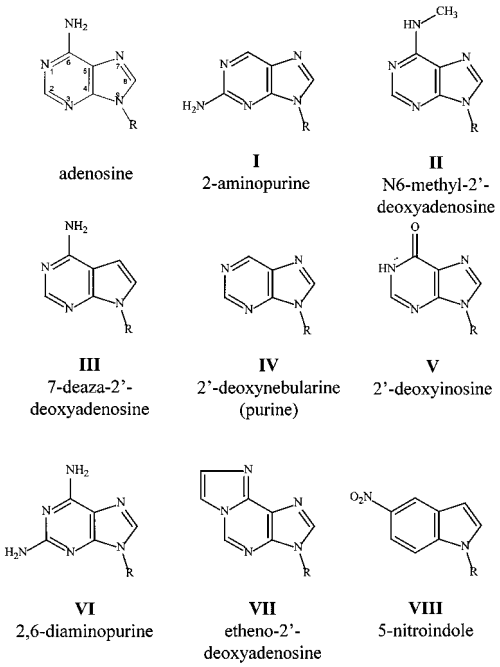

FIGURE 6: Adenine analogues that were incorporated into the −11 nontemplate position.

the −12 position. We did so because in the single-base substitution experiments shown in Figure 3 (in which the fork junction DNA had its last base pair at the −11 position), cytosine at the −11 nontemplate position was determined to be deleterious to binding by RNAP. As shown in Figure 7A, cytosine (−11 nontemplate) in the context of DNA with a fork junction at −12 was also deleterious to binding. We thus concluded that the DNA fork with the last base pair at position −12 was an appropriate model for the studies with adenine analogues at the −11 nontemplate position.

Adenine can be specifically recognized by a protein through formation of hydrogen bond(s) between the protein residue(s) and any of the potential hydrogen bond donors or acceptors present in adenine. The DNA fork with 5-nitroindole (VIII), a base analogue which was shown to be capable of stacking interactions similar to purine bases (*36*) but devoid of any hydrogen bond donors or acceptors present in adenine, bound RNA polymerase very poorly (Figure 7B). The adenine analogues used (Figure 6) were designed to eliminate or alter individual hydrogen bond donors or acceptors of the adenine at the −11 nontemplate position.

(i) N7. The nitrogen at position N7 is selectively eliminated in compound III. DNA fork with analogue III at position −11 binds RNA polymerase with an affinity similar to consensus fork DNA (Figure 7A). We concluded that nitrogen at position N7 was not essential for recognition of adenine at position −11.

(ii) N6. The N6 amino group is selectively eliminated in compound IV, and the DNA fork with this analogue at position −11 was able to bind polymerase with an affinity similar to the consensus fork (Figure 7A). Also, methylation of the N6 amino group (compound II) did not significantly reduce the affinity of the fork DNA for binding RNA polymerase (Figure 7B). We concluded that the N6 amino group was not essential for recognition of adenine at position −11.

(iii) N1. None of the commercially available adenine analogues perturb only the N1 position. However, the results of binding experiments with compounds I, V, VI, and VII when taken together allow the conclusions regarding the importance of this position. Analogues in which position N1 is covalently modified (VII) or in which its hydrogen bonding nature is changed from hydrogen bond acceptor to hydrogen bond donor (V) exhibited a severe defect in binding (Figure 7B). In both of these analogues, in addition to perturbation of position N1, the N6 amino group was either eliminated (V) or covalently modified (VII), which complicated analysis of these results. However, we have previously determined (see above) that the N6 amino group is dispensable for the binding. Thus, we conclude that the N1 nitrogen of adenine at position −11 is the major determinant for specific recognition of this base by RNA polymerase. In view of this conclusion, the observation that an additional amino group at position 2 (I and VI) perturbs somewhat the binding can be explained by steric hindrance imposed by an additional group in the nearest vicinity of the functional N1 position.

(iv) N3. None of the available analogues perturbed the N3 position. Thus, the available data cannot rule out the role of this position in specific recognition of adenine at position −11. However, perturbations of the N1 position result in a loss of binding affinity similar to that obtained when adenine is replaced with C, T, G, or VIII, suggesting that if N3 is important for specific recognition of adenine its role is probably minor compared to the N1 position.

The results of binding experiments with DNA forks with duplex terminating at positions −11 and −12 (Figure 7A) showed very little difference in affinity between fork junction DNA in which the adenine at position −11 was or was not base-paired. To further investigate the role of base-pairing of the base at the −11 nontemplate position, the binding experiments were performed with 2-aminopurine (I) and 2,6-diaminopurine (VI) incorporated into this position in fork junction DNA's whose duplex portions terminated at −11 or −12 (Figure 7C). Our rationale for using these analogues was that they were still able to form hydrogen bonds with thymine present in the template strand. The 2-aminopurine (I) is able to form two hydrogen bonds with thymine, but one of them is different than in a canonical Watson−Crick A·T base pair. 2,6-Diaminopurine can form three hydrogen bonds with thymine, producing a more stable base pair in comparison to the two hydrogen bonds normally present in an A·T base pair. The results of these binding experiments showed that within experimental error the affinity of the fork DNA with these analogues was very similar regardless of whether the base at position −11 was base-paired.

DISCUSSION

In this study, we have determined the relative importance of each position within the −10 hexamer sequence of promoter DNA (within the context of a fork junction DNA) for the binding to RNA polymerase holoenzyme. This fork junction DNA most likely mimics interactions involved in the formation of the open complex. We find that adenine at position −11 and thymine at position −7 of the nontemplate strand are the most important for sequence-specific binding to RNA polymerase. The replacement of consensus bases at these positions with any other base resulted in a loss of ∼1 kcal/mol. Figure 8 compares the base preference for binding the fork junction DNA determined from our data with the

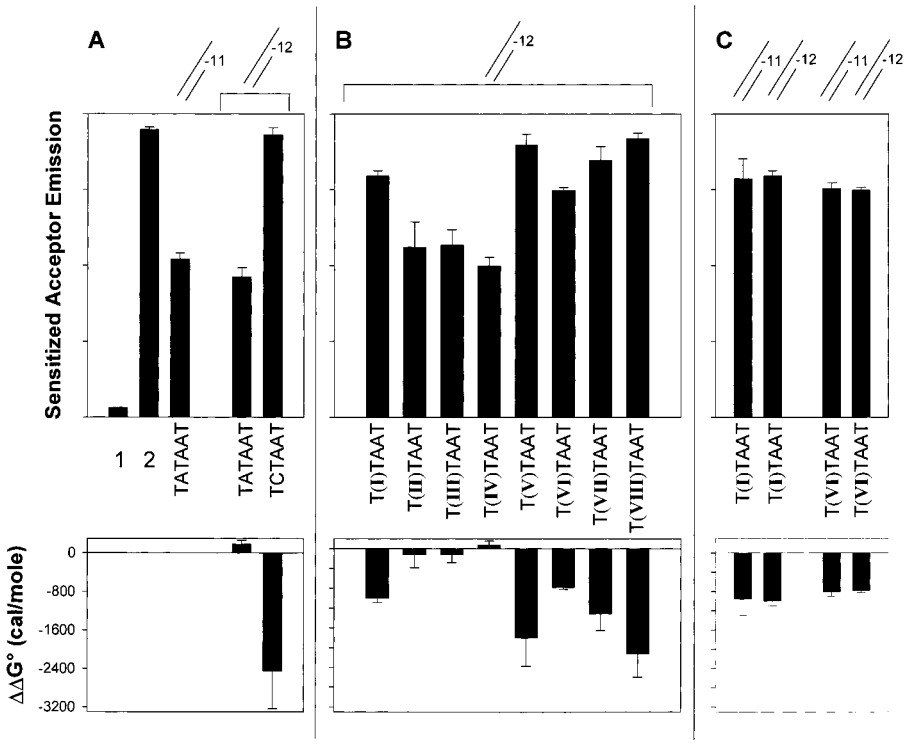

FIGURE 7: Results of using various adenine analogues at the −11 nontemplate position. The concentrations used were 5 nM donor-labeled RNAP, 15 nM acceptor-labeled consensus fork DNA, 15 nM unlabeled mutant fork DNA. (A) Ability of RNAP to bind with comparable affinity to model fork DNA when the last base pair is at the −11 or −12 position. Lanes 1 and 2 are as in Figure 3. (B) Effect of various adenine analogues (see Figure 6 for analogue structure) at the −11 nontemplate position on binding to RNAP. (C) Effect of selected adenine analogues at the −11 nontemplate position in fork DNA (whose duplex portion ended at either the −12 or the −11 position) on binding to RNAP.

Sequence conservation/genetics:	**TA**T**AA**T
ssDNA binding (Qiu & Helmann (1999) NAR 27, 4541-4546)	**TA**T**AA**T
fork DNA binding	**TA**T**AA**T

FIGURE 8: Comparison of base sequence preferences in the −10 hexamer region for RNAP-σ^{70} promoters as determined by sequence alignment analysis (*4*), ssDNA competition binding as determined by electrophoretic gel mobility shift assay (*36*), and fork DNA competition binding using a σ^{70} subunit specific luminescence resonance energy transfer methodology.

sequence conservation pattern in the −10 region and with previous studies in which a base preference for the binding of ss oligonucleotides corresponding to the −10 region of the nontemplate strand was determined (*37*). The base preference for binding RNA polymerase which has emerged from our data corresponds very closely to the sequence conservation of bases in the −10 region observed in nature. Positions −11 and −7 are the most conserved and are the most important for the binding whereas positions −9 and −8 exhibit the lowest degree of sequence conservation and appear to be not essential for the binding. Interestingly, quite different results were obtained when the base preference for binding ss oligonucleotides was studied [(*37*), Figure 8]. The −11 nontemplate position was also found to be important in these studies. However, thymine at position −7 was not essential whereas the identity of bases at positions −8 and −9 was critical. The experiments with fork junction DNA containing both −10 and −35 hexamers also did not suggest the important role of −7 T (*17*). On the other hand, recent kinetic experiments (*38*) suggested that the T at −7 is important for binding ss oligonucleotides. These apparent differences in base preferences observed for different DNA substrates may reflect different methodologies employed in these studies which could have emphasized different aspects of DNA recognition by the polymerase. For example, the affinities of fork junction DNA determined in this work were measured in solution using an equilibrium technique whereas binding of −10 and −35 element-containing fork junction DNA was determined using heparin-resistance of the complexes *(38)*. Heparin-resistance experiments may have emphasized the dissociation rate of the complexes whereas the solution equilibrium approach reported overall equilibrium binding properties. On the other hand, the observed base preference differences could be a reflection of differences in the structure between ss DNA, −10-containing fork junction DNA, and −10- and −35-containing fork junction DNA. This may have interesting mechanistic implications. One possibility could be that the ds/ss DNA junction is required for a proper presentation of the nontemplate strand in single-stranded form for sequence-specific interactions with RNA polymerase. In the absence of ds/ss junction, some "out of register" interactions would be possible, resulting in an apparent difference in base preference pattern observed with ss oligonucleotides. In this context, it is interesting to note that gross deviations from simple additivity observed with some double mutants (Figure 4) appear to involve mutants in which one of the mutations was in the duplex portion and the other mutation was in the single-stranded portion of the fork junction DNA. Double mutants in which both mutations were in a single-stranded portion of the fork junction DNA exhibited reasonable additivity of the effects of individual mutations (Figure 4). Such pattern of nonadditivity would be consistent with the described above function

of the duplex portion of the fork junction DNA.

It has been suggested previously that RNA polymerase has an ability to specifically recognize fork DNA containing a junction between ss and ds DNA (*17*). Thus, in addition to sequence-specific recognition, there should also be a DNA structure-specific component to the recognition of fork junction DNA. The comparison of the binding affinity between fork DNA and corresponding ss nontemplate strand oligonucleotide suggested only a very small preference for the binding of the fork DNA. This small preference was observed regardless if DNA contained consensus −10 sequences or if the −10 element was replaced with a nonspecific sequence. The difference in affinity was small enough that it could be explained by some trivial effects and not necessarily by a selective recognition of ss/ds DNA junction structure. For example, the fork DNA containing a stretch of duplex DNA should exhibit a higher degree of monovalent cation condensation compared to the corresponding ss oligonucleotide (*39*). Under the same buffer conditions, the binding of polymerase to the fork DNA would thus displace a higher amount of condensed cations compared to the ss oligonucleotide, making DNA−polymerase complex formation thermodynamically more favorable in the case of the fork DNA. A small difference between the binding affinity of fork junction DNA vs ss DNA suggests that for the specific recognition of fork DNA the sequence-specific interactions are much more important than recognition of the ss/ds DNA junction structure. As already mention, the role of the fork junction is probably not to enhance the overall affinity for the binding to the polymerase, but maybe to ensure a proper presentation of the single-stranded nontemplate DNA for sequence-specific interactions with the polymerase.

Adenine at position −11 of the nontemplate strand is one of the most conserved bases in the −10 region, is essential for the specific binding of ss nontemplate strand oligonucleotides, and is essential for specific binding of fork DNA (Figure 8). All of these observations underscore a special role of this residue in transcription initiation. It has been proposed that the probable role of this residue is in the nucleation of transcription bubble formation [ref (*40*) and references cited therein]. We have found that the N1 position of this residue is essential for its specific recognition by RNA polymerase. The N1 position is involved in Watson−Crick hydrogen bonding in the A-T base pair. Therefore, a simple mechanism for the nucleation of melting can be envisioned in which RNA polymerase would directly compete with the thymine of the A-T base pair for hydrogen bonding to the N1 position of adenine in the −11 nontemplate position. The critical role of the N1 position of adenine at −11 is consistent with the previously proposed mechanism of melting nucleation involving a "base-flipping" reaction of the −11 adenine followed by base-specific interaction with the polymerase (*41*). Specific interaction with the thymine at position −7 could function as a specific "check-point" to ensure downstream extension of duplex melting.

Specific recognition of the N1 position of −11 adenine suggests that fork DNA molecules in which adenine at position −11 is not base-paired should bind polymerase more tightly compared to DNA forks in which position −11 is base-paired. However, our results showed essentially no difference between fork DNA with the duplex DNA portion terminating at −11 or −12 regardless if the base at −11 was adenine (Figure 7), 2-aminopurine (Figure 7C), or 2,6-diaminopurine (Figure 7C). Our interpretation of this observation is that this is probably due to a very low stability of a terminal A-T base pair resulting in a low energy cost of disrupting the adenine N1−thymine N3 hydrogen bond in order to allow an adenine N1−polymerase interaction.

In summary, our results provide additional support for the role of sequence-specific interactions between RNA polymerase and the −10 region of promoter DNA in a partially ss conformation during the process of promoter melting. The discovery that the N1 position of adenine at the −11 nontemplate position is specifically recognized by RNA polymerase provides an important clue to the understanding of the nucleation of promoter melting.

ACKNOWLEDGMENT

D.L.M. thanks Dr. Craig T. Martin for a helpful discussion pertaining to the adenine analogues

REFERENCES

1. Chamberlin, M. (1976) *RNA Polymerase* (Losick, R., and Chamberlin, M., Eds.) pp 17−67, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
2. Ishihama, A. (1990) *Adv. Biophys. 26*, 19−31.
3. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 784−788.
4. Harley, C. B., and Reynolds, R. P. (1987) *Nucleic Acids Res. 15*, 2343−2361.
5. Lisser, S., and Margalit, H. (1993) *Nucleic Acids Res. 21*, 1507−1516.
6. Gralla, J. D. (1990) *Methods Enzymol. 185*, 37−54.
7. Burgess, R. R., Travers, A. A., Dunn, J. J., and Bautz, E. K. (1969) *Nature 221*, 43−46.
8. Fenton, M. S., Lee, S. J., and Gralla, J. D. (2000) *EMBO J. 19*, 1130−1137.
9. Savinkova, L. K., Baranova, L. V., Knorre, V. L., and Salganik, R. I. (1998) *Mol. Biol. (Engl. Transl.) 22*, 651−656.
10. Savinkova, L. K., Sokolenko, A. A., Kel, A. E., Tolukhonov, I. I., Kumarev, V. P., Baranova, V. P., Rar, V. A., and Salganik, R. I. (1996) *Mol. Biol. (Engl. Transl.) 30*, 188−191.
11. Severinova, E., Severinova, K., Fenyö, D., Marr, M., Brody, E. N., Roberts, J. W., Chait, B. T., and Darst, S. A. (1996) *J. Mol. Biol. 263*, 637−647.
12. Marr, M. T., and Roberts, J. W. (1997) *Science 276*, 1258−1260.
13. Huang, X., Lopez deSaro, F. J., and Helmann, J. D. (1997) *Nucleic Acids Res. 25*, 2603−2609.
14. Callaci, S., and Heyduk, T. (1998) *Biochemistry 37*, 3312−3320.
15. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R. H. (2000) *Cell 101*, 601−611.
16. Brodolin, K., Mustaev, A., Severinov, K., and Nikiforov, V. (2000) *J. Biol. Chem. 275*, 3661−3666.
17. Guo, Y., and Gralla, J. D. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11655−11660.
18. Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J., and Waggoner, A. S. (1993) *Bioconjugate Chem. 4*, 105−111.
19. Callaci, S., Heyduk, E., and Heyduk, T. (1998) *J. Biol. Chem. 273*, 32995−33001.
20. Burgess, R. R., and Jendrisak, J. J. (1975) *Biochemistry 14*, 4634−4638.
21. Polyakov, A., Severinova, E., and Darst, S. A. (1995) *Cell 83*, 365−373.
22. Hager, D. A., Jin, D. J., and Burgess, R. R. (1990) *Biochemistry 29*, 7890−7894.
23. Heyduk, E., and Heyduk, T. (1997) *Anal. Biochem. 248*, 216−227.

24. Heyduk, E., Heyduk, T., and Lee, J. C. (1992) *Biochemistry 31*, 3682−3688.
25. Heyduk, E., and Heyduk, T. (1999) *J. Biol. Chem. 274*, 3315−3322.
26. Callaci, S., Heyduk, E., and Heyduk, T. (1999) *Mol. Cell 3*, 229−238.
27. Suh, W. C., Ross, W., and Record, M. T., Jr (1993) *Science 259*, 358−361.
28. Craig, M. L., Suh, W. C., and Record, M. T., Jr (1995) *Biochemistry 34*, 15624−15632.
29. Forster, T. (1948) *Ann. Phys. 2*, 55−75.
30. Stryer, L. (1978) *Annu. Rev. Biochem. 47*, 819−846.
31. Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301−304.
32. Grana, D., Gardella, T., and Susskind, M. M. (1988) *Genetics 120*, 319−327.
33. Matlock, D. L., and Heyduk, T. (1999) *Anal. Biochem. 270*, 140−147.
34. Record, M. T., Reznikoff, W. S., Craig, M. L., McQuade, K. L., and Schlax, P. J. (1996) in *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhardt, F. C., Ed.) Vol. I, pp 792−821, American Society of Microbiology, Washington, DC.
35. Li, T., Ho, H. H., Maslak, M., Schick, C., and Martin, C. T. (1996) *Biochemistry 35*, 3722−3727.
36. Loakes, D., and Brown, D. M. (1994) *Nucleic Acids Res. 22*, 4039−4043.
37. Qiu, J., and Helmann, J. D. (1999) *Nucleic Acids Res. 27*, 4541−4546.
38. Fedoriw, A. M., Liu, H., Anderson, V. E., and deHaseth, P. L. (1998) *Biochemistry 37*, 11971−11979.
39. deHaseth, P. L., Lohman, T. M., Burgess, R. R., and Record, M. T., Jr (1978) *Biochemistry 17*, 1612−1622.
40. deHaseth, P. L., and Helmann, J. D. (1995) *Mol. Microbiol. 16*, 817−824.
41. Helmann, J. D., and deHaseth, P. L. (1999) *Biochemistry 38*, 5959−5967.

BI001433H