

Characterization of a novel RNA polymerase II arrest site which lacks a weak 3' RNA–DNA hybrid

Peter J. Hawryluk^{1,2}, Andrea Újvári¹ and Donal S. Luse^{1,2,*}¹Department of Molecular Biology, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA and ²Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

Received as resubmission February 12, 2004; Accepted March 4, 2004

ABSTRACT

Transcript elongation by RNA polymerase II is blocked at DNA sequences called arrest sites. An exceptionally weak RNA–DNA hybrid is often thought to be necessary at the point of arrest. We have identified an arrest site from the tyrosine hydroxylase (TH) gene which does not fit this pattern. Transcription of many sequence variants of this site shows that the RNA–DNA hybrid over the three bases immediately preceding the major arrest point may be strong (i.e. C:G) without interfering with arrest. However, arrest at the TH site requires the presence of a pyrimidine at the 3' end and arrest increases when the 3'-most segment is pyrimidine rich. We also demonstrated that arrest at the TH site is completely dependent on the presence of a purine-rich element immediately upstream of the RNA–DNA hybrid. Thus, the RNA polymerase II arrest element from the TH gene has several unanticipated characteristics: arrest is independent of a weak RNA–DNA hybrid at the 3' end of the transcript, but it requires both a pyrimidine at the 3' end and a polypurine element upstream of the RNA–DNA hybrid.

INTRODUCTION

RNA synthesis by RNA polymerase II is highly regulated in the cell. While it has become clear that transcript elongation is one of the regulated steps (see 1–4 for reviews), both the sequence signals and the *trans*-acting factors which control elongation are not yet well understood. Transcript elongation by RNA polymerase II *in vitro* does not proceed uniformly (5). The average rate of transcript elongation by RNA polymerase II on naked DNA templates, in the absence of any additional factors, is ~5 nt/s under optimal conditions (6). However, at some template positions, the polymerase may pause for as long as 1 min before continuing transcription (see, for example, 7). At other locations, called arrest sites, a fraction of transcribing polymerases are completely blocked and cannot continue RNA synthesis in a standard experimental

time course. Arrest occurs because the polymerase has translocated upstream on the template such that the 3' end of the transcript is displaced from the active site. Thus, arrested polymerases have not terminated RNA synthesis. The arrested complex remains stably associated with the DNA, maintaining an 8–9 bp RNA–DNA hybrid upstream of the transcript 3' end (8–10). To escape from arrest, the polymerase must either translocate forward on the template so that the transcript 3' end is once again aligned with the active site or break the RNA phosphodiester bond at the active site to generate a new 3' end at this position. Both of these events occur for arrested complexes *in vitro*, but in the absence of additional factors they may require up to several hours to do so (11). The elongation factor SII (TFIIS) greatly increases the rate of transcript cleavage, allowing the polymerase to elongate through arrest sites relatively quickly (i.e. in seconds) (12–15). Thus, sensitivity of an elongation block to SII can be used as an assay to define arrest.

Many studies of transcriptional arrest by RNA polymerase II have used an arrest site found in the histone H3.3 gene as a model (16). This site contains the sequence TTTTTCCTTTTTT in the non-template strand (see Fig. 1). We will refer to this as a T-space-T type arrest site. During transcription through this site, a certain proportion of polymerases fall into arrest after incorporating the fifth or sixth U residue (underlined) encoded by the first T run (17). This proportion can be manipulated *in vitro* by altering the temperature, NTP concentrations, reaction time and the presence or absence of elongation factors like SII (17,18).

There are two major interactions between the nucleic acids and polymerase that are thought to mediate arrest (reviewed in 2). Both of these interactions have been shown to be important in arrest at the H3.3 site. The strength of the RNA–DNA hybrid, especially the last few base pairs at the 3' end of the transcript, is a critical feature. The U:A base pair is the weakest of the Watson–Crick base pairs, so a poly(U) segment in the transcript would produce the least stable RNA–DNA hybrid (for a detailed discussion of this point, see 19). A stable interaction in the last base pair is presumably necessary in order to properly orient the 3' end with the incoming nucleotide for phosphodiester linkage to the transcript. Thus, if the hybrid at the 3' end of the transcript is weak, bond formation may be delayed, allowing the polymerase time to translocate upstream on the template. If translocation results in

*To whom correspondence should be addressed at Department of Molecular Biology, Mail Location NC20, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA. Tel: +1 216 445 7688; Fax: +1 216 444 0512; Email: lused@ccf.org

MATERIALS AND METHODS

Reagents

Ultrapure NTPs and dNTPs were obtained from Amersham Pharmacia Biotechnology, ApC dinucleotide from Sigma, ^{32}P -labeled NTPs from Perkin Elmer/New England Nuclear, Taq DNA polymerase from Fermentas, restriction enzymes from New England Biolabs, DNA oligonucleotides from Operon or IDT and streptavidin-coated paramagnetic beads from Promega Biotech.

Plasmids

All plasmids used in this study contain the adenovirus major late (AdML) promoter. The series of constructs based on the pML20-42 plasmid (25) were made either by replacing the XhoI–HindIII fragment (which begins 45 bp downstream of +1) with 50 bp of sequence from the TH gene or from other sources, as indicated, or by site-directed mutagenesis of existing constructs using the protocol in the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). Inserted fragments were made by annealing complementary DNA oligos containing the 50 bp of sequence to be assayed and overhanging ends appropriate to XhoI and HindIII cleavage, followed by ligation of these fragments into the pML20-42 plasmid. The pML23-100 series of constructs was made by replacing the StuI–XhoI fragment (which begins 22 bp downstream of +1) of pML20-23like3 (26) with 80 bp of the sequences indicated in the lower panel of Figure 3. Note that the pML20-42- and pML23-100-based constructs are very different in the sequences preceding the TH gene inserts.

Generation of templates for *in vitro* transcription

Transcription reactions used biotinylated PCR-generated fragments for attached template transcription, as described below. The upstream primer (relative to the direction of transcription), 5'-GGCATCAAGGAAGGTGATTG-3', is biotinylated at the 5' end, which is 96 bp upstream of the transcription start site. The 5' end of the downstream primer, 5'-CGACGGCCAGTGCCAAGC-3', is 114 (for the 20-42 series of constructs) or 171 bp (for the 23-100 series) downstream of transcription start. The resulting 210 (20-42) or 267 bp (23-100) biotinylated PCR fragments were purified using the Concert Rapid PCR Purification System (Marligen Biosciences).

In vitro transcription of attached templates

An aliquot of 1 μg of biotinylated PCR fragment was immobilized on 200 μg streptavidin-coated paramagnetic beads by incubating at room temperature for 10 min in 150 μl of BC100 buffer (20 mM Tris pH 7.9, 100 mM KCl, 20% glycerol and 0.2 mM EDTA). The beads were concentrated, resuspended in a final volume of 50 μl containing 25 μl of HeLa nuclear extract, 10 mM Tris–HCl pH 7.9, 8 mM MgCl_2 and 60 mM KCl and incubated at 30°C for 30 min. All *in vitro* transcription reactions used the same preparation of nuclear extract. The resulting preinitiation complexes were then washed three times with 200 μl of BC100MD (BC100 buffer with 8 mM MgCl_2 and 1 mM dithiothreitol) and resuspended in BC100MD. Complexes were incubated with either 100 μM ATP, 10 μM CTP and 1 μM [α - ^{32}P]UTP (for 20-42 based

constructs) or 1 mM ApC, 20 μM ATP, 50 μM dATP and 1 μM [α - ^{32}P]GTP (for 23-100 based constructs) at 30°C for 9 min. An aliquot of 10 μM cold UTP (for 20-42 based constructs) or GTP (for 23-100 based constructs) was added and the reactions were incubated for an additional 3 min at 30°C to ensure that all complexes reached position A23 (for 20-42 based constructs) or G17 (for 23-100 based constructs) (see Fig. 3 for initially transcribed sequences). Reactions were washed once at room temperature with 200 μl of 1% sarkosyl in MEMDM buffer (30 mM Tris pH 7.9, 10 mM β -glycerophosphate, 62 mM KCl, 0.5 mM EDTA, 1 mM dithiothreitol and 8 mM MgCl_2), twice with 200 μl of MEMDM and resuspended in MEMDM (48 μl for standard chase \pm SII). Fractions of 24 μl of each stalled complex were incubated with 1 mM NTPs with or without 34 ng/ μl SII at 37°C for 5 min, unless indicated otherwise. Chase reactions were stopped by the addition of EDTA to 50 mM, phenol/ CHCl_3 extracted and ethanol precipitated. Reactions for size markers were prepared identically to chase reactions from template immobilization through washing of the stalled complex and were resuspended in 200 μl of MEMDM. Fractions of 48 μl of the stalled complex were used for each ladder reaction. For the A marker lane, complexes were preincubated at 37°C for 30 s, followed by the addition of 10 μM ATP and 1 mM CTP, GTP and UTP. Samples of 12 μl of the reaction were removed after incubating with the NTPs for 30 s and 1, 2 and 3 min and immediately stopped with 50 mM EDTA. Time points were pooled, phenol/ CHCl_3 extracted, and ethanol precipitated. Reactions were repeated with limiting CTP, GTP or UTP (with the limiting base always at 10 μM). These reactions produce complexes stalled at positions on the template where the next base to be added is limiting, thus transcript sizes correspond to complexes stalled before each addition of the limiting NTP. Precipitated nucleic acid was resuspended in 50 or 95% formamide + dye mix, heated at 95°C for 2–5 min to denature the RNA and electrophoresed on 10 or 15% acrylamide gels containing 7 or 8 M urea. Gels were visualized and individual bands quantified using a PhosphorImager system and ImageQuant software (Amersham/Molecular Dynamics). For the quantifications in Figure 3, rectangles were drawn around the bands corresponding to the arrested and run-off products and the signal within each rectangle corresponding to an arrested product was divided by the total signal of all the arrested and run-off products in that lane. Note that this direct comparison of signal corresponding to each transcript accurately reflects the amount of each transcript produced because the transcripts are labeled only in the initial region before sarkosyl rinsing (i.e. each transcript larger than 23 nt contains the same amount of ^{32}P , on average, regardless of its length).

RESULTS

We searched for possible RNA polymerase II arrest sites within TH gene sequences using *in vitro* transcription. We constructed a set of template DNAs in which the AdML promoter and initially transcribed region from the pML20-42 plasmid (25) were fused to one of five 500 bp segments spanning the central 2.5 kb of the TH gene. These five original templates were named according to the TH exons which they contained; thus, the 7/8 arrest site (see below) was found in the

template spanning exons 7 and 8. Preinitiation complexes were assembled on these templates by incubation in HeLa cell nuclear extracts. The RNA polymerases were advanced to +23 with a subset of the NTPs, the transcription complexes were purified by washing with sarkosyl and RNA synthesis was continued by incubation at 37°C with all NTPs at 1 mM (see Materials and Methods for details). This procedure revealed three prominent arrest sites, whose sequences are given in Figure 1 (primary data not shown). The fraction of the total RNA polymerase which halted within each of the indicated sequence elements was at least 7% during our 5 min incubation period. The addition of SII (TFIIS) to the reaction eliminated the transcriptional block in all cases (primary data not shown; arrest levels given in Fig. 3). To determine arrest locations at single base resolution, as shown in Figure 1, we constructed and assayed a second round of templates in which 50 bp segments surrounding the three arrest sites were fused to the AdML promoter beginning at +45.

An initial inspection of the central 2.5 kb of the TH gene had revealed two sites with closely spaced T runs in the non-template strand (Fig. 1). One of these predicted arrest sites (TH site 1, Fig. 1) did function as an arrest site in our test tube assay (site 1 from the 10/11 template; see Fig. 1). The second predicted site, TH site 2, did not cause significant levels of arrest *in vitro*. In one of the other observed arrest sites, transcription halted within a T run (site 2 from the 10/11 construct). However, site 10/11 2 does not contain a second T run in the non-template strand.

The third observed arrest site, from the 7/8 template, has a very surprising sequence. This site does not contain any poly(T) segment. RNA polymerase halted at two positions, with most arrest occurring at the more downstream location (the C residue marked by the dot in Fig. 1). The RNA in RNA–DNA hybrid when the polymerase active site is at the major 7/8 arrest position is UAUAUGUC, assuming an 8 bp hybrid. While four of the eight bases in this sequence are U residues, the hybrid just upstream of the major arrest position contains two G:C base pairs out of the final three. If hybrid stability were the most critical feature of the arrest site, one would have expected to find the major arrest position at the more upstream location, where the last 5 bp are locally weakest. It should also be noted that the position of the 7/8 arrest site within the TH gene corresponds most closely, among the arrest sites observed *in vitro*, to the position of the block to transcript elongation proposed from the nuclear run-ons assays performed by Kroll *et al.* (24). Thus, because of its novel sequence and its potential biological importance, we decided to focus our efforts on characterization of the 7/8 arrest site.

The 7/8 site behaves as a conventional arrest site in that arrest is relieved by SII-mediated transcript cleavage (Fig. 2), accompanied by the release of a large (8–11 nt) 3' fragment (data not shown). Note that in all experiments in this study which use SII, the SII was added along with the NTPs to the RNA polymerases stalled at +23. The effect on arrest of reducing elongation rate via NTP limitation is consistent with the results of similar experiments performed on T-space-T type arrest sites (17,27). Briefly, the frequency of arrest at the 7/8 site increases as the substrate NTP concentration is reduced and the exact position at which arrest occurs can be manipulated by limiting one NTP and keeping the others at saturating levels (data not shown; see 28). However, from

inspection of the sequence, arrest at the 7/8 site does not seem to depend on an exceptionally weak RNA–DNA hybrid at the point of arrest and thus it is not clear from existing models why upstream translocation and arrest should take place at this site.

In order to better understand the mechanism of arrest at the 7/8 site, we constructed a series of sequence variants of this element. We altered the template immediately upstream of the major arrest position on the 7/8 DNA to assess the role of the RNA–DNA hybrid in arrest at this site. All of these initial mutants were made in the context of the 7/8 construct shown in Figure 2, which consists of 50 bp of TH sequence surrounding the arrest site fused downstream from the initial 45 bp of our AdML promoter construct. Note that on this template the major and minor arrest positions are at +75 and +72, respectively (Fig. 2; values for arrest levels given in Fig. 3). Arrest at both locations is relieved by the addition of SII to the transcription reaction (compare Fig. 2, lanes 2 and 3). In the T74C mutant, all three of the 3'-most base pairs in the RNA–DNA hybrid would be G:C base pairs, if arrest occurred at +75. Since this sequence change strengthens the RNA–DNA hybrid at the 3' end of the transcript, one would expect to see a reduced level of arrest at +75 in this case. However, arrest frequency was higher on the T74C template (compare Fig. 2, lanes 2 and 4). On the C75T construct, the base pair at the major arrest position was changed from C:dG to the weaker U:dA. A higher frequency of arrest was observed for this mutant than for the 7/8 template, as would be expected if hybrid weakness at the transcript 3' end were a determinant of arrest frequency (Fig. 2, lane 6). The sequence changes for the T/3/C construct strengthen the upstream segment of the RNA–DNA hybrid for RNA polymerases halted at +75, relative to the parental 7/8 template. We found that arrest on T/3/C was considerably reduced relative to the original 7/8 construct and the location at which polymerase halts was also altered (Fig. 2, lane 10).

Upon arrest at the 7/8 site, the active site of the RNA polymerase retreats primarily to positions 67 and 64 (as judged by SII cleavage assay; data not shown), where the RNA–DNA hybrid is relatively G:C rich. The G/3/A construct was designed to weaken the hybrid at these locations, to test whether the polymerase could be made to retreat farther upstream (to a location where the hybrid is stronger, for example at position 60). It was not expected to alter arrest frequency at the major arrest position, since when the active site is at that position, the sequence in the RNA–DNA hybrid is unchanged. However, the G/3/A template exhibited a decreased frequency of arrest at position 75 and an increased level at position 72 (compare Fig. 2, lane 2 with 12). When the polymerase has just added the nucleotide at position 72, two of the base changes between the G/3/A and 7/8 constructs are in the upstream part of the RNA–DNA hybrid. These replace two G:dC base pairs present in the 7/8 hybrid with weaker A:dT base pairs. Thus, the increased arrest at position 72 is consistent with weakness in the RNA–DNA hybrid contributing to arrest. A simple explanation for the reduced arrest at position 75 on the G/3/A mutant was not apparent.

Arrest at high NTP concentrations at the histone H3.3 arrest site and mutants of that site (17,21), at the 7/8 site and its mutants (this work) and at the pML5-4NR arrest site (27) always occurs with either a C or a U residue at the transcript 3' end. In addition, pause, arrest and termination sites in

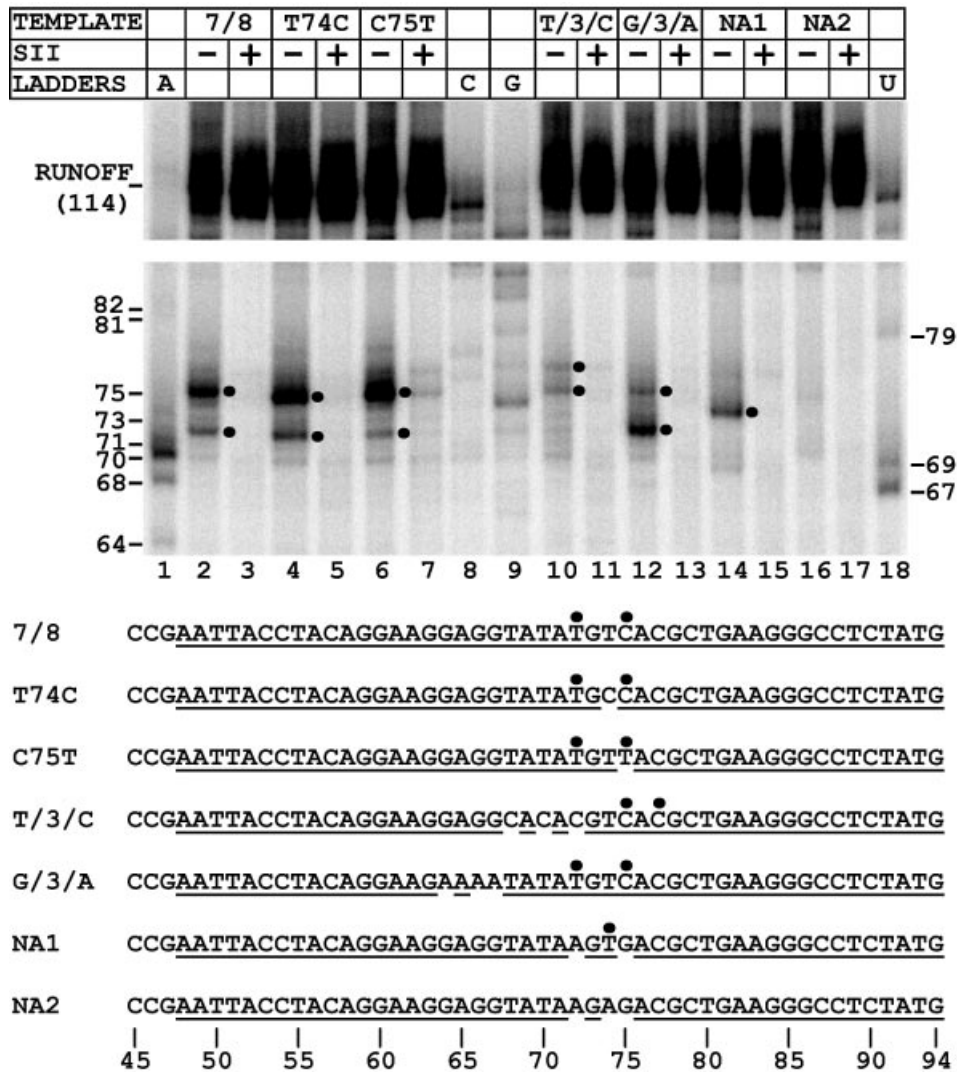


Figure 2. Analysis of point mutations at the 7/8 site. The 7/8 and mutant templates were transcribed as described in Materials and Methods, with SII included in the reactions as indicated. RNA lengths taken from the single NTP limiting marker reactions (generated on the NA2 template) are given on the left and right, respectively. Sequences listed below are the 50 bp fragments inserted in the pML20-42 plasmid to generate each template (see Materials and Methods). In this and the following figures, sequences from the TH gene are underlined, distances downstream of +1 are given below the sequences and arrest positions are indicated by dots above the last nucleotide to be incorporated into the transcript.

mammalian N- and *c-myc* genes have been precisely mapped and it was found that at most of these sites, the transcript ends with a pyrimidine (29). Therefore, we tested the effect of changing the bases at the two arrest positions in the 7/8 construct from pyrimidines to purines. The changes made in constructing this template (called NA1; see Fig. 2) resulted in an increase in the strength of the RNA–DNA hybrid at position 72 but no change in hybrid strength at the major arrest position, +75. The NA1 template did not support arrest at either position 72 or 75 (Fig. 2, lane 14). However, there was some arrest at position 74, after incorporation of the sole remaining pyrimidine residue in the vicinity of the original arrest site. When this position was changed to its corresponding purine in the NA2 construct, the template did not support a significant level of arrest at any position at 1 mM NTP concentrations (Fig. 2, lane 16).

Since maximal arrest at the histone H3.3 T-space-T site requires the sequence downstream of the first T run (17), we

wished to determine the role of the downstream sequence in arrest at the 7/8 site. Constructs were made (Fig. 4) in which blocks of sequence were exchanged among the 7/8 site, the strong arrest site from the histone H3.3 gene and sequence from the pML20-23 template (21). Earlier work had shown that complexes stalled at the end of the short T run on pML20-23 are fully elongation competent and not upstream translocated (21). Based on these results, we treated sequences from pML20-23 as ‘neutral’ sequences that do not contribute to arrest. The new constructs were designed so that the major arrest position from the 7/8 template (constructs 7/8, d30 and d23), the more downstream of the two major arrest positions from histone H3.3 (constructs TsT and 30d) and the third T of a T triplet from the 20-23 template (construct 23d) were aligned at position 75 (see Fig. 4).

For the d30 template, TH sequence downstream of the 7/8 site was replaced with the downstream sequence from the histone H3.3 site. Arrest occurred more frequently on this

20-42 CONSTRUCTS				
20-42 ITS	1	ACTCTCTCCCCTTCTCTTTAAAGGCCTTGGGAAACCCTCGAG	44	
TEMPLATE	45	SEQUENCE	94	ARREST POSITION %
7/8		CCGAATTACCTACAGGAAGGAGGTATA [•] GT [•] CACGCTGAAGGGCCTCTATG		72 3.8 75 9.2
10/11#1		ACAAGTCTCAGGAGAAAGGGATTTT [•] TGTGGGTTT [•] TGCTGCCACCTGCTGG		70 6.7
10/11#2		CTCAAGCACAGTGTCTAGAGAAGAGTCTTTT ^{••••} CA [•] TGCTCTTCTCCCCTTCTC		75 9.6 77 9.0
T74C		CCGAATTACCTACAGGAAGGAGGTATA [•] GC [•] CACGCTGAAGGGCCTCTATG		72 3.6 75 12.2
C75T		CCGAATTACCTACAGGAAGGAGGTATA [•] GT [•] TACGCTGAAGGGCCTCTATG		72 2.3 75 15.2
T/3/C		CCGAATTACCTACAGGAAGGAGGCACACGTCA [•] CGCTGAAGGGCCTCTATG		75 3.2 77 2.8
G/3/A		CCGAATTACCTACAGGAAGAAAATATA [•] TGT [•] CACGCTGAAGGGCCTCTATG		72 7.5 75 4.1
NA1		CCGAATTACCTACAGGAAGGAGGTATAAGT [•] GACGCTGAAGGGCCTCTATG		74 5.3
NA2		CCGAATTACCTACAGGAAGGAGGTATAAGAGACGCTGAAGGGCCTCTATG		NO ARREST
d30		CCGAATTACCTACAGGAAGGAGGTATA [•] GT [•] CACCTTTTTTGGAGAGGCG		72 4.5 75 14.7
d23		CCGAATTACCTACAGGAAGGAGGTATA [•] GT [•] CACGAAACCCTCGAGGTGTC		72 5.1 75 8.8
TsT		AGGAAGAAGCAGGAAAAGAGAGGCCTTTTT ^{••••} TCCTTTTTTGGAGAGGCG		73 9.5 74 7.7
30d		AGGAAGAAGCAGGAAAAGAGAGGCCTTTTT ^{••••} TGCTGAAGGGCCTCTATG		74 12.1 77 7.7
23d		CCGCGGCAGGAAGAGGAAGAAGCAGGCCTTGGGCTGAAGGGCCTCTATG		NO ARREST

23-100 CONSTRUCTS				
23-100 ITS	1	ACAGGAAGAGGAAGAAGCAGGCCTTTAAATTGATGAGAATATGGGAAAGGATGT	54	
TEMPLATE	55	SEQUENCE	104	ARREST POSITION %
7/8*		GGCAATTACCTACAGGAAGGAGGTATA [•] GT [•] CACGCTGAAGGGCCTCTATG		85 9.6 87 7.7
TH3		GGTTTGTAGTTAGTGAGGGAGGATATA ^{••••} TGT [•] CA [•] CGCTGAAGGGCCTCTATG		85 9.8 87 5.8
TH4		GGTTTGTAGTTAGTGAGGGAGGATATA ^{••••} TGT [•] CACGCTGAAGGGCCTCTATG		85 15.7 87 7.8
TH5		GGTTTGTACCTACAGGAAGGAGGTATA [•] GT [•] CA [•] CGCTGAAGGGCCTCTATG		85 6.7 87 6.2
NA2*		GGCAATTACCTACAGGAAGGAGGTATAAGAGACGCTGAAGGGCCTCTATG		NO ARREST
TH8		GGCAATTACCTACAGGAAGGAGGTATA [•] GT [•] CA [•] CGTGGTAGAAGGAGGCC		85 8.1 87 3.9
TH9		GGTTTGTACCTACAGGAAGGAGGTATA ^{••••} TGT [•] CA [•] CGTGGTAGAAGGAGGCC		85 8.8 87 4.5
TH10		GGTTTGTAGTTAGTGAGGGAGGATATA ^{••••} TGT [•] CA [•] CGTGGTAGAAGGAGGCC		82 5.1 85 14.4
TH11		GGTTTGTAGTTAGTGAGGGAGGATATA ^{••••} TGT [•] CA [•] CGTGGTAGAAGGAGGCC		83 8.2 85 12.5

Figure 3. Sequences of constructs tested and quantification of transcriptional arrest. Initially transcribed sequences (ITS) from +1 to the start of the 50 bp TH inserted fragments are shown at the top of the lists of pML20-42- and pML23-100-based constructs. Sequences from the TH gene are underlined and positions where arrest occurred are marked by dots. The major arrest positions are listed on the right for each template, followed by the fraction of complexes that arrested at each position. The percentage arrest was calculated as described in Materials and Methods and for most templates is the average of several reactions.

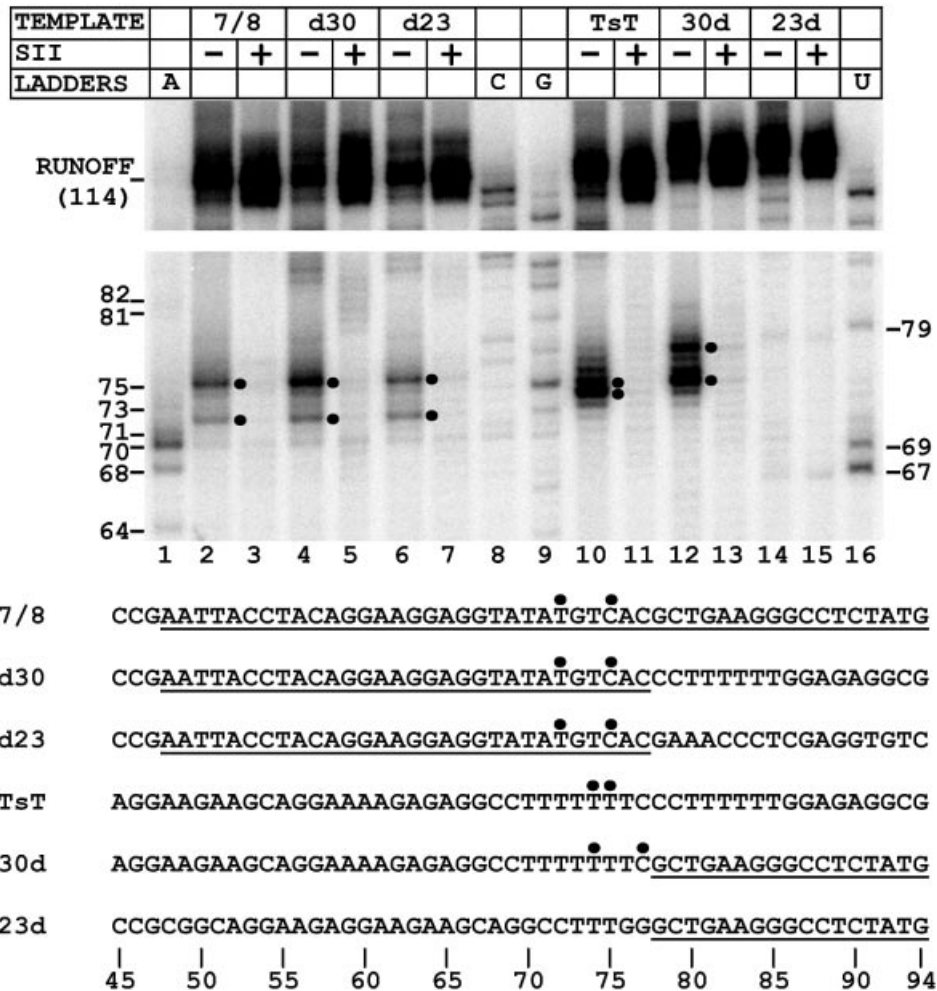


Figure 4. The effect of downstream sequence on arrest at the 7/8 site. Templates were transcribed and size markers were generated from the NA2 template as described in Materials and Methods. In the lower part of the figure, sequences from the TH gene are underlined, distances downstream of +1 are given below the sequences and arrest positions are indicated by dots above the last nucleotide to be incorporated into the transcript. In the TsT template (containing the arrest site of the H3.3 gene), a 1 bp deletion occurred during cloning in the sequence upstream of the 50 bp inserted fragment, so transcripts of this template are 1 nt shorter than the numbers listed below the sequences (e.g. the sizes of the two major arrested transcripts are 73 and 74 nt).

template than on the 7/8 template (Fig. 4, compare lane 2 with 4; see arrest levels in Fig. 3), indicating that the downstream H3.3 sequence contributes more highly to arrest than the downstream 7/8 sequence. When the 7/8 downstream sequence was replaced with sequence from 20-23 (construct d23), the frequency of arrest was only slightly reduced (compare Fig. 4, lane 2 with 6). Therefore, the downstream 7/8 sequence probably does not contribute highly to arrest at this site.

In our hands the native histone H3.3 element (TsT in Fig. 4) supported arrest at the same locations reported in earlier studies of this sequence (17). If the downstream sequence from the histone H3.3 site is a higher contributor to arrest than the downstream 7/8 sequence, replacing the sequence downstream of the H3.3 arrest site with downstream 7/8 sequence (the 30d template) should yield a reduced level of arrest. However, overall frequency of arrest was roughly the same for the TsT and 30d templates (compare Fig. 4, lane 10 with 12). It is interesting to note that while changing the sequence downstream of the arrest position on the 7/8 site did not

change the positions of arrest (compare lane 2 with lanes 4 and 6), changing the downstream sequence of the histone H3.3 site did alter the positions of arrest in its T run (compare lane 10 with 12).

The last block mutant, the 23d template, replaced sequence downstream of the T triplet in the pML20-23 template with sequence downstream of the 7/8 site. We constructed this template to ask whether the downstream 7/8 sequence could act independently to cause arrest in otherwise neutral sequence from pML20-23. No complexes fell into arrest when transcribing this template (Fig. 4, lane 14). Furthermore, when transcription complexes were walked to the end of the T triplet at position 75 on this template and stalled there, they chased completely upon addition of NTPs to the reaction mixture (data not shown).

For the histone H3.3 arrest site, the role of sequence upstream of the first T run is not entirely clear. It was originally reported that sequences in this region do modulate arrest (17), however, one particular swap of upstream segments did not alter arrest levels at the H3.3 site (21). If

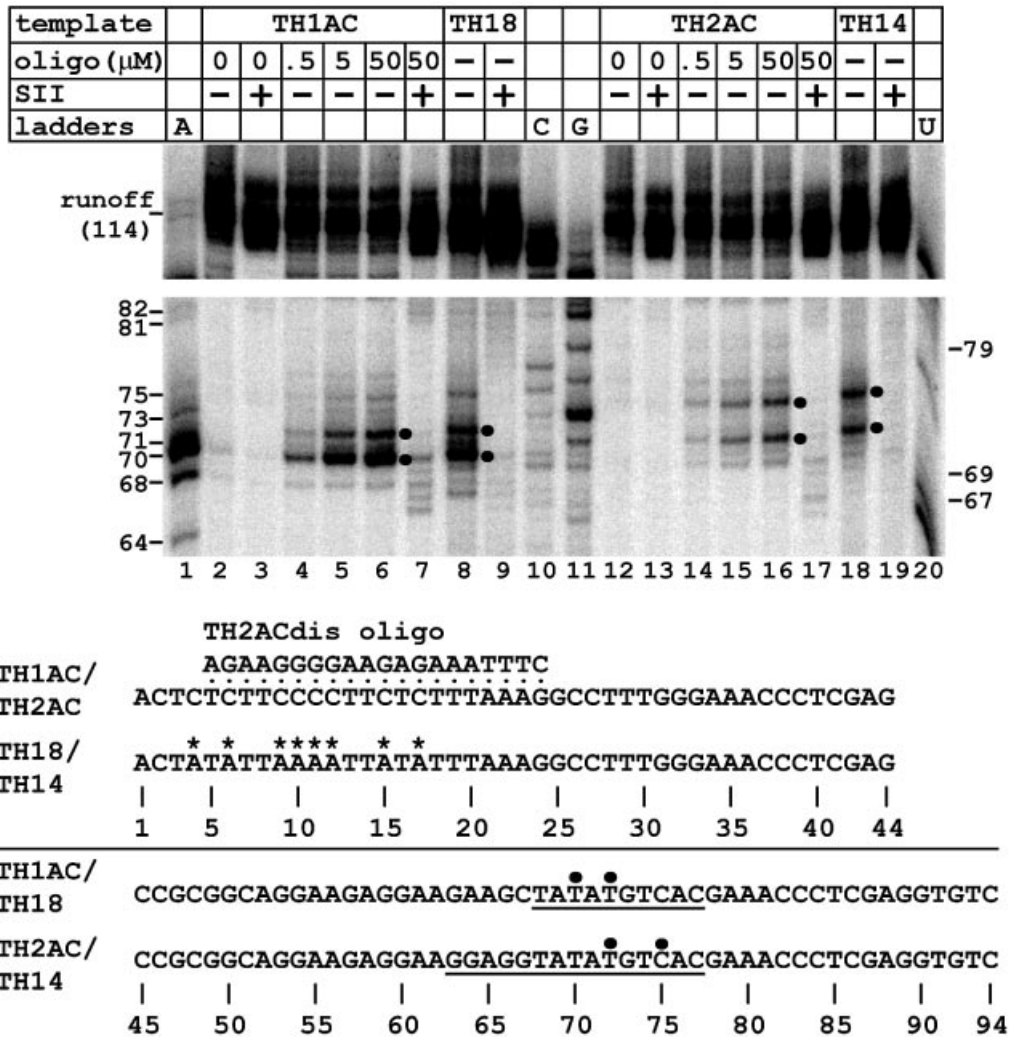


Figure 5. Arrest can be induced on templates which otherwise fail to support arrest by disrupting the formation of upstream secondary structure. Templates were transcribed and size markers were generated (from the NA2 template) as described in Materials and Methods. Sequences from transcription start through the 50 bp of inserted sequence are listed in the lower part of the figure for each template. The sequence of the TH2ACdis oligo is shown above the region of the TH1AC and TH2AC sequences to which it anneals. This oligo was added at the indicated concentrations in lanes 4–7 and 14–17. Asterisks mark the bases that have been changed between the TH1AC/TH2AC and TH18/TH14 initially transcribed sequences. Large dots above the sequences indicate the positions where arrest occurred.

upstream sequences do not play a major role in arrest at the 7/8 site then the functional arrest cassette for this site might be very short, given the relatively minor contribution of downstream elements (Fig. 4). In order to determine the minimum sequence that supports arrest, three new templates were constructed, TH1AC, TH2AC and TH7 (the TH1AC and TH2AC sequences are shown in Fig. 5). On these templates, increasing amounts of sequence upstream of the 7/8 arrest site were replaced by presumably neutral sequence from pML20-23. All of these templates retained pML20-23 sequence downstream of the arrest site. When we tested these constructs we found that the TH7 template, which retained all but 5 bp of the upstream TH sequence, supported arrest at the same locations and to the same extent as the 7/8 template (data not shown here; see 28). However, no arrest was observed with either TH1AC or TH2AC (Fig. 5). We were initially surprised by these results, but we subsequently realized that the design of TH1AC and TH2AC did not take into account the potential

for the G-rich pML20-23 RNA sequences to base pair with the C-rich RNA from the initially transcribed region (see Fig. 5). We searched for possible secondary structures in the transcripts of these templates using the mfold program (30,31) We found that the RNAs from the TH1AC and TH2AC templates, which did not support arrest, were predicted to fold into stable hairpins that paired the initially transcribed region with the transcript segment upstream of the TH arrest site (28). This double-stranded RNA segment was expected to extend to a point 17–20 bases upstream of the major TH arrest position. Based on the results of Reeder and Hawley (8), such secondary structure should block upstream translocation and prevent arrest. The transcript of the TH7 template, which did support arrest, was not predicted to form an analogous secondary structure.

If our explanation is correct, blocking the formation of secondary structure should restore arrest with the TH1AC and TH2AC constructs. We took two approaches to this problem.

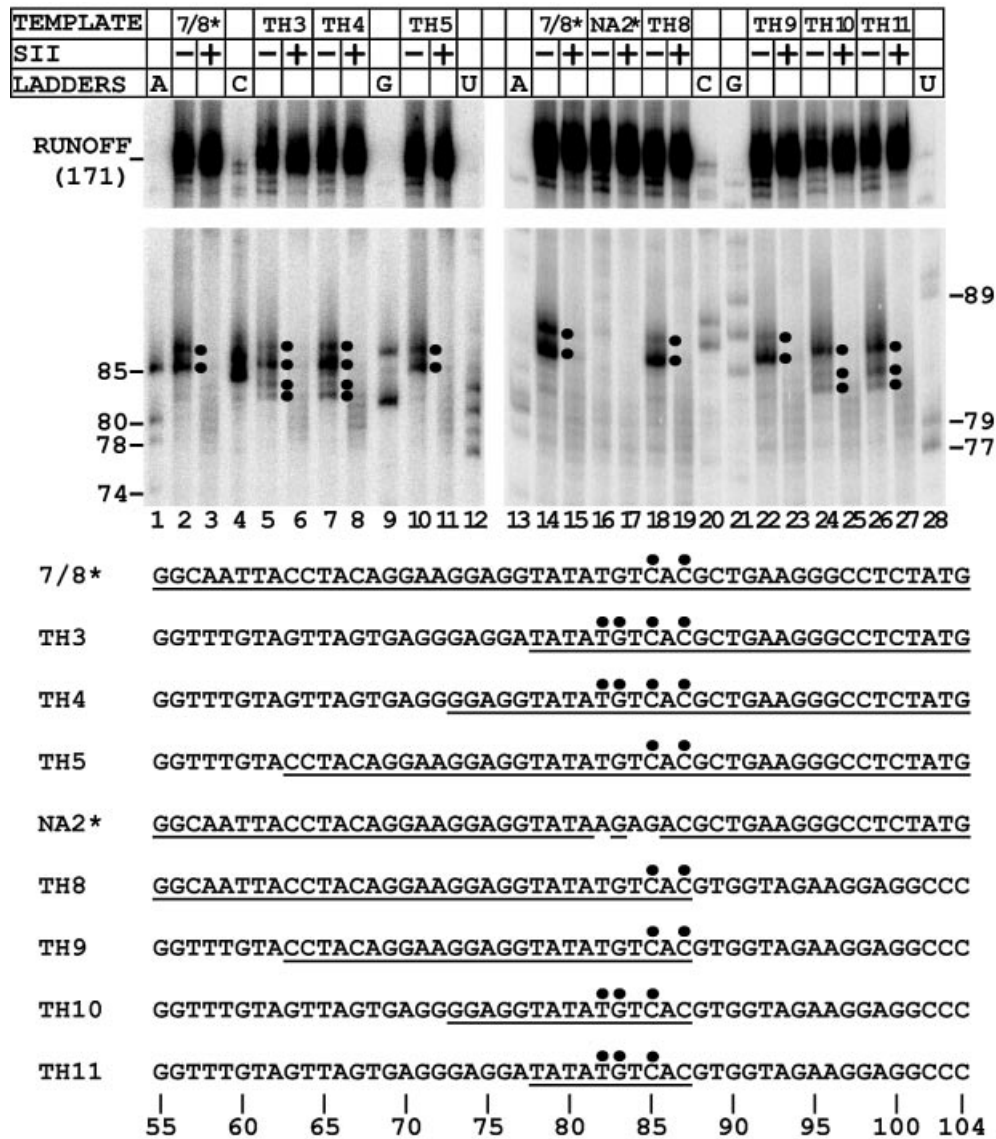


Figure 6. A 10 bp cassette of TH sequence from the 7/8 site is sufficient to cause arrest in one flanking sequence context. Templates were transcribed and size markers were generated (from the 7/8* template for the gel on the left and the NA2* template for the gel on the right) as described in Materials and Methods. In the lower part of the figure, sequences from the TH gene are underlined, distances downstream of +1 are given below the sequences and arrest positions are indicated by dots above the last nucleotide to be incorporated into the transcript.

First, we challenged transcription reactions on TH1AC and TH2AC with a 20 base oligonucleotide (TH2ACdis oligo; see Fig. 5) which should pair with bases 5–24 of the TH1AC and TH2AC transcripts and thus prevent these regions from base pairing with RNA immediately upstream of the arrest site. As can be seen in Figure 5, lanes 4–6 and 14–16, addition of the TH2ACdis oligo did restore arrest on TH1AC and TH2AC, in a dose-dependent fashion. We also altered the initially transcribed regions of the two templates, to create TH18 (derived from TH1AC) and TH14 (derived from TH2AC). In each case, the C residues between positions +4 and +18 were substituted with other bases, thus eliminating the predicted secondary structure upstream of the arrest site (see Fig. 5). As expected, these changes also restored arrest (Fig. 5, lanes 8 and 18). We concluded that the lack of arrest on the TH1AC

and TH2AC templates resulted from secondary structure in the RNA which blocked upstream translocation at the arrest site.

In order to examine the effects of the immediately upstream flanking sequence on arrest at the 7/8 site in the absence of transcript secondary structure, a different series of templates was constructed (Fig. 6). Templates in this second series (based on the pML23-100 construct; see Materials and Methods) retain the same AdML promoter upstream of +1. However, their initially transcribed regions are purine-rich, not pyrimidine-rich, and downstream of +23 they feature a C-free cassette extending from position 23 up to the first C in the sequence cassette from the 7/8 site (see Fig. 3). Therefore, since the transcript cannot form intramolecular G:C base pairs when the polymerase has begun transcribing TH sequence, one would not expect any stable transcript secondary structure

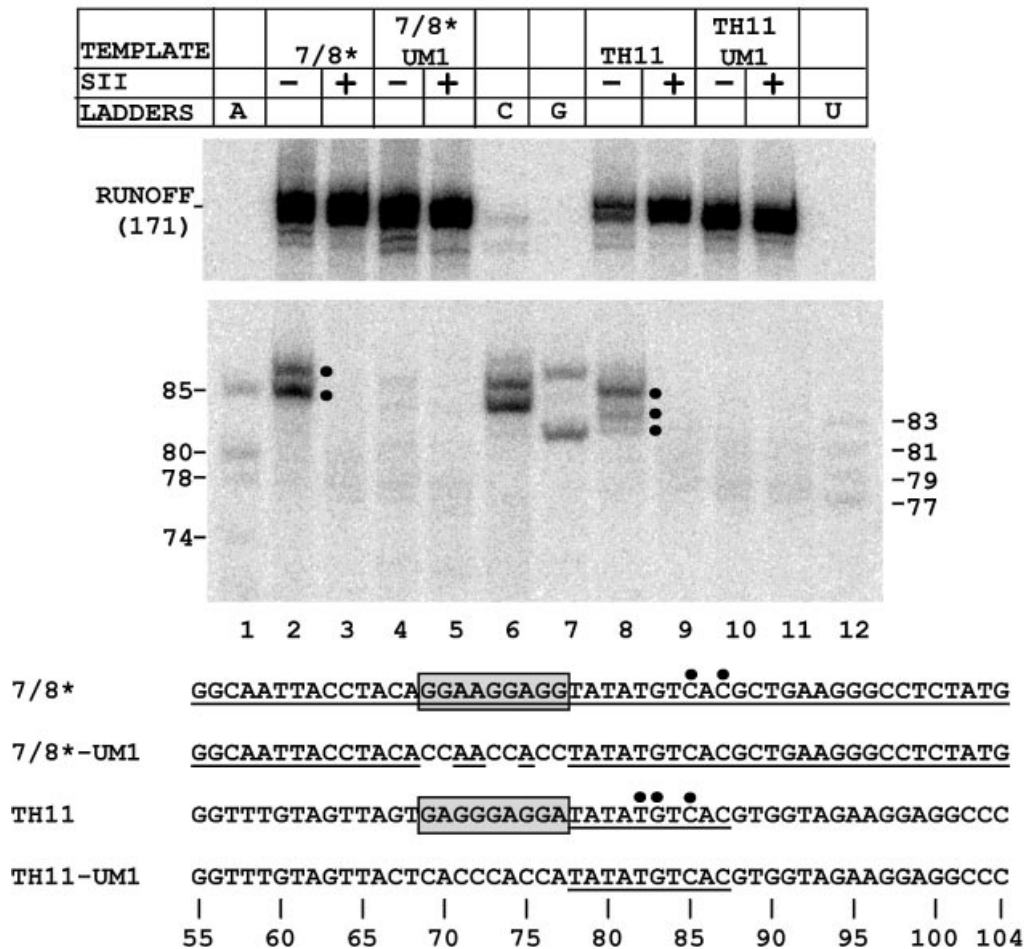


Figure 7. Arrest at the 7/8 site depends on the presence of an upstream polypurine element. Templates were transcribed and size markers were generated (from the 7/8* template) as described in Materials and Methods. In the lower part of the figure, sequences from the TH gene are underlined, distances downstream of +1 are given below the sequences and arrest positions are indicated by dots above the last nucleotide to be incorporated into the transcript. The upstream polypurine element is indicated by the shaded box for the 7/8* and TH11 constructs.

at this point, a prediction which was supported by analysis with the mfold program (30,31).

The 7/8* template is the pML23-100-based homolog to the 7/8 template discussed earlier. The 7/8* DNA also contains 50 bp of sequence from the 7/8 site. Levels of arrest on the 7/8* template were comparable to those on the 7/8 template (see Fig. 3) but one of the arrest positions was different. On the 7/8 template, arrest occurred primarily after incorporating the C at position 75 and, to a lesser extent, after the U at position 72 (Fig. 2). There are a further 10 bp of non-TH sequence upstream of each TH sequence cassette in the 23-100 series of constructs in comparison with the 20-42 series, placing the major arrest position on the 7/8* template at +85. Arrest frequency at this position is the same for both the 7/8 and 7/8* templates, but on the 7/8* template there is no significant arrest at position 82 and arrest occurred after incorporating the C at position 87 (Fig. 6).

The TH3, TH4 and TH5 templates, which are also 23-100 based, contain different amounts of TH sequence upstream of the arrest positions and the same 19 nt of downstream TH sequence as present in the 7/8 and 7/8* templates (Fig. 6). The TH8, TH9, TH10 and TH11 templates contain the same

amount of TH upstream sequence present in 7/8*, TH3, TH4 and TH5, respectively. In these constructs, however, the downstream TH sequence is replaced with additional sequence from the parental pML23-100 plasmid (Fig. 6). All of the pML23-100 based constructs supported arrest at roughly similar levels to the parental 7/8* construct (Fig. 6). This was true even for TH11, which retains only 10 bp of TH sequence. All of the constructs supported arrest at position 85, and those that contained the C-free cassette up to position -13 relative to the major arrest position (templates TH3, TH4, TH10 and TH11) also produced arrest at position 82 (Fig. 6, lanes 5, 7, 24 and 26). For the NA2* template the three pyrimidine bases at and near the major and minor arrest positions were changed to purines. These changes eliminated arrest (Fig. 6, lane 16), as observed for the analogous NA2 template (see Fig. 2).

It seemed reasonable to conclude that the 10 bp of TH gene sequence in the TH11 construct represents a minimal RNA polymerase II arrest cassette, since this segment could produce arrest in both its original sequence context and in the completely different pML23-100 background. However, the TH11 and 7/8* templates do have one feature in common in

addition to the 10 bp of original TH gene sequence. Immediately upstream of the 10 bp TH cassette, both templates have a polypurine segment of 9 (TH11) or 10 (7/8*) bases on the non-template strand (see Fig. 6). The potential importance of this sequence was suggested by the presence of a similar element immediately upstream of the arrest sites on the histone H3.3, AdML and adenosine deaminase genes, as well as in both of the other arrest sites which we observed within TH gene sequences (see Fig. 1). In order to test the potential importance of the polypurine elements, we created two new templates, 7/8*-UM1 (from 7/8*) and TH11-UM1 (from TH11) (see Fig. 7). The changes were made in the 9 bp segment immediately upstream of the 10 bp cassette of TH sequence common between TH11 and 7/8* (shaded boxes, Fig. 7). We reasoned that our modifications should preserve the original strength of the RNA–DNA hybrid, so all G residues on the non-template strand were changed to C but the A residues were left unaltered. The results of transcription of these templates were dramatic: neither 7/8*-UM1 nor TH11-UM1 supported significant levels of arrest (Fig. 7, compare lane 2 with 4 and lane 8 with 10). Thus, the minimal arrest cassette from the TH7/8 gene segment is the 10 bp sequence TATATGTC plus an upstream polypurine element.

DISCUSSION

At every RNA polymerase II arrest site which has been examined in sufficient detail, arrest has been found to be accompanied by an upstream translocation of the polymerase on the template which cannot be reversed in a short time period. The pathway to backtracking appears to pass through a paused, transcriptionally inactive intermediate state (7,32–34). The molecular basis for backtracking is not clear, but it is correlated either with mispairing (35,36) or very weak pairing of the 3' end of the transcript and the template. Existing theories of arrest require that the RNA–DNA hybrid at the point of arrest be very weak to initiate upstream translocation (reviewed in 2). The results we report for the 7/8 site do not seem to agree with this model, since a weak hybrid at the 3' end of the transcript is not needed to provoke arrest at this site. It is therefore important to consider how our findings may be accommodated within current theories on transcriptional processivity and arrest.

Some aspects of arrest at the 7/8 site are comparable to those seen at canonical T-space-T sites. Upstream translocation over a relatively long distance (8–11 nt for the 7/8 site) and the relief of arrest by SII-mediated cleavage are both observed at the 7/8 site. However, arrest at the 7/8 site is clearly independent of a weak hybrid at the 3' end of the transcript (note in particular the T74C construct, Fig. 2). Weakness in the upstream part of the RNA–DNA hybrid may be important (see Fig. 2, construct T/3/C), but other interpretations of the effect of the mutations in T/3/C are also possible, as will be discussed below.

Any model of arrest should account for the two general features of arrested polymerases: they have translocated relatively far (i.e. at least 5–6 nt) upstream of the arrest position and they have a very low probability of returning to the transcriptionally competent position over a time course of many minutes. This suggests that there are two functionally

distinct parts to an arrest site: an element which causes the initial pausing and backtracking and a second element which serves to prevent the polymerase from returning to the transcriptionally competent template location. One might ask why it is necessary to postulate a second element. Why would a strong signal to backtrack be insufficient to drive arrest? Our laboratory has documented a number of examples of paused RNA polymerase II complexes which are partially or completely backtracked and yet are fully elongation competent when chased (21,26). Thus, there is evidence that sequence elements capable of effectively driving upstream translocation of paused complexes can nevertheless fail to function as arrest sites. These observations support a two-part model for arrest, in which backtracking is a necessary but not sufficient initial event.

How do sequence elements provoke pausing and backtracking by RNA polymerase II? At T-space-T arrest sites, the long T run almost certainly serves to initiate backtracking because it encodes the weakest possible RNA–DNA hybrid. At the 7/8 site, the initial cause for upstream translocation is not clear. The particular configuration of the 3' end of the RNA–DNA hybrid at this template position must be in some way unfavorable, slowing the formation of the next bond and allowing the polymerase time to translocate upstream. This hypothetical structure at the 3' end does not require low hybrid strength at the arrest position but appears to be linked to the nature of the nucleotide incorporated at the 3' end. Eliminating all non-template strand pyrimidine residues around the major 7/8 arrest position also eliminated arrest (Fig. 2, NA2 construct; Fig. 6, NA2* construct). This is consistent with earlier data indicating that RNA polymerase II generally pauses with a pyrimidine at the 3' end (29). Our results also indicate that polypurine segments near the transcript 3' end facilitate elongation, perhaps because they bind particularly well to the polymerase near the active center.

The T/3/C mutation, in which the transcript sequence from 4 to 8 bases upstream of the major arrest position was changed from UAUAU to CACAC, considerably reduced but did not eliminate arrest (Fig. 2). The simplest interpretation of this finding is that a weak upstream hybrid in the presence of an unfavorable 3' end configuration makes upstream translocation much more likely. However, it is also possible that the signal for pausing and backtracking could be provided by the entire RNA–DNA hybrid or it could encompass the upstream region as well. In this context it is useful to mention the class II termination site for T7 RNA polymerase. The T7 polymerase terminates transcription 7–8 bp downstream of the sequence element (T/C)ATCTGTT in the non-template strand (37). Very recent results from Sousa and colleagues (38) suggest that T7 RNA polymerase changes configuration as it crosses the class II element, thereby opening up the DNA-binding clamp and poisoning the polymerase for termination. These observations could be germane to our results in two ways. First, they indicate that the transcriptional state of a RNA polymerase at a particular template location can be changed in response to a DNA sequence well upstream of the RNA–DNA hybrid. This is consistent with the importance of upstream sequences in arrest at the 7/8 site. Second, it is striking that the T7 class II termination site is only slightly different from the central segment of the arrest element in the 7/8 construct (TATATGTC; also, recall from Fig. 2 that the

C75T variant, TATATGTT, shows enhanced arrest). One must clearly be cautious in comparing the T7 and RNA polymerase II systems. The T7 class II site signals termination well downstream of itself and thus probably functions as double-stranded DNA, whereas the corresponding element in RNA polymerase II 7/8 arrest site signals arrest as soon as it is completely transcribed and thus is presumably working, at least initially, as an RNA–DNA hybrid. However, both the T7 and RNA polymerase II results would be consistent with the idea that the interaction of unusual nucleic acid structures with the RNA polymerase can misalign functional components of the enzyme and thus compromise processivity.

Once backtracking has taken place, arrest will occur only if the polymerase remains in the upstream position for an extended period. One could imagine that an element that ensures upstream trapping might not be necessary at T-space-T arrest sites. As a polymerase which has translocated upstream from such a site begins to slide back downstream, the RNA–DNA hybrid becomes progressively weaker. This would increasingly favor a retreat towards upstream locations, rather than the continued downstream translocation needed to reunite the active site and the transcript 3' end. However, if this model is correct it then becomes difficult to explain why all long T-runs are not arrest sites and why arrest takes place only in the first of the two T runs at T-space-T sites. As noted above, examination of the limited available data suggests another possibility. A relatively long (from 6 to 13 base) uninterrupted polypurine segment (non-template strand) lies immediately upstream of the arrest sites on the histone H3.3, AdML and adenosine deaminase genes (Fig. 1). Furthermore, both of the actual arrest sites from the TH10/11 segment have a long polypurine segment just upstream of the arrest positions (9 and 12 bases, respectively), but TH site 2, which did not support detectable arrest, lacks such an element (again, see Fig. 1).

As shown in Figure 7, a polypurine segment seems to be a necessary part of the arrest function at the 7/8 site. The 10 bp of native TH sequence in TH11 can cause arrest in two completely different sequence contexts (7/8* and TH11; see Fig. 6). Note that only 3 of the 9 bases upstream of the 10 bp TH sequence in TH11 are the same between the 7/8* and TH11 constructs, but all 9 of these bases are purines on both templates (see shaded sequence in Fig. 7). The functional significance of this similarity was demonstrated by the complete elimination of arrest on the 7/8*-UM1 and TH11-UM1 templates, which lack the polypurine segments but are otherwise identical to the parental templates. Very recently, we have assembled three new templates in which we have substituted the sequence AGGAAAGAGG, AAGGGGAGGA or GAAAGAGGGG for the original 7/8 polypurine element. We found that all of these new templates support arrest (L.Steele and D.Luse, unpublished results), further emphasizing the importance of this upstream element.

Given that the polypurine element is a functional part of the 7/8 arrest site, what is its specific role in the arrest process? As noted above, it could be part of the initial pausing/upstream translocation step. Since it is located in the template segment within which the polymerase active site resides after backtracking, it is attractive to suggest that the polypurine element is also involved in retaining the polymerase in the upstream location (although such a role does not exclude participation

earlier in the arrest process as well). In particular, it is tempting to speculate that the role of the polypurine tract correlates with the resistance to arrest of transcription complexes with purine residues at the 3' end of the RNA. If polypurine sequences in the transcript bind especially well with the active center of the polymerase during normal transcript elongation, thereby facilitating continued transcription, perhaps these sequences also interact unusually well with the same polymerase segment after upstream translocation, thereby making a return to the transcriptionally competent downstream location much less likely. Thus, the apparent contribution of purines to processivity at the 3' end of the RNA could also contribute to their role as upstream components of arrest sites.

The central importance of an upstream polypurine element in the 7/8 arrest site provides a significant and novel addition to our understanding of the mechanism of transcriptional processivity and arrest for RNA polymerase II. However, it should be emphasized that this polypurine element has not been fully characterized. We chose to make changes over the entire 9 base polypurine region common to 7/8* and TH11 in creating the UM1 templates, but it is possible that the functional polypurine element is shorter; for example, note that the region upstream of the arrest site in the adenosine deaminase gene has only six consecutive purine residues (Fig. 1). Also, while the potential role in arrest of an upstream polypurine element was suggested by comparison of the sequence of the 7/8 site with other arrest sites, the effect on arrest of eliminating the polypurine sequences at these other sites has not been systematically tested. Finally, it is clear that some RNA polymerase II arrest sites simply lack any obvious upstream polypurine element (see, for example, the pML5-4NR site in Fig. 1). Thus, there must also be other, unrelated sequence motifs which can serve the upstream element function.

As a final comment, we note that while it is important to refine rules for the sequence of arrest elements and to attempt to fit these data into mechanistic models, it is necessary to remember that the processes in question are extremely complex. The inability to comprehensively predict the presence of RNA polymerase II arrest sites or to explain behavior at known sites with simple models is a reminder that multiple pathways to arrest may exist. A useful analogy may be made with termination by bacterial RNA polymerase, where it has proven very difficult to develop rules which explain all examples of factor-independent termination (see, for example, 39).

In summary, while all RNA polymerase II arrest events appear to require backtracking as a central mechanistic feature, there is clearly a diversity of sequences which can function as arrest sites. Further investigation will be needed to provide a more complete picture of the molecular mechanisms involved in transcriptional arrest and in the maintenance of transcriptional competence by RNA polymerase II. This effort would be aided by the establishment of a more complete collection of RNA polymerase II arrest sites.

ACKNOWLEDGEMENTS

We thank Maria Czyzyk-Krzeska for providing the tyrosine hydroxylase gene plasmids and Richard Padgett, Donna

Driscoll, Pieter de Haseth and members of the Luse laboratory for advice and stimulating discussions during the course of these experiments. This work was supported by grant GM-29487 from the National Institutes of Health.

REFERENCES

- Landick,R. (1999) Transcription—shifting RNA polymerase into overdrive. *Science*, **284**, 598–599.
- Nudler,E. (1999) Transcription elongation: structural basis and mechanisms. *J. Mol. Biol.*, **288**, 1–12.
- Reines,D., Conaway,R.C. and Conaway,J.W. (1999) Mechanism and regulation of transcriptional elongation by RNA polymerase II. *Curr. Opin. Cell Biol.*, **11**, 342–346.
- Uptain,S.M., Kane,C.M. and Chamberlin,M.J. (1997) Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.*, **66**, 117–172.
- Conaway,J.W., Shilatifard,A., Dvir,A. and Conaway,R.C. (2000) Control of elongation by RNA polymerase II. *Trends Biochem. Sci.*, **25**, 375–380.
- Izban,M.G. and Luse,D.S. (1992) Factor-stimulated RNA polymerase-II transcripts at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J. Biol. Chem.*, **267**, 13647–13655.
- Palangat,M. and Landick,R. (2001) Roles of RNA:DNA hybrid stability, RNA structure and active site conformation in pausing by human RNA polymerase II. *J. Mol. Biol.*, **311**, 265–282.
- Reeder,T.C. and Hawley,D.K. (1996) Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell*, **87**, 767–777.
- Nudler,E., Mustaev,A., Lukhtanov,E. and Goldfarb,A. (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, **89**, 33–41.
- Kireeva,M.L., Komissarova,N., Waugh,D.S. and Kashlev,M. (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.*, **275**, 6530–6536.
- Rudd,M.D. and Luse,D.S. (1996) Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J. Biol. Chem.*, **271**, 21549–21558.
- Reines,D., Chamberlin,M.J. and Kane,C.M. (1989) Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene *in vitro*. *J. Biol. Chem.*, **264**, 10799–10809.
- Reines,D. (1992) Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J. Biol. Chem.*, **267**, 3795–3800.
- Izban,M.G. and Luse,D.S. (1992) The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'→5' direction in the presence of elongation factor SII. *Genes Dev.*, **6**, 1342–1356.
- Rudd,M.D., Izban,M.G. and Luse,D.S. (1994) The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc. Natl Acad. Sci. USA*, **91**, 8057–8061.
- Reines,D., Wells,D., Chamberlin,M.J. and Kane,C.M. (1987) Identification of intrinsic termination sites *in vitro* for RNA polymerase II within eukaryotic gene sequences. *J. Mol. Biol.*, **196**, 299–312.
- Kerppola,T.K. and Kane,C.M. (1990) Analysis of the signals for transcription termination by purified RNA polymerase II. *Biochemistry*, **29**, 269–278.
- Gu,W. and Reines,D. (1995) Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. *J. Biol. Chem.*, **270**, 11238–11244.
- Yager,T.D. and von Hippel,P.H. (1991) A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry*, **30**, 1097–1118.
- Gu,W., Powell,W., Mote,J., Jr and Reines,D. (1993) Nascent RNA cleavage by arrested RNA polymerase II does not require upstream translocation of the elongation complex on DNA. *J. Biol. Chem.*, **268**, 25604–25616.
- Samkurashvili,I. and Luse,D.S. (1996) Translocation and transcriptional arrest during transcript elongation by RNA polymerase II. *J. Biol. Chem.*, **271**, 23495–23505.
- Kash,S.F., Innis,J.W., Jackson,A.U. and Kellems,R.E. (1993) Functional analysis of a stable transcription arrest site in the 1st intron of the murine adenosine deaminase gene. *Mol. Cell. Biol.*, **13**, 2718–2729.
- Kash,S.F. and Kellems,R.E. (1994) Control of transcription arrest in intron 1 of the murine adenosine deaminase gene. *Mol. Cell. Biol.*, **14**, 6198–6207.
- Kroll,S.L., Paulding,W.R., Schnell,P.O., Barton,M.C., Conaway,J.W., Conaway,R.C. and Czyzyk-Krzeska,M.F. (1999) von Hippel-Lindau protein induces hypoxia-regulated arrest of tyrosine hydroxylase transcript elongation in pheochromocytoma cells. *J. Biol. Chem.*, **274**, 30109–30114.
- Samkurashvili,I. and Luse,D.S. (1998) Structural changes in the RNA polymerase II transcription complex during transition from initiation to elongation. *Mol. Cell. Biol.*, **18**, 5343–5354.
- Pal,M., McKean,D. and Luse,D.S. (2001) Promoter clearance by RNA polymerase II is an extended, multistep process strongly affected by sequence. *Mol. Cell. Biol.*, **21**, 5815–5825.
- Izban,M.G. and Luse,D.S. (1991) Transcription on nucleosomal templates by RNA polymerase II *in vitro*: Inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev.*, **5**, 683–696.
- Hawryluk,P.J. (2002) Characterization of an unusual RNA polymerase II arrest site. MS thesis, Case Western Reserve University.
- Keene,R.G., Mueller,A., Landick,R. and London,L. (1999) Transcriptional pause, arrest and termination sites for RNA polymerase II in mammalian N- and c-myc genes. *Nucleic Acids Res.*, **27**, 3173–3182.
- Matthews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
- Zuker,M., Matthews,D.H. and Turner,D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In Barciszewski,J. and Clark,B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 11–43.
- Neuman,K.C., Abbondanzieri,E.A., Landick,R., Gelles,J. and Block,S.M. (2003) Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell*, **115**, 437–447.
- Erie,D.A. (2002) The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim. Biophys. Acta Gene Struct. Expr.*, **1577**, 224–239.
- Zhang,C.F., Yan,H.G. and Burton,Z.F. (2003) Combinatorial control of human RNA polymerase II (RNAP II) pausing and transcript cleavage by transcription factor IIF, hepatitis delta antigen and stimulatory factor II. *J. Biol. Chem.*, **278**, 50101–50111.
- Thomas,M.J., Platas,A.A. and Hawley,D.K. (1998) Transcriptional fidelity and proofreading by RNA polymerase II. *Cell*, **93**, 627–637.
- Shaevitz,J.W., Abbondanzieri,E.A., Landick,R. and Block,S.M. (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, **426**, 684–687.
- Macdonald,L.E., Durbin,R.K., Dunn,J.J. and McAllister,W.T. (1994) Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. *J. Mol. Biol.*, **238**, 145–158.
- Mukherjee,S., Brieba,L.G. and Sousa,R. (2003) Discontinuous movement and conformational change during pausing and termination by T7 RNA polymerase. *EMBO J.*, **22**, 6483–6493.
- Reynolds,R. and Chamberlin,M.J. (1992) Parameters affecting transcription termination by *Escherichia coli* RNA polymerase 2. Construction and analysis of hybrid terminators. *J. Mol. Biol.*, **224**, 53–63.
- Kerppola,T.K. and Kane,C.M. (1988) Intrinsic sites of transcription termination and pausing in the c-myc gene. *Mol. Cell. Biol.*, **8**, 4389–4394.
- Wiest,D.K. and Hawley,D.K. (1990) *In vitro* analysis of a transcription termination site for RNA polymerase II. *Mol. Cell. Biol.*, **10**, 5782–5795.