Role of T7 RNA Polymerase His784 in Start Site Selection and Initial Transcription†

Luis G. Brieba, Robert Padilla, and Rui Sousa*

Department of Biochemistry, University of Texas Health Sciences Center, 7703 Floyd Curl Drive, San Antonio, Texas 78284-7760

Received December 18, 2001; Revised Manuscript Received February 12, 2002

ABSTRACT: The role of steric constraints vs sequence preference in start site selection by T7 RNA polymerase was investigated by using a series of synthetic promoters in which the preferred template strand 'CC' initiation sequence was moved away from its normal position relative to the −17 to −6 element of the T7 promoter. It was found that the CC sequence directs efficient initiation if placed 1 or 2 nt downstream of its normal position, but not if placed upstream, or more than 2 nt downstream, of +1. Mutagenesis revealed that part of the bias to initiate with GTP is due to an interaction between histidine 784 and the 2-amino group of a guanosine bound in the initiating triphosphate position. This interaction is also important for holding short transcripts within the transcription complex during initial transcription.

The 23 base pair T7 promoter has a tripartite structure, with distinct elements important for specific binding of T7 RNAP,[1] for promoter opening, and for facilitating the initiation and extension of transcripts to a length which allows formation of a stable elongation complex (EC). The sequence from −17 to −6 is critical for sequence-specific binding of the polymerase (*1*−*9*). Mutations in this region reduce promoter−polymerase affinity, and the crystal structure of an RNAP−promoter complex reveals numerous interactions with bases in this region (*10, 11*). Mutations in the −4 to −1 'TATA' sequence reduce promoter activity, but can be compensated for by negative supercoiling (*3*, *4*), suggesting that the role of this element is to facilitate promoter opening. In fact, while the −17 to −6 sequence must be duplex to direct specific binding of the polymerase, the promoter can be single-stranded downstream of −5 and still be fully active (*12*). Crystal structures (*10*, *11*) and permanganate reactivity (*13, 14*) confirm that the −1 to −4 sequence is melted during initiation, but that the upstream sequence remains duplex. The third part of the promoter is formed by the sequence immediately downstream of the transcription start site (+1 to +6; the initially transcribed sequence or ITS). The nontemplate (NT) strand ITS of the most active T7 promoters is 'GGGAGA' (*15*). Changes to this sequence reduce initiation rates, increase apparent NTP K_s values, and increase dissociation of short transcripts from the transcription complex during initial transcription, with changes at +1 or +2 having the most severe effects (*6*, *16*−*19*). The ITS also appears to be important in specifying the transcription start site. Though start site selection is determined primarily by steric constraints—transcription usually begins 6 nt downstream of the −17 to −6 element—there is a strong preference for initiating with 'GG'. In fact, if the T strand base at +1 is changed to G, A, or T, but C occurs at +2, then transcription will start from +2 (*6*, *20*).

The effects of changes to the ITS are similar both in fully duplex promoters and in those lacking an NT strand downstream of −5, indicating that the NT strand is unlikely to play a major role in ITS function. The role of the ITS in directing efficient initiation and in stabilizing the association of short transcripts with the initial transcription complex (ITC) may involve multiple mechanisms. When starting with the canonical 'GGGAGA' sequence, purine−purine stacking and G−C hydrogen bonding may enhance binding of the initiating NTPs, as well as transcript−template and transcript−NTP interactions. However, such mechanisms should operate with any RNAP, but not all RNAPs display a strong preference for initiating with guanosines.

Therefore, it is likely that T7 RNAP active site side chains also make base-specific interactions with either the T strand or the NTPs. To identify such interactions, we examined the structure of a T7 RNAP ITC and identified two side chains likely to make base-specific interactions important either for start site selection or for binding the transcript during initial transcription. Mutation of one of these (R425) severely reduces RNAP activity, and mutations at this position were not further characterized. However, the effects of mutations of H784 reveal that this residue makes an H-bond to the 2-amino group of the initiating guanosine triphosphate and the 3′-guanosine of the transcript. Thus, this interaction is important both for directing initiation with guanosines and for holding G-terminated short RNAs within the ITC during initial transcription.

MATERIALS AND METHODS

Mutant construction, enzyme expression, and purification were as described (*21*). Modified NTPs were from Trilink

† Supported by the Welch Foundation, by NIH Grant GM 52522 (to R.S.), and by a Fulbright-CONACYT Fellowship (to L.G.B.).

* To whom correspondence should be addressed. Phone: 210-567-8782. Fax: 210-567-8778. E-mail: sousa@biochem.uthscsa.edu.

[1] Abbreviations: RNAP, RNA polymerase; EC, elongation complex; ITS, initially transcribed sequence; NT, nontemplate; T, template; ITC, initial transcription complex; wt, wild type; TBE, Tris−borate−EDTA; nt, nucleotide(s).

10.1021/bi016057v CCC: $22.00
Published on Web 03/23/2002

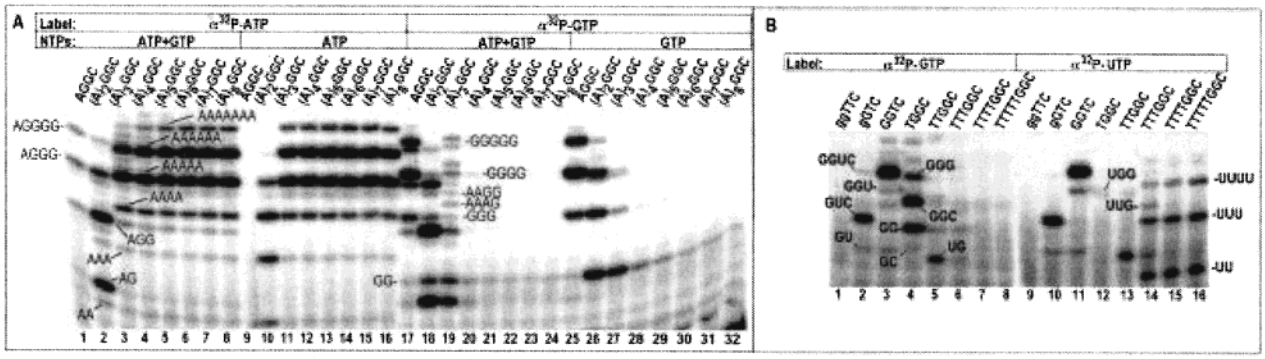

FIGURE 1: (A) T7 promoters transcribed with either ATP and GTP (lanes 1−8, 17−24), ATP only (lanes 9−16), or GTP only (lanes 25−32) with either [α-^{32}P]ATP (lanes 1−16) or GTP (lanes 17−32). The nontemplate strand promoter sequences starting at +1 are given above each lane. (B) T7 promoters transcribed with UTP, GTP, 3′-dCTP using either [α-^{32}P]GTP (lanes 1−8) or [α-^{32}P]UTP (lanes 9−16). Nontemplate strand promoter sequences are given above each gel lane with bases downstream of −1 in upper case and the −2, −1 bases in lower case.

Biotechnologies. Promoters with consensus −23 to −1 sequences and varying initially transcribed sequences (as described in the text and figure legends) were prepared using partially single-stranded 'hairpin' synthetic promoters (*22*). In these promoters, the −23 to −5 region is designed to form a duplex region upon intramolecular annealing, whereas the downstream template strand remains single-stranded. Promoter assembly was performed by heating the synthetic oligonucleotides at 85 °C for 10 min and slow-cooling for 30 min in 10 mM Tris-HCl, pH 8, 1 mM EDTA, and 50 mM NaCl. Transcription reactions were carried out in 40 mM Tris, pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$, and 5 mM DTT with NTPs as indicated in the figure legends. Reactions were carried out at 37 °C with polymerase at 3.3×10^{-7} M and promoter at 1×10^{-7} M, and were terminated after 5−15 min with an equal volume of 95% formamide, 25 mM EDTA, 0.01% xylene cyanol. Reactions were resolved by electrophoresis on 20% polyacrylamide, 1% bis-acrylamide, 6 M urea, 1×TBE gels, and imaged with a Molecular Dynamics Phosphorimager. Runoff transcription reactions were carried out with *Hin*dIII-cut pT7-5 (*23*), *Dra*I-cut p1.1.B (*24*), or *Hin*dIII-cut pT7-GA with RNAPs at 3.3×10^{-8} M and plasmid at 1×10^{-8} M in standard transcription buffer containing 0.5 mM of each nucleotide or nucleotide analogue, as indicated in the figure legends. pT7-GA was prepared by mutagenizing the initially transcribed sequence of pT7-5.

RESULTS

Sequence-Directed Initiation at +2/+3. We were first interested in determining the extent to which sequence bias in start site selection could overcome the steric constraints that normally cause initiation to start 6 nt downstream of the −17 to −6 binding element. A series of synthetic promoters were constructed in which the preferred CC initiation site on the template was moved away from the normal +1 position by interposing varying numbers of T or A bases between the TATA element and the preferred CC initiation sequence on the T strand. Figure 1A reveals how the polymerase initiates on a promoter in which varying numbers of Ts are inserted at +1. In the presence of ATP and GTP (lanes 1−8 and 17−24), initiation on a promoter containing a T at +1 occurs primarily at +2, as revealed by the weak labeling of the transcripts in reactions with [α-^{32}P]-ATP (lane 1), and the fact that most of the transcripts made in reactions with ATP, GTP, and [α-^{32}P]GTP (lane 17) migrate with the oligo-G transcripts made in the presence of GTP only (lane 5). If two Ts are inserted at +1, initiation occurs primarily at +2 and +3 as revealed by the following: (1) the weak 'AA' and oligo-A synthesis in reactions containing GTP, ATP, and [α-^{32}P]ATP (lane 2; the AA and oligo-A transcripts were identified by their comigration with transcripts made in lane 10 where only ATP is present); and (2) the presence in reactions with ATP, GTP, and [α-^{32}P]-GTP (lane 18) of both oligo-G transcripts (identified by comigration with transcripts in lane 26) and transcripts which comigrated with those in the ATP-labeled reaction in lane 2. Moving the CC site even further from +1 by introducing three Ts (lanes 3, 11, 20, 27) into the template shifts most initiation back to +1 so that oligo-A transcripts are the predominant products even if both ATP and GTP are present (note that most of the transcripts in lane 3 comigrate with the transcripts made in lane 11 where only ATP is present). Moving the CC sequence three bases away from +1 also suppresses oligo-G synthesis in reactions with GTP only (compare lane 27 to lanes 26 and 25), and inserting more Ts into the template further suppresses initiation at CC even when only GTP is present (note the decrease in GG synthesis in lanes 27−32) and leads to predominant oligo-A transcript synthesis even when both GTP and ATP are present (lanes 3−8). Similar experiments were done with promoters in which As were introduced at +1 (Figure 1B; shown are reactions with GTP, ATP, and CTP present where the label was either GTP (lanes 1−8) or [α-^{32}P]UTP (lanes 9−18)). We also tried moving the preferred CC initiation site closer to the upstream binding site (to the −2 or −1 positions; lanes 1, 2, 9, 10 of Figure 1B) while introducing an unfavorable initiation site at +1 (AA or CA) to see if the polymerase could initiate at what are normally the −1 or −2 positions. No initiation at −1 or −2 could be detected in such a case (note absence of transcripts in lanes 1and 9). The effects of introducing As at +1 to move the CC site downstream were similar to those seen when the CC was shifted by introducing Ts. When the CC was at +2 or +3, initiation was directed to +2 or to +2 and +3, respectively, though the suppression of initiation at +1 was stronger in this case (note the absence of U-labeled transcripts in lane 12 of Figure 1B, while in lane 1 of Figure 1A we see low levels of A-labeled transcripts). When the CC was shifted even further down-

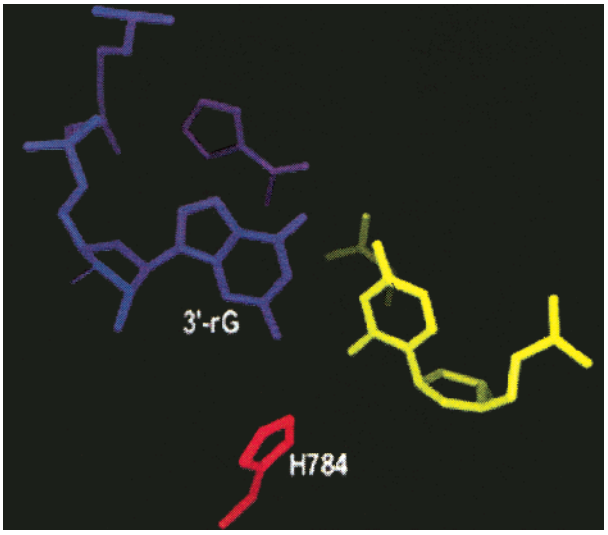

FIGURE 2: Crystal structure of a T7 RNAP ITC containing a 3 nt RNA (pdb accession no. 1QLN) reveals that the Nϵ2 of H784 is within 3.7 Å of the 2-amino group of the 3′-guanosine of the transcript.

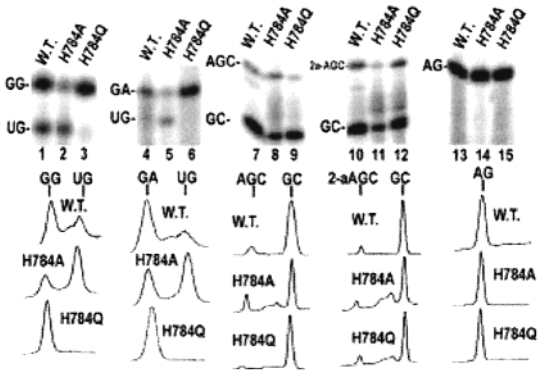

FIGURE 3: Transcription reactions run with the indicated RNAPs and promoters that initiate with either UGG (lanes 1−3), UGA (lanes 4−6), AGC (lanes 7−12), or UAG (lanes 13−15) in the presence of either 1 mM UTP, 0.05 mM GTP, and [α-^{32}P]GTP (lanes 1−3); 1 mM UTP, 0.1 mM GTP, 0.1 mM ATP, and [α-^{32}P]-GTP (lanes 4−6); 0.5 mM GTP, 0.5 mM CTP, [α-^{32}P]GTP, and either 0.5 mM ATP (lanes 7−9), 0.5 mM 2-amino-ATP (lanes 10−12), or 1 mM UTP, 0.1 mM GTP, 0.1 mM ATP, and [α-^{32}P]ATP (lanes 13−15). Scans of the gels are presented below each set of lanes, with the peaks normalized to the highest peak in each plot. Absolute intensities should not be compared between plots.

stream, initiation shifted back to +1, and oligo-U transcripts were the predominant products (note the lack of G-labeled transcripts in lanes 6−8 and the presence of oligo-U transcripts in lanes 14−16). Thus, the strong preference for initiation at a 'CC' can direct initiation to occur no further than 2 nt downstream of the usual start site, but cannot direct initiation to occur upstream of the normal start site.

Effect of H784 Mutations on Start Site Selection. Inspection of a T7 RNAP ITC structure containing a 3 nt RNA with the sequence 'GGG' (*11*) reveals that the Nϵ2 of H784 is 3.7 Å from the 2-amino group of the transcript 3′-guanosine bound in the initiating NTP binding site (the *i* site; Figure 2). While this appears far for an H-bond, the degree of precision in this 3 Å structure and the possibility of conformational changes in the active site, especially upon NTP binding, mean that an H-bond between H784 and the 2-amino group of a guanosine bound in the *i* site is a possibility. An alanine substitution at position 784 should eliminate this putative H-bond, while an H784Q substitution should place a more electronegative H-bond acceptor closer to the guanosine 2-amino group, and should therefore strengthen this H-bond.

We therefore characterized the effects of H784A and H784Q mutations on transcription start site selection (Figure 3). In a reaction containing only GTP and UTP with a promoter whose +1 to +4 T strand sequence is 'ACCG', the wt enzyme initiates predominantly at +2 and synthesizes GG dinucleotides (lane 1 of Figure 3). With H784A, initiation at +2 is reduced so that initiation at +1 and synthesis of UG dinucleotides predominate (lane 2 of Figure 3), while the H784Q substitution has the opposite effect and shifts all initiation to +2 (lane 3). A similar pattern is seen with a promoter whose +1 to +4 T strand sequence is 'ACTG' (lanes 4−6). The wt enzyme initiates predominantly at +2 to synthesize GA dinucleotides, though some initiation at +1 is evident (lane 4). The H784A mutant shows a shift to initiation at +1 and reduced utilization of guanosine as the initiating NTP (lane 5), while H784Q initiates exclusively at +2. Eliminating the H-bonding potential of the 784 side chain therefore reduces the wt preference for initiating with G, while enhancing the H-bonding potential at 784 increases this preference. Since in these experiments the choice was between starting with U or G, these effects could be specific for guanosine, or they could reflect a preference for initiating with a purine vs a pyrimidine. To distinguish these possibilities, we used a promoter that started as 'TCGA', so that initiation at +1 vs +2 involves a choice between use of A vs G as the initiating nucleotide (Figure 3, lanes 7−9). Reactions were done in the presence of ATP, GTP, and CTP, so that initiation at +1 results in synthesis of an AGC trinucleotide, while initiation at +2 produces a GC dinucleotide. With the wt enzyme, the strong preference for starting with G directs most initiation to +2 (lane 7). H784A shows reduced initiation at +2 and enhanced initiation at +1, leading to a ~6-fold increase in the ratio of +1 to +2 initiation (lane 8), while H784Q shows the opposite effect (lane 9). These observations are most simply explained in terms of the potential for the side chain at 784 to form H-bonds with the 2-amino group of a guanosine bound in the *i* site. To test this, we did similar experiments with 2-amino-ATP substituted for ATP (Figure 3, lanes 10−12). With the wt enzyme, use of 2-amino-ATP did not significantly change the amount of initiation at +1 relative to +2 (lane 10). With H784A, use of 2-amino-ATP reduced the amount of initiation at +1 relative to +2 (lane 11), while with H784Q use of 2-amino-ATP enhanced initiation at +1 (lane 12). The net effect of use of 2-amino-ATP was therefore to cause the ratio of initiation at +1 vs +2 to become similar for all three enzymes. Finally, to see if the H784 substitutions had any effect on the preference for use of G as the second nucleotide in initiation, we assessed start site preference on a promoter whose +1 to +4 T strand sequence was 'ATCG', so that initiation at +1 leads to synthesis of UA and UAG, while initiation at +2 results in an AG dinucleotide (CTP was omitted from the reactions). On this promoter, all three enzymes initiated almost exclusively at +2 (Figure 3, lanes 13−15), suggesting that substitutions at 784 do not affect the choice of the elongating (*i* + 1) nucleotide.

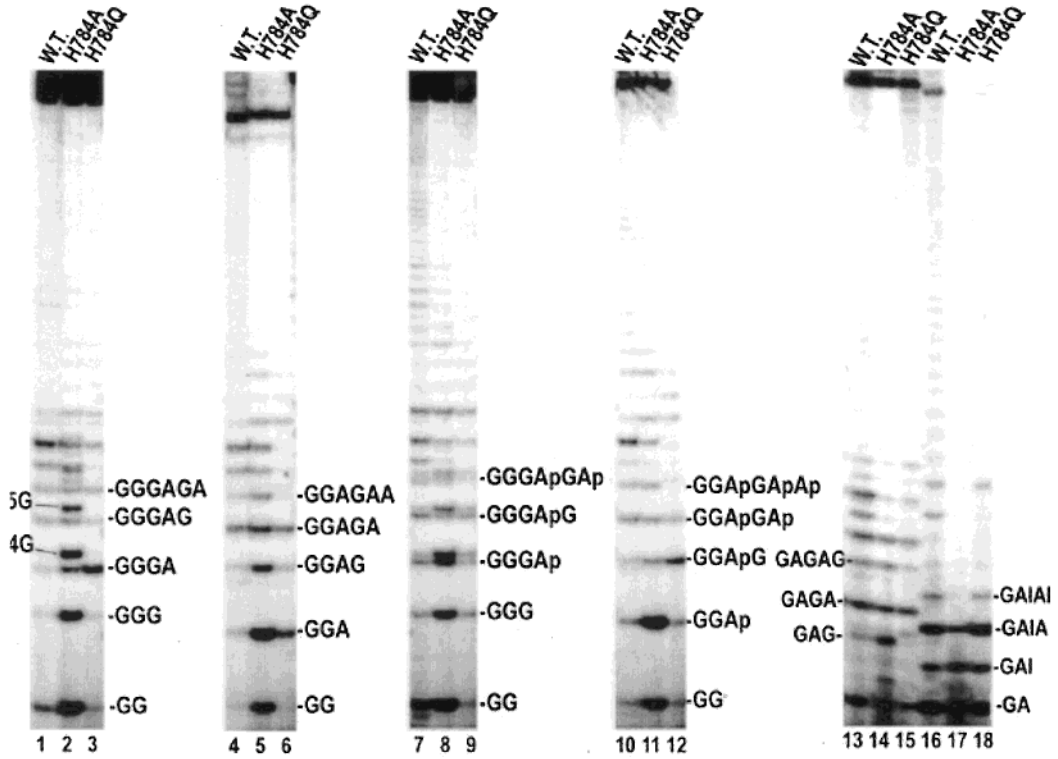

FIGURE 4: Transcription reactions run with pT7-5 (lanes 1−3, 7−9), p1.1B (lanes 4−6, 10−12), or pT7-GA (lanes 13−18) in the presence of 0.5 mM UTP, 0.5 mM CTP, and either 0.5 mM ATP, 0.5 mM GTP, and 0.1 mCi/mL [α-^{32}P]GTP (lanes 1−6); 0.5 mM 2-amino-ATP, 0.5 mM GTP, and 0.1 mCi/mL [α-^{32}P]GTP (lanes 7−12); 0.5 mM ATP, 0.5 mM GTP, 1 mM GMP, and 0.1 mCi/mL [α-^{32}P]ATP (lanes 13−15); 0.5 mM ATP, 0.5 mM ITP, 1 mM GMP, and 0.1 mCi/mL [α-^{32}P]ATP (lanes 16−18). Polymerases used in the reactions are indicated above each lane. The +1 to +6 ITS of pT7-5 is GGGAGA, of p1.1B is GGAGAA, and of pT7-GA is GAGAGA. Transcript sequences are as given in the figure, and oligo-G transcripts ('4G', '5G') are indicated in lane 2.

These experiments showed that the side chain at 784 affects the preference for using G as the initiating nucleotide, probably by H-bonding to the guanosine 2-amino group. Since the initiating NTP and the 3′-NMP of the transcript are presumed to occupy the same position in the active site when the transcript is in the post-translocated position, substitutions at 784 could also affect interactions with transcript. We therefore assessed the effects of the H784A and H784Q substitutions on the abortive transcription properties of the polymerase on a promoter whose ITS is GGGAGA (Figure 4, lanes 1−3). Compared to the wt enzyme (lane 1), the H784A substituted enzyme (lane 2) showed substantially increased transcription abortion, particularly for the shortest RNAs (2−4 nt), and increased oligo-G synthesis ('4G', '5G' transcripts in lane 2). Oligo-G synthesis occurs when the transcript slips during initial transcription, resulting in addition of multiple G bases. Both increased transcript slippage (leading to increased oligo-G synthesis) and increased transcript dissociation (leading to increased abortion) could reflect a weakened interaction with the 3′-end of the transcript in the H784A substituted enzyme. In contrast, the H784Q substituted enzyme (lane 3) displays similar or reduced transcription abortion, relative to the wt enzyme, for short transcripts which contain a 3′-guanosine ('GG', 'GGG', 'GGGAG'). Abortion after synthesis of the 'GGGA' transcript, however, is substantially greater than with the wt enzyme. To determine if this variation in the degree of abortion with H784Q actually reflects the identity of the transcript 3′-base, rather than the transcript length, we used a promoter whose ITS is GGAGAA (lanes 4−6). On this promoter, H784A (lane 5) again displays increased abortion for short (2−4 nt) transcripts relative to wt (lane 4), while H784Q (lane 6) displays increased abortion after synthesis of the 'GGA' transcript, but similar or reduced abortion for transcripts with a 3′-guanosine ('GG', 'GGAG'). Thus, the increased abortion with H784Q after synthesis of a 4mer in lane 3, or a 3mer in lane 6, reflects the identity of the transcript 3′-base. To determine if the increased abortion with H784Q and transcripts with a 3′-adenosine was due to the lack of a 2-amino group on the adenosine, we carried out similar experiments with 2-amino-ATP instead of ATP (lanes 7−12). The H784A substituted enzyme continues to exhibit increased abortion for short RNAs (lanes 8, 11) and increased oligo-G synthesis (lane 8) with 2-amino-ATP present. However, the 2-amino group on the adenosine base completely eliminates the specific increases in abortion for the GGGA and GGA transcripts seen in reactions with ATP and H784Q (lanes 9, 12). We also examined transcription with ITP substituted for GTP. For these experiments, we used a promoter whose ITS was GAGAGA and included GMP to act as the initiating nucleotide (initiation in the presence of ITP alone was very poor). With GTP in the reaction (lanes 13−15), H784Q displayed the expected pattern of reduced abortion for short, guanosine-terminated transcripts (lane 15), a pattern of abortive transcription that was clearly distinct from that of the wt (lane 13) or H784A enzymes (lane 14). However, if ITP was substituted for GTP (lanes 16−18), abortive transcription was greatly increased, and the abortive transcript patterns of the wt and H784Q enzymes became essentially identical (at least for transcripts <7 nt in length). The ability of 2-amino-ATP and ITP substitution to eliminate the effects of the terminal transcript base on abortive

transcription by H784Q indicates that these effects involve a specific interaction between the Gln side chain and the 2-amino group.

DISCUSSION

Roles of Sequence Bias vs Steric Constraints in Start Site Selection. Steric constraints play an important role in transcription start site selection. RNAP II of higher eukaryotes initiates 25−30 nucleotides downstream of the TATA box (*25*), and *E. coli* RNAP holoenzyme initiates 8−9 nucleotides downstream of the −10 element (*26*). It is believed that these steric constraints reflect the distances between the upstream DNA or transcription factor contacts which recruit the polymerase to the promoter, and the position of the active site on the promoter-bound RNAP (*27*).

Similarly, T7 RNAP appears to be constrained to initiate a set distance downstream of the −17 to −6 contacts which anchor it to its promoter. However, this steric limitation can be at least partially overcome by a bias to initiate with guanosines, so that T7 RNAP will initiate from a C at +2 if the template base at +1 is A, G, or T (*6*, *20*). We were interested in determining the number of nucleotides that could be interposed between the upstream binding element and the preferred CC initiation site before initiation at this site was abrogated. We found that efficient initiation can occur as far downstream as +3, but that initiation at the CC sequence drops precipitously if it is moved further downstream than this (Figure 1). This was not simply due to competition from initiation at upstream sites, since if we omitted all NTPs except GTP from the initiation reactions we still observed a sharp drop in initiation at the CC sequence when it was moved more than 3 nt away from −1 (lanes 25−32 of Figure 1A). While there is therefore a strong bias for T7 RNAP to initiate with a G, steric constraints appear to limit initiation to occur within 2 nt downstream of the normal start site. Presumably this reflects a constraint on the ability of the polymerase to 'stretch', or of the DNA to loop-out or 'scrunch', so as to allow the active site of the enzyme to reach further downstream than is normally the case. However, it has been shown that if the preferred initiation sequence is connected to −1 by a nonnucleosidic linker, then the linker length can be longer than the equivalent of 2 nt of DNA while still directing T7 RNAP to initiate at the CC (*28*). It is possible that looping-out or scrunching of such a linker is more facile than scrunching of the bulkier and stiffer DNA molecule, and it may also be that the active site has a higher affinity for the intervening DNA than for a nonnucleosidic linker, so that the intervening DNA competes more effectively with the CC sequence for the active site.

Unexpectedly, we found oligo-A or oligo-U transcripts could be made with high efficiency if the CC sequence was moved far enough downstream (lanes 4−8 and 14−16 of Figure 1A,B, respectively), though it has been reported that T7 RNAP initiates very poorly with A or U (*6*, *19*). Though we have not looked at whether A- or U-initiated transcripts can be extended to allow formation of an elongation complex, these observations indicate that utilization of a template T or A as an initiation site can be enhanced if all Cs are moved far enough downstream so that they no longer compete for the active site.

Role of H784 in Start Site Selection. Inspection of a T7 RNAP ITC structure (*10*) suggests that part of the preference for initiating with a G might be due to an H-bond between the H784 side chain and the 2-amino group of a guanosine nucleotide bound in the *i* site. The effects of an H784A mutation on start site selection support this hypothesis. When given a choice between initiating with a G, A, or U, an H784A mutant showed a reduction in the preference for initiating with G. This was largely due to a decrease in the amount of G initiation, while the amount of A or U initiation remained constant or increased modestly (Figure 3, lanes 1−9). This is consistent with the hypothesis that the H784A substitution eliminates an interaction specific to the guanosine base. Conversely, a Gln side chain at 784 should make a stronger H-bond with a guanosine 2-amino group. Consistent with this, we found that an H784Q mutant showed an enhanced preference, relative to the wt enzyme, for initiating with a G (Figure 3, lanes 1−9). To determine if the relative preferences of the different enzymes for initiating with G are due to varying interactions with the guanosine 2-amino group, we tested the effects of replacing ATP with 2-amino-ATP during initiation. Such an experiment is sensitive to two confounding effects. The first is due to a change in the interaction with template, since 2-amino-ATP can make an additional H-bond with the template base and therefore can form a more stable base pair. The other is that subtle stereochemical conflicts with the RNAP active site may impede utilization of the unnatural analogue. Thus, while the naive expectation is that the wt enzyme should show a preference for initiating with 2-amino-ATP vs ATP (since both pairing with template and interaction with H784 should be strengthened), the actual result is that initiation with 2-amino-A vs A is similar for the wt enzyme (compare lanes 7 and 10 of Figure 3). This may reflect the net outcome of favorable interactions with 2-amino-ATP (more stable base pair, H-bond between H784 and the 2-amino group), and unfavorable effects (poorer utilization of the unnatural analogue for stereochemical reasons). However, by using three different mutant enzymes, we can separate the effects of the interaction between the 784 side chain and the 2-amino group from these confounding effects. Specifically, the prediction is that if the varying preferences of the mutant enzymes for G vs A in initiation are due to varying interactions with the 2-amino group, then the mutants should all show *identical* preferences for initiating with 2-amino-A vs G, since these two bases are *identical* at the position presumed to be discriminated by the 784 side chain. What we observe is that the +1/+2 initiation ratio seen with ATP and the wt enzyme (Figure 3, lane 7) is similar to that seen with 2-amino-ATP (Figure 3, lane 10). Presumably the favorable H784−2-amino interaction is negated by unfavorable stereochemical effects with the unnatural analogue. Use of 2-amino-A *decreases* initiation at +1 for H784A (Figure 3, compare lanes 8 and 11). Presumably, this is due to unfavorable stereochemical effects with no compensatory interaction with the 2-amino group. Last, use of 2-amino-A *increases* initiation at +1 for H784Q (Figure 3, compare lanes 9 and 12). Presumably this reflects especially favorable interactions with the 2-amino group which overcompensate for any unfavorable effects. The result is exactly as expected if the 784 side chain is discriminating the H-bonding potential at the 2-position of an NTP bound in the *i* site: all three enzymes show similar ratios of 2-amino-A vs G initiation (Figure 3, lanes 10−12).

Role of H784 in Productive Initiation. Since the initiating NTP and the 3′-NMP of the transcript bind the same site in the polymerase, interactions made with the initiating NTP may also be made with the 3′-NMP of the transcript. Such interactions may be especially important for holding onto the transcript during the early stages of transcription initiation, when the transcript is short and interactions neither with the template nor with the polymerase are sufficient to hold it stably within the transcription complex. The large increases in the amount of abortive transcription seen with H784A, especially for transcripts <6 nt in length (Figure 4), are therefore consistent with loss of an interaction with the 3′-base of the transcript. The increased synthesis of oligo-G transcripts (Figure 4, lanes 2, 8), which are made when the transcript slips, is also consistent with looser binding of the transcript in the H784A mutant. However, the amount of transcription abortion at any given transcript length is the net outcome of the rate at which the transcript is extended and the rate at which it dissociates. Since any active site mutation will tend to reduce transcript extension rates, the effects of the H784A mutation may reflect reduced extension rates as well as increased transcript dissociation. However, the distinctive effects of the H784Q mutation on abortive transcription reveal the presence of base-specific interaction with the 3′-NMP of the transcript during initial transcription. This mutant shows increased abortion (relative to wt) of short (<6 nt) transcripts terminated with A, but shows similar or decreased abortion of transcripts terminated with G (Figure 4, lanes 3, 6). That this reflects an interaction with the guanosine 2-amino group is shown by experiments with 2-amino-ATP and ITP (lanes 10−12 and 13−18, respectively). Both of these analogues eliminate the 3′-base-specific effects on abortion seen with H784Q and transcripts 2−5 nt in length. Interestingly, the effects of the mutations on abortive transcript patterns differ for transcripts 2−5 vs 6−8 nt in length. Over the 6−8 nt range, H784Q generally shows less abortion than the wt enzyme, while H784A shows a similar level. It may be that differences in the nature of the interactions with the 3′-NMP of the transcript become less important as the transcript grows longer and makes more interactions with the template and polymerase. Alternatively, polymerase−transcript interactions may change as the transcript is extended beyond ∼5 nt, possibly because of conformational changes in the enzyme.

The first reports of the T7 RNAP−promoter complex and ITC structures proposed that H784 discriminated between rNTPs and dNTPs by making an H-bond to the ribose 2′-substituent (*10*, *11*). Subsequent mutagenesis experiments disproved this (*21*), and the results reported here show that this residue instead makes a specific interaction which is important for binding the initiating GTP, for selecting the transcription start site, and for holding short transcripts within the transcription complex during initial transcription. Such observations caution against drawing functional conclusions solely from inspection of crystal structures, particularly for enzymes such as polymerases which undergo extensive conformational changes and induced fit transitions during their reaction cycles.

REFERENCES

1. Klement, J. F., Moorfield, M. B., Jorgensen, E. D., Brown, J. E., Risman, S., and McAllister, W. T. (1990) *J. Mol. Biol. 215*, 21−29.
2. Bailey, J. N., Klement, J. F., and McAllister, W. T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2814−2818.
3. Chapman, K. A., and Burgess, R. R. (1987) *Nucleic Acids Res. 15*, 5413−5432.
4. Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R., and Burgess, R. R. (1988) *Nucleic Acids Res. 16*, 4511−4524.
5. Raskin, C. A., Diaz, G. A., Joho, K., and McAllister, W. T. (1993) *J. Mol. Biol. 229*, 805−811.
6. Imburgio, D., Rong, M., Ma, K., and McAllister, W. T. (2000) *Biochemistry 39*, 10419−10430.
7. Li, T., Ho, H. H., Maslak, M., Schick, C., and Martin, C. T. (1996) *Biochemistry 35*, 3722−3727.
8. Ujvari, A., and Martin, C. T. (1997) *J. Mol. Bio*l. *273*, 775−781.
9. Rong, M., He, B., McAllister, W. T., and Durbin, R. K. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 515−519.
10. Cheetham, G., Jeruzalmi, D., and Steitz, T. A. (1999) *Nature 399*, 80−83.
11. Cheetham, G., and Steitz, T. A., (1999) *Science 286*, 2305−2309.
12. Maslak, M., and Martin, C. T. (1993) *Biochemistry 32*, 4281−4285.
13. Place, C., Oddos, J., Buc, H., McAllister, W. T., and Buckle, M. (1999) *Biochemistry 38*, 4948−4957.
14. Brieba, L. G., and Sousa, R. (2001) *Biochemistry 40*, 3882−3890.
15. Dunn, J. J., and Studier, F. W. (1983) *J. Mol. Biol. 166*, 477−453.
16. Guajardo, R., Lopez, P., Dreyfus, M., and Sousa, R. (1998) *J. Mol. Biol. 281*, 777−792.
17. Ikeda, R. A. (1992) *J. Biol. Che*m. *267*, 11322−11328.
18. Lopez, P., Guillerez, J., Sousa, R., and Dreyfus, M. (1997) *J. Mol. Biol. 269*, 49−51.
19. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
20. Jiang, M., Rong, M., Martin, C., and McAllister, W. T. (2001) *J. Mol. Biol. 310*, 509−522.
21. Brieba, L. G., and Sousa, R. (2000) *Biochemistry 39*, 919−923.
22. Diaz, G. A., Rong, M., and McAllister, W. T. (1996) *Biochemistry 35*, 10837−10843.
23. Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074−1078.
24. Ikeda, R. A., Lin, A. C., and Clarke, J. (1992) *J. Biol. Chem. 267*, 2640−2649.
25. Hampsey, M. (1998) *Microbiol. Mol. Biol. Rev. 62*, 465−503.
26. Frederick, K., and Helmann, J. D. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 4982−4987.
27. Leuther, K. K., Bushnell, D. A., and Kornberg, R. D. (1996) *Cell 85*, 773−779.
28. Weston, B. F., Kuzmine, I., and Martin, C. T. (1997) *J. Mol. Biol. 272*, 21−30.

BI016057V