Effects of Arabinosylcytosine-Substituted DNA on DNA/RNA Hybrid Stability and Transcription by T7 RNA Polymerase†

Thomas Mikita‡ and G. Peter Beardsley*,§

Departments of Molecular Biophysics and Biochemistry, Pediatrics, and Pharmacology, Yale University School of Medicine, New Haven, Connecticut 06510

Received February 8, 1994; Revised Manuscript Received May 10, 1994®

ABSTRACT: Cytosine arabinoside (araC) is a potent antileukemic agent which interferes with DNA replication both as a dNTP competitive inhibitor as well as after its misincorporation into DNA. We previously developed a chemical methodology for the synthesis of DNA oligomers containing araC which allowed us to study its site specific effects on duplex stability and chemical reactivity [Beardsley, G. P., Mikita, T., Klaus, M., & Nussbaum, A. (1988) *Nucleic Acids Res. 16,* 9165], as well as its effects on DNA ligase and DNA polymerase activity [Mikita, T., & Beardsley, G. P. (1988) *Biochemistry 27,* 4698]. The DNA polymerase studies, in addition to other observations, showed that araC in DNA templates could have an inhibitory effect on polymerase bypass. As a template lesion, there exists the potential for interference with other aspects of DNA metabolism, such as transcription. We have characterized a DNA/RNA hybrid containing an araC-G base pair, comparing thermal stability, chemical cleavage rates, and duplex gel mobility to an identically sequenced DNA duplex. We find that the A-form DNA/RNA hybrid and the B-form DNA duplex are nearly identical in the extent their thermal stability is affected by an araC-G(dG) base pair. Substitutions of araC for dC were made at various positions in a series of DNA duplex substrates containing a T7 RNA polymerase promoter with variable length coding strands. These were used to probe the effect of araC on promoter recognition, initiation, and elongation by T7 RNA polymerase in vitro. Substitutions in the central promoter region had no observable effect on RNA polymerase binding, initiation rate, or transcriptional output. Coding strand substitutions defined an area of high sensitivity in the initiation region where miss-starts, primer slippage, and an inability to escape from abortive cycling occur depending on the position substituted. Substitutions after position 10 had little effect on transcription output. These highly variable, position dependent effects indicate a narrow window of vulnerability where transcription output is severely reduced (~100-fold) by a subtle DNA lesion that has little or no consequence when situated elsewhere in these small coding units.

Cytosine arabinoside (1-β-D-arabinofuranosylcytosine, araC) is a potent antileukemic agent as well as a general inhibitor of DNA replication (Frei et al., 1969; Cozzarelli, 1977; Heintz & Hamlin, 1983; Wright & Brown, 1990). Its mechanism of action involves conversion to araCTP, competitive inhibition of DNA polymerase (Furth & Cohen, 1968; Dicioccio & Srivastava, 1977), and misincorporation into DNA, where it functions as an inhibitor of further chain elongation and subsequently as a template lesion (Townsend & Cheng, 1987; Mikita & Beardsley, 1988; Perrino & Mekosh, 1992). AraC (Figure 1) is an altered form of the normal cytosine nucleosides. It differs from dC at $C_{2'}$ of the sugar moiety where it carries an additional -OH *trans* to the -OH at $C_{3'}$. As such, it is a configurational isomer of rC. In araC, the major structural elements of normal nucleosides including the base, pentose ring, and 3′ and 5′ hydroxyls are all present, yet the anomaly in the sugar ring often has profound effects on DNA function. We were interested in the structural and functional alterations araC introduces into DNA from the standpoint of both elucidating in fine detail the molecular mechanism(s) of this antileukemic agent as well as investigating its usefulness as a perturbation probe of protein/DNA interactions.

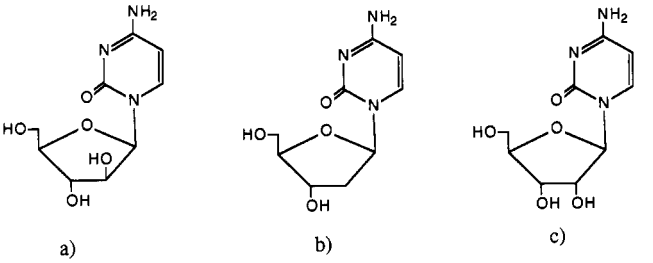


FIGURE 1: Chemical structures of cytosine nucleosides: (a) cytosine arabinoside, (b) deoxycytidine, and (c) cytidine.

We previously developed a chemical methodology for the synthesis of DNA oligomers containing araC which allowed us to study its site specific effects on duplex stability and chemical reactivity (Beardsley et al., 1988), as well as its effects on DNA ligase and DNA polymerase activity (Mikita & Beardsley, 1988). This approach allows focus on structural and functional effects which result from araC misincorporation separately from its impact as a nucleoside triphosphate. The DNA polymerase studies, in addition to other observations, showed that araC in DNA templates could have an inhibitory effect on polymerase bypass. As a template lesion, there exists the potential for interference with other aspects of DNA metabolism, such as transcription.

† Supported by NIH/NCI Grant CA 42300.

* Address correspondence to this author at the Department of Pediatrics, Yale University School of Medicine, P.O. Box 208064, New Haven, CT 06520-8064.

‡ Department of Molecular Biophysics and Biochemistry.

§ Deparments of Pediatrics and Pharmacology.

® Abstract published in *Advance ACS Abstracts,* July 15, 1994.

0006-2960/94/0433-9195$04.50/0

Furthermore, an X-ray crystallographic and molecular modeling study showed that while araC-G base pairs could be accommodated into a B-form DNA duplex, strong steric clashes occurred when araC-G base pairs were built into an A-form helix environment (Gao et al., 1991). Since the DNA/RNA hybrids that are generated during transcription are believed to be of an A-form, this provided an additional basis for believing that araC might have an effect on transcription. In addition, it has often been observed that the functional consequences of a DNA lesion can vary with local sequence context (Clark & Beardsley, 1989; Evans et al., 1993; Maccabee et al., 1994). It was therefore reasonable to consider that different helix types, with their different attendant steric constraints, might also modulate the functional consequences of a particular DNA lesion. Finally, the substantial literature on the effects of DNA damage upon replication, recombination, and repair has contributed greatly to mechanistic understanding of these processes. Studies of the effects of DNA structural lesions on transcription, though currently far less extensive, also promise important mechanistic insights into the complex processes involved.

In work described in this paper, we have characterized a DNA/RNA hybrid containing an araC-G base pair, comparing stability effects and helix type to a DNA duplex of identical sequence. We have also made araC substitutions in DNA duplexes containing a T7 promoter sequence and an artificial coding strand. These substrates were used to study the effects of araC substitution at different positions on promoter recognition, initiation, and elongation by T7 RNA polymerase in vitro. Each of these stages of the transcription process is experimentally accessible and defines distinct steps which RNA polymerase must carry out to successfully transcribe RNA.

MATERIALS AND METHODS

Chemicals. Cytosine-β-D-arabinoside (araC) was purchased from Sigma. 1,10-Phenanthroline (1,10-OP), cupric sulfate ($CuSO_4$), 3-mercaptopropionic acid (MPA), and 2,9-dimethylphenanthroline (DMOP) were purchased from Aldrich. All other chemicals were reagent grade or better. Unlabeled nucleotides were from Sigma and checked by HPLC for purity (>95%) prior to use. [γ-^{32}P]GTP and -ATP at 5000 Ci/mmol and [α-^{32}P]CTP, -UTP, and -ATP at 400 Ci/mmol were purchased from Amersham.

Enzymes. T7 RNA polymerase was purchased from Stratagene as a solution of 50 000 units/mL with a specific activity of 75 000 units/mg. A unit is defined as the amount of T7 RNA polymerase which incorporates 1 nM of nucleoside triphosphate into higher molecular weight material in 60 min. T4 polynucleotide kinase, RNase H, and T4 DNA ligase were purchased from New England Biolabs.

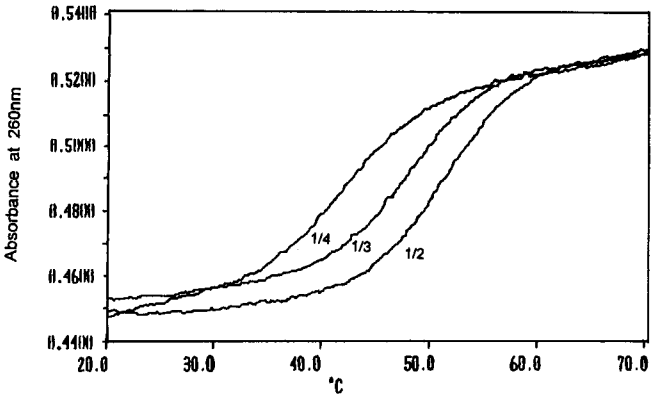
DNA Substrates. DNA and RNA oligonucleotides, the sequences of which are given in Figures 2 and 4, were synthesized at the Protein and Nucleic Acid facility at Yale University. The DNA oligomers containing araC substitutions required a protected araC phosphoramidite, the chemical synthesis of which was performed in our laboratory and has been described previously (Beardsley et al., 1988). The oligonucleotides were purified to single-base resolution on denaturing poly(acrylamide) gels followed by column desalting and then storage in the appropriate buffer. DNA concentrations were determined by measuring the UV absorbance at 260 nm.

Melting Temperature Studies. The absorbance vs temperature profiles for the DNA/RNA and DNA/DNA duplexes were performed on a Perkin-Elmer Lambda 2 UV/vis spectrometer equipped with water-jacketed cuvette holders. Temperature ramping was controlled with a Perkin-Elmer Peltier system. For each sample, approximately 0.5 OD unit of the appropriate oligomers was annealed in a 1 to 1 ratio in a buffer of 100 mM Tris, 100 mM NaCl to produce the duplexes which are shown in Figure 2. The absorbance at 260 nm was measured for each sample over a range from 20 to 70 °C. The temperature was increased linearly at a rate of 2 °C/min. The melting temperatures (T_m) where determined from a derivative plot of the absorbance vs temperature profiles.

Copper Phenanthroline Reactions. Stocks (20×) of the chemical cleaving reagents (1,10-OP, $CuSO_4$, MPA) were prepared in 100 mM Tris, pH 7.5. The stop reagent (DMOP) was 20× in a 50/50 solution of 100 mM Tris/100% EtOH. The reaction conditions were typically 0.6 μM in duplex DNA or RNA/DNA hybrid (one strand ^{32}P labeled) which had been annealed in 100 mM Tris, pH 7.5. 1,10-OP and $CuSO_4$ were present at 200 and 40 μM, respectively, and the reaction was initiated by the addition of MPA at a concentration of 4 mM. Aliquots were taken at various time points, and the reaction was stopped by addition of DMOP at a concentration of 2 mM. The reactions were normally carried out at 21 °C (although a full range of temperatures from 14 to 60 °C was tested to convince us of the utility of this reagent as a tool for probing helix variation within duplex nucleic acids). The products of the cleavage reactions were then denatured by heating to 90 °C for 3 min prior to fractionation on a denaturing poly(acrylamide) gel.

In Vitro Transcription Reactions. The standard transcription reaction was carried out in a 25 μL volume with 0.6 μM DNA. The buffer conditions for the reaction were 40 mM Tris, pH 8, 8 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 30 mM DTT, and 0.2 μg/μL BSA. The rNTPs were present at 800 μM. In general, the initiating nucleotide, GTP, carried the ^{32}P label in these reactions. GTP was labeled at the γ position so that only a single label was incorporated per message no matter how long or short, since only the initiating nucleotide of each message retains the γ phosphate. In some reactions, the label was switched to the α position of CTP, ATP, or UTP to assist in the identification of particular abortive transcripts (see individual figure legends for when this occurred). T7 RNA polymerase, 50 units (0.3 μM), was added to start the reactions. All transcription reactions were carried out at 37 °C. Reactions comparing activity on the various substrates in a given figure were done under identical conditions at the same time. Aliquots were taken at times indicated in individual figure legends and added to a dye/EDTA/formamide mix to stop the reactions. These samples were then heated to 90 °C for 4 min prior to fractionation on a denaturing poly(acrylamide) gel.

Electrophoresis and Densitometry. Products analyzed by denaturing poly(acrylamide) electrophoresis were separated on 0.6 mm 18% poly(acrylamide) [19:1 (w/w) acrylamide: *N,N'*-methylenebis(acrylamide)] gel containing 7 M urea. Denaturing gels were generally run at 1000 V for 3–4 h. Products analyzed under native conditions were separated on the same percentage and thickness gels but without urea. Sample loading buffer for native gels contained no formamide. The native gels were run at low voltage (200–400 V) to keep the gel temperature below 30 °C. Some native gels were run in the cold room to ensure even lower gel temperature (<20 °C). Gels were typically not dried prior to autoradiography due to the short exposure time (20–45 min) the reaction



	DNA-RNA HYBRIDS	Tm		DNA DUPLEXES	Tm
(1/2)	5' CGCCAGUACCG 3' GCGGTCATGGC	51.6 C	(5/2)	5' CGCCAGTACCG 3' GCGGTCATGGC	52.2° C
(1/3)	5' CGCCA GUACCG 3' GCGGT^{ara}CATGGC	48.1° C	(5/3)	5' CGCCA GTACCG 3' GCGGT^{ara}CATGGC	48.0° C
(1/4)	5' CGCCAGUACCG 3' GCGGTTATGGC	41.8° C	(5/4)	5' CGCCAGTACCG 3' GCGGTTATGGC	40.0° C

FIGURE 2: Sequence and melting temperatures (T_m) for the control and substituted DNA/RNA hybrids and DNA duplexes used in the experiments described here and in Figure 3. Conditions for measuring the absorbance vs temperature profiles from which the T_ms were calculated are given in Materials and Methods. The numbers in parentheses to the left of each hybrid or duplex identify them by a number assigned to each individual strand of the duplex. The first number in the parentheses identifies the top strand; the second number identifies the bottom strand. Absorbance vs temperature profiles are shown for the hybrids only, as the values for the DNA duplexes came from a larger study on DNA duplex effects which will be published elsewhere.

conditions permitted. Quantitation of product bands on exposed XAR-5 film was determined by densitometric scanning and integration using an LKB ultrascan laser densitometer.

RESULTS

Duplex Stability. To assess the effects of araC substitution in two helically different duplex environments, we first investigated the effect of an araC-G base pair on DNA/RNA hybrid vs DNA duplex stability. Figure 2 shows the sequences of the DNA/RNA hybrids and DNA duplexes and lists the melting temperatures (T_m) determined from the absorbance (260 nm) vs temperature profiles. The central base pair in these two sets of otherwise identical 11-mer duplexes is either dC-G, araC-G, or dT-G (where G is rG in the hybrid RNA strand and dG in the identically sequenced DNA duplex strand). What is immediately apparent is the similarity in T_m for both the parent duplexes as well as the substituted variants. If the araC-G base pair introduced a greater distortion into the hybrid compared to the DNA duplex, we would expect to see a greater depression in T_m for the hybrid. Comparison to the drop in T_m recorded with the dT-G(dG) mispair-containing duplexes indicates that the araC-G(dG) base pairs have a destabilizing effect that is $\sim^1/_3$ of that introduced by the mispairs. These results indicate that both the DNA duplex and the DNA/RNA hybrid can accommodate an araC-G(dG) base pair with only a moderate and equivalent loss of stability.

Helix Type: Chemical Cleavage. To be certain that the DNA duplexes and DNA/RNA hybrids were actually of different helix types, we next investigated their susceptibility to the nucleic acid cleavage reagent copper phenanthroline ($Cu(OP)_2$). The activity of $Cu(OP)_2$ toward nucleic acids has been shown to be sensitive to overall helix geometry (Sigman, 1986). As such, it has been used successfully to discriminate B- from A-form helices. The proposed reaction mechanism involves the reagent cleaving DNA or RNA by attacking the sugar moiety via the minor groove of the duplex. It is believed that the changing shape of the minor groove with helix type and the resultant difference in diffusable access and coordination of the precleavage complex are responsible for cleavage of a B-form helix 3–5 times more efficiently than an A-form helix (Sigman et al., 1993).

Figure 3 top, shows the results of $Cu(OP)_2$ cleavage reactions performed on the duplexes shown in Figure 2. The cleaved duplexes have been fractionated on a denaturing poly(acrylamide) gel. For these duplexes, the rate of cleavage for the labeled DNA strand should vary if there is a difference in helix type (Sigman et al., 1993). The figure clearly shows that the DNA strand common to both duplexes varies in the degree to which it is degraded, depending on whether it is paired with a DNA strand (strong cleavage) or an RNA strand (weak cleavage). That the DNA/DNA duplexes are readily cleaved, while the DNA/RNA duplexes are more resistant to cleavage, is consistent with the DNA duplexes being of the B-form and the DNA/RNA hybrids being of the A-form.

Helix Type: Gel Mobilities. Others have noted the reduced mobility of A-form DNA/RNA hybrids relative to identically sequenced B-form DNA duplexes on non-denaturing poly(acrylamide) gels (Bhattacharyya et al., 1990; Roberts & Crothers, 1992). These mobility differences have been shown to correlate with circular dichroism determinations of the A- and B-helix forms. As Figure 3, bottom, shows, we also observe mobility differences for the DNA/RNA hybrids compared to the identically sequenced DNA duplexes. The more slowly migrating bands in the hybrid lanes were also susceptible to RNase H cleavage (data not shown), further indicating that the observed mobility shift was indeed due to differences in helix type, rather than to other possibilities such as triplex formation. These data also support the conclusion that our DNA/RNA duplexes are in an A-conformation while the DNA/DNA duplexes are in a B-conformation.

In Vitro Transcription: 38-mer Duplexes. Figure 4 shows the sequence of the DNA duplexes used in our in vitro transcription assays with T7 RNA polymerase. These substrates contain a T7 promoter and a 15-, 16-, or 21-nucleotide double-stranded coding region. The arrows show where single araC substitutions have been made at positions in either the promoter or the coding strand. The duplexes are designated as to their length, with the position of araC substitution shown in parentheses. These substrates allow us to place the araC substitution at sites along the DNA which correspond to the functional stages of promoter recognition, initiation, and the start of processive elongation.

Control. Figure 5, 38(C) lanes, shows the results of a steady state in vitro transcription reaction using T7 RNA polymerase and the control 38-mer duplex. Products generated using the control substrate, 38(C), clearly illustrate the early dissociative phase of synthesis, characterized by a large accumulation of dimer, trimer, tetramer, and pentamer abortive products. There is a doublet of dimer abortive products which results from initiation occurring at two different start positions. The upper band of this doublet is a GG dimer that results from the correct +1 start. The lower band is a GA dimer that results from a +2 start (i.e., initiation occurring from position 2). A small amount of +2 trimer is also seen, but beyond that, there are almost no larger products which have originated from this

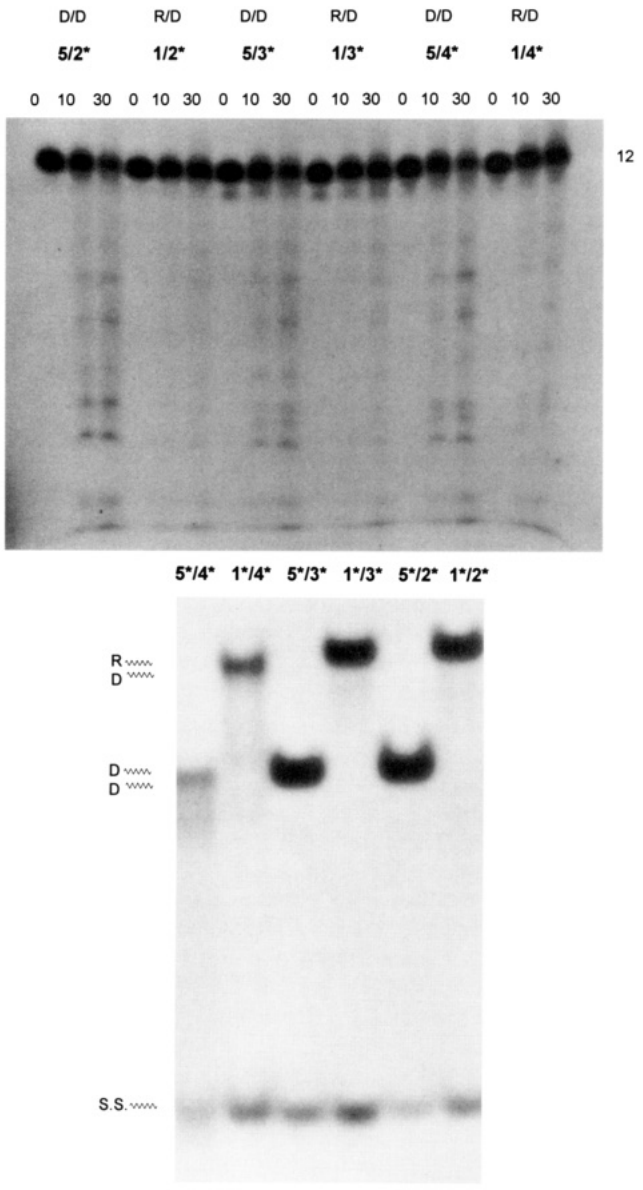


FIGURE 3: (top) Products of the $Cu(OP)_2$ cleavage reactions for the RNA/DNA hybrids and DNA duplexes, the sequences and identity of which are shown in Figure 2. Only one strand in each duplex is labeled with ^{32}P (marked with an asterisk). D/D designates a DNA duplex; R/D designates an RNA/DNA hybrid. Aliquots were taken at 0, 10, and 30 min, denatured, and fractionated on a poly(acrylamide) gel containing urea. Reaction conditions, electrophoresis, and autoradiography were as described in Materials and Methods. (bottom) Comparison of the duplex mobilities of the RNA/DNA hybrids and the identically sequenced DNA duplexes run out on a non-denaturing poly(acrylamide) gel. Both strands of each duplex are ^{32}P-end labeled (*). The bands are identified as to whether they correspond to single strand (S,S), DNA duplex (DD), or RNA/DNA hybrid (RD). The bands corresponding to the mismatch-containing duplexes (5*/4* and 1*/4*) appear fainter because these less stable duplexes are in equilibrium with partial duplex forms as indicated by the smearing below the duplex bands (running the gel at sufficiently lower temperature eliminates the smearing). The temperature of this gel during electrophoresis was approximately 30 °C. Additional information is given in Materials and Methods.

alternative, low-frequency initiation pathway. Other experiments show that such larger products originating from a +2 start would be resolvable on a gel of this type (see Figure 6, for example). At the end of abortive cycling, there is a switch to a more processive phase of synthesis such that the next products found to accumulate are full length runoff transcripts. These transcripts are a combination of the coded 16-mer message plus a more abundant 17-mer which is most likely the result of promiscuous end addition to the 16-mer message. This heterogeneity in runoff transcripts is often seen in T7 in vitro transcription assays (Martin et al., 1988; Milligan et al., 1987).

Since each product (abortive or full length) is the result of a single initiation event, this means that the bands less than full length are dead end products which will not be further extended. The ratio of the amount of product observed at each position, as determined by densitometry, to the sum of all products equal to or greater in length than that product describes the frequency of termination at each position of the message (Martin & Coleman, 1988). Table 1 lists the termination frequencies for 38(C) as well as for many of the other substrates used in this study. With 38(C), we found that 22% of the trimer product was not further extended. The termination frequencies for the tetramer and pentamer products were 38% and 60%, respectively, but only 13% and 4%, respectively, for the hexamer and heptamer products. For this substrate, the switch from abortive cycling to processive synthesis occurs after the insertion of the sixth or seventh nucleotide. The full length transcripts constituted only 15% of the originally initiated material. This low level of full length production is not unusual for a T7 promoter in assays of this type (Milligan et al., 1987; Martin et al., 1988).

Promoter Region Substitutions. Figure 5 also shows the results of in vitro transcription reactions using duplexes containing araC substitutions in the –12 to –5 region of the promoter. The sites of dC to araC substitutions in this region of the promoter were chosen on the basis of results from both footprinting (Muller et al., 1989) and interference experiments (Jorgenson et al., 1991) which suggest DNA/protein contacts at these locations. Figure 5 shows that the promoter substitutions at positions –12, –9, and –5 had no observable effect on promoter utilization as measured by the initiation rate, the pattern of short abortive products generated, or the production of full length products. We also made a dC to araC substitution at position –7 when more recent studies indicated this position to be an important determinant of promoter binding and activity (Ikeda et al., 1992; Diaz et al., 1993). Like the other sites of araC substitution, this site also failed to have any observable effect on activity as measured in this type of assay (data not shown).

Coding Strand Substitutions. Figure 6, top, shows the results of the steady state in vitro transcription reactions using the 38-mer duplexes containing araC substitutions in the coding strand. Products generated with the 38(1) duplex showed a change in band pattern and intensity among the shorter abortive products as well as the full length runoff products. This anomalous band pattern could arise from either misincorporation of incorrect nucleotides and/or initiation occurring from a different position(s). In order to distinguish between these possibilities, we first identified the sequence of these new bands by consecutively using ^{32}P-labeled GTP, ATP, CTP, and UTP in separate experiments (Figure 6, bottom left). By observing the position at which the different labeled nucleotides were incorporated, we were able to identify the sequence of all the bands up to the heptamer level. This is possible because of the excellent separation of material in this region of the gel. For example, CTP and UTP were incorporated as the fourth and seventh nucleotide, respectively, with the 38(C) duplex. But with the 38(1) duplex, CTP and UTP were respectively the third and sixth nucleotide incorporated. ATP is the third nucleotide incorporated with 38-

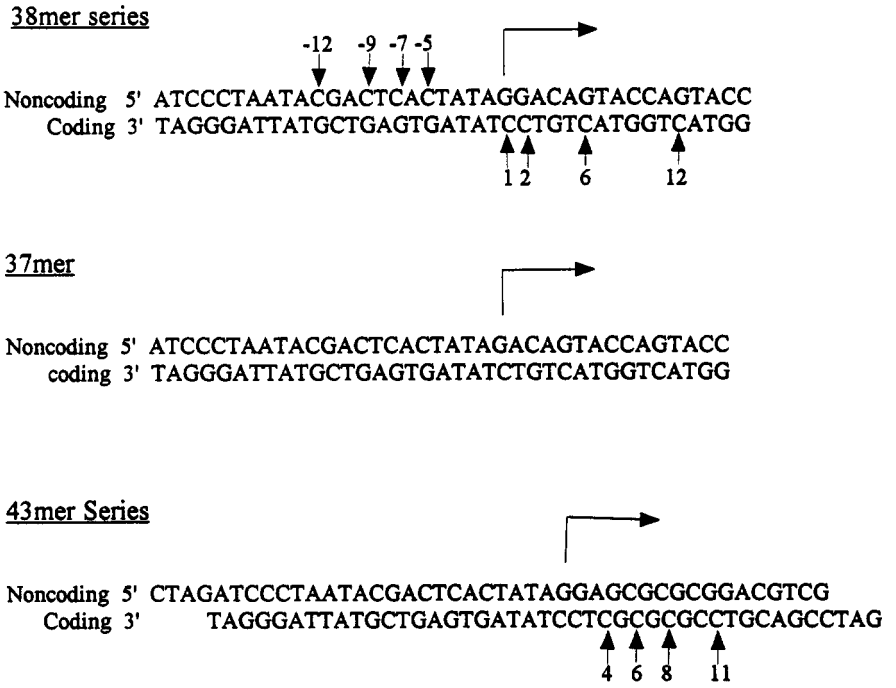


FIGURE 4: Sequence and sites of substitution for the DNA substrates used in the transcription experiments. The numbered arrows indicate the positions where dC to araC substitutions were made. For example, substitution at −12 in the 38-mer substrate is designated 38(−12), substitution at +4 in the 43-mer substrate is designated 43(4), and so on. Only one substitution was made per duplex. The bent arrow indicates the normal +1 start position and direction of transcription.

(C), but it is the second nucleotide incorporated with 38(1). These data indicate that it is not through misincorporation of incorrect nucleotides that the anomalous band pattern arises with 38(1) but rather from the polymerase initiating the message almost exclusively from position 2(+2 start) of the coding strand. This explanation accounts for the single nucleotide shift in size and sequence of the abortive products. A low frequency of the normal +1 start accounts for the rest of the bands.

Initiation rate, abortive transcript profile, and full length message yields have been shown to be highly sensitive to the sequence of the first few nucleotides of the coding strand (Milligan et al., 1987; Martin et al., 1988). For example, a message that begins with GGG will result in different abortive product amounts and yields of full length message than a message that begins with GCC. Since the +1 start initiates with GGA... and the +2 start initiates with GAC..., it was important to determine if the termination frequencies of the abortive products observed along this now dominant +2 start pathway arose because of further downstream perturbations introduced from having araC at position 1 or were simply the result of the different initiating sequence associated with the +2 start. The data in Figure 6, bottom right, resolve this issue. The figure shows a single-lane comparison between the 38(C), 38(1), 38(2), and 37(C) duplexes. The 37(C) duplex differs from 38(C) in that it lacks the first nucleotide of the 38(C) coding sequence (see Figure 4 for sequences). As such, the +1 start sequence on 37(C) is the same as the +2 start sequence with the 38(C) duplex (i.e., both messages initiate with GAC...). If we compare the pattern and intensity of the trimer to hexamer abortive products from the 37(C) duplex to the trimer to hexamer products that are generated only from 38(1) +2 start pathway, we see that they are essentially identical. This is further illustrated by the values listed in Table 1 which show that the 38(1) +2 start abortive product frequencies are essentially the same as the abortive product frequencies with the 37(C) duplex. This shows that the termination frequencies at positions along the 38(1) +2 start pathway are due to the new initiating sequence and not because of additional downstream effects resulting from araC at position 1. It is also clear from the data in Figure 6, bottom right, that this start sequence is less efficient at producing full length material than the 38(C) +1 start sequence. Much of the reduction in full length product observed with 38(1) is therefore the result of the polymerase being forced to initiate from this less productive +2 start sequence. The greater heterogeneity in full length products seen with the 38(1) duplex is a result of some +1-initiated material combined with the +2 start material. Generation of full length products from 38(1) is decreased about 3-fold compared to that from 38(C).

Products generated with the 38(2) duplex (Figure 6) included products that result from both +1 and +2 starts, though the intensity of the +2 start trimer and tetramer abortive products indicates that the majority of starts occur along the +2 pathway. However, determining the termination frequencies for the early abortive products along the +1 and +2 start pathways is difficult with this substrate because of overlaps in the gel pattern at the pentamer. Accordingly, no values are assigned in Table 1. It is therefore difficult to determine if one or both of these start pathways have local changes in termination frequencies. However the low abundance of the +1 trimer relative to the +1 tetramer observed with both the 38(2) and 38(1) substrates suggests that termination frequencies along the +1 pathway, at least, are altered by both these araC substitutions.

Because substitution at position 1 resulted in a shift to a +2 start, we wondered if araC substitution at position 2 would result in +3 starts. To check this, we utilized γ-^{32}P-labeled ATP, since the initiating nucleotide for a +3 start would be ATP. The result of this experiment showed that there were essentially no +3 starts (<1%) with 38(C), a low level of <5% for both 37(C) and 38(1), and an increase to about 10% +3 starts with the 38(2) substrate (data not shown). There is a reduced level of initiation that is consistently seen with the 38(2) substrate compared to the other substrates in this series. A 10–12-fold reduction in the production of full length material

Table 1: Termination Frequencies for Individual Duplex Products[a]

abortive products	38(C)	38(1) +1 start	38(1) +2 start	37(C)	38(6)	38(12)	43(C)	43(6)	43(8)
trimer	22%	nd[b]	61%	61%	21%	23%	21%	20%	16%
tetramer	38%	53%	59%	58%	42%	37%	22%	14%	20%
pentamer	59%	34%	50%[c]	48%	>90%	54%	43%	62%	41%
hexamer	13%	16%	35%	32%	>90%	11%	51%	70%	34%
heptamer	4%	17%	12%	16%		3%	5%	nd[b]	10%
octamer							13%	45%	25%
nonamer							15%	80%	73%
undecamer	1.5%					32%			
dodecamer	1%					36%			

[a] This table lists the percent dissociation at each position along the message for many of the substrates used in this study. These values are the ratio of the amount of product observed at each position to the sum of all products equal to or greater in length than that product. These values were determined by densitometry of the autoradiograms shown in Figures 6, top, and 7, top. Densitometry was performed as described in Materials and Methods.[b] Value not determined. [c] Because of overlap of the +1 and +2 pentamers, the assumption is made that 80% of this material results from the +2 start. This is a reasonable assumption because of the low level of +1 starts with the 38(1) substrate.

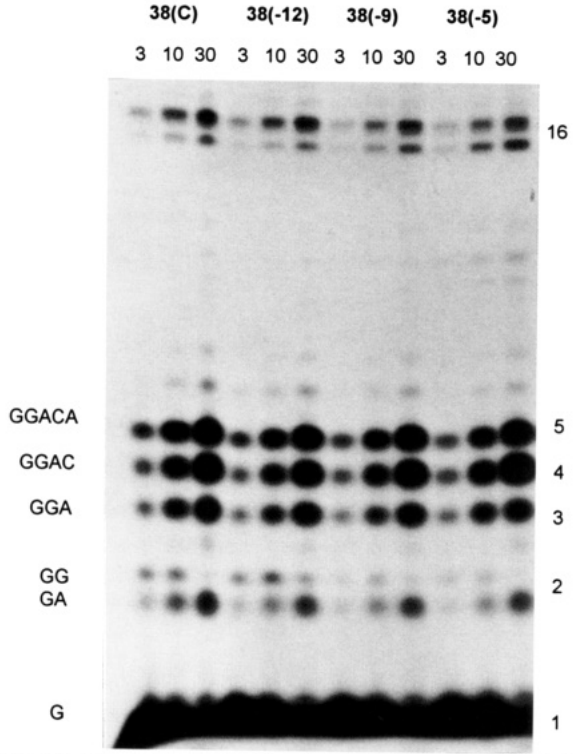


FIGURE 5: Promoter substitutions have no observable effect. The gel shows the products of the steady state transcription experiments for the 38(C), 38(–12), 38(–9), and 38(–5) substrates. RNA synthesis off of these templates initiates with the introduction of GTP. For these and subsequent transcription reactions, the label is at the γ position of GTP unless otherwise indicated. In this way, a single label is incorporated per message no matter how long or short since only the initiating nucleotide possesses the γ phosphate. Aliquots were taken at 3, 10, and 30 min for each substrate. The numbers on the right of the autoradiogram indicate the size of the major abortive products, the sequences of which are shown on the left. Reactions were carried out under standard conditions described in Materials and Methods.

is observed. These full length products are a mix of +1- and +2-initiated material.

When substitution is at position 6, the polymerase initiates normally but fails to escape from abortive cycling (Figure 6, top). While the qualitative pattern of short abortive products is the same as with the control, full length message production is reduced 100-fold. This same level of polymerase arrest was seen when either UTP or CTP (incorporated at position 4 and 7, respectively) was omitted from the reaction mixture with the control substrate (data not shown). As Table 1 shows, the termination frequencies for the 38(6) pentamer and hexamer abortive products were markedly increased compared to those for the control. That the abortive frequencies for the trimer and tetramer products were unaltered from the control values indicates a strong local effect centered at position 5 and 6. Misincorporation of incorrect nucleotides does not appear to play a role in this inhibition as the mobility of the pentamer and hexamer abortive products is unaltered from that of the control. Given the resolution of the gel in this region, a misincorporated nucleotide at position 5 or 6 in the abortive product would observably change the gel mobility of the abortive transcript.[1]

The last set of lanes in Figure 6, top, shows the result of substitution at position 12. When the polymerase has reached this position of the coding strand, it has ceased abortive cycling and has begun the transition to a more processive mode of synthesis characterized by few intermediate bands between the last major abortive product (the pentamer) and the runoff transcripts. To control for sequence context effects, we made the 3′ and 5′ nearest neighbors around position 12 the same as around position 6. The polymerase initiated correctly with the 38(12) substrate, but there are two new abortive products seen at position 11 and 12. Because the termination frequencies at these positions on the control are normally very low (1.5% and 1%, respectively), the 15–30-fold increase in termination frequency observed here has the effect of reducing runoff transcripts by only 2-fold compared to the 38(C) yield.

In Vitro Transcription: 43-mer Duplexes. Given the varied response of the polymerase to these different araC coding strand substitutions, we made another substrate series in order to investigate this variable sensitivity more closely. The sequence of the coding strand was changed to probe other sites of dC to araC substitution. As shown in Figure 4, substitutions were made at positions 4, 6, 8, and 11 in the 43-mer duplex. In addition to the new substitution sites, substitution at position 6 was made again to see if the change in local sequence context of this substrate would alter the strong polymerase arrest seen with the previous coding strand sequence. Figure 7, top, shows the products of the steady state transcription reactions for the control and substituted 43-mer duplexes.

Control. If we first compare the products from 43(C) with those from 38(C) shown in Figure 6, top, we see the same

[1] The ability to resolve abortive products of the same size but different sequence in this region of the gel is clearly demonstrated by our ability to resolve the +1 and +2 start abortive products from the 38(1) and 38(2) duplexes. A low percentage (<10%) of misincorporation, however, cannot be ruled out because the resultant minor bands next to the dominant pentamer and hexamer bands would likely not be resolvable into observable doublets.

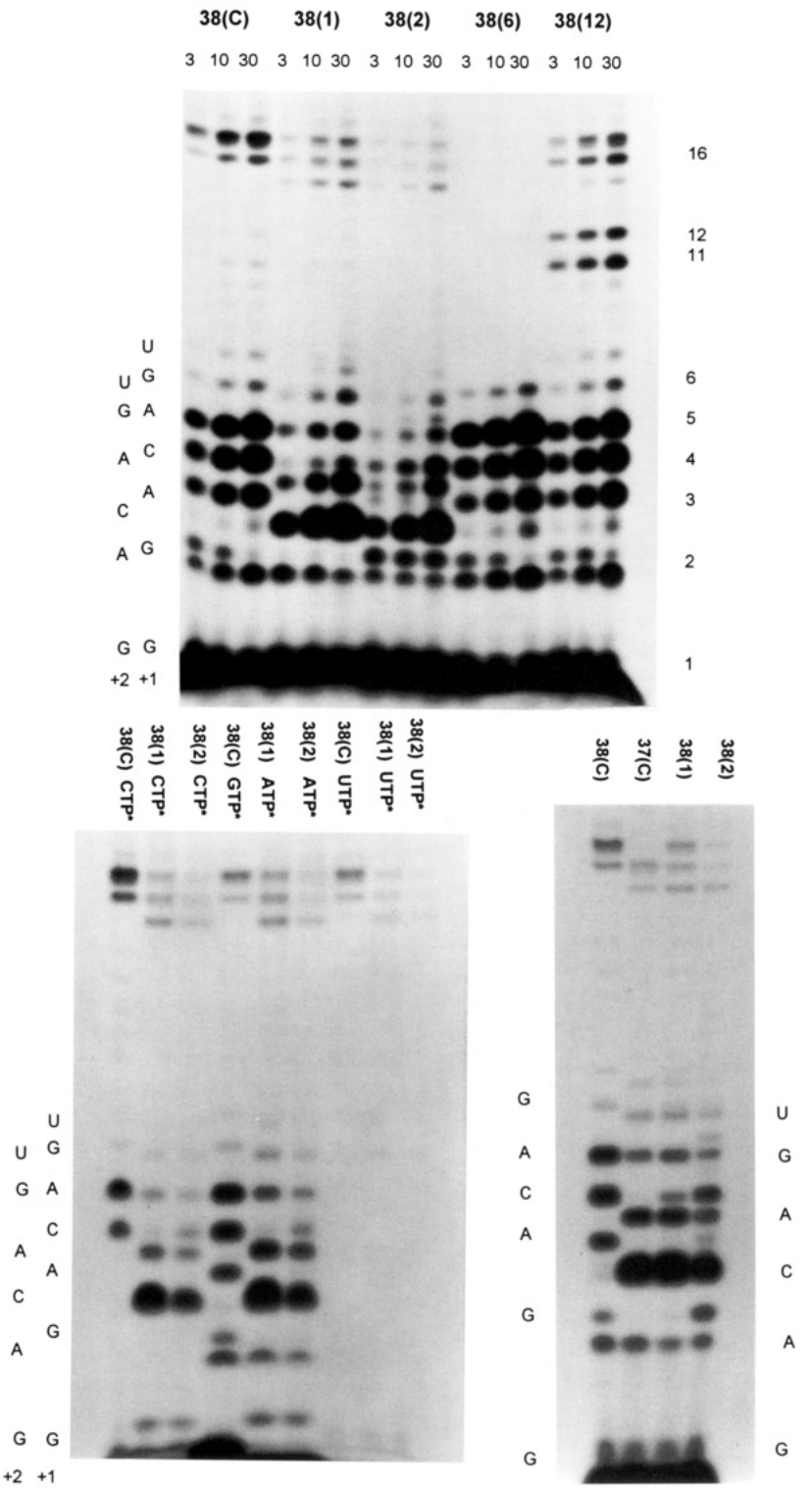


FIGURE 6: (top) Effects of araC coding strand substitutions are highly position dependent. This panel shows the products of the steady state transcription reactions carried out as described in Materials and Methods. Aliquots were taken at 3, 10, and 30 min for each substrate. The letters on the left show the sequence of abortive products along both the +1 and +2 pathways. For example, the +1 start trimer is GGA, the +2 start trimer is GAC, and so on. The numbers on the right refer to the size of the products that have initiated by the +1 start pathway. Electrophoresis and autoradiography were as stated in Materials and Methods. (bottom left) Successively changing the labeled NTP reveals that initiation occurs predominantly from position 2 with the 38(1) and 38(2) substrates. The panel shows a single-time point comparison of the transcription products with the 38(C), 38(1), and 38(2) substrates. The ^{32}P label is successively placed on the γ position of GTP or the α position of ATP, CTP, or UTP as so designated at the top of each lane. The sequence of abortive products along both the +1 and +2 start pathways is shown on the left-hand side. With the exception of the label switching, the reactions were carried out under the standard conditions listed in Materials and Methods. (bottom right) Altered downstream abortive product frequencies are the result of the +2 initiation sequence and not due to downstream effects of araC. This panel shows a single-time point comparison of the products generated from the 38(C), 37(C), 38(1), and 38(2) substrates. The 37(C) substrate +1 pathway is identical in sequence to the 38(C) +2 pathway. Standard reaction conditions were followed with a single aliquot taken at 15 min. The letters on the left of the autoradiogram indicate the sequence of the abortive products originating from a +1 start with the 38(C). The sequence on the right refers to the +1 start abortive products from 37(C), which are the same as the +2 start abortive products from 38(1) and 38(2).

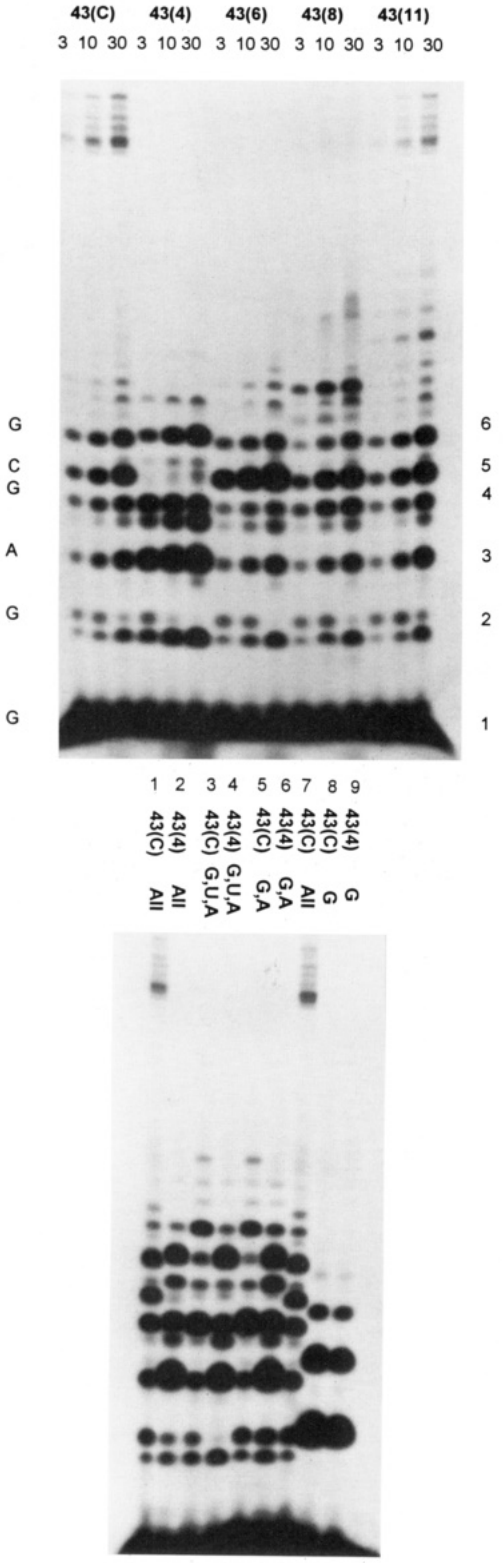
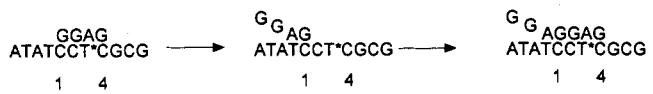


FIGURE 7: (top) Coding strand substitutions in the 43-mer duplex series. This panel shows the products of the steady state transcription reactions carried out as described. Aliquots were taken at 3, 10, and 30 min. The numbers on the right designate the size of the abortive products for the +1 start pathway. The sequence of the abortive products are indicated on the left. Electrophoresis and autoradiography were as described in Materials and Methods. (bottom) By leaving out selective nucleotides, it is shown that for the 43(4) substrate the products larger than the tetramer are the result of primer slippage. This panel shows a single-time point comparison between the 43(C) and 43(4) substrates under conditions of full and selectively depleted NTPs as indicated at the top of each lane.

general features of an abortive cycling phase followed by a transition to a more processive mode of synthesis. However, because the coding strand sequences differ after position 3, this comparison illustrates once again of how changes of sequence in the initiation region of the coding strand can alter the abortive frequency of these short transcripts (compare the termination frequencies for the 37(C), 38(C), and 43(C) tetramer to heptamer products listed in Table 1). The 43(C) duplex also produces a more heterogeneous pattern of runoff transcripts, many of which are larger than that dictated by the coding strand length. Unlike with the 38(C) duplex where it was easy to count the well-resolved individual bands up to the end of the message, this is a more difficult task with the products of the 43(C) duplex. Between the abortive hexamer transcript and the first major runoff transcript, there should be a difference of 17 incorporated nucleotides. Though the bands in this region of the gel are very faint even after longer periods of film exposure, we often count 18–19 bands between these two positions. Given the alternating CGCG content of the coding strand, it seemed likely that these extra bands would be the result of internal transcript slippage. However, leaving UTP out of the transcription reaction (data not shown) results in a single dark band (signifying polymerase arrest) at/near position 16 (the first site for coded UTP insertion) and two more bands at/near positions 19 and 20 (position 19 is the second site for coded UTP insertion). That there is heterogeneity in full length material but only a single band at/near position 16 in the minus UTP experiment suggests that most of this end heterogeneity arises because of end slippage or promiscuous addition (Milligan et al., 1987; Martin et al., 1988). The doublet at/near position 19 (which is close to the end of the message) in this same experiment provides support for such an end effect. However, a missing nucleotide experiment should result in an arrest band one nucleotide before the missing nucleotide insertion site. That would be position 15 on our coding strand, and the gel should show a 15-mer band. Because it is not clear in our experiments if this band corresponds to a 15-mer or 16-mer, it is possible that one nucleotide is also added internally because of a primer slippage event brought about by the alternating CG content of the central coding region. Full length or longer material constitutes only 14% of initiated message from the 43(C) duplex.

Substituted 43-mer Duplexes: Coding Strand Effects. As seen in Figure 7, top, substitution of araC at position 4 resulted in a complicated change in abortive products. While the accumulation of abortive products larger than the tetramer suggested that the polymerase traverses the site of araC substitution, the unusual mobilities of these products (suggesting some repetitive pattern) as well as the results from selective NTP depletion experiments indicated that these larger products were the result of araC-induced transcript slippage. The data supporting this conclusion are shown in Figure 7, bottom. When either CTP (the fifth nucleotide dictated by the coding strand) or both CTP and UTP (a possible misincorporated nucleotide that might account for the unusual mobilities) were left out of the reaction with the 43(4) duplex, there was no effect on abortive products generated. This shows that the abortive products larger than the tetramer were composed of only G and A. These products can be reconciled with the coding sequence if we assume the slippage mechanism shown in Scheme 1. Interestingly, omitting CTP or CTP and UTP from the reaction (Figure 7, bottom) with the control 43(C) substrate resulted in an abortive product pattern very similar to that produced using the 43(4) substrate, providing further support for this primer slippage mechanism. There is also some evidence that the +2 initiation pathway may contribute to some of the products (the +2 pathway dimer

Scheme 1: Slippage Mechanism[a]



[a] After incorporation of G opposite araC at position 4, the polymerase is unable to extend the message past this position of the coding strand. Because of a coincidence of coding strand sequence, the 3′ end of the tetramer abortive product is able to pair at −1 and +1 to initiate another round.

GA is more abundant than normal), though this is more difficult to resolve due to potential overlaps beyond the dimer. Since the larger products can be accounted for by the slippage mechanism, there is no evidence to suggest that the polymerase traverses position 4 by any start pathway. Densitometry shows that full length message production is decreased 100-fold.

Substitution at position 6 of this substrate had the same effect on full length message production as substitution at this position had with the 38-mer duplex. The polymerase initiates correctly, but the production of full length material was again decreased ~100-fold. As with the 38-mer substrate, there was no significant increase in +2 starts, nor was there any evidence of misincorporation of incorrect nucleotides at/near position 6. However, there were differences in local termination frequency near the substitution site as revealed in Figure 7, top, and the values listed in Table 1. The more gradual effect on polymerase arrest seen with the 43(6) substrate is likely the result of the different sequence context surrounding araC.

When substitution is made at position 8, in addition to an increased incidence of transcript termination at/near this site, the majority (60%) of material which is extended past position 8 aborted several nucleotides later. The reason for this is unclear. There may be a delayed derailment of synthesis, or it may be that the repetitive nature of the sequence in this part of the template combined with the perturbation introduced by araC results in local transcript slippage at/near position 8, giving the appearance of larger than expected abortive transcripts. There were no changes in the pattern of early abortive products. Full length products are decreased 12-fold.

The last template substitution is at position 11. There are no changes in the pattern of early abortive products, although there is a local increase in termination frequency at/near the substitution site. Unlike the position 8 substitution, this substitution did not induce (or give the appearance of inducing) a further derailment of products several nucleotides later. A 3-fold reduction in full length products was seen with this duplex.

Larger Substituted Constructs. The 43-mer series was also designed with 5′ overhangs at both ends so they could be ligated to larger coding strands. We were concerned that the small size of the coding strands might be contributing to the intensity of the substitution effects we were observing, possibly through the loss of stabilizing downstream contacts. The results from experiments with substrates where the coding strand had been lengthened to ~100 nucleotides showed the same response to the araC substitutions as was visible with the 43-mer substrates (data not shown). This indicated that the size of our substrates was not modulating the effects observed at araC substitution sites.

DISCUSSION

Our previous work has focused on the effects of araC misincorporation on DNA duplex stability and chemical reactivity (Beardsley et al., 1988), as well as its effects on DNA ligase and DNA polymerase activity (Mikita & Beardsley, 1988). In addition to other observations, the DNA polymerase studies showed that araC in a DNA template could have an inhibitory effect on polymerase bypass. This established a new component to the well-established inhibitory effects of araC as a competitive inhibitor with dCTP (Furth & Cohen, 1968; Dicioccio & Srivastava, 1977), as well as its "leaky" chain terminator effects upon incorporation at DNA primers (Mikita & Beardsley, 1988; Wright & Brown, 1990; Perrino & Mekosh, 1992). As a template lesion, there exists the potential for araC to interfere with other aspects of DNA metabolism, such as transcription.

This template effect was interesting from a structural standpoint because our earlier stability studies suggested only a mild destabilization of the DNA duplex (Beardsley et al., 1988). This view was further supported by an X-ray study (Gao et al., 1991) of an araC-containing B-form DNA duplex which revealed small but interesting duplex changes. In the X-ray study, araC was seen to pair normally with dG and the araC residue was conformationally superimposable with the dC residue of the control duplex. However, the araC-dG base pair was displaced slightly from the helical axis, resulting in a loss of stacking with neighboring base pairs. This loss of stacking energy is most likely what is being reflected by the drop in melting temperature observed in our experiments. In addition, there were also small changes in local torsion angles as well as twisting of the internucleotide phosphate linkage to araC. The steric clashes driving these changes appeared centered in the sugar–phosphate backbone. Like the dC residue in the control duplex, the arabinose sugar conformation was concluded to be $C_{2'}$ endo. In addition, the $C_{2'}$-OH of the arabinose sugar was seen to project into the major groove of the helix at a site that is normally quite hydrophobic. NMR studies (Pieters et al., 1989, 1990; Schweitzer et al., 1994) are largely consistent with the crystallographic study but reached different conclusions about the arabinose sugar conformation.

In summary, these structural studies on B-form DNA all indicate that the perturbations introduced by araC are slight to moderate in their effect, particularly compared to those changes introduced by many of the more bulky modifications of the DNA bases which grossly disturb duplex structure and stability (Basu & Essigmann, 1988). However, its inhibitory effects on DNA polymerase can be equal to or greater than some of these more aggressive structural lesions (Mikita & Beardsley, 1988). This combination of features makes the araC DNA structural lesion a potentially valuable probe for investigating small but critical DNA recognition elements that are utilized by DNA and RNA polymerase to carry out the template-directed synthesis of nucleic acids.

While there has been no X-ray or NMR structure of an araC-containing duplex demonstrating an A-form, a recent molecular modeling study (Gao et al., 1991) indicated that strong helix specific perturbations would arise if an araC-dG base pair were to exist in an A-form helix environment (all sugars $C_{3'}$ endo). These perturbations appeared to be much greater than those introduced by araC in B-form DNA. Since it has been observed that araC is not misincorporated into RNA to any appreciable degree (Kufe et al., 1980; Kufe & Spriggs, 1985), the authors of the modeling study hypothesized that steric effects may partially or wholly account for this exclusion, i.e., araCTP would not be introduced into RNA because it could not be made to fit into the A-form RNA/DNA hybrid which is believed to exist at the site of catalysis within the RNA polymerase ternary complex.

Initial consideration of the structural elements and functional groups involved in these clashes leads us to conclude that the same steric constraints would be operative if araC was in either strand of the hybrid (none of the clashes involved the ribose $C_{2'}$-OH). This is contingent on the assumption that the sugars of both strands are in the $C_{3'}$ endo conformation. However, the long-standing view that both strands of a RNA/DNA hybrid have the $C_{3'}$ endo sugar conformation has recently been called into question. While numerous X-ray structures (Arnott et al., 1968; Saenger, 1984; Wang et al., 1982; Egli et al., 1992) support the view that both strands of the hybrid are A-form (all sugars $C_{3'}$ endo), this view has become increasingly challenged by NMR (Chou et al., 1989; Jaishree et al., 1993; Federoff et al., 1993), circular dichroism (Steely et al., 1986; Roberts & Crothers, 1992), and gel mobility studies (Bhattacharyya et al., 1990; Roberts & Crothers, 1992) which indicate that the RNA strand is A-form but the DNA strand has a conformation in solution closer to the B-form, or at least intermediate between the B- and A-forms (i.e., sugars neither $C_{2'}$ nor $C_{3'}$ endo). This has become an area of considerable interest because of the role hybrids play in reverse transcription, transcription, and DNA replication (primase function).

We decided to directly assess the impact of an araC-G base pair on DNA/RNA hybrid stability. We introduced araC into the DNA strand of the hybrid, rather than the RNA strand, because our interest is in the structural and functional effects arising from the therapeutically relevant context of araC misincorporation into DNA. The results of these thermal stability measurements failed to indicate any difference in the magnitude of the perturbation an araC-G(dG) base pair introduces into these identically sequenced but helically different duplexes. The dT-G(dG) mispair was chosen for comparison because it is one of the most stable mispairs (Ke & Wartell, 1993; Modrich, 1987) and provides a reference for the relative impact of an araC-G base pair. This impact, by comparison, is moderate. However, the structural basis of duplex destabilization is different for these two base pairs. The mispair destabilization is centered in altered hydrogen bonding at the interface of the T-G(dG) bases, whereas the araC-G(dG) defect is related to a structural difference in the sugar–phosphate linkage which introduces local torsional angle changes as well as a loss in local base stacking energy, but it does not alter the hydrogen-bonding potential of the cytosine base (Gao et al., 1991; Schweitzer et al., 1994).

Our chemical probing of the different duplexes with $Cu(OP)_2$ as well as the mobility shifts observed on native gels is consistent with the hybrids being of an A-form and the DNA duplexes being of a B-form. But this level of probing can not resolve the fine structural detail relating to the shape of the DNA strand of the hybrid. It merely indicates that our findings are consistent with the duplexes being of different helix types. Since our melting temperature studies show that the presence of an araC-G(dG) base pair within these two helix contexts has a moderate, and equivalent, destabilizing effect, these results are more consistent with a structure for the hybrid where the sugars of the DNA strand do not have the $C_{3'}$ endo conformation.

Since the magnitude of the araC-introduced destabilization of the DNA/RNA hybrid was similar to, although not larger than, that of the DNA duplex, we felt it was appropriate to investigate the effects of araC-containing DNA on transcription since our previous studies had shown that such modest perturbations can have relatively profound biochemical effects. The moderate effect on base pair stability, the unaltered base moiety, and the small change in sugar structure that leaves the 3′ and 5′-OH groups unaltered all combine to make araC a potentially valuable perturbation probe of DNA structure and function in a region of the duplex (the sugar–phosphate backbone) where information is considerably lacking compared with the wealth of information which exists on the structure and function of the DNA bases. In addition, the complex, multistage aspect of transcription made us wonder if some stages would be more vulnerable than others to the same degree of perturbation. In this context, araC would be used as a perturbation probe that would be moved down the course of the DNA to sites where these different stages of transcription occur.

T7 RNA polymerase is a well-studied single polypeptide enzyme which carries out all aspects of transcription in vitro without accessory proteins (Chamberlin & Ryan, 1983). Considerable work has been done by others in elucidating how it carries out the various stages of transcription (Martin et al., 1988; Mookhtiar et al., 1991; Sousa et al., 1992; Gross et al., 1992; Maslak et al., 1993). In addition, the crystal structure of T7 RNA polymerase has recently been determined to 3.3 Å resolution (Sousa et al., 1993). Along with *Escherichia coli* RNA polymerase (Yager & von Hipple, 1987), it serves as a model system to investigate complicated mechanistic questions believed to be relevant to transcription in general.

Our promoter substitutions were limited to dC residues because we did not want to introduce any sequence changes with our sugar substitutions. These substitutions at −12, −9, −7, and −5 were at, or adjacent to, sites that others have shown to be critical for promoter recognition (Chapman & Burgess, 1987; Muller et al., 1989; Klement et al., 1990; Jorgensen et al., 1991; Diaz et al., 1993) primarily by a direct readout mechanism, although an indirect readout component may also be operative (Raskin et al., 1993). These substitutions did not significantly disturb the stability of the duplex, nor did they introduce any changes in the functional groups involved in sequence specific recognition.

No difference in promoter utilization was observed in terms of initiating nucleotide consumed or in the pattern and quantity of products generated using these promoter-substituted duplexes. In performing these steady state kinetic experiments to assay for differences in promoter utilization, we are aware that the rate-limiting step in initiation occurs after binding and promoter opening, most likely at the first chemical step(s) of phosphodiester bond formation (Martin & Coleman, 1987; Maslak & Martin, 1993). In such a case, the fast steps of polymerase–promoter binding and open complex formation may be affected by our substitutions and yet not be detected in this type of assay. To assay for binding directly, we also performed band shift experiments. These also showed no observable difference between substituted and unsubstituted promoters (data not shown). However, only large differences would have been detectable because of the relatively low level of specific binding T7 RNA polymerase exhibits toward its promoters (Gunderson et al., 1987; Bonner et al., 1992). From these experiments, we can therefore only conclude that substitution of araC at these promoter sites has no gross effect on either polymerase promoter binding or the rate of initiation and the type of products generated.

Once T7 RNA polymerase has bound, opened the promoter, and aligned itself at the start position for nucleotide incorporation, it ceases to be a sequence specific DNA-binding protein and becomes a nonsequence specific DNA-binding protein capable of rapidly translocating along the DNA as it

catalyzes the specific incorporation of nucleotides into the growing RNA message. While many of the mechanistic details of this transition are unknown, it is clear that the initiation phase is characterized by a high dissociation rate of small aborted RNA products (Martin et al., 1988). This phase ends after the incorporation of 6–10 nucleotides, whereupon the polymerase enters the elongation phase, characterized by tighter binding, a faster rate of synthesis, and near total processivity (Martin et al., 1988; Ling et al., 1989; Bonner et al., 1992).

In the case of the multi-subunit *E. coli* RNA polymerase, where the mechanism of this transition has received considerable study [reviewed in Von Hippel et al. (1984)], the enzyme apparently needs to break strong promoter contacts before it can become fully processive. Until these promoter contacts are cleared, most likely through the release of the σ subunit (Straney & Crothers, 1987), the enzyme can not transcribe freely down the coding strand, and so multiple initiations occur and abort until the polymerase breaks free. This transition is likely facilitated by the stabilizing effect of the growing RNA/DNA hybrid.

The single polypeptide T7 RNA polymerase exhibits strong abortive cycling, but it is not clear to what degree it stays associated with its promoter during this stage of initiation (Martin et al., 1988). Preliminary trap experiments performed in our lab with the control substrates suggest that the polymerase often dissociates during abortive cycling. Under conditions where the polymerase was limited to abortive cycling on the initial substrate due to selective nucleotide depletion, subsequent addition of a high-concentration trap substrate showed little if any activity continuing on the initial substrate. Rather, activity was seen to have continued only on the trap substrate when products were analyzed after several minutes of incubation (Mikita and Beardsley, unpublished results).

At the end of abortive cycling, the ternary complex must be stabilized in some way because elongation becomes highly processive (Martin et al., 1988; Ling et al., 1989; Bonner et al., 1992). While it is likely that there is a stabilizing contribution from the growing DNA/RNA hybrid, an increasing amount of data suggests that there is a binding site for the emerging single-stranded RNA on T7 RNA polymerase and that it is this binding site which is the key to the transition from abortive cycling to processive elongation (Muller et al., 1989; Gross et al., 1992). According to this model, the transition occurs when a sufficient length of message has been laid down such that the 5′ end contacts a RNA binding site on the polymerase which results in a stabilized and processive complex (Muller et al., 1989). A conformational change in the enzyme may also be associated with this transition as well (Ikeda & Richardson, 1986).

The effects introduced by the araC coding strand substitutions need to be considered in the context of the complex transition from abortive initiation to processive elongation. The results from both substrate series will be grouped together, where appropriate, to simplify the discussion. It should be kept in mind that even though there are such large differences in processivity between the abortive initiation phase and the processive elongation phase, it is still the case that at each position in the message the polymerase is at the center of similar competing kinetic pathways. The one results in the incorporation of a nucleotide and a message that has grown to length $n+1$; the other results in termination of synthesis and an abortive or stalled transcript of length n.

At the start of initiation, the first and second nucleotide are joined to begin the message. This requires that both these nucleotides hydrogen bond to the coding strand and maintain the proper Watson–Crick geometry long enough for the polymerase to catalyze phosphodiester bond formation. On the control substrates, the polymerase initiates at high frequency from the +1 position of the coding strand. A low frequency of +2 starts is also seen, but these starts result in only short abortive products. When araC is at position 1, the polymerase responds by bypassing araC and initiating from position 2. It is likely that having araC at this position interferes with the stable pairing (possibly through a loss of stabilizing stacking energy) of the first two nucleotides of a +1 start. Instead, the normally low-frequency +2 start becomes the kinetically dominant pathway, most likely because the stable pairing of nucleotides opposite positions 2 and 3 of the coding strand is not affected by araC at position 1.

This ability to initiate from position 2 shares a similarity with the reutilization of the +1 dimer with the 38(C) and 43(C) substrates. It can be observed that the +1 dimer, in contrast to all other products, does not accumulate over the time course of these steady state reactions (see Figure 6, top). Lack of accumulation of the dimer has been described by another group (Moroney & Piccirilli, 1991) who showed that the dimer is not a dead end product but rather can be used to initiate a new message. These authors noted that this utilization of the dimer for initiation suggests a flexibility in the positioning of the catalytic site for the start of transcription, i.e., when GTP initiates transcription, the first phosphodiester bond is made between positions 1 and 2, and when the dimer is used, the first phosphodiester bond is made between positions 2 and 3. It is possible that it is this flexibility which allows the polymerase to bypass araC and initiate from position 2 without a further negative effect on transcription.

When araC is at position 2, both +1 and +2 starts occur, but both initiation pathways have lower than normal yields. This is most likely the result of araC destabilizing the first and/or second nucleotide along either of these start pathways. In both cases, the residence time of one or both of the first two nucleotides would be reduced, thereby reducing the likelihood of phosphodiester bond formation taking place. Furthermore, the ability of the polymerase to switch to a +3 start is limited, further reducing the ability of the polymerase to initiate transcription with this substrate. This difficulty in making a +3 start contrasts with the ease of the switch to a +2 start on the 38(1) substrate and may therefore mark the limit of flexibility in positioning the catalytic site for initiation.

The primer slippage that occurred with the 43(4) substrate results from the polymerase being blocked after incorporation of G opposite araC. With this substrate, incorporation of the next nucleotide to produce the pentamer cannot compete with transcript termination. Larger products than the tetramer are formed because a coincidence of sequence allows the tetramer to slip back, base pair, and be used for another initiation round (Scheme 1). This shares some similarity with the previously described formation of G ladders (Martin et al., 1988). G ladders occur when substrates that code for the insertion of two or more G residues at the start of transcription are incubated with GTP only (see Figure 7, bottom, lanes 8 and 9). A primer slippage mechanism results in the formation of long runs of G residues, particularly in the case of substrates that code for the initiating sequence GGG (Martin et al., 1988). All these instances of primer slippage and reinitiation result in larger transcripts than that dictated by the position of the polymerase on the coding strand alone. Though such incidences of primer slippage appear largely limited to the initiation stage of synthesis, there is evidence that transcript

slippage can occur in homopolymeric stretches beyond the initiation region as well (MacDonald et al., 1993). The occurrence of primer slippage in the presence of all four nucleotides with the 43(4) substrate indicates that substitution of araC at position 4 is a complete block to further productive synthesis.

There is a 100-fold reduction in full length products seen with both the 38(6) and 43(6) duplexes. In both cases, the increase in product termination frequency starts at the pentamer product. This is what would be expected if incorporation of G opposite araC is slow, i.e., it increases the likelihood that termination will occur prior to nucleotide addition. The two substrates differ to some degree in what happens next. With 38(6), a small amount of G is incorporated to make the hexamer but this is followed by essentially no further extension. With 43(6), a small amount of G is incorporated opposite araC followed by the addition of three more nucleotides before extension terminates. The more drawn out path to transcript termination seen with the 43(6) substrate is likely the result of the dfferent sequence context of the 43(6) duplex.

It is interesting that the normal termination frequency for the hexamer product with 43(C) is relatively high (51%) while the termination frequency for this same product with 38(C) is comparatively low (13%). However, no extension beyond the hexamer abortive product is seen with 38(6), while some of the hexamer with 43(6) is extended to the heptamer, octamer, and nonamer before synthesis terminates. This cautions against thinking of the perturbation introduced by araC as having a set magnitude that simply needs to be factored into the normal termination frequency of a given substitution site. Certainly, the sites in the initiation region (which have high termination frequencies) are the most affected by araC substitution. Conversely, sites in the elongation region (which have much lower termination frequencies) are far less affected by substitution. However, other factors, such as sequence context, clearly influence the termination frequency at a particular substitution site. In addition, the positional flexibility allowed at the start of transcription, as well as possible interference with a polymerase conformational change (Ikeda & Richardson, 1986) and/or binding of the emerging message to another site on the enzyme (Muller et al., 1989; Gross et al., 1992), may also modulate response to araC substitution. All these could occur in a way that is not reflected in the normal termination frequencies along the unsubstituted coding strand, particularly in the region where the complex transition from abortive cycling to processive elongation takes place.

In addition to the local increase in termination frequency seen with the 43(8) substrate, there also occurs an increase in abortive products several nucleotides beyond the site of substitution. It is possible that some type of primer slippage mechanism, due to the repetitive GC sequence surrounding position 8, is responsible for these products, particularly given the primer slippage we observed with the position 4 substitution. However, a measurable amount of material is extended to full length and appears no more heterogeneous than that material produced from the control. Thus, there may well be some downstream effect which results from substituting araC at this position, although it is not clear what might account for such a delayed derailment of RNA synthesis. It is possible that two (or more) kinetically different extension pathways become partitioned as a result of araC at position 8. This would share a similarity to the +1 and +2 starts that were both seen to occur, and which yielded different products, with the 38(2) substrate. In the two-pathway case, some of the transcripts would slowly be extended past araC to the normal transcript length. The transcripts seen to accumulate several nucleotides after position 8 would be formed on the second pathway and would represent a population of terminated transcripts arising from an araC-induced slippage mechanism or a delayed termination effect due to transcript destabilization. The occurrence of branched kinetic pathways has recently been observed with *E. coli* RNA polymerase and was interpreted as indicating the existence of multiple conformational states for the polymerase (Erie et al., 1993). It was postulated that the ability of the polymerase to adopt multiple conformations may play a role in the fidelity and regulation of transcription, particularly in cases where mismatches and polymerase pausing occur (Erie et al., 1993).

When araC is moved out to position 11 on the 43-mer duplex and position 12 on the 38-mer duplex, the effects on transcription are significantly reduced. In both cases, the effects are local in that the major increase in termination frequency occurs at/near the substitution site. At both these positions, the polymerase has begun the transition to processive elongation, and therefore the normal termination frequency for these positions is quite low. Consequently, a 15–30-fold increase in termination frequency at a site that normally has only a 1%–2% termination frequency has the effect of decreasing a full length message by only a factor of 2 or 3.

If we combine our observations from all the substrates used in the transcription studies and consider the consequences of araC substitution as a form of DNA damage, we see that no observable effects occurred when araC for dC substitutions were made in the duplex recognition region of the central promoter. As we moved the araC substitutions down the coding strand, we observed an interference with transcription output that increased steeply through the initiation phase. This region of strong inhibition was characterized by miss-starts, primer slippage, and an inability to escape from abortive cycling. The severity of these coding strand effects were seen to drop off quickly as the substitutions were moved farther downstream to positions where the polymerase had already begun the transition to processive elongation. While sequence context is clearly important in the modulation of these effects, the particular stage of transcription appears far more deterministic of the magnitude of the inhibition that is seen. These highly variable, position dependent effects indicate a narrow window of vulnerability where transcription output is severely reduced by a DNA lesion that has little or no consequence when situated elsewhere in these synthetic coding units.

Thus, in the central promoter region where sequence recognition appears largely mediated through a direct readout mechanism between polymerase and base specific functional groups, it is not surprising that the sugar-centered change that araC introduces has little effect on this recognition mechanism. In addition, promoter recognition is followed by a polymerase-driven duplex melting process that may actually benefit from the slight destabilization which the araC-dG base pair introduces into the duplex. During the abortive cycling stage of initiation where termination often successfully competes with the incorporation of the next nucleotide, it is not surprising that anything which further destabilizes an already unstable transcription complex would result in increases in local termination frequencies at some positions to the point where productive synthesis is defeated altogether. Indeed, it has been shown by others that the incorporation of U opposite A in this region of the coding strand can often increase the termination frequency at the next position because

of the presumed weaker stability of the A-U base pair (Martin et al., 1988). Though the selection mechanism for correct nucleotide addition during abortive initiation and processive elongation is most likely governed by identical active site requirements of tight Watson–Crick base pair geometry and stability, the fact that termination normally competes so poorly with nucleotide addition during processive elongation makes the destabilization introduced by araC in this region of the coding strand far less likely to terminate synthesis.

While we have considered the effects of araC-containing DNA on transcription from the standpoint of local base pair destabilization, it is possible that sites of araC substitution also increase the local termination frequency because of interference with the polymerase translocation step. It is possible that the additional -OH group which the arabinose sugar carries interferes in some way with the ability of the polymerase to freely track along the sugar–phosphate backbone of the DNA. This would introduce a lag in polymerase bypass of these sites which would likely have the effect of allowing termination to compete more effectively with nucleotide addition. However, because so little is known about the mechanism of polymerase translocation, the above possibility remains conjectural.

Studies by others involving the effects of DNA damage upon transcription have established a link between transcription and DNA repair through the observation of the selective repair of DNA coding strands (Mellon et al., 1987). This has been observed in both mammalian cells (Mellon et al., 1987) and *E. coli* (Mellon & Hanawalt, 1989; Kunala & Brash, 1992). In vitro studies show that bulky adducts which arrest RNA polymerase do so only when the damage occurs on the coding strand, not on the noncoding strand (Nath & Romano, 1991; Chen et al., 1991; Chen & Bogenhagen, 1993). More recently, a direct link between damage arrested RNA polymerase and repair has been established with the cloning of the *E. coli mfd* gene and the purification and characterization of its gene product (Selby & Sancar, 1991, 1993). This protein has been shown to be directly involved in displacing a stalled RNA polymerase and then recruiting DNA repair machinery to remove the adduct (Selby & Sancar, 1993). That such machinery has evolved indicates the selective pressure to avoid the significant impact DNA damage can have on transcription output and fidelity. Indeed, a recent study on the effects of abasic sites suggests that these noncoding lesions can be mutagenic at the level of transcription as well as during DNA replication (Zhou & Doetsch, 1993).

The differential sensitivity to araC substitutions that we observed for the different stages of transcription illustrates the unequal consequences of damage at different positions along a DNA duplex given the functional context of transcription. It is possible that there are other multistage processes involved in DNA metabolism which are stage specifically vulnerable, because of variations in the way a protein or enzyme associates with DNA upon binding or as it travels along the DNA, and that the consequences of DNA damage will sometimes need to be considered in this context.

ACKNOWLEDGMENT

We thank the Grindley lab for use of the densitometer and Drs. Kasim Mookhtiar, B. J. Rao, Joseph Coleman, and Nigel Grindley, as well as members of the Beardsley lab, for helpful discussion and criticism. We thank Lisa Gardner for a critical reading of an earlier version of this manuscript.

REFERENCES

Arnott, S., Fuller, W., Hodgson, A., & Prutton, I. (1968) *Nature (London) 220*, 561–564.
Basu, A. K., & Essigmann, J. M. (1988) *Chem. Res. Toxicol. 1*, 1–18.
Beardsley, G. P., Mikita, T., Klaus, M., & Nussbaum, A. (1988) *Nucleic Acids Res. 16*, 9165–9176.
Bhattacharyya, A., Murchie, A. I. H., & Lilley, D. M. (1990) *Nature 343*, 484–487.
Bonner, G., Patra, D., Lafer, E. M., & Sousa, R. (1992) *EMBO J. 11*, 3767–3775.
Chamberlin, M., & Ryan, T. (1983) in *The Enzymes* (Boyer, P. D., Ed.) Vol. 15, pp 87–108, Academic Press, New York.
Chapman, K. A., & Burgess, R. R. (1987) *Nucl. Acids Res. 15*, 5413–5426.
Chen, Y.-H., & Bogenhagen, D. F. (1993) *J. Biol. Chem. 268*, 5849–5855.
Chen, Y.-H., Matsumoto, Y., Shibutani, S., & Bogenhagen, D. F. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9583–9587.
Chou, S.-H., Flynn, P., & Reid, B. (1989) *Biochemistry 28*, 2435–2443.
Clark, J. M., & Beardsley, G. P. (1989) *Biochemistry 28*, 775–779.
Cozzarelli, N. R. (1977) *Annu. Rev. Biochem. 46*, 641–684.
Diaz, G. A., Raskin, C. A., & McAllister, W. T. (1993) *J. Mol. Biol. 229*, 805–811.
Dicioccio, R. A., & Srivastava, B. I. S. (1977) *Eur. J. Biochem. 79*, 411–418.
Egli, M., Usman, N., Zhang, S., & Rich, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 534–538.
Erie, D. A., Hajiseyedjavadi, O., Young, M. C., & von Hipple, P. H. (1993) *Science 262*, 867–873.
Evans, J., Maccabee, M., Hatahet, Z., Courcelle, J., Bockrath, R., Ide, H., & Wallace, S. (1993) *Mutation Res. 299*, 147–156.
Fedoroff, O. Y., Salazar, M., & Reid, B. R. (1993) *J. Mol. Biol. 233*, 509–523.
Frei, E., III, Bickers, J., Lane, M., Leardy, W., & Tally, R. (1969) *Cancer Res. 29*, 1325–1352.
Furth, J. J., & Cohen, S. S. (1968) *Cancer Res. 28*, 2061–2067.
Gao, Y.-G., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1991) *Biochemistry 30*, 9922–9931.
Gross, L., Chen, W.-J., & McAllister, W. T. (1992) *J. Mol. Biol. 227*, 488–505.
Gunderson, S. I., Chapman, K. A., & Burgess, R. R. (1987) *Biochemistry 26*, 1539–1546.
Heintz, N. H., & Hamlin, J. L. (1983) *Biochemistry 22*, 3557–3562.
Ikeda, R. A., & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3614–3618.
Ikeda, R. A., Warshamana, G. S., & Chang, L. L. (1992) *Biochemistry 31*, 9073–9080.
Jaishree, T. N., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1993) *Biochemistry 32*, 4903–4911.
Jorgenson, E. D., Durbin, R. K., Risman, S. S., & McAllister, W. T. (1991) *J. Biol. Chem. 266*, 645–651.
Ke, S.-H., & Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137–5143.
Klement, J. F., Moorefield, M. B., Jorgenson, E., Brown, J. E., Risman, S., & McAllister, W. T. (1990) *J. Mol. Biol. 215*, 21–29.
Kufe, D. W., & Spriggs, D. R. (1985) *Semin. Oncol.* 12, 34–38.
Kufe, D. W., Major, P. P., Egan, E. M., & Beardsley, G. P. (1980) *J. Biol. Chem. 255*, 8997–9000.
Kunala, S., & Brash, D. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11031–11035.
Ling, M. L., Risman, S. S., Klement, J. F., McGraw, N., & McAllister, W. T. (1989) *Nucleic Acids Res. 17*, 1605–1618.
Maccabee, M., Evans, J. S., Glackin, M. P., Hatahet, Z., & Wallace, S. S. (1994) *J. Mol. Biol. 236*, 514–530.

Macdonald, L. E., Zhou, Y., & McAllister, W. T. (1993) *J. Mol. Biol. 232*, 1030–1047.
Martin, C. T., & Coleman, J. E. (1987) *Biochemistry 26*, 2690–2696.
Martin, C. T., Muller, D. K., & Coleman, J. E. (1988) *Biochemistry 27*, 3966–3974.
Maslak, M., & Martin, C. T. (1993) *Biochemistry 32*, 4281–4285.
Maslak, M., Jaworski, M. D., & Martin, C. T. (1993) *Biochemistry 32*, 4270–4274.
Mellon, I., & Hanawalt, P. C. (1989) *Nature 342*, 95–98.
Mellon, I., Spivak, G., & Hanawalt, P. C. (1987) *Cell 51*, 241–249.
Mikita, T., & Beardsley, G. P. (1988) *Biochemistry 27*, 4698–4705.
Milligan, J. F., Groebbe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.
Modrich, P. (1987) *Annu. Rev. Biochem. 56*, 435–466.
Mookhtiar, K. A., Pelusa, P. S., Muller, D. K., Dunn, J. J., & Coleman, J. E. (1991) *Biochemistry 30*, 6305–6313.
Moroney, S. E., & Piccirilli, J. A. (1991) *Biochemistry 30*, 10343–10349.
Muller, D. K., Martin, C. T., & Coleman, J. E. (1989) *Biochemistry 28*, 3306–3313.
Nath, S. T., & Romana, L. J. (1991) *Carcinogenesis 12*, 973–976.
Perrino, F. W., & Mekosh, H. L. (1992) *J. Biol. Chem. 267*, 23043–23051.
Pieters, J. M., de Vroom, E., van der Marel, G. A., van Boom, J. H., & Altona, C. (1989) *Eur. J. Biochem. 184*, 415–425.
Pieters, J. M., de Vroom, E., van der Marel, G. A., van Boom, J. H., Konig, T. M., Kaptein, R., & Altona, C. (1990) *Biochemistry 29*, 788–799.
Raskin, C. A., Diaz, G. A., & McAllister, W. T. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3147–3151.
Roberts, R. W., & Crothers, D. M. (1992) *Science 258*, 1463–1466.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, 1st ed., pp 277–282, Springer-Verlag, New York.
Schweitzer, B. I., Mikita, T., Kellog, G. W., & Beardsley, G. P. (1994) *Biochemistry* (in press).
Selby, C. P., & Sancar, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 8232–8236.
Selby, C. P., & Sancar, A. (1993) *Science 260*, 53–57.
Sigman, D. S. (1986) *Acc. Chem. Res. 19*, 180–186.
Sigman, D. S., Mazumder, A., & Perrin, D. M. (1993) *Chem. Rev. 93*, 2295–2316.
Sousa, R., Patra, D., & Lafer, E. M. (1992) *J. Mol. Biol. 224*, 319–334.
Sousa, R., Chung, Y. J., Rose, J. P., & Wang, B.-C. (1993) *Nature 364*, 593–599.
Steely, H. T., Jr., Gray, D. M., & Ratliff, R. L. (1986) *Nucl. Acids Res. 24*, 10071–10090.
Straney, D. C., & Crothers, D. M. (1987) *J. Mol. Biol. 193*, 267–278.
Townsend, A. J., & Cheng, Y.-C. (1987) *Mol. Pharmacol. 32*, 330–339.
von Hippel, P. H., Bear, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.
Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., & Rich, A. (1982) *Nature 299*, 601–604.
Wright, G. E., & Brown, N. C. (1990) *Pharmacol. Ther. 47*, 447–497.
Yager, T. D., & von Hippel, P. H. (1987) in *Escerichia coli and Salmonella typhimurium* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanick, B., Schaecter, M., & Umbarger, H. E., Eds.) pp 1241–1275, Am. Soc. Microbiol., Washington, DC.
Zhou, W., & Doetsch, P. W. (1993) *Proc. Natl. Acad. Sci. U.S.A 90*, 6601–6605.