2774–2781 *Nucleic Acids Research, 1996, Vol. 24, No. 14*

Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus

Daniel S. Mytelka[1,2] and Michael J. Chamberlin[2,*]

[1]Graduate Group in Genetics and [2]Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, Berkeley, CA 94720, USA

Received February 27, 1996; Revised and Accepted June 4, 1996

ABSTRACT

We have studied a novel class of DNA sequences that cause DNA polymerases to pause. These sequences have the central consensus Py-G-C and are not necessarily adjacent to hairpins in the DNA template. Since most consensus sequences do not cause pauses under standard conditions, additional template features must exist that make it difficult to incorporate nucleotides at these positions. We believe that these pauses result from constraints that make the conformation change involved in nucleotide selection more difficult. These pauses can obscure parts of DNA sequencing ladders and prevent DNA amplification by the polymerase chain reaction. The addition of betaine, and some related compounds, relieves these pauses.

INTRODUCTION

DNA polymerases need to elongate rapidly and accurately to function effectively *in vivo* and *in vitro*, yet certain DNA regions appear to interfere with their progress. One common problem is pause sites, at which DNA polymerase molecules cease elongation for varying lengths of time. Many strong DNA polymerase pauses are at the beginnings of regions of strong secondary structure such as template hairpins. These pauses can be eliminated by methods that destabilize hairpins, but other pauses are not affected by these modifications (1–9).

Most of the studies on elongation by DNA polymerases have focused on the normal elongation process rather than the situations that lead to pauses. Evidence from studies on various DNA polymerases and HIV reverse transcriptase suggest that all of these enzymes follow a similar six-step elongation cycle consisting of nucleic acid binding, nucleotide binding, a conformation change that aids in nucleotide discrimination, phosphodiester bond formation, a conformation change that permits pyrophosphate release, and translocation or nucleic acid release. Although the initial conformation change is typically the slowest step during processive elongation, almost any of these steps can become rate limiting under appropriate conditions (reviewed in 10).

A variety of studies have been done on T7 DNA polymerase that make it a particularly useful system for studying pausing. T7 DNA polymerase consists of two proteins, the virally encoded gene 5 product and the host thioredoxin protein. Thioredoxin confers processivity on the gene 5 product, allowing it to synthesize thousands of nucleotides without releasing the template (11). By comparison, estimates for the processivity of the Klenow fragment of *Escherichia coli* DNA polymerase I range from tens to hundreds of nucleotides (10–12), giving it a higher background of incompletely extended products.

Modified versions of T7 DNA polymerase have been produced that lack the normal 3′→5′ exonuclease activity (12–13). This activity usually provides an extra opportunity for removing misincorporated nucleotides, but also inconveniently removes chain-terminating dideoxynucleotides (ddNTPs) during DNA sequencing, and would complicate analysis of polymerase pauses. Neither the mutated enzyme nor the native version has a strong nucleotide bias, accepting ddNTPs almost as well as dNTPs; this leads to relatively uniform band intensities on sequencing gels and reduced pausing at minor pause sites (13). Remaining pause positions appear as bands in all four lanes on a sequencing gel, since fragments of the lengths corresponding to the distance of pause positions from the primer accumulate regardless of whether a chain-terminating ddNTP has been incorporated. This system provides an excellent opportunity to study the mechanism of DNA polymerase pausing in a situation with potential practical applications.

MATERIALS AND METHODS

Betaine monohydrate, trimethylamine *N*-oxide (TMANO), *N*,*N*-dimethylglycine, sarcosine, tetraethylammonium acetate (TEAAc) and torula yeast RNA were purchased from Sigma (St Louis, MO). Betaine was stored at –20°C as a 5.5 M stock (pH 7.0). Nucleotide mixes and T7 DNA polymerase (Sequenase 2.0, genetically modified to lack exonuclease activity) were purchased from US Biochemicals (Cleveland, OH). The Klenow fragment of *E.coli* DNA polymerase I was purchased from Pharmacia (Piscataway, NJ).

Supercoiled DNA sequencing

Sequencing procedures generally followed the protocol of Del Sal *et al.* (14). Approximately 1 pmol of supercoiled double-stranded plasmid DNA [purified using a QIAGEN (Chatsworth, CA) column] was mixed with 2–4 pmol of an oligonucleotide primer in a volume of 10 µl containing 0.1 N NaOH. After 10 min at 68°C, the reaction was moved to room temperature and 4 µl TDMN (200 mM NaCl, 50 mM DTT, 120 mM HCl, 280 mM TES) were added. After an additional 10 min, 2 µl labeling mix

* To whom correspondence should be addressed

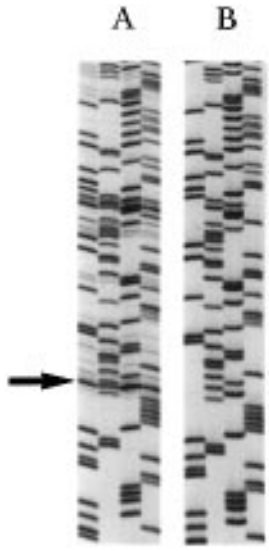

Figure 1. T7 DNA polymerase pauses. DNA sequencing was performed in the absence (**A**) or presence (**B**) of 2 M betaine. Lanes contain reactions terminated by ddATP, ddCTP, ddGTP and ddTTP respectively from left to right. The arrow indicates pause 25.

(7.5 µM dCTP, 7.5 µM dGTP, 7.5 µM dTTP) and 5 µCi [^{35}S]dATP were added, followed by 3 U T7 DNA polymerase. This mix was incubated for 5 min at room temperature, then 3.5 µl aliquots were added to four tubes preheated to 37°C containing 2.5 µl of one of the termination mixes (80 µM each dNTP, 8 µM one ddNTP, 50 mM NaCl). When testing pause suppressers, additional chemicals were added to these tubes before preheating. For example, 3.5 µl of 5.5 M betaine were added to achieve a final concentration of 2 M. After 5 min at 37°C, the reactions were stopped with 4 µl stop solution (80% deionized formamide, 1× TBE, 0.05% xylene cyanol, 0.05% bromphenol blue) and electrophoresed on a 6% polyacrylamide–7 M urea–1× TBE gel.

For the experiments on impurities in DNA preparations, plasmid DNA was prepared by a standard 'mini-prep' procedure involving lysing the cells in SDS/alkali then precipitating the DNA with isopropanol, without RNAse digestion or additional treatment (15).

T7 DNA polymerase stability assay

Forty units of T7 DNA polymerase were incubated in a volume of 80 µl of diluent (10 mM Tris–HCl pH 7.5, 5 mM DTT, 0.5 mg/ml BSA) containing 2 M betaine, 2 M proline or no osmoprotectant. After various periods at 37°C, 10 µl samples were removed and stored on ice. The remaining DNA polymerase activity in each sample was measured by adding 90 µl assay mix (which varied to equalize the osmoprotectant concentration) to obtain reactions containing 300 µM each dNTP, 40 mM Tris–HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 200 mM proline, 200 mM betaine, 75 c.p.m./pmol [^{32}P]dATP and 1.5 pmol of plasmid DNA with annealed primer. The reactions were incubated for 3 min at 37°C, then stopped with 200 µl YEP (0.5 mg/ml torula yeast RNA, 50 mM EDTA, 50 mM NaPPi). TCA-precipitable counts were determined and compared with a standard curve to determine remaining units (the DNA–polymerase ratio approached 1:1 at the highest polymerase concentrations, so this curve was not linear, but was

Pause #	Sequence	Distance
1	GATATCGAATTCCTGCAGCCC	65
2	AGCCATTTCTTCTGCGGACAT	136
3	GCTACAGGCGGGCGGCATTGG	171
4	CCGTGGCGTTGCGGCCAAGTT	54
5	CCAGCACCGCGCCAATCTCTT	122
6	ACCCAATGCCGCCCGCCTGTA	71
Mild	-c--------kCc--c-----	
7	TCGAAACGTTGCCGGAGCGCG	179
8	AGAGATTGGCGCGGTGCTGGA	108
9	CTACAGGCGGGCGGCATTGGG	172
10	TGCTCCAGTTGCCCTATTGCC	77
11	CCCAATGCCGCCCGCCTGTAG	72
Moderate	----A--m--GCmG----G--	
12	TAGAGCGGCCGCCACCGCGTG	103
13	CCGCCACCGCGTGGAGCTCAT	111
14	ATGCATTTTTCCTTAAATTTG	192
15	CGTTGCCGGAGCGCGAAAAAC	185
16	AGATTGGCGCGGTGCTGGAGG	110
17	AAACGGTTACGCACTAAACTG	172
18	GTCAGTAAATGCCGCACTTTA	206
19	TGCTACAGGCGGGCGGCATTG	170
20	ACAGGCGGGCGGCATTGGGTT	174
21	GCGTATCCGTGGCGCTATGCT	258
22	CGTATCCGTGGCGCTATGCTG	259
23	GATGAACTTCGCAGCCGTGAC	30
24	TGGGTGCCGCGCAGCGTGCGA	51
25	CCAGTTTTTCGCGCTCCGGCA	179
26	CCAATGCCGCCCGCCTGTAGC	73
Severe	-------c-GCGC-mc----t-	
Overall	-M-----MgYGCMMc--Tg--	
Prob.	-6-----6-00020-------	

Figure 2. Sites at which T7 DNA polymerase pauses. Sequences (5′→3′) are centered around individual pause sites, with all pause positions in bold. Distance is from the 5′ end of the primer. K = G or T, M = G or C, Y = C or T. Lower case consensus bases would occur at the observed frequency 10–20% of the time by chance. Upper case consensus bases would occur at the observed frequency <10% of the time by chance (for the overall consensus, this percentage is noted on the bottom line).

reproducible). The osmoprotectants had no detectable effect on overall enzyme activity, as determined by comparing units present in their presence and absence without preincubation. Most incorporation occurred during the first 1–2 min, so the 3 min incubation allowed completion of one round of synthesis per polymerase (11).

Single-stranded DNA sequencing

Single-stranded circular DNA was prepared according to the protocol of Russel *et al.* (16) as modified by Promega (17). Sequencing was performed as described for double-stranded DNA.

Sequencing with the Klenow fragment

Sequencing was performed according to the above protocol with the following modifications. Labeling mix contained 20 µCi [^{35}S]dATP and 25 mM DTT. Five units of Klenow polymerase were used in place of T7 DNA polymerase. Instead of performing complete sets of sequencing reactions, all reactions were terminated with the 'A' termination mix (250 µM ddATP, 25 µM

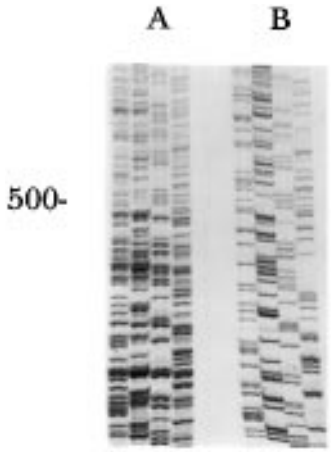

Figure 3. Effects of primer distance on T7 DNA polymerase pausing. Sequencing was performed in the absence (**A**) or presence (**B**) of 2 M betaine. The gel was run for 12 h. Lanes are as in Figure 1. The size of a 500 nt product is shown.

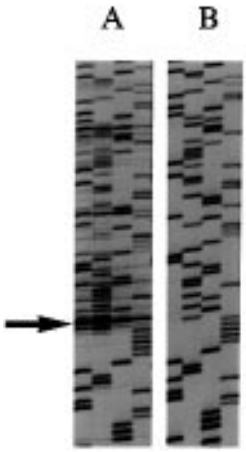

Figure 4. Effects of template purity on T7 DNA polymerase pausing. Sequencing was performed on mini-prepped DNA in the absence (**A**) or presence (**B**) of 2 M betaine. Lanes and arrow are as in Figure 1.

dATP, 250 µM dCTP, 250 µM dGTP, 250 µM dTTP). Chases were performed with 1 µl chase mix (250 µM each dNTP) for 10 min at 37°C.

RESULTS

An analysis of a set of sequences at which T7 DNA polymerase tends to pause

During supercoiled double-stranded DNA sequencing reactions using a standard protocol (see Materials and Methods), T7 DNA polymerase pauses at certain positions, leading to bands in all four lanes on the resulting denaturing polyacrylamide gel (Fig. 1A). We have selected 26 examples of pause sites at random, including pauses on several different templates at a variety of distances from the primer. We classified these pause sequences as mild, moderate or severe based on the extent of pausing: mild pauses have significant bands visible in multiple lanes, moderate pauses have at least one other band approximately as intense as the band corresponding to the nucleotide actually present at that position, and severe pauses have strong bands in all four lanes, making it impossible to determine what nucleotide is actually present at the corresponding template position.

Since DNA polymerase pause sites often occur at positions where the DNA polymerase first encounters a hairpin, we first searched the regions containing these pause sites for secondary structures using the MFOLD program from the Genetics Computer Group (Madison, WI) software package. We found that pause sites are often located near putative hairpins, but that there is no consistent positioning of the pauses: some are located at the beginnings of stems, some in loops and some at the ends of stems (data not shown). These secondary structures may contribute to elongation problems, but the lack of a consistent relationship suggests that they are not uniquely responsible. This conclusion is supported by the existence of positions at which T7 DNA polymerase pauses on both strands: if the polymerase is pausing on one strand as it first encounters a stable secondary structure, it must be just completing the synthesis of DNA complementary to that secondary structure as it elongates on the other strand.

We next attempted to locate common features in the primary structure of these pauses. As can be seen from an alignment of these sequences (Fig. 2), most of the pauses occur just after the incorporation of deoxyguanosine, in the middle of a sequence which is most commonly Py-G-C, and which essentially never differs at more than one position from this consensus. The region surrounding this trinucleotide tends to be GC-rich, with a particular tendency for a G or C to appear 3 nt after the last incorporated nucleotide. The only pause that differs by >1 nt from this consensus is pause 11, which is in a cluster of pauses in the middle of an extremely GC-rich region. There appears to be an overall tendency for pauses to occur in the vicinity of other pauses (Fig. 2; refs 4,8,9), suggesting some global or cumulative impediment to elongation in these regions.

Other factors that affect pausing by T7 DNA polymerase

The short consensus sequence for pauses described in the above section occurs every 32 base pairs by chance, yet major pauses are far less common in practice. While the consensus sequence may determine the exact position of a pause, other factors clearly affect the likelihood of pausing and the severity of pauses. As noted above, the region surrounding the consensus tends to be GC-rich (58%), but there is not a direct correlation between GC-richness and pause severity. In fact, the least GC-rich pause (pause 14) is a severe pause. However, this pause has a run of five thymidine nucleotides that might contribute to DNA bending, which may also exacerbate pauses (pauses 15 and 25 have similar runs, an abnormally high frequency for a sample of this size). There is also a tendency for pauses to occur with higher frequency farther from primers, ultimately completely obscuring the DNA sequence (Fig. 3A), but pauses can occur at any distance from the primer (Fig. 2).

Since there are no obvious specific causes of most of these pauses, we decided to look at external factors that influence the frequency and severity of pauses in general. The effects of these factors have been reduced over the last 20 years as DNA sequencing conditions have been optimized. When sequencing reactions were run at 30°C instead of 37°C, there was a substantial increase in the severity of some pauses (data not

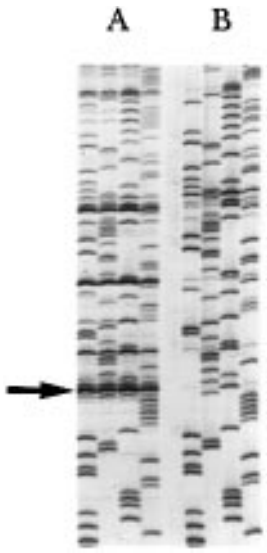

Figure 5. Effects of dITP on T7 DNA polymerase pausing. Sequencing was performed using dITP labeling and termination mixes (US Biochemicals) in the absence (**A**) or presence (**B**) of 2 M betaine. Lanes and arrow are as in Figure 1.

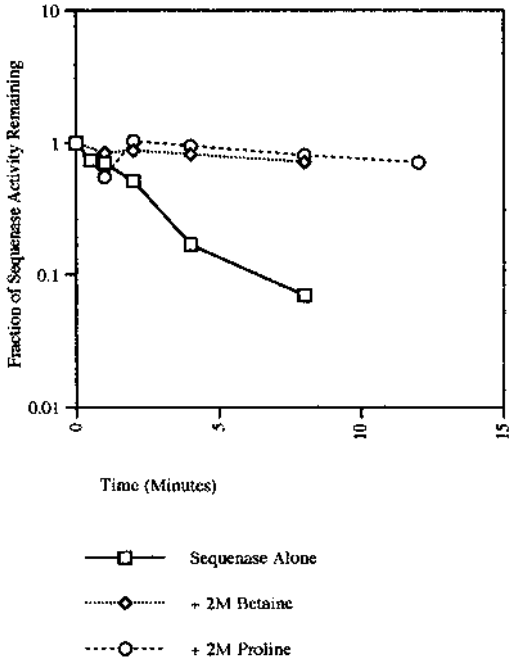

Figure 6. Betaine and proline stabilize T7 DNA polymerase equivalently. The half-life of T7 DNA polymerase was determined in the presence of 2 M betaine, 2 M proline or no osmoprotectant as described in Materials and Methods. Calculated half-lives were 21 min, 20 min and 2 h 5 min respectively.

shown). Similarly, using DNA templates that had been prepared by a mini-prep protocol (see Materials and Methods) instead of a more thorough purification led to an increase in both the frequency and severity of pauses (Fig. 4A). Finally, substituting the nucleotide analog deoxyinosine for deoxyguanosine also made pauses more numerous and more severe (Fig. 5A).

The experiments described above suggest that pauses may be caused by difficulties in the conformation change that precedes phosphodiester bond formation. Since this is normally a fairly slow step, factors that slow it further can easily affect the overall reaction rate. This is also the major step during processive elongation that can be influenced by factors both upstream and downstream from the position of the polymerase active site: upstream mismatches (such as deoxyinosine for deoxyguanosine in the trinucleotide consensus) can make this transition far less favorable, and sequences downstream have been seen to influence nucleotide selection (18) and pausing (19) by HIV reverse transcriptase. Furthermore, this transition is likely to be far slower at lower temperatures, and the presence of polynucleotides in an impure template preparation would speed the competing reaction, displacement of the polymerase. Other elongation steps are unlikely to be affected by one or more of these factors.

Betaine virtually eliminates pausing by T7 DNA polymerase

Since it is clear that some feature of the template is having a significant affect on whether T7 DNA polymerase pauses, we experimented with the inclusion of betaine (*N*,*N*,*N*-trimethyl glycine) in our sequencing reactions. Betaine is a zwitterionic osmoprotectant found in many halophilic organisms (20) that has been found to alter DNA stability such that GC-rich regions melt at temperatures more similar to AT-rich regions (21). We felt that this compound might alter the structure of the DNA at pause sites, possibly affecting the tendency of DNA polymerases to pause.

Betaine makes GC-rich and AT-rich DNA melt at the same temperature when present at a concentration of 5.2 M (21). We tried concentrations ranging from 5 mM to 3 M, adding betaine only during the elongation–termination step to minimize destabilization of the annealed primer–template; the duplex region is already somewhat lengthened by the preceding (labeling) step. We found that the addition of betaine at a concentration of 2 M eliminated virtually all Py-G-C pauses at this step, whether they were ones that occurred in normal sequencing reactions (Fig. 1) or ones exacerbated by distance from the primer (Fig. 3), low temperature (data not shown) or impurities in the template (Fig. 4). Although there are fewer pauses on single-stranded templates, those that do occur are in similar positions to the ones we have described and betaine suppresses these pauses as well (data not shown). However, pauses caused by the substitution of dITP for dGTP were not wholly eliminated by the addition of betaine (Fig. 5). Higher concentrations of betaine had no additional effect, and lower concentrations had progressively lesser effects, with concentrations <0.5 M being ineffective. One interpretation for the incomplete suppression of pauses in the presence of dITP is that the mismatch at the last incorporated position may actually be making the conformation change following nucleotide binding thermodynamically less favorable, while the other two changes are merely altering the relative kinetics for the conformation change versus polymerase displacement.

Some pauses, particularly those caused by template hairpins, can be suppressed by increasing the effective concentration of DNA polymerase, either by adding more polymerase (1) or by stabilizing the polymerase in the reaction (such as by the addition of glycerol; 22). Since betaine can stabilize some proteins (23,24), we felt that it would be important to test whether stabilization is the primary mechanism by which betaine overcomes pauses. To this end, we compared the stabilization and pause

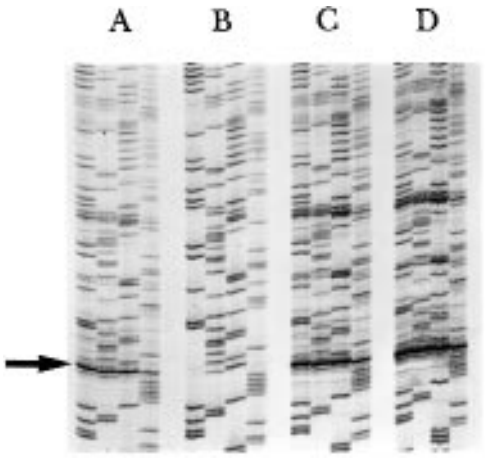

Figure 7. Proline cannot eliminate T7 DNA polymerase pauses. DNA sequencing was performed in the absence of added compound (**A**), or in the presence of 2 M betaine (**B**), 1 M proline (**C**) or 2 M proline (**D**). Lanes and arrow are as in Figure 1.

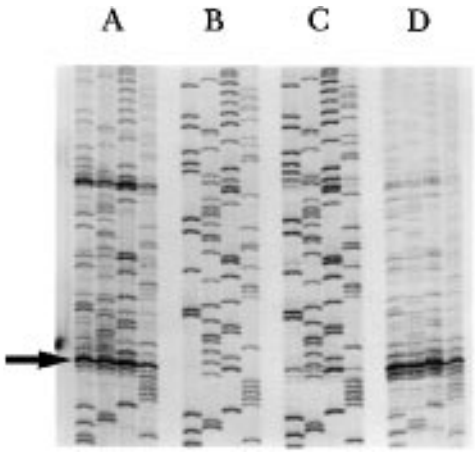

Figure 8. Betaine can chase most paused complexes. DNA sequencing was performed in the absence (**A**) or presence of 2 M betaine (**B**). A reaction identical to the A reaction was chased by adding betaine to 2 M (**C**) or by adding a comparable volume of water (**D**). Lanes and arrow are as in Figure 1.

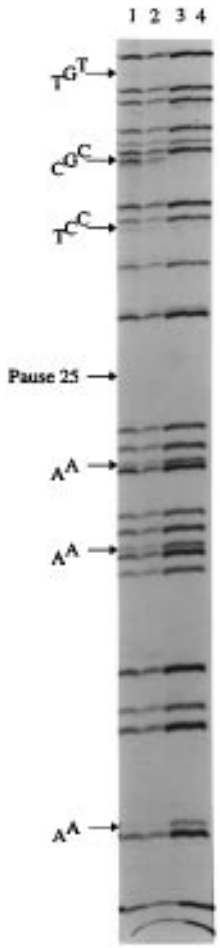

Figure 9. Pausing by the Klenow polymerase. Sequencing was carried out in the absence (1–2) or presence (3–4) of 2 M betaine. Reactions 2 and 4 were chased (see Materials and Methods). Only reaction terminated with ddATP are shown. Arrows indicate pauses, T7 DNA polymerase pause 25 and positions where dideoxynucleotide incorporation frequency changes.

suppression abilities of betaine and the imino acid proline, another osmoprotectant. We found that both chemicals stabilize T7 DNA polymerase ~10-fold against thermal denaturation (Fig. 6), but that proline has no effect on pausing by the polymerase (Fig. 7). Thus, protein stabilization is not sufficient to prevent the polymerase from pausing at these sites.

In this study, we have defined a pause rather broadly as a place at which the DNA polymerase ceases elongation; it may or may not remain associated with the resulting paused complex, but it should be at least capable of reassociating. If the consensus we have defined really causes pausing, betaine should be able to chase complexes that have been permitted to pause at a pause site. We tested this by allowing pauses to accumulate for an initial 5 min termination reaction in the absence of betaine, then adding betaine or water and incubating for an additional 5 min. The additional incubation with water had no effect on the pauses, but the addition of betaine permitted most paused complexes to resume elongation, almost as well as if betaine had been present throughout the termination reaction (Fig. 8). This result demonstrates that the phenomenon we have studied is caused by pausing by the DNA polymerase. This also confirms the above conclusion that betaine is not acting simply by stabilizing the polymerase: stabilization is not necessary for pause suppression, since the amount of polymerase present at the end of 5 min is sufficient to chase pauses in the presence of betaine, but the greater amount initially present is not sufficient in its absence. Thus, stabilization in neither necessary nor sufficient for the suppression of pausing.

Generality of pauses and pause suppression

In order to determine whether the results we have described above are generally true of DNA polymerases, we attempted to extend our results by examining the abilities of other DNA polymerases to pause while replicating DNA, and the effect of betaine on these other polymerases. Our initial studies used the Klenow fragment of *E.coli* DNA polymerase I, since this enzyme has been used in pausing studies in the past. Pausing by this polymerase is not identical to pausing by T7 DNA polymerase (Fig. 9). The Klenow fragment does not pause at pause 25, which is a severe pause for

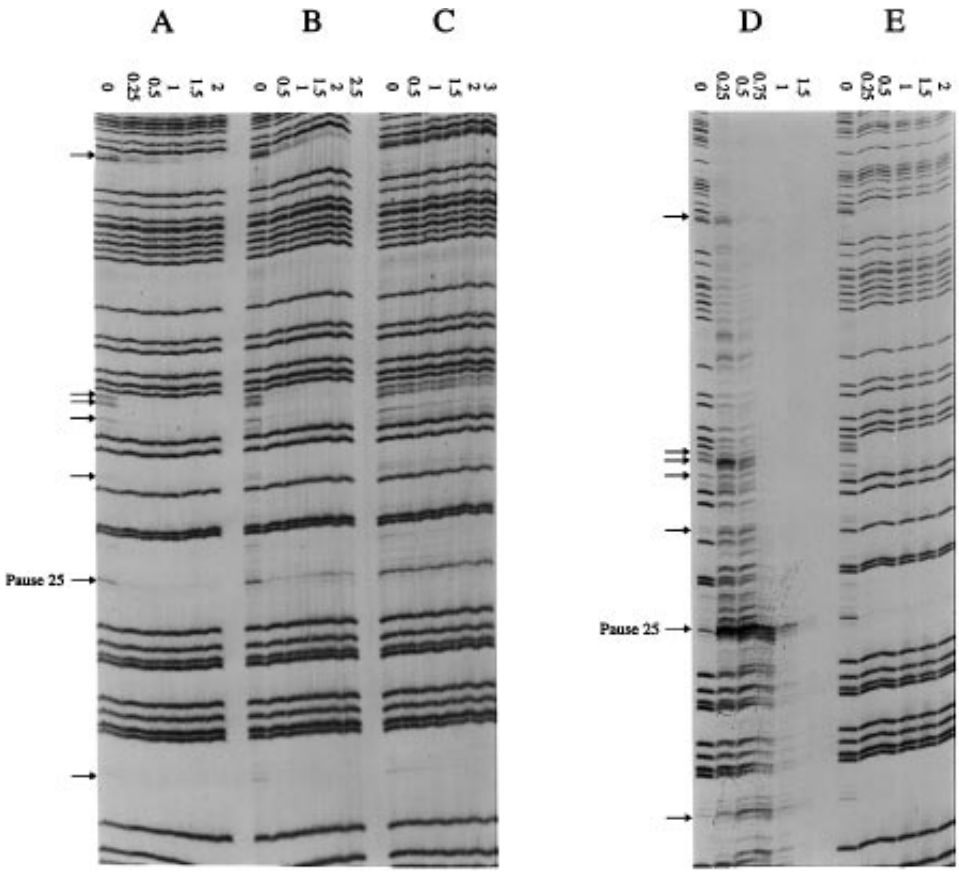

Figure 10. Effects of other chemicals on T7 DNA polymerase pausing. Sequencing was carried out in the presence of the indicated molar concentration of betaine $(CH_3)_3NCH_2COOH$, (**A**); dimethylglycine $(CH_3)_2NHCH_2COOH$, (**B**); sarcosine $CH_3NH_2CH_2COOH$, (**C**); TEAAc $(CH_3CH_2)_4N{\cdot}Ac$, (**D**); or TMANO $(CH_3)_3NO$, (**E**). Only reactions terminated with ddATP are shown; arrows indicate pauses.

T7 DNA polymerase. It does, however, pause at three pauses that are weak pauses by T7 DNA polymerase. These three pauses are all within 1 nt of the consensus sequence described above: TCC, CGC and TGT. In each case, the pause is not chased by the addition of more nucleotides (Fig. 9, lane 2), but is prevented by the presence of 2 M betaine (lane 3). This data suggests that Klenow stops at a similar consensus, but that it has somewhat different regions that it has difficulties elongating through. The nature of the pauses is probably similar, however, since betaine still prevents them. Intriguingly, betaine also seems to alter nucleotide selection, at least at some positions (making it easier, for example, to incorporate a dideoxyadenosine in the second position of the sequence AA; Fig. 9).

While initiating trial experiments with *Taq* DNA polymerase, we discovered that the Griffith group were working on a template that provided a perfect natural test of pausing by this polymerase at the Py-G-C consensus. Wang and Griffith (University of North Carolina at Chapel Hill, unpublished data) were examining a region of the myotonic dystrophy gene containing 75 repeats of the triplet TGC. They were unable to amplify a 500 base pair region containing this sequence using normal PCR conditions, but were able to do so using our modified conditions containing 2 M betaine. We have also found higher yields of some longer fragments amplified in the presence of 1–1.5 M betaine (data not shown). These experiments suggest that *Taq* DNA polymerase will also stop at the consensus sequence, and that this pausing can also be prevented by the addition of betaine, demonstrating the generality of this type of pause.

Chemicals related to betaine can also affect pausing

In order to learn more about pausing and mechanisms of relieving pausing, we examined how betaine might be affecting the elongation process by testing other chemicals that might have similar effects, to determine how betaine is functioning and which of its properties are critical. We looked at two classes of chemicals: those that are physically similar to betaine and those that have some functional similarity to betaine.

The two chemicals most structurally similar to betaine that we tested were *N*,*N*-dimethylglycine and *N*-monomethylglycine (sarcosine). Like betaine, sarcosine has some DNA melting ability (25), and dimethylglycine is probably intermediate between the two chemicals in this regard. All three of these chemicals have comparable abilities to stabilize proteins (24), though betaine is best at overcoming the sensitivity of cells to high osmolarity and sarcosine is ineffective (26). When we examined the effects of these chemicals on T7 DNA polymerase pausing, we found that dimethylglycine was less effective than betaine, and that sarcosine was largely ineffective (Fig. 10B and C). This supports our previous hypothesis that betaine-like compounds are acting on

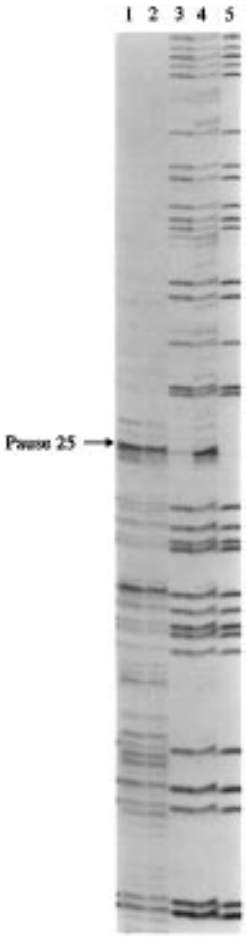

Figure 11. Betaine can partially overcome pausing caused by TEAAc. Sequencing was carried out in the absence (1–2) or presence (3–5) of 1 M betaine. Reactions 1 and 3 also contain 0.5 M TEAAc; reactions 2 and 4 contain 0.75 M TEAAc. Only reactions terminated with ddATP are shown.

the DNA (which these compounds affect differentially) rather than on the enzyme (on which they seem to act similarly). Furthermore, it suggests that the trimethylamine group is an important functional group for pause suppression (and perhaps for osmotolerance).

Numerous chemicals are known to affect the melting of DNA, but most of them are ionic and tend to interfere with enzyme activity (27). Nevertheless, we examined the effects of TEAAc on T7 DNA polymerase pausing at concentrations ranging from 0.25 to 2 M. At lower concentrations, there is substantially more pausing (particularly at normal pause sites), with higher concentrations completely blocking elongation (Fig. 10D). When betaine was added in addition to TEAAc, much of the pausing was overcome (Fig. 11), indicating that betaine and this DNA melting agent are acting inversely, but on similar targets. Proline, which appears to act solely as a polymerase stabilizer, was not able to counteract the effect of TEAAc (data not shown). These experiments indicate that interactions with DNA are important for pause suppression, supporting our prior conclusions that there are additional features of the DNA that help determine the location of pauses.

Finally, we tried the osmoprotectant TMANO. In addition to stabilizing proteins (28), TMANO has the trimethylamine group that appeared important for betaine (compared with less amino-substituted derivatives), and has been shown to affect the melting of DNA (29). TMANO proved capable of eliminating the pauses by T7 DNA polymerase in our standard system (Fig. 10E), confirming that the trimethylamine moiety is a critical part of this molecule.

DISCUSSION

In this report, we have examined the tendency of DNA polymerases to pause at nonclassic pause sites and the ability of a class of chemicals to overcome these pauses. The pauses we examined have a central consensus Py-G-C, but other factors must also be involved, since not all Py-G-C sites are pauses. One possible explanation might be that anything that slows the polymerase, including both intrinsic features of the template and factors like temperature and contaminants, leads to increases in pausing at positions where nucleotide incorporation is difficult; this would explain why pauses often occur in clusters, and why the pause sites differ somewhat from polymerase to polymerase.

Abbotts *et al.* (19) have recently examined the tendency of reverse transcriptases to 'terminate' during processive elongation on a single-stranded M13mp2 template. They quantitated termination at every position over a 255 nt region, and looked for patterns in the positions with the highest termination frequencies, finding some polymerase-dependent tendencies that differ from those we have reported here, but little that was predictable. However, many of the sites they included in their analyses were very weak pauses, and may constitute a different class or classes of pause sequence; we would expect the class of pauses we have described here to occur at a far lower frequency than the ones they analyzed. If one only examines the strongest three sites in their study, which are on average 3-fold stronger than the next strongest site and 15-fold stronger than some other pauses included as strong pauses, one find that they form a cluster at the farthest point from the primer (~140 nt away) in which the first site is at a TGC sequence and the more distant two sites are 1 nt past TGC and CGC sites respectively, and may be mislocated. These sites may be of the same sort that we have examined in this manuscript.

Our evidence suggests that the problem leading to pauses occurs at the conformation change preceding phosphodiester bond formation. The fact that betaine alters nucleotide incorporation [*Taq* DNA polymerase: C. Robinett (University of California, Berkeley), unpublished observation; Klenow: Fig. 9; T7 DNA polymerase, AMV and HIV reverse transcriptases: 30] indicates an effect on either nucleotide binding or the conformation change involved in proofreading, and the more pronounced differences with G and C (21,30; nucleotides that have less in common structurally than, e.g., purines or pyrimidines, but function together as a base pair) suggest that the effect occurs at the geometric analysis of the base pair that accompanies the conformation change rather than at the chemical interactions with the polymerase and the template that occurs during nucleotide binding. Published links between misincorporation and pausing (31–33) further implicate this part of the elongation cycle, and the significant effect of a moderate temperature change suggests a problem at the conformation change.

Although pausing appears to be a problem at the polymerase active site, many of the problems that exacerbate pausing primarily affect the surrounding template structure. However, Boyer *et al.* have seen that nucleotides up to six positions away from the active

site can affect nucleotide selection, indicating a profound effect of the overall template structure (18). Similar results have been obtained with *E.coli* RNA polymerase, where surrounding sequences can significantly affect the ease of nucleotide incorporation at specific positions (34). Betaine also appears to be affecting nucleotide incorporation and polymerase pausing by altering the overall DNA template conformation.

Rees and von Hippel have suggested that betaine acts by contacting A–T base pairs in the major groove, particularly the methyl group of thymine (21). However, the interaction with that group was specifically precluded by earlier studies that they based their conclusions on (27,35). We feel that the differences we observed when dITP was substituted for dGTP suggest that the major effect may be in the minor groove, and that betaine may be altering the hydration of this region (the water spine) to affect the local structure of the DNA molecule (36); previous studies have demonstrated that betaine is able to alter hydration in other systems (24,37). It is known that AT-rich DNA is normally more hydrated (38), and a change that increases the hydration of GC-rich regions may also increase their flexibility, since GC-rich regions are generally more rigid (39). This increased flexibility could make the conformation change involved in nucleotide selectivity more favorable.

Betaine has numerous practical applications in the laboratory. It can improve sequencing reactions, permitting longer and more accurate reads under a variety of conditions [including in cycle sequencing; N. Salama and C. Robinett (University of California, Berkeley), personal communications]. It can also improve PCR amplification of some difficult sequences, and may be of some use for long PCR. Trial experiments have suggested that betaine may even allow direct sequencing of cell pellets lysed by the addition of NaOH, which would normally yield no readable sequence.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM12010 to M.J.C. and an NSF graduate fellowship to D.S.M. Some of this work is covered by pending US or foreign patents. We thank Charles Richardson, Zoè Weaver and Karen Christie for helpful comments on drafts of this paper.

REFERENCES

1 Sherman, L.A. and Gefter, M.L. (1976) *J. Mol. Biol.* **103**, 61–76.
2 Hacker, K.J. and Alberts, B.M. (1994) *J. Biol. Chem.* **269**, 24221–24228.
3 Challberg, M.D. and Englund, P.T. (1979) *J. Biol. Chem.* **254**, 7820–7826.
4 Myers, T.W. and Romano, L.J. (1988) *J. Biol. Chem.* **263**, 17006–17015.
5 LaDuca, R.J., Fay, P.J., Chuang, C., McHenry, C.S. and Bambara, R.A. (1983) *Biochemistry,* **22**, 5177–5188.
6 Reckmann, B., Grosse, F., Urbanke, C., Frank, R., Blocker, H. and Krauss, G. (1985) *Eur. J. Biochem.* **152**, 633–643.
7 Weaver, D.T. and DePamphilis, M.L. (1984) *J. Mol. Biol.* **180**, 961–986.
8 Weaver, D.T. and DePamphilis, M.L. (1982) *J. Biol. Chem.* **257**, 2075–2086.
9 Bedinger, P., Munn, M. and Alberts, B.M. (1989) *J. Biol. Chem.* **264**, 16880–16886.
10 Johnson, K.A. (1993) *Annu. Rev. Biochem.* **62**, 685–713.
11 Tabor, S., Huber, H.E. and Richardson, C.C. (1987) *J. Biol. Chem.* **262**, 16212–16223.
12 Tabor, S. and Richardson, C.C. (1987) *Proc. Natl Acad. Sci. USA* **84**, 4767–4771.
13 Tabor, S. and Richardson, C.C. (1989) *J. Biol. Chem.* **264**, 6447–6458.
14 Del Sal, G., Manfioletti, G. and Schneider, C. (1989) *BioTechniques* **7**, 514–520.
15 Silhavy, T.S., Berman, M.L. and Enquist, L.W. (1984) *Experiments with Gene Fusions.* Cold Spring Harbor Press, Cold Spring Harbor, New York.
16 Russel, M., Kidd, S. and Kelley, M.R. (1986) *Gene* **45**, 333–338.
17 Promega (1991) *Promega Protocols and Applications Guide.*
18 Boyer, P.L., Tantillo, C., Jacobo-Molina, A., Nanni, R.G., Ding, J., Arnold, E. and Hughes, S.H. (1994) *Proc. Natl Acad. Sci. USA* **91**, 4882–4886.
19 Abbotts, J., Bebenek, K., Kunkel, T.A. and Wilson, S.H. (1993) *J. Biol. Chem.* **268**, 10312–10323.
20 Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D. and Somero, G.N. (1982) *Science* **217**, 1214–1222.
21 Rees, W.A., Yager, T.D., Korte, J. and von Hippel, P.H. (1993) *Biochemistry* **32**, 137–144.
22 Pisa-Williamson, D. and Fuller, C. (1992) *USB Comments* **19**, 29–36.
23 Winzor, C.L., Winzor, D.J., Paleg, L.G., Jones, G.P. and Naidu, B.P. (1992) *Arch. Biochem. Biophys.* **296**, 102–107.
24 Santoro, M.M., Liu, Y., Khan, S.M.A., Hou, L.-X. and Bolen, D.W. (1992) *Biochemistry* **31**, 5278–5283.
25 Thakar, M., Bilenko, A. and Becktel, W.J. (1994) *Biochemistry* **33**, 12255–12259.
26 Sauvage, D., Hamelin, J. and Larher, F. (1983) *Plant Sci. Lett.* **31**, 291–302.
27 Melchior, W.B., Jr and von Hippel, P.H. (1973) *Proc. Natl Acad. Sci. USA* **70**, 298–302.
28 Lin, T.-Y. and Timasheff, S.N. (1994) *Biochemistry* **33**, 12695–12701.
29 Buche, A., Colson, P. and Houssier, C. (1993) *J. Biomol. Struct. Dynam.* **11**, 95–119.
30 Mytelka, D.S. (1995) Ph.D. Thesis, University of California, Berkeley.
31 Barnes, W.M. (1994) *Proc. Natl Acad. Sci. USA* **91**, 2216–2220.
32 Suzuki, M., Izuta, S. and Yoshida, S. (1994) *J. Biol. Chem.* **269**, 10225–10228.
33 Fry, M. and Loeb, L.A. (1992) *Proc. Natl Acad. Sci. USA* **89**, 763–767.
34 Levin, J.R. and Chamberlin, M.J. (1987) *J. Mol. Biol.* **196**, 61–84.
35 Shapiro, J.T., Leng, M. and Felsenfeld, G. (1969) *Biochemistry* **8**, 3219–3232.
36 Drew, H.R. and Dickerson, R.E. (1981) *J. Mol. Biol.* **151**, 535–556.
37 Arakawa, T. and Timasheff, S.N. (1983) *Arch. Biochem. Biophys.* **224**, 169–177.
38 Chuprina, V.P., Heinemann, U., Nurislamov, A.A., Zielenkiewicz, P., Dickerson, R.E. and Saenger, W. (1991) *Proc. Natl Acad. Sci. USA* **88**, 593–597.
39 Hogan, M., LeGrange, J. and Austin, B. (1983) *Nature* **304**, 752–754.