RNA Chain Initiation by *Escherichia coli* RNA Polymerase. Structural Transitions of the Enzyme in Early Ternary Complexes†

Barbara Krummel and Michael J. Chamberlin*
Department of Biochemistry, University of California, Berkeley, California 94720
Received March 27, 1989; Revised Manuscript Received June 6, 1989

ABSTRACT: We have studied the properties and structures of a series of *Escherichia coli* RNA polymerase ternary complexes formed during the initial steps of RNA chain initiation and elongation. Five different templates were used that contained the bacteriophage T7 A1 promoter or the *E. coli* Tac or the *lac* UV5 promoter, as well as variant templates with alterations in the initial transcribed regions. The majority of ternary complexes bearing short transcripts (from two to nine nucleotides) are highly unstable and cannot be easily studied. This includes transcripts from the phage T7 A1 promoter, for which the stability of complexes bearing transcripts as short as four nucleotides has previously been postulated. However, with one Tac promoter template, RNA polymerase forms ternary complexes with transcripts as short as five nucleotides that are stable enough for biochemical study. We describe several approaches to identifying and isolating such stable complexes and show that stringent criteria are needed in carrying out such experiments if the results are to be meaningful. Deoxyribonuclease I (DNase I) footprinting has been used to probe the general structure of the stable ternary complexes formed as the polymerase begins transcription and moves away from the start site. The enzyme undergoes a sequence of structural changes during initiation and transition to an elongating complex. Complexes with five to eight nucleotide transcripts, designated initial transcribing complexes (ITC), have identical footprints; they all retain the σ factor and have a slightly extended DNase I footprint (−57 to +24) as compared to the open promoter complex (−57 to +20). ITC complexes all show a region of marked DNase I hypersensitivity in the −25 region that may reflect bending or distortion of the DNA template. Complexes with 10 or 11 nucleotide transcripts, designated initial elongating complexes (IEC), have lost the σ factor and have a slightly reduced and shifted DNase I footprint (−32 to +30). However, these IEC have not yet achieved the much smaller footprint (~30 bp) reported as characteristic of elongating ternary complexes bearing longer RNA chains. During the initial phase of transcription, the RNA polymerase does not move monotonically along the DNA template as RNA chains are extended, but instead, the upstream and downstream contacts remain more or less fixed as the nascent transcript is elongated up to about eight nucleotides in length. Only after incorporation of 10 nucleotides is there significant movement of the enzyme away from the promoter region and a commitment to elongation.

RNA polymerases are unique among enzymes of nucleic acid synthesis in being able to initiate polynucleotide chains de novo by coupling two nucleoside triphosphate substrates. Initially, it was assumed that formation of the initial phosphodiester bond—to form a dinucleoside tetraphosphate—completed the initiation reaction, with subsequent additions of nucleoside monophosphate residues representing the initial steps in RNA chain elongation (Chamberlin, 1974).

Subsequently, important studies by several groups demonstrated that nascent RNA transcripts up to nine nucleotides (nt)[1] in length can be released from the RNA polymerase–promoter complex without dissociation of the polymerase from the promoter (Carpouis & Gralla, 1980; Munson & Reznikoff, 1981). Several cycles of this abortive initiation can occur for each RNA chain that is initiated. Abortive initiation has been seen at all *Escherichia coli* promoters that have been surveyed (Levin et al., 1987) and is probably a general feature of the initiation process for both prokaryotic and eukaryotic (Luse & Jacob, 1987; Hinrichsen et al., 1985; Mosig et al., 1985) RNA polymerases. Hence, initiation is not complete after formation of the first bond but is a phase of transcription that must extend through addition of the first 8–10 nt.

The rate of productive RNA chain initiation may vary for different promoters, depending on the efficiency with which RNA polymerase can pass through this initiation region (Gralla et al., 1980; Munson & Reznikoff, 1981), and there is convincing evidence that different sequences in the initial transcribed region can profoundly affect the strength of particular promoters, both in vitro and in vivo (Brunner & Bujard, 1987; Deuschle et al., 1986; Kammerer et al., 1986). Therefore, there are important reasons for wanting to understand the structure and properties of the RNA polymerase complex as it passes through this transition region.

The general structural features of *E. coli* RNA polymerase in the open promoter complex have been well-defined for several promoters (Siebenlist et al., 1980; von Hippel et al., 1984); this complex is the active intermediate in RNA chain initiation (Chamberlin, 1974). However, the structural features and properties of the complexes involved in the initiation phase—which we will designate initial transcribing complexes (ITC)—are poorly defined.

Several factors contribute to this lack of information about ITC. It has been established that some ITC are very unstable and rapidly dissociate to release the RNA transcript and reform the original open promoter complex (Carpousis & Gralla, 1985; Grachev & Zaychikov, 1980). This is also the most plausible conclusion to be drawn from the extensive formation

† This work was supported by a research grant (GM12010) and a training grant (GM07232) from the National Institute of General Medical Sciences.

[1] Abbreviations: nt, nucleotide; NTP, nucleoside triphosphate; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; SDS, sodium dodecyl sulfate.

0006-2960/89/0428-7829$01.50/0

PROMOTER/ PLASMID	TEMPLATE #	-50 -40 -30 -20 -10 +1 +10 DNA/***RNA*** SEQUENCE
T7 A1/pAR1707	1	TATCAAAAAG AGTATTGACT TAAAGTCTAA CCTATAGGAT ACTTACAGCC ***AUCGAGAGG GACACGGCGAA***
T7 A1/pA T7	2	TATCAAAAAG AGTATTGACT TAAAGTCTAA CCTATAGGAT ACTTACAGCC ***AUGGGCACC CAUCGUUCGUA***
Tac/pT T7	3	ATTCTGAAAT GAGCTGTTGA CAATTAATCA TCGGCTCGTA TAATGTGTGG ***AUGGGCACC CAUCGUUCGUA***
Tac/pKK177-3	4	ATTCTGAAAT GAGCTGTTGA CAATTAATCA TCGGCTCGTA TAATGTGTGG ***AAUUGUGAGC GGAUAACAAU***
lac UV5/pLAC	5	TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA TAATGTGTGG ***AAUUGUGAGC GGAUAACAAU***

FIGURE 1: Sequences of promoters and initial transcribed regions used. The DNA sequence of the −50 to −1 promoter recognition region (PRR) of the nontemplate strand and the RNA sequence of the first 20 nucleotides (bold type) which we refer to as the initial transcribed region (ITR) are shown. Positions are numbered relative to the transcription start site of +1. The standard names of the promoter units and plasmids from which the DNA fragments were obtained are also listed. For simplicity, we refer to the DNA templates by the numbers 1–5. For additional information on the DNA templates, see Experimental Procedures.

of abortive transcripts from 2 to 9 nt in normal initiation. In addition, many methods of structural analysis are not applicable to short-lived complexes. Finally, there are serious technical problems when one tries to halt the polymerase at a particular point in transcription. This is usually done by using limited subsets of nucleoside triphosphate substrates, often together with an initiating di- or trinucleotide primer. Analysis shows that in such reactions the transcripts are often elongated far past the expected point, due to contamination of the reaction with the missing substrates (Levin et al., 1987). This contamination may be due to traces of the missing nucleotides in triphosphate preparations; pyrophosphorolysis of primers may also contribute (Kassavetis et al., 1986).

However, there are many reports in which RNA polymerase complexes, purported to bear short RNA chains or even single nucleotides, have been studied. At the current time, it seems quite difficult to judge the validity of such studies, given the conflicting information available. To attempt to resolve this difficult situation, we have set out to define conditions and procedures that allow formation, identification, and isolation of ternary RNA polymerase complexes positioned at unique sites along the transcription unit (Levin et al., 1987). With such methods it should be possible to develop stringent criteria for the preparation and study of such complexes as the polymerase proceeds through the early phases of the transcription reaction and to critically evaluate studies of this sort.

Experimental Procedures

Reagents. RNA polymerase holoenzyme was purified from *E. coli* DG156 by the method of Gonzalez et al. (1977). Holoenzyme preparations used in this study contained 40–65% active molecules, measured as described by Chamberlin et al. (1979). The following reagents were purchased from the sources noted: α-^{32}P-labeled nucleoside triphosphates (NEN) except [α-^{32}P]CTP (Amersham); [γ-^{32}P]ATP (ICN or Amersham); highly purified nucleoside triphosphates (HPLC grade), nucleoside triphosphate derivatives methylated at the 3'-position, and Sephadex G-50 (Pharmacia); dinucleotides ApU and ApA (Sigma); T4 polynucleotide kinase (NEB). Bovine pancreatic deoxyribonuclease I (DNase I) was purchased from Worthington Biochemicals and was treated with Bentonite (VWR) to remove ribonuclease (RNase; Maniatis et al., 1982).

Buffers. Buffer compositions are as follows: transcription buffer [44 mM Tris-HCl, pH 8.0, 14 mM $MgCl_2$, 14 mM 2-mercaptoethanol, 20 mM NaCl, 2% (v/v) glycerol, 0.04 mM EDTA, 40 μg of acetylated bovine serum albumin/mL; see Levin et al. (1987)]; TBE buffer (89 mM Tris base, 89 mM boric acid, 2.5 mM EDTA); sucrose loading buffer (60% sucrose, 0.01% bromophenol blue, 0.01% xylene cyanol); RNA formamide loading buffer (80% deionized formamide, 1× TBE, 0.05% bromophenol blue, 0.05% xylene cyanol); DNA formamide loading buffer (95% deionized formamide, 8 mM NaOH, 2 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol).

DNA Templates. Plasmids pA T7 and pT T7 (Telesnitsky & Chamberlin, 1989) are pKK177-3 derivatives that contain the promoter recognition region from either the T7 A1 promoter (pA T7) or the Tac promoter (pT T7; see Figure 1) and have identical transcribed sequences. Other plasmids included pAR1707 (Studier & Rosenberg, 1981), pKK177-3 (Brosius & Holy, 1984), and a pBR322 derivative that contains five copies of an *Eco*RI fragment bearing the *lac* L8:UV5 promoter (Straney & Crothers, 1985) referred to here as pLAC.

Promoter DNA fragments for transcription were prepared by restriction enzyme digestion (Maniatis et al., 1982) or by the polymerase chain reaction (PCR; Higuchi et al., 1988). Plasmid DNA was used as the starting material for both. The method of DNA fragment preparation, restriction sites where appropriate, and end-point coordinates relative to the transcription start site of +1 were as follows: pAR1707 (PCR; −149/+153), pA T7 (PCR; −149/+91), pT T7 (PCR; −157/+91), pKK177-3 (restriction enzyme *Bam*HI; −220/+45), and pLAC (restriction enzyme *Eco*RI; fragment length is 203 bp, and exact end points were not determined). For simplicity, we refer to the DNA fragments used in these studies as templates 1–5, respectively (Figure 1).

Preparation of Ternary Complexes. To prepare a collection of ternary complexes with short RNA transcripts (3–10 nt) initiated at various promoters, *E. coli* RNA polymerase (10–25 μg/mL; ~18–45 nM) was incubated with DNA template (18 nM promoter) for 10–20 min at 30 °C in transcription buffer to form open promoter complexes (EP_o). Transcription was initiated by the addition of a dinucleotide primer at 50 μM and a mixture of nucleoside triphosphates (see Figure 3 legend). The first nucleotide incorporated after the dinucleotide

was an [α-^{32}P]NTP at 50 000–100 000 cpm/pmol. Therefore, the smallest labeled transcript synthesized from any template is always the trinucleotide. Ternary complexes are referred to by the transcript length and the 3′-terminal RNA nucleotide; for example, "A20 complex" has a 20-nucleotide transcript with an AMP as the 3′-terminal nucleotide.

As one measure of ternary complex stability, after formation of a particular ternary complex and removal of a sample for gel electrophoresis, all four NTPs were added to the above reaction to a final concentration of 1 mM and incubated for 30 s (a "chase" reaction). The products of the reaction before and after the chase were then analyzed by gel electrophoresis as described below.

To form a homogeneous population of the ternary complexes C6 and A11 on template 3, RNA polymerase and DNA were incubated as described above. The dinucleotide ApU (50 μM) and 5 μM each of GTP (20 000–50 000 cpm/pmol) and CTP (to form C6) or GTP, CTP, and ATP (to form A11) were added, and transcription was allowed for 5–10 min. To remove aborted transcripts and unincorporated NTPs, complexes were placed on centrifugal filtration columns as described below.

Reactions were stopped by the addition of an equal volume of RNA formamide loading buffer and loaded directly onto a gel (0.3 mm thick, 40 cm long) containing 7 M urea, 20% acrylamide, and 3% bis(acrylamide). Electrophoresis was carried out at 1000–1200 V until the xylene cyanol migrated 15–20 cm from the wells (~16 h). Gels were left on the glass plate and covered with plastic wrap, and autoradiographs were made with or without a screen at −70 °C.

Transcript sizes and sequences were determined by nearest-neighbor analysis [see Levin et al. (1987)]. The high-percentage gel is necessary to obtain resolution of very short RNA chains (1–10 nt) on the basis of size.

Synthesis of a 4-nt Transcript Formed from the T7 A1 Promoter. We have used several different methods to determine whether stable complexes with 4-nt transcripts can be formed at the T7 A1 promoter. Initial experiments employed our normal transcription conditions as described under Preparation of Ternary Complexes except that the DNA template was T7 ΔD111 and only CTP (50 000 cpm/pmol) and GTP were present at 5 μM each, along with the dinucleotide primer ApU. The NTPs were purchased from Pharmacia as HPLC purified. Incubation was at 30 °C for 3 and 10 min followed by a chase with 50 μM of the missing NTPs. These conditions give good synthesis of the 4-nt product AUCG, but there is no formation of stable complexes as judged by the inability of the 4-nt RNA to be chased in a subsequent reaction. We have also attempted such syntheses using the transcription conditions described by Ruetsch and Dennis (1987). The reaction volume was 50 μL, and the transcription buffer contained 40 mM Tris-HCl, pH 7.9, 80 mM KCl, 1 mM dithiothreitol, and 10 mM $MgCl_2$. RNA polymerase and T7 ΔD111 concentrations were 40 nM and 4.0 nM, respectively. The dinucleotide ApU was present at 50 or 500 μM, and CTP (2000 cpm/pmol) and GTP were present at 300 μM each. The NTPs were purchased from ICN as HPLC purified. Incubation was at 30 °C for 3 and 10 min followed by a chase with 200 μM of the missing NTPs.

Centrifugal Gel Filtration Chromatography. To obtain ternary complexes free of NTPs rapidly and without dilution, centrifugal gel filtration was used. This procedure also provides a measure of complex stability; ternary complexes are found in the excluded volume while unincorporated NTPs and released transcripts remain in the column. Stable ternary complexes are defined as those that can be subsequently elongated upon addition of all four NTPs. Reactions (100–200 μL) were chilled for 3–5 min on ice and then applied to a "spin column" [0.7 mL of packed Sephadex G-50 in a 1-cm^3 tuberculin syringe equilibrated in transcription buffer without BSA at 4 °C; see Levin et al. (1987)] and centrifuged at low speed for 3.5 min in an IEC clinical centrifuge at 4 °C. The flow-through fractions were then analyzed by gel electrophoresis before and after elongation in the presence of all four NTP substrates.

Isolation of Transcription Complexes by Nondenaturing PAGE. To isolate transcription complexes for biochemical analyses, we used nondenaturing polyacrylamide gel electrophoresis (PAGE). This procedure separates binary and ternary complexes and free DNA into discrete bands (Straney & Crothers, 1985; Levin et al., 1987) and provides another measure of ternary complex stability. Standard transcription reactions were mixed with a one-tenth volume of sucrose loading buffer and loaded into the wells of an acrylamide gel [37.5:1 acrylamide:bis(acrylamide); 0.7 mm thick, 23 cm long] in 1× TBE. Electrophoresis was carried out at 4 °C at a voltage of 200 V until the xylene cyanol was 10–12 cm from the wells (4–5 h). To analyze the RNA components of various complexes, the complexes were identified by autoradiography and excised. Gel slices were soaked in 10–15 μL of RNA formamide loading buffer at room temperature for 1–3 h. Gel slices were then placed in the wells of a 1.4 mm thick, 40 cm long, 20% acrylamide/3% bis(acrylamide) gel in 1× TBE. Electrophoresis and autoradiography were as for 0.3 mm thick gels.

"Walking" Ternary Complexes Formed on Template 3. To form ternary complexes between C6 and A11, additional procedures were required that involved stepwise elongation of the C6 ternary complex. C6 ternary complexes were formed as above by the addition of ApU, CTP, and GTP (50 000 cpm/pmol) to the open promoter complex. After a 10-min incubation, the entire reaction (100–200 μL) was loaded onto a Sephadex G-50 spin column (as described above). Two consecutive columns were sometimes necessary to remove all NTPs. To elongate to position A7, the eluate was collected, and BSA (final concentration 40 μg/mL) and ATP (final concentration 5 μM) were added. Incubation was for 30 s; if incubation was continued for 5 min, all of the A7 complex was converted to the G5 complex. To elongate to position C8 exclusively, 5 μM ATP and 50 μM 3′OMeCTP were added to the C6 eluate and incubated for 5 min. To elongate to position C10, the A7 ternary complex was chromatographed once on a spin column and the eluate incubated with 5 μM CTP for 2 min. Finally, to obtain some enzyme paused at positions C8, C9, and C10, a C6 reaction (complexes formed with 1 μM GTP and CTP) was used directly (no centrifugal gel filtration). After C6 formation, ATP (5 μM) was added with 3′OMeCTP (40 μM), and incubation was continued for 5 min.

DNase I Footprinting of Template 3 Complexes. To obtain DNase I footprints of template 3 transcription complexes, the following steps were always performed in sequence: (1) formation of ternary complexes; (2) DNase I digestion; (3) separation of complexes and free DNA by nondenaturing PAGE; (4) denaturing PAGE of electroeluted DNA. Complexes were formed as described above. The DNA template used was a uniquely 5′-end-labeled 270-bp fragment prepared with the polymerase chain reaction (see DNA Templates). A one-fourth volume of "carrier mix" (transcription buffer with 8 mM $CaCl_2$ and 0.34 mg of sonicated salmon sperm DNA per milliliter) was added, followed by the addition of DNase

TEMPLATES	SUBSTRATES	EXPECTED PRODUCTS
1	ApU + CTP	AUC (3)
	ApU + CTP + GTP	AUCG (4)
	ApU + CTP + GTP + ATP	AUCGAGAGGGACACGGCGAA (20)
2 or 3	ApU + GTP	AUGGG (5)
	ApU + GTP + CTP	AUGGGC (6)
	ApU + GTP + CTP + ATP	AUGGGCACCCA (11)
4 OR 5	ApA + UTP	AAUU (4)
	ApA + UTP + GTP	AAUUGUG (7)
	ApA + UTP + GTP + ATP	AAUUGUGAG (9)

FIGURE 2: RNA products expected with different sets of substrates in transcription reactions. Promoter sequences for templates 1–5 are shown in Figure 1. The specific combination of dinucleotide and nucleoside triphosphates added ("substrates") determines the size of the expected transcript(s). Transcript sequences are shown with sizes indicated in parentheses.

I to a final concentration of 0.01 μg/mL (0.03 μg/mL for A7 complexes). Reactions were incubated for 1.5 min (30 s for A7 complexes) at 30 °C and quenched by the addition of a 30-fold excess of sonicated salmon sperm DNA. A one-tenth volume of sucrose loading buffer was added, and the reactions were loaded immediately onto a 4% acrylamide nondenaturing gel and electrophoresed as described above. The complexes were identified by autoradiography and electroeluted from gel slices in 0.2× TBE buffer. The eluate was extracted once with a phenol–chloroform mixture (1:1 v/v) and then extracted once with chloroform. The RNA and DNA were precipitated with ethanol, then washed with 70% ethanol, and resuspended in 5–10 μL of DNA formamide loading buffer. Immediately before loading onto the gel, samples were heated for 10 min at 90 °C and cooled on ice. Electrophoresis was carried out in gels (0.3 mm thick, 40 cm long) containing 8.3 M urea and 6% acrylamide [19:1 acrylamide:bis(acrylamide)] in 1× TBE at 1600–1800 V. Chemical sequencing was by the method of Maxam and Gilbert (1977).

Analysis of σ Content of Template 3 Complexes. To determine the presence of σ factor in a transcription complex, we analyzed the protein subunits present in complexes isolated by gel electrophoresis. Reactions were carried out in transcription buffer as described above, except that the DNA and RNA polymerase concentrations were 110 nM and 400 nM, respectively (~5 μg of enzyme per reaction). After 10 min at 30 °C, NaCl was added to a final concentration of 0.35 M to ternary complex reactions only. Reactions were incubated on ice for 5 min followed by the addition of heparin (100 μg/mL final concentration). To prevent anomalous migration due to the presence of high-salt concentrations, reactions were diluted to approximately 0.1 M NaCl by the addition of TE or the salt was removed by centrifugal gel filtration. The high salt concentration disrupts binary complexes, and heparin prevents rebinding of displaced holoenzyme. Reactions were loaded onto a 4% nondenaturing gel (as described above) into 2 cm wide wells. Complexes were identified by autoradiography of the wet gel, excised, soaked in 30 μL of 1% SDS plus 5% 2-mercaptoethanol, and placed in the wells (2 cm wide) of a 1.4 mm thick SDS protein gel (Straney & Crothers, 1985). The protein bands were visualized by silver staining.

RESULTS

Rationale. The primary goal of our studies was to obtain structural information about the ternary complex of RNA polymerase, DNA template, and nascent RNA as the enzyme passes through the initial transcribed region (ITR), which we loosely define as sequences between +1 and +20. Such complexes (designated initial transcribing complexes, or ITC, and initial elongation complexes, or IEC) can be formed easily by use of a dinucleotide and limited subsets of NTP, or 3′-*O*-methyl-NTP, to halt transcription at defined sites. Previous studies indicated that most ITC would be too unstable to study by current techniques. Fortunately, we found that a particular ITC bearing a transcript of six nucleotides was relatively stable and could be easily isolated for study. This allowed us to test general methods for detecting and characterizing ITC, to identify those stable enough for study. A summary of the templates tested and the products expected in various reaction procedures is shown in Figures 1 and 2, respectively.

Note that previous studies have shown that the expected products of limited reactions are often quite different from those actually observed (Levin et al., 1987). An expected product may not be formed at all, there may be formation of shorter products due to "pausing" effects, or there may be read-through to give longer products due to contamination of the substrates or due to low levels of pyrophosphorolysis of primer. Because of these complexities, the products of a particular reaction can change dramatically with different sets of reagents, or with time of incubation. Hence, the products of such a reaction cannot be predicted exactly but must be analyzed each time, most simply by gel electrophoresis. In addition, short transcripts (less than 10 nt) can show aberrant migration on polyacrylamide gels (Levin et al., 1987; Cai & Luse, 1987; Krummel, unpublished experiments) and must be sequenced or analyzed by nearest-neighbor ^{32}P transfer (Josse et al., 1961) to provide unequivocal identification.

Once the products of a particular reaction are known, it is possible to determine the stability of the individual ternary complexes bearing each transcript. We have used three criteria:

(A) Ability of the RNA chains in a complex to continue elongation ("chase"). If the ternary complexes formed in a particular reaction are presented with a mixture of all of the substrates needed, transcripts associated with stable ternary complexes will disappear, since they can be elongated to longer chains. Transcripts from unstable complexes have been released and will remain unchanged.

(B) Isolation of the complexes by passage through a gel exclusion sizing column that retards free RNA transcripts.

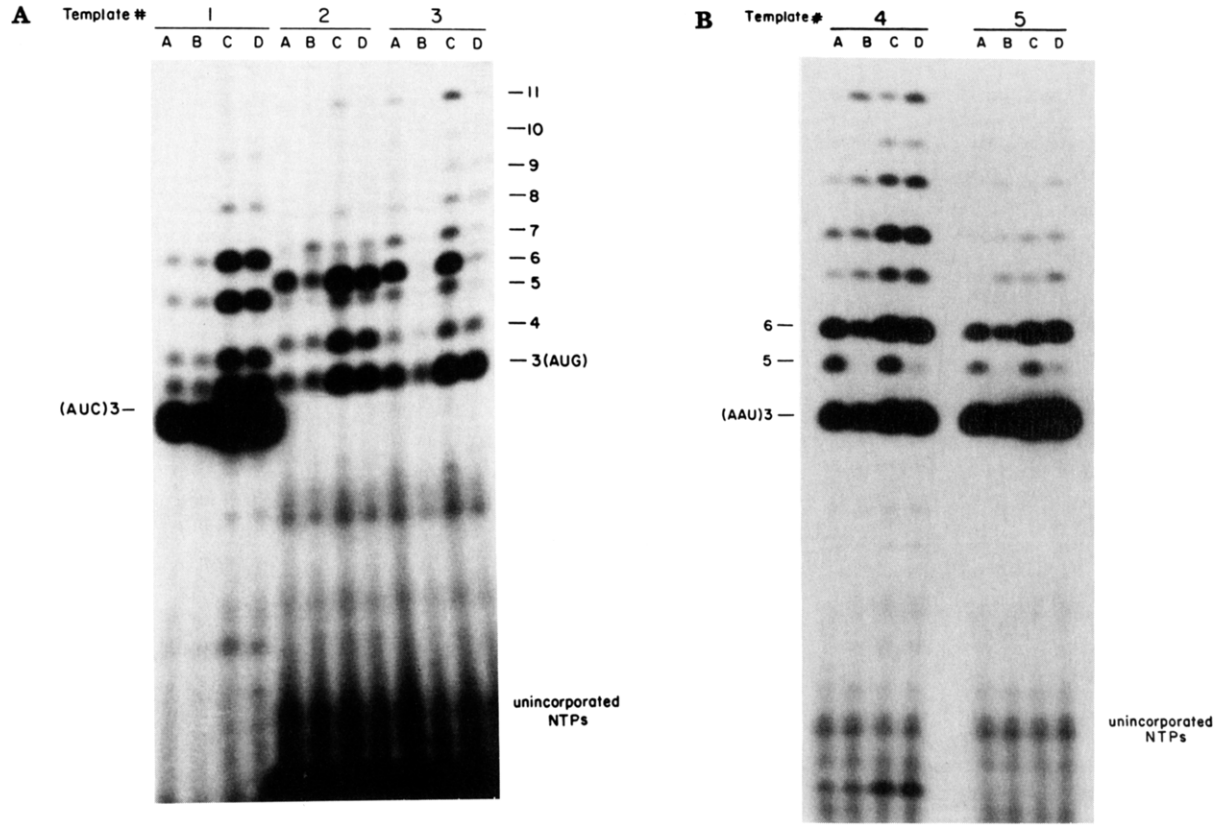

FIGURE 3: Survey of ITC stability with five different templates. The templates used are described in Figure 1 and are referred to by template numbers 1–5. Transcription conditions were as described under Experimental Procedures, and NTPs used were as follows. Panel A: template 1 (50 μM ApU, 1 μM CTP*, 0.1 μM ATP and GTP); templates 2 and 3 (50 μM ApU, 1 μM GTP*, 0.1 μM ATP and CTP). Panel B: templates 4 and 5 (50 μM ApA, 1 μM UTP*, ATP, CTP, and GTP). An asterisk indicates an α-^{32}P-labeled NTP. Lanes A and C: 1- and 5-min incubation, respectively, with limiting NTPs. Lanes B and D: 30-s chase of reactions in lanes A and C, respectively, with 1 mM each of all four NTPs. Transcript sizes are indicated on the right for templates 2 and 3 and on the left for templates 4 and 5. Note that the specific sequence of small RNAs significantly affects their mobility (e.g., AUC vs AUG).

Stable complexes flow through such a column in the excluded volume, and the RNA component can be elongated when NTPs are added.

(C) Isolation of the complexes by electrophoresis in polyacrylamide gels under nondenaturing conditions. Stable complexes have characteristic mobilities and can be isolated free from open promoter complexes and other elongating ternary complexes.

Stability of Initial Transcribing Complexes (ITC). Our initial studies and those of others led us to expect that ternary complexes with transcripts shorter than 10 nt would not be stable (Levin et al., 1987; Carpousis & Gralla, 1985; D. C. Straney & Crothers, 1987). However, some reports suggested that ternary complexes with transcripts as short as 3 nt were stable (Schulz & Zillig, 1981). To address the question more systematically, we screened a variety of linear DNA templates for the ability to form stable ternary complexes bearing short transcripts (3–10 nt). The general form of the experiment was to assay the ability of ternary complexes formed under limiting NTP conditions to be elongated upon addition of the missing NTPs at high concentrations. The transcripts are readily resolved by high-percentage polyacrylamide gel electrophoresis (PAGE). The five templates studied represent a continuum of related promoter and transcribed sequences (Figure 1).

The majority of complexes with transcripts shorter than 10 nt cannot be elongated further even after only 1 min of incubation with limiting NTPs (Figure 3, lanes B). This instability can be compared to the general stability of complexes with longer transcripts in the same reaction that are elongated with 100% efficiency even after 5 min; note, for example, the quantitative disappearance of the 11-nt transcript formed with templates 2 and 3 (Figure 3A, compare lanes C and D). This measure of stability is extremely permissive in that the only requirement is the ability of chains to be elongated after the complexes are formed in solution. There is no challenge from potentially destabilizing reagents (e.g., high salt or heparin), and no isolation is involved that might disrupt or damage the complex. Therefore, it is clear that most of these ternary complexes with transcripts shorter than 10 nt do not have half-lives long enough to allow isolation and characterization by normal biochemical techniques. Note particularly the failure to detect formation of stable ternary complexes with template 1, a DNA fragment that contains the entire T7 A1 promoter with its normal downstream region. With this template, transcripts of sizes 3–9 nt are formed in detectable quantities, yet none of these transcripts can be elongated in the chase reaction.

In contrast with the general instability of most ITC, ternary complexes with transcripts as short as 5 nt formed on templates 2 and 3 and a single complex with a 5-nt transcript on templates 4 and 5 are relatively stable as judged by their ability to be elongated even after 5 min (Figures 3 and 4; the autoradiogram in Figure 3A is overexposed relative to Figure 4). Nearest-neighbor analysis of these transcripts showed them to have the expected sequences. We will refer to these complexes by the 3′-terminal nucleotide and length of the transcript. For example, the stable complex with a 6-nt transcript formed on templates 2 and 3 is C6.

To show conclusively that these transcripts were part of a ternary complex, we tested their ability to remain associated

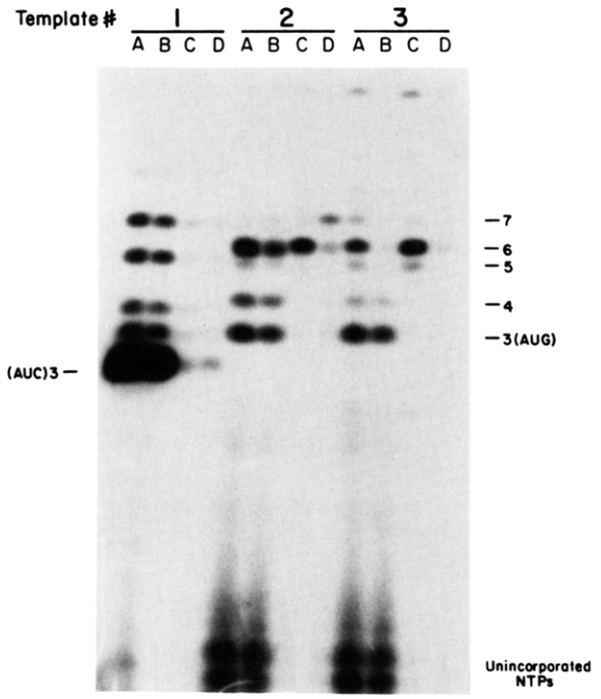

FIGURE 4: Isolation of stable ITC by centrifugal filtration chromatography. Reactions were as described in Figure 3. Lanes A: 5-min incubation with limiting NTPs. Lanes B: 30-s chase of reactions in lanes A with 1 mM each of all four NTPs. Lanes C: centrifugal filtration column excluded volume. Lanes D: 30-s chase of reactions in lanes C with 1 mM each of all four NTPs.

with the RNA polymerase and DNA template during centrifugal gel filtration chromatography (Levin et al., 1987). RNA that is part of a ternary complex is found in the excluded volume while released RNAs and unincorporated NTPs remain in the included volume. In fact, the majority of the 5- and 6-nt RNA formed with template 3 and the 5-nt RNA of templates 4 and 5 was found in the excluded volume after filtration (template 3, Figure 4, lane C; templates 4 and 5, data not shown). In addition, all of this excluded RNA was elongated when NTPs were added (template 3, Figure 4, lane D). These results confirm that these transcripts are part of stable ternary complexes that can be isolated in good yield.

Although there is some recovery of the template 2 C6 complex, the lower yield and reduced ability to be chased indicates a much lower stability compared to the template 3 C6 complex (Figure 4, lanes B and C). This is interesting because template 2 shares the identical transcribed region with template 3 (Figure 1). These results demonstrate that the RNA sequence alone does not dictate the stability of a particular ternary complex. Rather, the upstream promoter sequences play a significant role as well.

To investigate the role of the promoter recognition region (PRR; see Figure 1) in determining ITC stability, we compared ITC formation on templates that had identical PRR but different initial transcribed regions (ITR). Template 1 shares the T7 A1 PRR with template 2 but has the ITR normally found in the T7 A1 transcription unit. There is no formation of stable ternary complexes bearing transcripts shorter than 8–9 nt on template 1, as judged by the inability to chase such transcripts (Figure 3A) and also by the lack of appearance of significant levels of these transcripts in ternary complex fractions after gel exclusion chromatography (Figure 4, lane C). Therefore, such complexes are not suitable for further study using this template. Template 4 (Tac PRR, lac ITR; Figure 1) shares the Tac PRR with template 3, and yet with template 4 there is only the single stable ITC with a 5-nt transcript described above. These results show that the stability of complexes in this early transcribed region varies significantly among the templates analyzed and appears to be dictated by both the promoter and the initial transcribed sequences.

Synthesis and Stability of Individual ITC Complexes Formed with Template 3. Since the C6 ITC formed on template 3 can be isolated in active form, we were able to study its stability and properties in some detail. The C6 complex was prepared by the addition of ApU, GTP, and CTP (Figure 2) and was isolated by gel exclusion chromatography. We followed the stability of the complex by incubating the isolated complex for increasing times under different conditions and measuring the ability of the RNA to be chased into longer products. The C6 complex is significantly more stable at 4 °C than at 30 °C ($t_{1/2} > 2$ h vs $t_{1/2} = 45$ min). It is also stable to treatment at elevated salt concentrations (0.35 M NaCl for 5 min) and resistant to treatment with rifampicin (100 μg/mL for 2 min) and heparin (100 μg/mL for 5 min; data not shown). These properties resemble those of the longer ternary complexes formed from this and other promoters (Levin et al., 1987) and are quite different from open promoter complexes (EP_o) and the majority of ITC. EP_o complexes are disrupted in the presence of 0.35M NaCl and destabilized at low temperatures. This unusual stability of the C6 ITC allowed us to carry out structural studies on this complex.

To isolate and study a series of ternary complexes from template 3 that differed by one nucleotide in their transcript lengths, we elongated the enzyme from C6 to discrete points along the DNA template using multiple rounds of gel filtration followed by incubation with limited substrates, a process that we refer to as "walking". The goal of these walking experiments from C6 was to obtain homogeneous populations of paused ternary complexes A7 to C10 for study. For example, if ATP is added to C6 after column isolation, it is expected that C6 will be elongated to A7; while if ATP and 3′OMeCTP are added, C6 should elongate to C8. C8 would not be further elongated because the 3′OMeCTP is a chain-terminating molecule (see Figure 2).

In practice, the products of such walking experiments are often quite different from those predicted (Figure 5). Consider the reaction in which ATP is added to the C6 complex. Simple theory predicts synthesis of the A7 transcript and longer transcripts if some contaminating NTPs are present. Quite unexpectedly, one of the major products was a transcript 2 nt shorter than A7—a G5 (Figure 5, lane 2). What is the source of this unexpected product? We believe it to be a result of specific pyrophosphorolysis although other alternatives such as hydrolysis have not been ruled out. This enzyme-catalyzed process allows removal of the 3′-terminal NMP by the addition of pyrophosphate (PP_i) to form the free NTP plus an enzyme-associated transcript shortened by one nucleotide (Maitra & Hurwitz, 1967; Rozovskaya et al., 1982). If 2 mol of PP_i is added successively, then the transcript can be shortened by two nucleotides as A7 was shortened to G5. It is clear that when A7 is incubated for increasing times the major product is G5, and there is little C6 (Figure 6, lane 3). The conversion of A7 to G5 occurs very rapidly and can be counteracted by the addition of a low concentration of CTP (10 nM, Figure 6, lanes 4–6). The addition of CTP appears to allow elongation of G5 to regenerate A7 (Figure 6, compare lanes 1–3 to lanes 4–6). If 100 nM CTP is added, however, A7 is elongated to A11 with some pausing at intermediate positions (Figure 6, lanes 8 and 9).

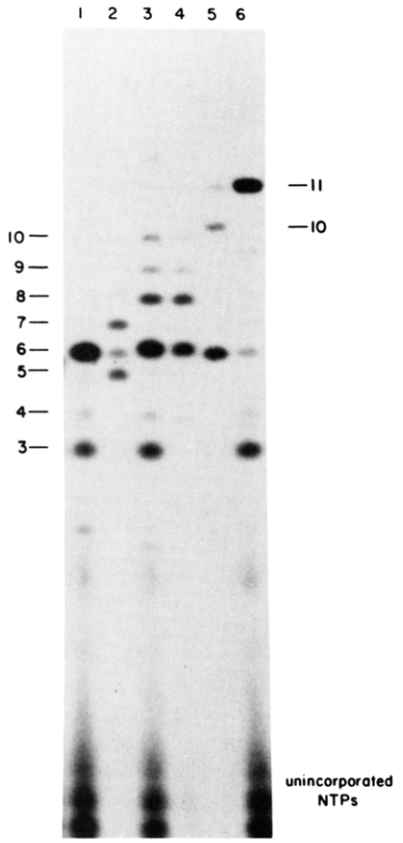

FIGURE 5: Formation of ITC and IEC RNA transcripts with template 3. Reaction conditions are described under Experimental Procedures. The RNA transcript sequence is $^{+1}$AUGGGCACCCAU^{+12}. Lane 1: C6 reaction before centrifugal filtration chromatography. Lane 2: elongation of C6 to form A7. Lane 3: elongation of C6 by the addition of ATP and 3′OMeCTP in the presence of low CTP (40 μM 3′OMeCTP, 1 μM CTP) to form C8, C9, and C10. Lane 4: elongation of C6 by the addition of ATP and 3′OMeCTP to form C8; no CTP was present. Lane 5: elongation of A7 (lane 2) by the addition of CTP to form C10. Lane 6: A11 reaction before centrifugal filtration chromatography.

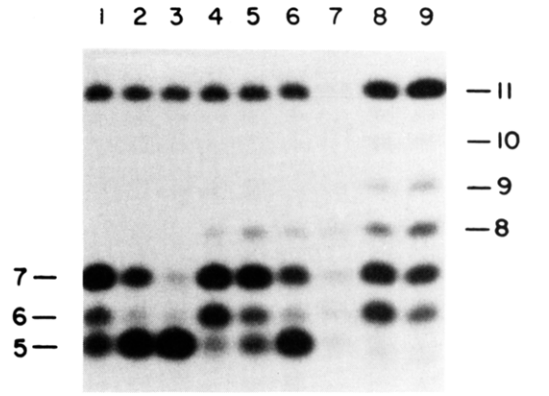

FIGURE 6: Pyrophosphorolysis of A7. C6 ternary complexes were formed from template 3 as described under Experimental Procedures and separated from unincorporated NTPs by two consecutive centrifugal filtration columns. C6 complexes were then elongated in the presence of 1 mM ATP and the stated concentration of CTP for various times. Lanes 1–3: elongation with ATP only for 45 s, 2 min, and 5 min, respectively. Lanes 4–6: elongation with ATP and 10 nM CTP for 45 s, 2 min, and 5 min, respectively. Lane 7: 30-s chase of reaction in lane 6 with 1 mM each of all four NTPs. Lanes 8 and 9: elongation with ATP and 100 nM CTP for 45 s and 2 min, respectively. Transcript sizes are indicated.

These results show an unexpected sequence dependence of product formation and stability. Since we do not understand the factors that determine whether pyrophosphorolysis or elongation will predominate at a particular step under limiting NTP conditions, we cannot predict the exact transcription products even in the highly defined reactions described here. This is clearly an obstacle to determining which complexes can be prepared and isolated. Again, the important conclusion is that only analysis of the reaction products by denaturing gel electrophoresis for each particular experiment can demonstrate exactly which complexes are formed under specific reaction conditions.

Isolation of Stable ITC by Nondenaturing Gel Electrophoresis. In order to investigate the structural characteristics of the various transcription complexes, we have used nondenaturing PAGE as a preparative method for separating binary and ternary transcription complexes and free DNA [Figure 7; see also Straney and Crothers (1985) and Levin et al. (1987)]. To detect the binary (EP_o) complexes, the DNA template must be ^{32}P-labeled at the ends. However, ternary complexes can be detected when either the DNA template or the RNA transcript is ^{32}P-labeled. After electrophoresis, the complexes can be excised from the gel and the RNA, DNA, or protein components analyzed.

In addition to serving as a preparative procedure, nondenaturing PAGE has also provided some interesting information on stability and mobility differences among the ITC described above and complexes with RNA transcripts of 10 and 11 nt, referred to here as initial elongation complexes (IEC). The A11 IEC is formed by incubating EP_o with ApU, ATP, GTP, and CTP (Figure 5, lane 6). It has a mobility significantly faster than that of the open promoter complex (Figure 7A). After nondenaturing PAGE, the gel slices containing the complexes were excised and the RNA components analyzed by denaturing PAGE (Figure 7). The RNA components of the reactions before nondenaturing PAGE (Figure 7, lanes R) can be compared to those RNA transcripts in the upper (lanes U) or lower (lanes L) mobility complexes from the nondenaturing gel. The stability of the ternary complex can be determined by comparing the amount of RNA in lane R to that in lanes U or L. By this measurement, the A11 ternary complex is very stable (>50%) to electrophoresis for 4–5 h in 1× TBE at room temperature. The C10 complex is similar in mobility and stability to A11 (Figure 7C,D).

The ITC complexes have properties quite different from those of the IEC under standard electrophoresis conditions. First, the ITC comigrate with the open promoter complex (Figure 7A). Second, when these ITC—G5, C6, and A7—labeled in the RNA are electrophoresed in a 1× TBE gel at room temperature, they cannot be detected (data not shown). However, if the electrophoresis is performed at 4 °C approximately 10% of the RNA loaded onto the gel is recovered in the complex gel slice (Figure 7). The exact amount of RNA recovered is somewhat variable. The stability of these complexes is further enhanced by electrophoresis in buffers containing much lower salt concentrations, TE (10 mM Tris, 1 mM EDTA) versus TBE (89 mM Tris, 89 mM borate, 2.5 mM EDTA) (data not shown).

We wished to address the nature of the ITC instability during nondenaturing PAGE. That is, was the loss of RNA content during nondenaturing PAGE of the C6 complex due to release of both the RNA and enzyme from the DNA template or to release of the RNA only? In experiments where the DNA was ^{32}P labeled, we isolated the C6 complex by electrophoresis after disruption of residual EP_o (see below). The DNA present in the C6 complex was quantitated and compared to the amount of DNA present in the much stabler A11 complex. The amount of DNA in the two complexes was very similar, while the difference in RNA content described

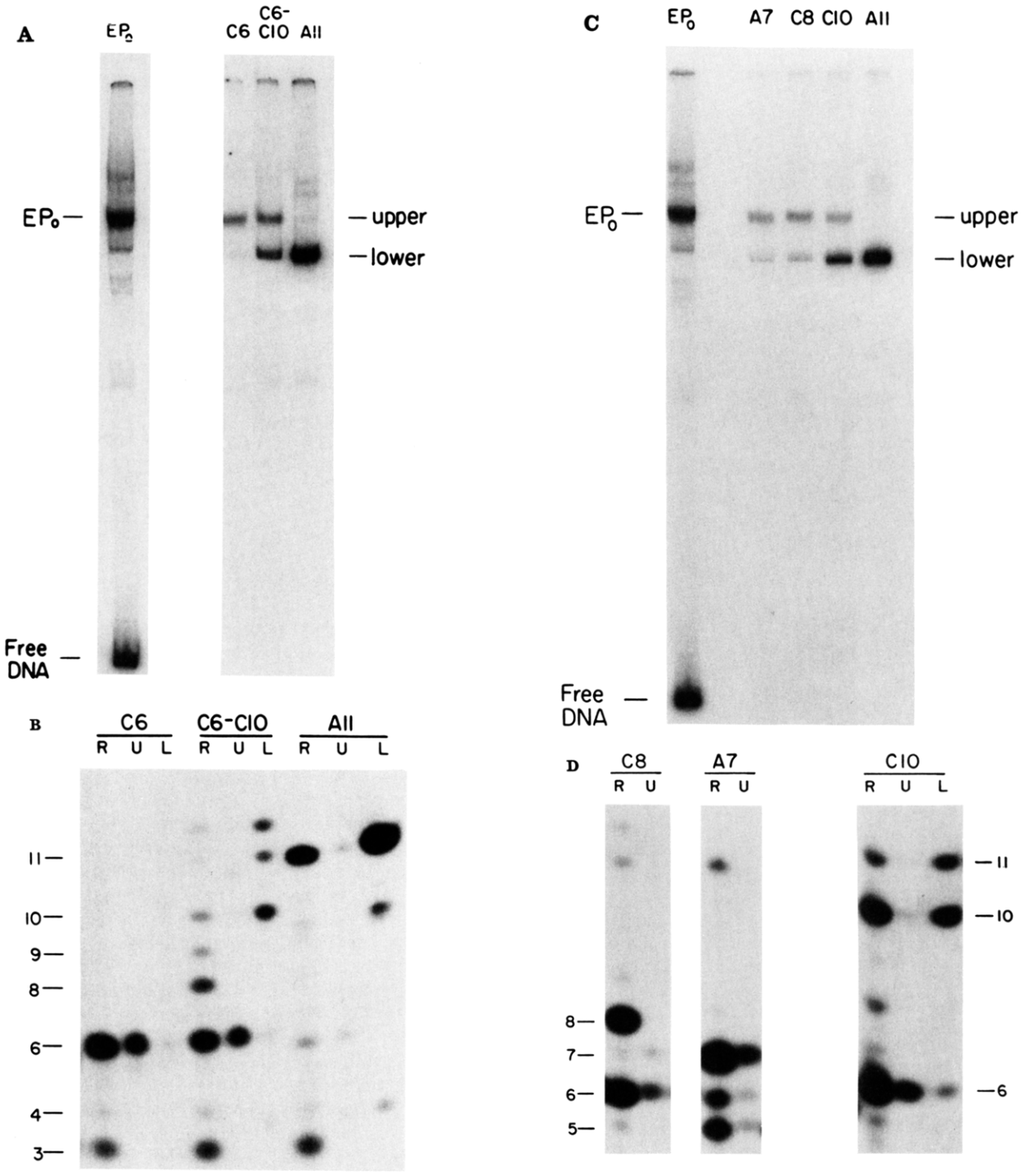

FIGURE 7: Isolation of ITC and IEC by nondenaturing PAGE. All ternary complexes were formed from template 3 as described under Experimental Procedures. Panels A and C show the mobility of binary and ternary complexes on a 4% nondenaturing polyacrylamide gel. Only the open promoter complex (EP_o) had ^{32}P-labeled DNA. All ternary complexes were [α-^{32}P]RNA-labeled. Electrophoresis was performed in 1× TBE at 4 °C. Panels B and D: RNA components of the ternary complexes excised from the nondenaturing gels shown in panels A and C were analyzed by denaturing PAGE. Ternary complexes are indicated. Lanes R: ternary complex reactions before nondenaturing PAGE. Lanes U: RNA from upper band. Lanes L: RNA from lower band. The amount of reaction loaded in lanes R is one-seventh and one-half of the total amount loaded onto the nondenaturing gel for panels B and D, respectively.

above remained (data not shown). Therefore during electrophoresis of the ITC, the RNA is released, but the enzyme remains associated with the DNA in a complex characteristic of EP_o. This result is consistent with the currently accepted mechanism of abortive initiation where the RNA transcripts are released but the holoenzyme remains bound to the promoter.

Although the G5 and A7 complexes behave similarly to the C6 complex during nondenaturing PAGE, the RNA of a C8 complex cannot be detected after electrophoresis under any condition. This is not surprising considering that the C8 complex half-life in solution is also significantly lower (~5 min; data not shown). However, as is the case for the C6 complex, the enzyme remains bound to the DNA template after the RNA has been released. This has allowed us to footprint the C8 complex and draw some likely conclusions as to its σ content (see below).

Comigration of the ITC and EP_o presents a problem because approximately 10–20% of the open promoter complexes do not form C6 under our conditions, probably because the enzyme

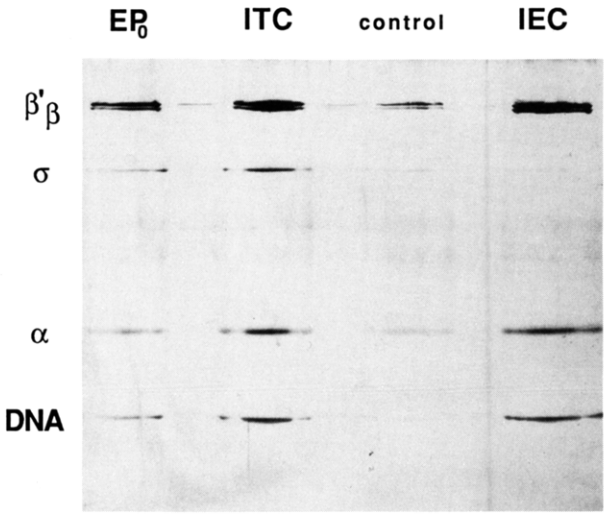

FIGURE 8: Assay for σ factor in transcription complexes. Transcription complexes were formed from template 3, separated by nondenaturing PAGE, excised, and electrophoresed on an SDS denaturing gel, and protein subunits were identified by silver staining (see Experimental Procedures). Analysis of the open promoter complex (EP_o), C6 (ITC), and A11 (IEC) is shown. The control lane shows residual protein present in a gel slice excised from the A11 lane at the upper mobility position where EP_o migrates. The mobilities of the RNA polymerase holoenzyme subunits β', β, σ, and α are indicated. The DNA template is also silver stained.

in these complexes is catalytically inactive (Levin et al., 1987). To obtain C6 complexes free of EP_o, we took advantage of the resistance of C6 complexes to high-salt treatment and heparin. After C6 formation (in transcription buffer containing 20 mM NaCl), the salt concentration was raised to 0.35 M NaCl for 5 min. Heparin (100 μg/mL) was then added to prevent any enzyme that had been displaced from rebinding the template before reactions were diluted or filtered. This treatment yields the C6 complex completely free of EP_o (data not shown).

A clear transition in mobility occurs between complexes A7 and C10. This change may be due to the loss of σ factor, a change in enzyme conformation, or both. Additionally, the instability of the C8 and C9 complexes to nondenaturing electrophoresis (Figure 7B) may be characteristic of a transition complex, perhaps in the process of losing σ factor. These points will be addressed below.

σ Content of ITC and IEC. σ factor release from core enzyme during transcription has been studied in a variety of ways (Bernhard & Meares, 1986; Shimamoto et al., 1986; Hansen & McClure, 1980). In none of the studies, however, has a ternary complex with an RNA transcript shorter than 11 nt been isolated and analyzed for σ content. We have used the nondenaturing PAGE described above to isolate the C6, A7, C10, and A11 complexes and to determine the presence or absence of σ factor in each isolated complex (Figure 8). Contaminating EP_o were removed from all ternary complex reactions by high-salt and heparin treatment as described above. Gel slices containing the complexes were placed in the wells of an SDS–polyacrylamide gel. After electrophoresis, RNA polymerase subunits were visualized by silver staining. The silver-stained gels show that σ factor is part of the C6 and A7 complexes while it is no longer detectable above background in the C10 or A11 complexes (Figure 8; data for A7 and C10 not shown). Note that the staining intensity of the σ subunit band (one 70-kDa subunit) relative to the α subunit band (two 36-kDa subunits) in the C6 complex is quite comparable to that in the EP_o, indicating that there is a stoichiometric amount of σ in the ITC.

Although we cannot isolate the C8 complex using nondenaturing polyacrylamide gel electrophoresis, we believe that it must also contain σ factor, since the C8 RNA is released during electrophoresis but the enzyme–DNA complex remains and re-forms the EP_o complex. If σ had been absent in the C8 complex, we would expect a core polymerase–DNA complex to form or a complete loss of DNA–enzyme complexes, since σ and C8 complexes should separate rapidly in the gel. Core RNA polymerase forms DNA complexes with mobilities different from those of EP_o (Karen Arndt and M. J. Chamberlin, unpublished observations). Our inability to isolate significant quantities of the C9 complex (Figure 7B) prevents the determination of the precise nucleotide at which σ factor is released. However, it is clear that this transition must occur between positions 8 and 9 or 9 and 10.

Table I: Structural Features of Early Transcription Complexes[a]

complex	transcript sizes	DNase I footprint (bp)	σ
EP_o	none	−57 to +20 = 77	+
ITC	5–8	−57 to + 24 = 81	+
IEC	10 and 11	−32 to +30 = 62 (−heparin)	−
		−10 to + 30 = 40 (+heparin)	

[a] A summary of DNase I footprints and the presence of σ factor is presented. (+) indicates the presence of σ, and (−) indicates its absence.

Structural Analysis of ITC and IEC: DNase I Footprinting. Due to the instability of most ternary complexes in the early transcribed region (+1 to +10) the only structural information on complexes in this region has been obtained previously from abortively transcribing enzymes. Abortive initiation is a rapid process in which the promoter-bound enzyme synthesizes and releases short oligonucleotides. After release of the RNA transcript, the enzyme remains bound to the template and can reinitiate synthesis. The cyclic nature of the process and the instability of the ternary complex intermediates prevent characterization of a known and discrete complex. In contrast, the stability of the ITC complexes we have described has allowed us to obtain the DNase I footprints of a collection of well-defined complexes from G5 to A11. The work presented below demonstrates that the enzyme undergoes a sequence of structural changes during initiation and the transition to an elongating complex.

Unique footprints of the EP_o, G5, C6, and A11 can be obtained by taking advantage of differential stabilities of the EP_o and C6 and mobility differences of the C6 and A11 on nondenaturing gels as described above (see Experimental Procedures). Footprints of ternary complexes A7, C8 and C10 can also be obtained, but for A7, G5 is present, for C8, C6 is present, and for C10, A11 is present (Figure 7). Thus for these complexes a change in footprint would be detected over the known G5, C6, or A11 footprint. In all footprinting experiments, the following steps were performed in sequence: (1) formation of ternary complexes; (2) DNase I digestion; (3) separation of complexes and free DNA by nondenaturing PAGE; (4) denaturing PAGE of electroeluted DNA.

Footprinting these complexes shows two distinct transitions in the early transcribed region (Figures 9 and 11 and Table I). The EP_o protects 77 bp (−57 to +20) with a hypersensitive site at −37 on the template strand. Incorporation of six nucleotides (ITC) leads to a 4-bp increase in protection at the downstream edge (−57 to +24) and a strong new hypersensitive site between −20 to −25 on both strands, suggesting a bending of the template as a result of a conformational change in the enzyme. The other upstream contacts (−30 to −50) remain identical to those of EP_o (Figure 9). Complexes G5, A7, and C8 share these characteristics with no detectable differences (data not shown). The second transition occurs between C8

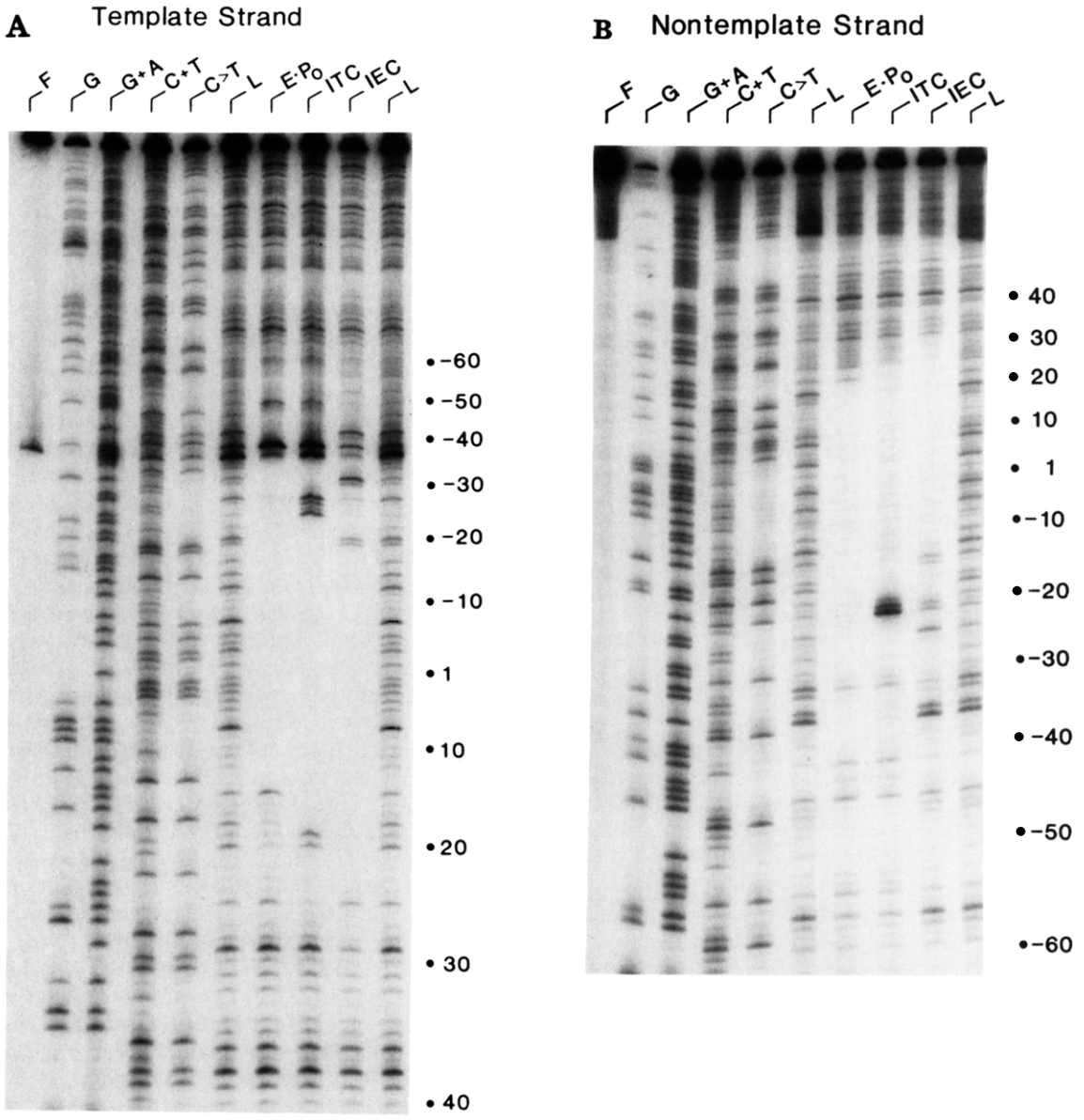

FIGURE 9: DNase I footprinting of transcription complexes. Transcription complexes were formed as described on template 3, DNase I treated, separated by nondenaturing PAGE, excised, and electroeluted, and the DNA was analyzed by denaturing PAGE as shown here. Panel A: template strand (lower strand). Panel B: nontemplate strand (upper strand). Lanes F: untreated fragment. Lanes G, G+A, C+T, C>T: chemical sequencing reactions. Lanes L: DNase-treated fragment minus enzyme. Lane EP_o: open promoter complex. The ITC shown is C6, and the IEC shown is A11. DNase I cleavage positions are numbered relative to the start of transcription.

(ITC) and C10 (IEC) and correlates with the loss of σ factor. The majority of change occurs in the upstream region of the footprint. All protection upstream of −35 is lost and protection extending from −10 to −30 is not complete. In contrast, protection of the transcribed region has increased only slightly to +30. Thus a comparison of the region of strong protection between C6 and A11 shows a footprint of 60 bp versus 40 bp, respectively, with the loss of a major hypersensitive site. The region of partial protection (−10 to −30) seen in the IEC footprints can be disrupted by the addition of heparin (100 μg/mL, Figure 10). No other region of the IEC footprint is affected by heparin treatment, and neither the EP_o (Figure 10) nor the C6 footprint (data not shown) is disrupted in any region.

Discussion

We have demonstrated that specific RNA polymerase ternary complexes, bearing nascent RNA transcripts as short as 5 nt, can be isolated in active and stable form. This has allowed, for the first time, structural analysis of these important intermediates in RNA chain initiation. We believe our structural analysis of the ITC (G5 to C8) and the IEC (C10 and A11) to be the most accurate to date for two reasons. First, we have disrupted any residual EP_o in the case of ITC studies or have separated them from IEC by nondenaturing PAGE so there is no possibility of the binary complex contributing to the ternary complex footprints. Second, in every footprinting experiment we have simultaneously labeled the RNA transcript to confirm the presence and relative abundance of the expected transcripts (Figure 7). Finally, in each case we have determined the stability of the complex involved and known that the footprint is that of a stable and active RNA polymerase complex, not a mixture of undefined composition.

Initial structural studies have been carried out with template 3 by use of DNase I footprinting of stable complexes bearing RNA chains up to 11 nt to gain a general picture of upstream and downstream boundaries of enzyme–DNA contacts for these complexes. For the complexes we designate ITC, bearing chains up to 8 nt, there is no change in contacts upstream in

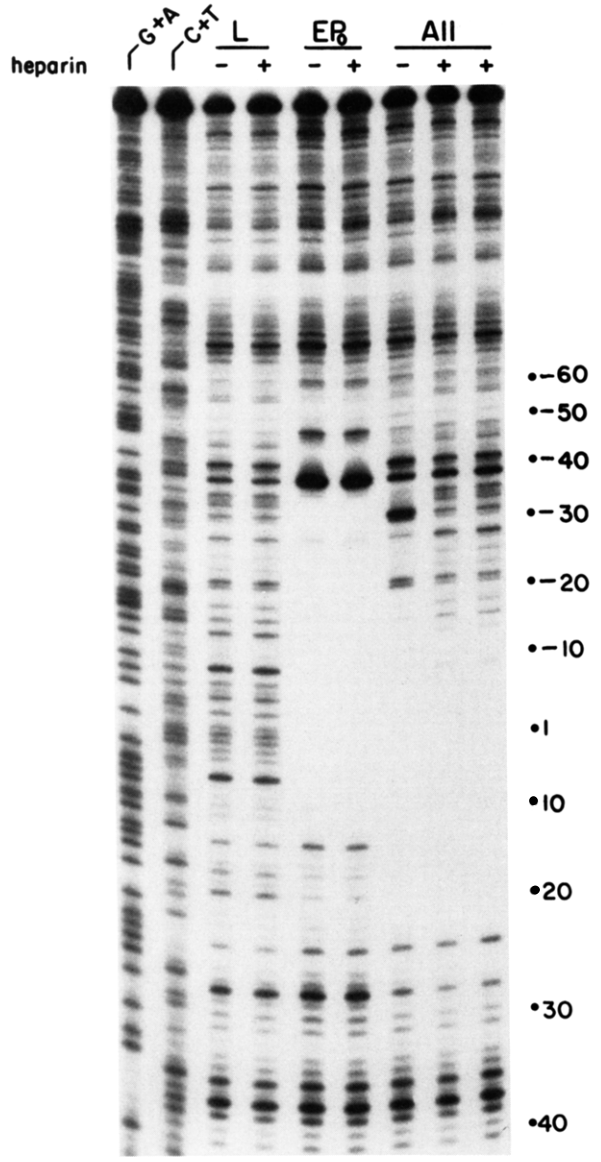

FIGURE 10: Effects of heparin on the A11 DNase I footprint. DNase I footprinting was as described in Figure 9. Lanes G+A and C+T: chemical sequencing reactions. Lane L: DNase I treated fragment minus enzyme. Lane EP_o: open promoter complex. Lane A11: ternary complex with 11-nt transcript. Complexes were formed and incubated for 5 min at 30 °C in the absence (−) or presence (+) of 100 μg/mL heparin before DNase I treatment. Addition of heparin to DNA alone (lanes L) shows that heparin does not affect the DNase I cleavage pattern. For the A11 complexes, heparin was added before or after transcript synthesis [left and right (+) lanes, respectively].

the −30 to −60 region of the promoter, as compared with the open promoter complex. However, the downstream contacts appear to be extended about 4 bp along the DNA, and there is the appearance of a dramatic region of DNase I hypersensitivity in the −25 region on both DNA strands. This hypersensitive region may signal a DNA bend or distortion, possibly brought about by stresses involved in the initial phase of transcription. This region is well upstream of the DNA sequences thought to be denatured in open promoter complex formation (Siebenlist et al., 1979).

Surprisingly, individual ITC bearing RNA chains from 5 to 8 nt all show identical DNase I footprints, suggesting that, within the resolution of the DNase I footprinting method, there is little or no translocation of the upstream and downstream contacts of the enzyme along the DNA as the RNA chain is elongated through this size range. Since there is DNase I cleavage at almost every position along this particular DNA template, including near the footprint boundaries, we should have been able to detect movement of the enzyme by even a few base pairs. If it is true that the upstream and downstream contacts of the enzyme on the duplex are not changed during the synthesis of the initial transcripts from 5 up to 8 nt, then either the DNA within the complex must move along the active site or the active site itself must move.

While it was not possible to obtain sufficient quantities of isolated complexes bearing chains of 9 nt directly, the 10-nt complex has a footprint substantially different from that of the ITC, and we have designated it an initial elongating complex (IEC) accordingly. This change is consistent with a translocation having occurred for both upstream and downstream contacts in the conversion of ITC to IEC. In this transition, the footprint shrinks to cover about 62 bp, as compared to 77 bp in the open promoter complex and 81 bp in the ITC. Carpousis and Gralla (1985) have shown that *lac* UV5 promoter complexes bearing 16- or 17-nt transcripts have footprints of about 30 bp and defined these as "elongation complexes". The IEC we have characterized are significantly different from such elongation complexes. Preliminary footprinting studies on isolated template 3 complexes bearing longer transcripts indicate a transition to an elongation complex comparable to that described by Carpousis and Gralla only after incorporation of about 19 nt (B. Krummel, unpublished results).

Individual ITC of 6 and 7 nt retain the σ subunit, which is lost in the 10-nt IEC. Although the 8-nt ITC is not stable enough to isolate for subunit analysis, it dissociates during nondenaturing polyacrylamide gel electrophoresis to form the open promoter complex, providing strong evidence that the σ subunit is still present. These findings rule out the model of Shimamoto et al. (1986), who suggested that the release of σ was triggered by formation of RNAs from 4 to 6 nt in length but that σ release was slow (~5 s) and could be delayed until longer chains had formed. Our C6 and A7 complexes are incubated for at least several minutes in solution before electrophoresis. According to Shimamoto's model, this would allow ample time for σ release. In addition, we have shown that during 4 h of electrophoresis these complexes retain σ or revert to open promoter complexes that contain stoichiometric amounts of σ factor. Hence, no triggering of σ release has taken place. Instead, our data fit well with the prior studies of Hansen and McClure (1980), who suggested a transition leading to σ release at about 8–9 nt. It will be important to measure the point of this transition with other templates, since cross-linking studies show a template dependence of the length of RNA that can interact with σ (Stackhouse et al., 1989).

Our results lead us to suggest that the primary steps involved in promoter binding and RNA chain initiation (Chamberlin, 1974; McClure, 1985) should be expanded to include the ITC to IEC transition. This sequence would then be

$$E + P \rightleftharpoons EP_c \rightleftharpoons EP_o \rightleftharpoons E_{ITC} \xrightarrow{\sigma} E_{IEC}$$

E is the enzyme, and P is the promoter DNA. EP_c and EP_o refer to the closed and open promoter complexes, respectively. Only the final step, involving the loss of σ subunit, is truly irreversible.

Can factors that affect the ITC to IEC transition play a regulatory role in transcription? S. B. Straney and Crothers (1987) have reported that the *lac* repressor can inhibit transcription without occluding the binding of RNA polymerase to the promoter. They suggested that the *lac* repressor acts

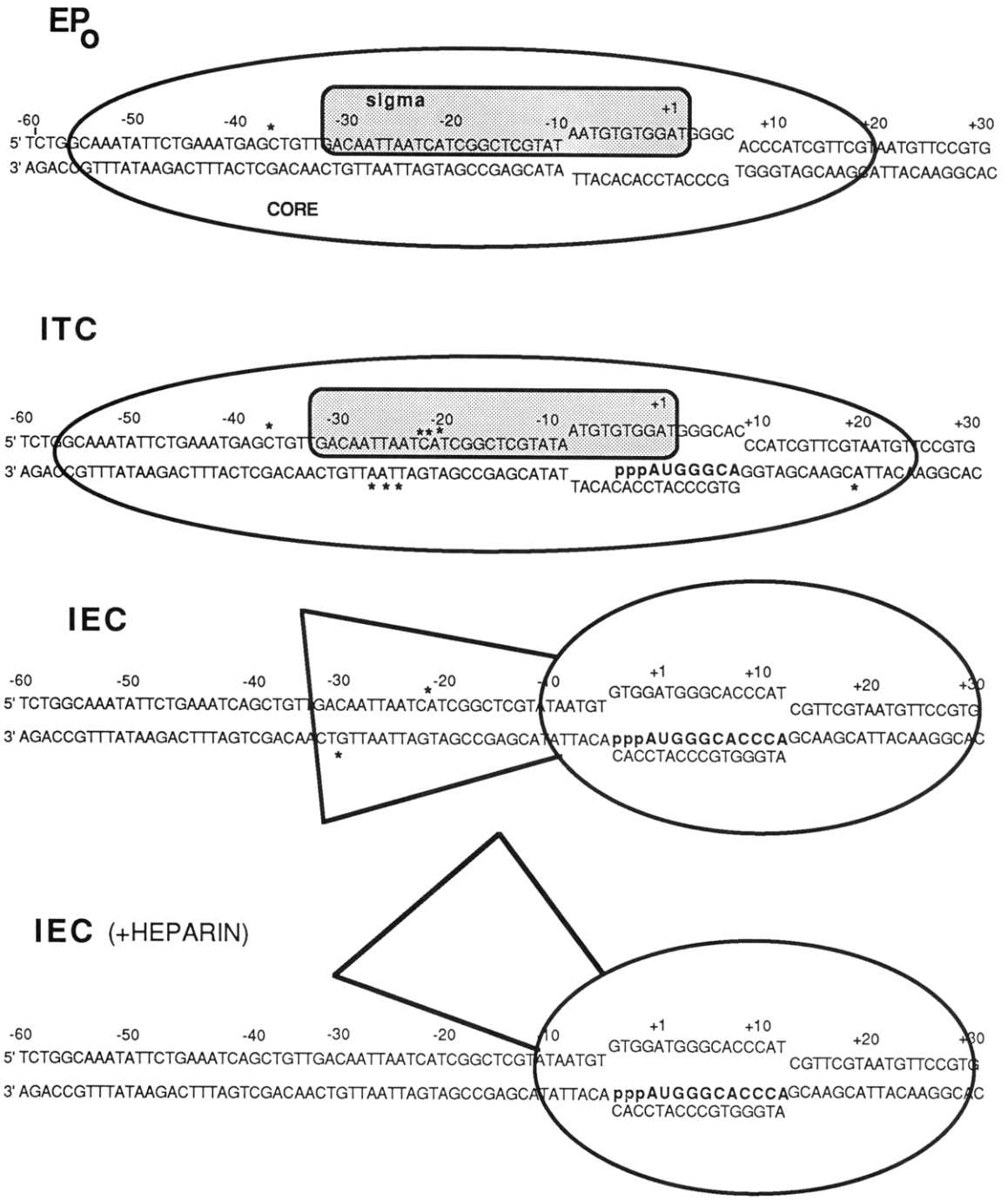

FIGURE 11: Schematic representation of RNA polymerase–DNA contacts in the template 3 transcription complexes. The regions shown as protected include the areas of partial protection (see Figure 9). The region of template unwinding is indicated by additional separation of the strands and is only an approximation. RNA transcripts are shown in bold type. The asterisks indicate DNase I hypersensitive sites. The addition of heparin to the IEC complex completely disrupts protection from −10 to −30 as shown to the bottom figure by the movement of the "tail".

to block the closed (EP_c) to open (EP_o) complex conversion. However, their experiments show that, in the presence of repressor, there is significant synthesis of aborted transcripts (6 and 8 nt) with no synthesis of longer transcripts (11 nt). Hence, the block to transcription in this case may actually occur at the ITC to IEC transition. A similar mechanism of regulation is suggested for the galactose operon P2 promoter, where the *gal* repressor blocks synthesis of long chains but not of aborted transcripts (DiLauro et al., 1979; Adhya & Miller, 1979).

An important question involves how generally the structural information we find for the template 3 ITC is likely to apply to ITC formed with other promoters. This is particularly important since, as we have stressed, most other ITC are quite unstable and therefore must differ in some way from these stable complexes we have studied. While this point will only be clarified with studies using other templates, we think it likely that the general features of the ITC footprints we have determined may well prove characteristic of ITC in general. Thus, while there are great differences in promoter strengths and rates of RNA chain initiation, the general footprint of the open promoter complex is quite similar for all such complexes (von Hippel et al., 1984).

Our results are consistent with those of D. C. Straney and Crothers (1987), who studied the structural differences among the *E. coli* RNA polymerase open promoter, abortive, and

elongating transcription complexes initiated from the *lac* UV5 promoter. For complexes synthesizing an aborted 8-nt transcript, they observed a slight increase in DNase I protection downstream with little change in the upstream region, as compared to EP_o complexes. After synthesis of 11 nt, σ factor was lost, the transcript remained stably associated, and there was a large decrease in the extent of protection upstream.

Changes in the enzyme–DNA contacts during early transcription by the phage T7 RNA polymerase and the mitochondrial RNA polymerase of *Saccharomyces cerevisiae* also appear similar to what we have shown for the *E. coli* enzyme. In both cases, during synthesis of approximately the first 10 nt, the enzyme appears to "reach" into the transcribed region before the loss of interactions in the upstream region (Ikeda & Richardson, 1986; Schinkel et al., 1988). It may be that a common pathway is followed among the bacterial, phage, and yeast mitochondrial RNA polymerases in which the enzyme maintains strong contacts with the stabilizing promoter elements during the initiation phase, prior to the transition to elongation.

Isolation of ITC that are stable enough for biochemical study depends on the exact combination of promoter recognition region and initial transcribed regions that comprise the DNA template. In our studies, only template 3 (Figure 1) allowed formation of highly stable ternary complexes with transcripts shorter than 10 nt, although it may be that conditions can be found for isolation of stable 5-nt complexes with templates 4 and 5 as well. [In fact, preliminary studies suggest that supercoiled DNA templates may significantly stabilize some ITC (B. Krummel, unpublished observations).] We have shown that this unusual stability is not dictated solely by the particular promoter used or only by the strength of the RNA–DNA hybrid. Instead, ternary complex stability within the first 10 nt transcribed must involve specific interactions of the enzyme with the DNA template throughout the promoter and initial transcribed sequences that determine enzyme conformation.

An important element of uncertainty in formation and study of particular complexes is introduced by the problems of substrate contamination and pyrophosphorolysis. The former factor, in particular, can lead to quite different RNA transcripts being synthesized in apparently identical reactions. In a significant proportion of our experiments, the products found included not only those expected but also longer chains. We believe that this kind of result is almost certainly due to contamination of the reactions with missing substrates as also shown by Levin et al (1987) for a comparable reaction. Such contamination is often found in commercial NTP preparations of purported purity but can also arise from slow deamination of ATP and CTP to give ITP and UTP (Shapiro & Klein, 1966) or by simple contamination of reagents during use. An alternative possibility is that misincorporation is taking place at the position of the missing NTP; Kahn and Hearst (1989) have reported such misincorporation; however, their reactions involved much longer times (90 min) and higher NTP concentrations (50 μM) than we used here. In either case, it is often necessary to repeat an experiment several times before the desired complex is obtained. We feel that this variability of individual experiments mandates direct structural determination for each experiment in which a particular complex is to be studied.

A troublesome result of our studies is the finding that ITC bearing chains up to 8 nt that are formed with a linear DNA template that contains the phage T7 A1 promoter are not stable under any conditions we have tested nor by any criterion we have used. Such complexes have been the subject of several reports in which detailed structural studies were reported (Kinsella et al., 1982; Panka & Dennis, 1985; Ruetsch & Dennis, 1987). It is important to note that in all of these reports it was assumed that the complexes formed in particular reactions were those predicted from the substrates added. It was never demonstrated directly by analytical gel electrophoresis that the relevant complexes were formed, or were stable, either by isolation or by the ability to be further elongated.

These studies employed an indirect criterion of "salt resistance" as a measure of stable complex formation. This involves measurement of the amount of total RNA synthesis in the reaction after addition of elevated salt concentrations and all substrates. It was assumed that stable ternary complexes were resistant to such conditions and would carry out RNA synthesis. However, in the event of contamination of the NTP used in such reactions, there would be formation of complexes over 10 nt and not simply the predicted 4-nt product. Complexes with transcripts over 10 nt in length are, of course, long enough to be stable, and such complexes would then give salt-resistant transcription. We have attempted to repeat the synthesis of a T7 A1 tetramer using the reaction conditions reported by Ruetsch and Dennis (1987). We observed extensive synthesis of RNA transcripts over 10 nt in size and saw no evidence for formation of a stable tetramer complex (see Experimental Procedures).

We think it most likely that the studies of Dennis' group were carried out with impure NTP preparations and that there was no formation of stable ITC in these experiments. Rather, the "stable complexes" they used were probably mixtures of elongating complexes longer than 10 nt and open promoter complexes in the process of abortive initiation. Even if the studies carried out by Dennis' group had some unusual element that allowed formation of stable ITC with the T7 A1 promoter, their significance cannot be evaluated because there was no unequivocal identification of the complexes actually formed in their reactions. Given the unknown quality of the nucleotides used in the particular experiments, the products of their reactions cannot be known exactly.

References

Adhya, S., & Miller, W. (1979) *Nature 279*, 492–494.

Bernhard, S. L., & Meares, C. F. (1986) *Biochemistry 25*, 5914–5919.

Brosius, J., & Holy, A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6929–6933.

Brunner, M., & Bujard, H. (1987) *EMBO J. 6*, 3139–3144.

Cai, H., & Luse, D. S. (1987) *J. Biol. Chem. 262*, 298–304.

Carpousis, A. J., & Gralla, J. D. (1980) *Biochemistry 19*, 3245–3253.

Carpousis, A. J., & Gralla, J. D. (1985) *J. Mol. Biol. 183*, 165–177.

Chamberlin, M. J. (1974) *Annu. Rev. Biochem. 43*, 721–775.

Chamberlin, M. J., Nierman, W. C., Wiggs, J., & Neff, N. (1979) *J. Biol. Chem. 254*, 10061–10069.

Deuschle, U., Kammerer, W., Gentz, R., & Bujard, H. (1986) *EMBO J. 5*, 2987–2994.

DiLauro, R., Taniguchi, T., Musso, R., & deCrombrugghe, B. (1979) *Nature 279*, 494–500.

Gonzalez, N., Wiggs, J., & Chamberlin, M. J. (1977) *Arch. Biochem. Biophys. 182*, 404–408.

Grachev, M. A., & Zaychikov, E. F. (1980) *FEBS Lett. 115*, 23–26.

Gralla, J. D., Carpousis, A. J., & Stefano, J. E. (1980) *Biochemistry 19*, 5864–5869.

Hansen, U. M., & McClure, W. R. (1980) *J. Biol. Chem. 255*, 9564-9570.
Higuchi, R., Krummel, B., & Saiki, R. K. (1988) *Nucleic Acids Res. 16*, 7351-7367.
Hinrichsen, A. I., Ortner, I. M., & Hartmann, G. R. (1985) *FEBS Lett. 193*, 199-202.
Ikeda, R. A., & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3614-3618.
Josse, J., Kaiser, A. D., & Kornberg, A. (1961) *J. Biol. Chem. 236*, 864-875.
Kahn, J. D., & Hearst, J. E. (1989) *J. Mol. Biol. 205*, 291-314.
Kammerer, W., Deuschle, U., Gentz, R., & Bujard, H. (1986) *EMBO J. 5*, 2995-3000.
Kassavetis, G. A., Zentner, P. G., & Geiduschek, E. P. (1986) *J. Biol. Chem. 261*, 14256-14265.
Kinsella, L., Hsu, C. J., Schulz, W., & Dennis, D. (1982) *Biochemistry 21*, 2719-2723.
Levin, J. R., Krummel, B., & Chamberlin, M. J. (1987) *J. Mol. Biol. 196*, 85-100.
Luse, D. S., & Jacob, G. A. (1987) *J. Biol. Chem. 262*, 14990-14997.
Maitra, U., & Hurwitz, J. (1967) *J. Biol. Chem. 242*, 4897-4907.
Maniatis, T., Fritsch, E., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560-564.
McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171-204.
Mosig, H., Schaffner, A. R., Sieber, H., & Hartmann, G. R. (1985) *Eur. J. Biochem. 149*, 337-343.
Munson, L. M., & Reznikoff, W. S. (1981) *Biochemistry 20*, 2081-2085.
Panka, D., & Dennis, D. (1985) *J. Biol. Chem. 260*, 1427-1431.
Rozovskaya, T. A., Chenchik, A. A., & Beabealashvilli, R. S. (1982) *FEBS Lett. 137*, 100-104.
Ruetsch, N. R., & Dennis, D. (1987) *J. Biol. Chem. 262*, 1674-1679.
Schinkel, A. H., Koerkamp, M. J. A. G., & Tabak, H. F. (1988) *EMBO J. 7*, 3255-3262.
Schulz, W., & Zillig, W. (1981) *Nucleic Acids Res. 99*, 6889-6905.
Shapiro, R., & Klein, R. (1966) *Biochemistry 5*, 2358-2362.
Shimamoto, N., Kamigochi, T., & Utiyama, H. (1986) *J. Biol. Chem. 261*, 11859-11865.
Siebenlist, U. (1979) *Nature 279*, 651-652.
Siebenlist, U., Simpson, R., & Gilbert, W. (1980) *Cell 20*, 269-281.
Stackhouse, T. M., Telesnitsky, A. P., & Meares, C. F. (1989) *Biochemistry* (in press).
Straney, D. C., & Crothers, D. M. (1985) *Cell 43*, 449-459.
Straney, D. C., & Crothers, D. M. (1987) *J. Mol. Biol. 193*, 267-278.
Straney, S. B., & Crothers, D. M. (1987) *Cell 51*, 699-707.
Studier, F. W., & Rosenberg, A. (1981) *J. Mol. Biol. 153*, 503-527.
Telesnitsky, A. P. W., & Chamberlin, M. J. (1989) *J. Mol. Biol. 205*, 315-330.
von Hippel, P. H., Bear, G. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389-446.

Sequence Dependence of DNA Conformational Flexibility†

A. Sarai,‡,§ J. Mazur,‖ R. Nussinov,⊥ and R. L. Jernigan*,‡

Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Advanced Scientific Computing Laboratory, Program Resource Inc., NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, and Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv, 69978 Israel

Received April 10, 1989; Revised Manuscript Received May 31, 1989

ABSTRACT: By using conformational free energy calculations, we have studied the sequence dependence of flexibility and its anisotropy along various conformational variables of DNA base pairs. The results show the AT base step to be very flexible along the twist coordinate. On the other hand, homonucleotide steps, GG(CC) and AA(TT), are among the most rigid sequences. For the roll motion that would correspond to a bend, the TA step is most flexible, while the GG(CC) step is least flexible. The flexibility of roll is quite anisotropic; the ratio of fluctuations toward the major and minor grooves is the largest for the GC step and the smallest for the AA(TT) and CG steps. Propeller twisting of base pairs is quite flexible, especially of A·T base pairs; propeller twist can reach 19° by thermal fluctuation. We discuss the effect of electrostatic parameters, comparison with available experimental results, and biological relevance of these results.

Various experimental studies indicate that DNA is a flexible molecule and that the flexibility is sequence dependent. The conformational flexibility can be manifested in various motions such as bending and twisting. The sequence dependence of the flexibility is observed in various experiments: DNAs with different base composition show different characteristic decay constants in optical measurements (Hogan et al., 1983). Poly(A)·poly(T) and poly(G)·poly(C) do not bind to nucleosome core proteins (Rhodes, 1979; Simpson & Kunzler, 1979; Kunkel & Martinson, 1981; Prunell, 1982). Certain types of sequences exhibit abnormal gel mobility, which has been attributed to the bending of DNA (Koo & Crothers, 1986; Hagerman, 1986). The flexibility of DNA appears to affect

† Research sponsored, at least in part, by the NCI, DHHS, under Contract NO1-CO-74102 with PRI.
‡ National Cancer Institute.
§ Present address: RIKEN Life Science Center, 3-1-1 Koyadai, Tsukuba-city, Ibaraki-ken, 305 Japan.
‖ NCI-Frederick Cancer Research Facility.
⊥ Tel-Aviv University.