The EMBO Journal vol.11 no.10 pp.3767-3775, 1992

Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure

Gary Bonner, Debabrata Patra, Eileen M.Lafer and Rui Sousa[1]

A502 Langley Hall, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

[1]Corresponding author

Communicated by D.Moras

In order to test the proposal that most nucleotide polymerases share a common active site structure and folding topology, we have generated 22 mutations of residues within motifs A, B and C of T7 RNA polymerase (RNAP). Characterization of these T7 RNAP mutants showed the following: (i) most of the mutations resulted in moderate to drastic reductions in T7 RNAP transcriptional activity supporting the idea that motifs A, B and C identify part of the polymerase active site; (ii) the degree of conservation of an amino acid within these motifs correlated with the degree to which mutation of that amino acid reduced transcriptional activity, supporting the predictive ability of this alignment in identifying the most functionally critical residues; (iii) a comparison of DNAP I and T7 RNAP mutants revealed similarities (as well as differences) between corresponding mutant phenotypes; (iv) the Klenow fragment structure is shown to provide a reasonable basis for interpretation of the differential effects of mutating different amino acids within motifs A, B and C in T7 RNAP. These observations support the proposal that these polymerase active sites have similar three-dimensional structures.

Key words: DNA polymerase/polymerases/polymerase fold/ T7 RNA polymerase

Introduction

The growth in the number of known nucleotide polymerase sequences has made it possible to identify ever more distant evolutionary relationships within this group of enzymes. The most ambitious of these attempts proposed a scheme for the structural unification of most of the known nucleotide polymerases (with the exception of the multi-subunit DNA-directed RNAPs). This proposal was based on the identification of limited amino acid sequence conservation in three motifs—dubbed A, B and C—in most of the nucleotide polymerases (Delarue *et al.*, 1990). In addition to the sequence similarity itself the scheme proposed by these investigators was supported by these observations: (i) in the crystal structure of the Klenow fragment (KF) of DNAP I (Ollis *et al.*, 1985), motifs A, B and C are close together in three-dimensional space so that they might form a catalytic site within the larger DNA-binding cleft of this enzyme; (ii) genetic and biochemical studies support the idea that residues located within these three motifs are catalytically critical in DNAP I (Mullen *et al.*, 1989; Catalano *et al.*, 1990; Polesky *et al.*, 1990; Rush and Konigsberg, 1990); (iii) at 4 Å resolution the T7 RNAP structure revealed similarities in the shape of its putative DNA binding cleft to the KF structure (Chung *et al.*, 1990a,b).

In further support of the Delarue *et al.* (1990) proposal it has been reported recently that the superficial similarity in shape between parts of T7 RNAP and KF may be a consequence of an extensive similarity in the tertiary structure of parts of these enzymes (Sousa, 1991; Sousa *et al.*, 1991). In this study, we have examined the validity of the proposal for a common active site structure in DNAP I and T7 RNAP by using targeted mutagenesis to change residues within motifs A, B and C of T7 RNAP that are conserved between DNAP I and T7 RNAP. Our results support the proposal for a common polymerase active site structure. Recently, Sankar and Porter (1992) have published a report describing a mutagenesis study of an RNA-dependent RNA polymerase that also supports the idea of a common polymerase active site structure.

Results

Targeted mutagenesis of T7 RNAP

Figure 1 shows the sequences of motifs A, B and C in DNAP I and T7 RNAP aligned according to Delarue *et al.* (1990). Above the DNAP I sequence the mutations that have been generated within these motifs in DNAP I by Polesky *et al.* (1990) are indicated. Below the T7 RNAP sequence the mutations generated in T7 RNAP and characterized in this study are indicated. The mutations fall into two classes: one group of mutations duplicates the mutations generated previously in DNAP I. This was done in order to allow a direct comparison to be made between putatively corresponding mutant phenotypes. In addition, every residue identified as conserved between T7 RNAP and DNAP I was also mutated. Because this represented a large number of residues, only one, or in some cases two, mutations were generated at each of these residue positions. In this latter set of experiments charged or polar residues were mutated to serine and non-polar residues were mutated to alanine. These particular mutations were selected because they maintain the polar or non-polar character of the side-chain in question and replace it with a small to moderately sized side-chain, i.e. they represent an economical way to probe the relative effects of mutating different residues and would usually be expected to cause only localized disturbance to the native structure of the enzyme. In cases where the wild type residue was a serine or alanine, the mutations indicated in Figure 1 were so that the residue changes would be minimally disruptive structurally.

Expression and purification of mutant enzymes

For *in vitro* characterization, mutant proteins were purified by chromatography on phosphocellulose after some initial ammonium sulfate and polymin P fractionation. For mutant

© Oxford University Press

```
                                                                882
                                        766                      |
                                         |                       N
                                         A                       S
                                         F                       A
                                         S                      A E
                                         |                    S \ /
     705                                 |                     \||
      |                                  |                     |||
/--------MOTIF A--------//------------MOTIF B----------//---MOTIF C---/
/VIVSADYSQIELRIMAHLSRDKGL//EQRRSAKAINFGLIYGMSAFGLARQLNI//RMINQVHDELVFEV/
      | |           ||       |   |     | ||   ||   |           ||
/LPLAFDGSCSGIQHFSAMLRDEVG//VTRSVTKRSVMTLAYGSKEFGFRQQVLE//ESFALIHDSFGTIP/
      | |           ||       |   |     | ||   ||   |           |||
      | |           ||       |   |     | ||   ||   |           /||
      | |           ||       |   |     | |A   ||   |          S / \
      E A           SS       S   S     A A    AA   S            A E
      S                                  S                      S A
      |                                  F                        S
     537                                 |                        N
                                        639                       |
                                                                 812
```

Fig. 1. Alignment of Delarue *et al.* (1990) motifs A, B and C in DNAP I (upper sequence) and T7 RNAP (lower sequence). Residues conserved in both enzymes are indicated by vertical lines; residues invariant in most polymerases are underlined. Above the DNAP I sequence the site-directed mutants characterized in DNAP I by Polesky *et al.* (1990) are shown. Below the T7 RNAP sequence the mutants characterized in this study are indicated.

construction, the T7 RNAP gene from the pBR322 derived plasmid pAR1219 was subcloned into a high copy number phagemid. The increase in copy number of the phagemid vector increased the expression level of the T7 RNAP protein so that the protein was sufficiently homogeneous for *in vitro* characterization after a single step of chromatography on phosphocellulose (Figure 2A). Examination of the relative expression levels of the mutant and wild type enzymes revealed that most expressed identically high levels of full-length T7 RNAP, which was soluble under non-denaturing conditions. However, mutants D537A, D537S, S539A, D552S, K631S and L637A were expressed at considerably lower levels. Gel analysis of cell extracts loaded immediately after lysis by boiling in SDS sample buffer revealed that this was due to increased protease sensitivity of these enzymes (Figure 2B). Even wild type T7 RNAP itself is quite susceptible to proteolysis (Davanloo *et al.*, 1984; Tabor and Richardson, 1985) and in a previous study, utilizing a collection of T7 RNAP mutants made by single amino acid insertion/deletion, almost all of the mutants were found to express at lower levels than the wild type enzyme (Patra *et al.*, 1992). This reduced expression was presumed to be due to the structurally disruptive effects of the amino acid insertion or deletion. We therefore interpret the high expression levels of most of the point mutants characterized here as indicating that these mutations are not significantly disruptive of the overall structure of the enzyme. This suggests that the observed phenotypes can be attributed to either 'direct effects' (removal of catalytically important side-chain features, obstruction of substrate binding sites, for example), or localized structural disruption of catalytically important parts of the enzyme. The increased protease sensitivity of some of these point mutants must be kept in mind as a caveat in interpreting the phenotypes of these mutants. However, even in these cases, we do not believe that the enhanced protease susceptibility is indicative of grossly altered structure: all of these mutants retain some transcriptional activity and the L637A mutant, for example, displays full wild type activity despite its increased protease sensitivity. However, because of their exceptional protease sensitivity, it was impossible to purify these six mutant enzymes to homogeneity. The *in vitro* characterization of these six enzymes was therefore carried out with freshly prepared cell lysates with rifampicin added to inhibit endogenous *Escherichia coli* RNAP. Estimates of the intact polymerase concentration in these lysates were carried out by visual comparison of Western blots of dilutions of these lysates and dilutions of a preparation of pure T7 RNAP of known concentration, as described in Patra *et al.* (1992). This concentration estimate is necessarily approximate (we also do not know if the proteolyzed enzyme retains any activity) and these factors must be considered in assessing the reliability of the activity measurements of this subset of mutant enzymes relative to the measurements carried out with the pure preparations of the stable mutant enzymes.

Promoter binding

Column seven of Table I summarizes the promoter binding properties of the mutants, as determined from gel retardation experiments. T7 RNAP is a poor sequence specific DNA binding protein with a low association constant for its promoter (in the 10^7 M range), which is only 100- to 1000-fold higher than its association constant for non-promoter DNA (10^4-10^5 M range; Gunderson *et al.*, 1987). These features make characterization of the thermodynamic constants of promoter binding problematical because even with short, synthetic promoter oligomers, non-specific and promoter-specific binding both contribute significantly to the total measured retention of a DNA molecule in gel retardation or nitrocellulose filter binding assays (Ikeda and Richardson, 1986; Gunderson *et al.*, 1987; Muller *et al.*, 1988). Accordingly, we did not attempt to determine if there were shifts in these association constants in the mutant enzymes; these gel retardation experiments were done to determine whether there were any drastic changes in the affinity of the mutant enzymes for the promoter that abolished promoter binding or placed the K_m for promoter utilization above the concentration used in the assays to measure kinetic parameters. It was found that all of the mutant enzymes had promoter and non-promoter DNA

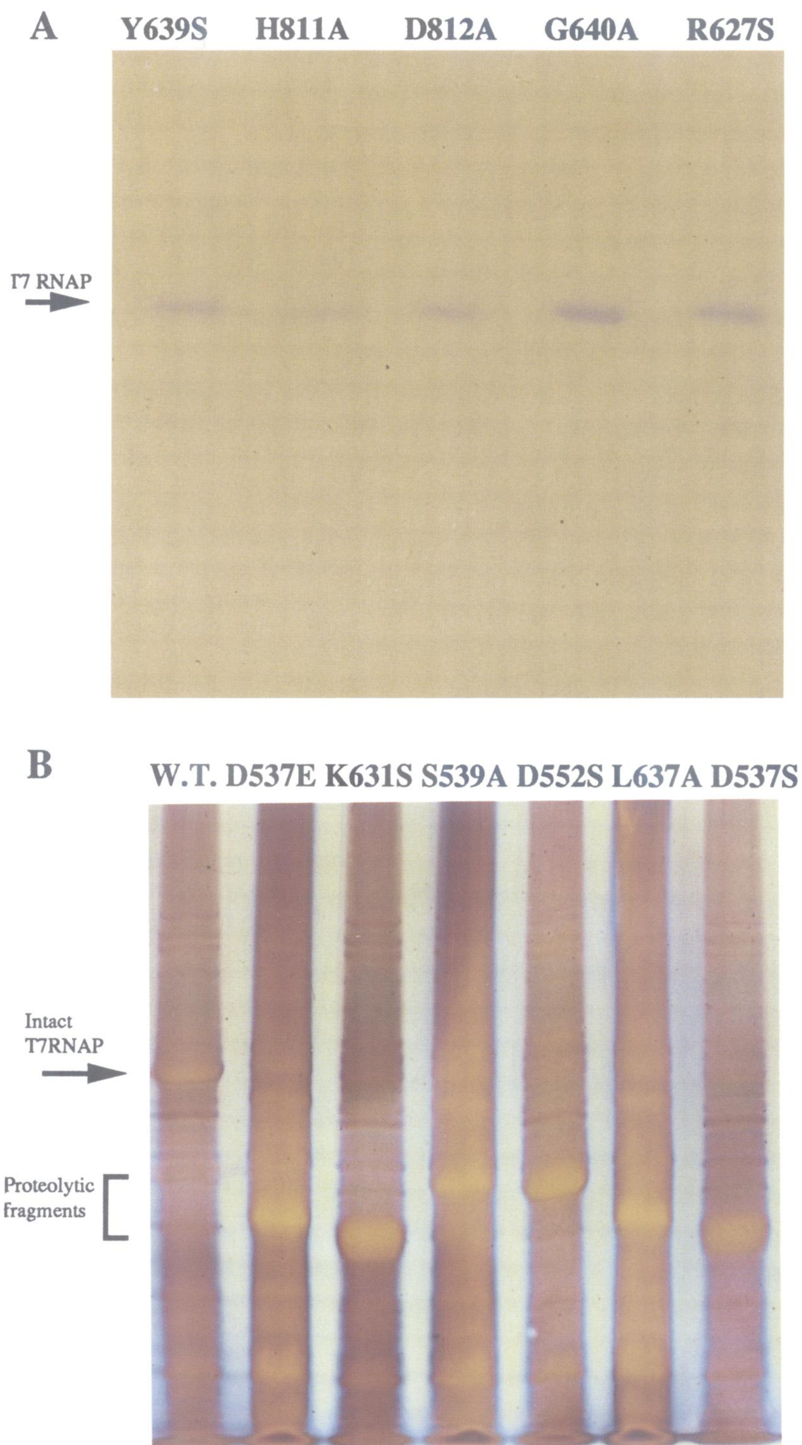

Fig. 2. (**A**) Purification of mutant T7 RNAPs. A representative Coomassie blue stained 7.5% polyacrylamide gel of mutant and wild type T7 RNAPs after purification by ammonium sulfate–polymin P fractionation and phosphocellulose chromatography. (**B**) Increased protease sensitivity of some of the T7 RNAP mutants is shown in a silver stained 7.5% polyacrylamide gel of cell lysates expressing the indicated mutant enzymes. The full-length polymerase and the major proteolytic fractions are indicated.

binding activity similar to that of the wild type enzyme. This indicates that the observed mutant phenotypes are not due to loss of promoter binding ability (Table I). As a further check, we increased the plasmid concentration in these activity assays 6-fold to determine whether any of the enzymes would be stimulated by an increase in template concentration; no stimulaton was seen indicating that the template concentration was above the K_m for promoter utilization in these assays.

Kinetic constants

Columns 1–5 of Table I summarize the kinetic constants of these enzymes. With only two exceptions, the NTP K_m values for all of these mutants are within a factor of two of the wild type levels. In two cases where a significant effect on NTP K_m is observed (G645A, Q649S), the effect is small: a 2- to 3-fold increase in NTP K_m is found. On the other hand, the effects on relative V_{max} are much more drastic: some mutants show reductions in transcriptional

Table I. Kinetic constants of T7 RNAP mutants

Mutation	K_m[1]				Relative V_{max}[2] (s)	P/NP[3]
	ATP	GTP	CTP	UTP		
Wild type	30–45	120–170	40–75	45–85	1000	3–4
D537 → S	30–60	120–170	30–75	35–85	2–4[4]	4–5
D537 → E	25–65	130–200	35–55	45–90	30–40[4]	3–4
S539 → A	40–55	120–180	30–70	55–80	30–50[4]	2–4
R551 → S	45–60	140–180	55–70	45–80	140–200	2–3
D552 → S	25–55	120–160	50–65	40–85	280–350[4]	3–4
R627 → S	30–50	150–180	40–80	35–90	30–55	2–4
K631 → S	35–60	140–190	45–75	50–90	8–10[4]	2–3
L637 → A	25–55	140–180	50–65	35–80	1000[4]	2–4
Y639 → S	40–60	130–170	30–65	55–80	0.4–0.6	3–4
Y639 → F	30–50	150–160	45–80	45–80	1000	3–4
Y639 → A	35–65	130–180	55–70	35–80	0.8–1.8	2–3
G640 → A	25–55	120–180	40–70	40–85	3.2–4.8	3–4
F644 → A	45–55	120–170	30–80	45–90	310–360	2–3
G645 → A	55–95	250–320	85–140	80–180	70–100	2–4
Q649 → S	40–65	270–360	45–80	50–85	150–190	3–4
I810 → S	25–45	120–160	30–55	35–80	12–18	2–4
H811 → S	30–60	150–190	50–75	40–80	0.07–0.17	2–3
H811 → A	50–55	130–170	45–70	50–90	0.3–0.6	2–3
D812 → A	25–55	140–180	30–70	45–85	0.1–0.2	3–4
D812 → E	35–65	120–160	40–60	50–85	0.2–0.4	2–3
D812 → S	25–55	130–170	55–70	40–85	0.01–0.03	3–4
D812 → N	40–55	150–190	55–80	55–90	0.02–0.07	3–4

[1]NTP K_m values give the ranges from three separate experiments.
[2]Relative V_{max} values are ±2 SD about the mean of 12 experiments (three experiments for each NTP). A V_{max} of 1000 corresponds to an incorporation rate of 110 (110 M NTP)/(M RNAP·s).
[3]The P/NP values give the ratio of binding to 24 bp promoter (P) and non-promoter (NP) DNAs in gel retardation experiments (ranges from three separate experiments).
[4]These enzymes are protease sensitive: the consequent poor expression of intact enzyme means that these numbers may be less reliable.

Table II. Correlation between the disruptive effects of mutating an amino acid within motifs A, B and C and the degree of conservation of that amino acid within the Delarue *et al.* alignment

Mutation	Fold reduction in activity	Percentage conservation of amino acid in the Delarue *et al.* alignment
D812 → S	50 000	100
H811 → S	8333	27
Y639 → A	1111	100
D537 → S	333	100
G640 → A	250	100
I80 → S	67	3 (F, Y, I and V: 88%)[1]
S539 → A	25	59
R627 → A	24	50
G645 → A	12	20
R551 → S	5.9	12
Q649 → S	5.9	15
F644 → A	3	25
D552 → S	2.7	23
L637 → A	1	15

[1]One of the indicated hydrophobic residues occurs at this position 88% of the time.

activity in the 1000- to 50 000-fold range. In those cases where multiple mutations were created at single residue positions, structurally conservative mutations generally result in smaller effects on relative V_{max} than non-conservative mutations. Mutants Y639F, D812E and D537E, for example, are more active than mutants where these same residues have been changed to more structurally dissimilar amino acids.

Table II offers a different presentation of some of the data from Table I. In Table II the relative decrease in V_{max} values for mutation of different residues within motifs A, B and C of T7 RNAP is presented along with the degree of conservation of each of these residues in the alignment of polymerase sequences presented in Delarue *et al.* (1990). The fold reduction in V_{max} obtained by mutating a particular residue and the degree of conservation of that residue between different polymerases correlate to a high degree.

Effects on processivity

In discussing RNAP processivity we must distinguish between processivity during two distinct phases of transcription: the poorly processive 'abortive phase' that takes place while the nascent RNA is less than 10 bases in length and the highly processive 'processive phase' that the RNAP enters into when the nascent RNA reaches a length of more than 9 bases. In order to assess the effects of these mutations on processivity, transcription reactions carried out with various mutant polymerases were analyzed by denaturing gel electrophoresis. In Figure 3, this gel analysis of transcript lengths is presented in order of decreasing mutant enzyme activity, as assessed by measuring the incorporation rate of NTP into long mRNA molecules (i.e. relative activity as defined in Table I). Because mutant enzyme activity varied over a range of approximately four orders of magnitude, it was necessary to carry out the transcription reactions with less active enzymes at high polymerase concentrations and with the more active enzymes at lower polymerase concentrations. In Figure 3 the dilution factor used for each enzyme is given; a factor of 50 means that the enzyme was used at a 50-fold lower concentration than when a factor of 1 is given {1 corresponds to a polymerase concentration of 1 μM in the transcription reaction, for some poorly active or poorly expressed polymerases [S539A (lane 16), D537E (lane 17), K631S (lane 20) and D537S (lane 23)] a lower than optimal concentration was necessary due to their very poor expression}. In addition, multiple exposures of the autoradiograph of 1, 4, 10 and 40 h (as indicated in Figure 3) were made. Figure 3 was then assembled from these multiple exposures in order to provide the clearest illustration of the relative processivities of these mutant enzymes. In examining Figure 3 these facts and a number of other considerations must be kept in mind: with exposures of 10 h or longer a number of 'spots' appear on the autoradiographs in the same size range as the abortive transcripts. These spots are starred in lane 4 of Figure 3. These spots represent minor contaminants in the radioactive label; they are present with equal intensity in all the transcription reactions and are present in the absence of polymerase and/or template (not shown). These contaminants are generally not problematical because they can be readily distinguished from the genuine abortive transcripts (labelled according to transcript size in lanes 14, 25 and 33). Even in the 40 h exposure autoradiographs, genuine abortive transcripts can be distinguished in lane 11, are clear in lane 13 and are overexposed in lanes 15, 29 and 35. However, in transcription reactions carried out with poorly active or poorly expressed enzymes (lanes 16, 17, 20, 23, 36 and 37) it is difficult to detect abortive transcripts against this background radioactivity. For these

Fig. 3. Processivity of wild type and mutant T7 RNAPs on a supercoiled plasmid template containing a T7 promoter. The degree of processivity of the wild type and mutant T7 RNAPs can be qualitatively assessed by denaturing PAGE analysis of the transcription reactions. In this figure the mutants used in each transcription reaction are indicated above each gel lane. The gel lanes are presented in order of decreasing mutant enzyme activity so that the correlation between decreasing enzyme activity (expressed in terms of the rates of incorporation of NTP into long transcripts as defined by the relative V_{max} values given in Table I) and decreased processivity is clear. Further details regarding this figure are given in Results.

mutants it is difficult to assess processivity from the analysis presented in Figure 3. The spots visible in lanes 3 and 4 are therefore not indicative of increased abortive transcript synthesis: Y639F and L637A display processivity and activity equal to that of the wild type enzyme. Processivity, of course, must be assessed in terms of the proportionate amounts of different sized transcripts synthesized: a 4 h exposure of the wild type reaction (lane 2) reveals a small amount of abortive transcripts while a 4 h exposure of the F644A reaction (lane 5) does not. This does not indicate increased processivity for mutant F644A because the wild type enzyme synthesizes proportionally more long transcripts as well: this is clear from the 1 h exposure of the wild type reaction (lane 1). For many mutants we show two different exposures of the autoradiograph in order to illustrate more clearly the differential processivity of these mutants during abortive and processive transcription. For example, the 4 h exposures in lanes 24, 26, 29, 30, 32 and 34 show the relative processivity of these mutants during abortive transcription (Y639A > Y639S > D812E > D812A > H811A > H811S) as revealed by the skewing towards shorter abortive transcript lengths with the less processive enzymes. The 10 and 40 h exposures of these same transcription reactions reveal that these mutant enzymes can progress beyond abortive transcription to synthesize longer transcripts, though it is clear that they continue to display reduced processivity during processive transcription.

The primary point illustrated by Figure 3 is therefore that decreasing mutant enzyme activity (assessed by measuring the incorporation rate of NTP into long mRNAs) consistently correlates with decreased processivity during both the abortive and processive stages of transcription. Decreased processivity during abortive transcription manifests itself as an increase in the amounts of the very short transcripts produced during this stage of transcription; increased levels of abortive transcripts are seen in the Q649S reaction (lane 11) and, relative to long transcript levels, increase as mutant enzyme activity decreases until only a small fraction of the transcripts are extended beyond the abortive transcript stage (as in lanes 24–35). It should be kept in mind that since the radioactive NTP is α-^{32}P-labelled, the molar excess of abortive transcripts relative to long transcripts is much greater than is immediately apparent from Figure 3. Decreased processivity during processive transcription manifests itself as a ladder of longer terminated transcripts extending up to the length of the full-length product (lanes 15, 18, 19, 21 and 22) or, in the case of the less active mutants (lanes 22 and 24–35), extending beyond the abortive transcription range, but terminating predominantly within a CG rich region of the template that lies between +19 and +34 (indicated in lane 25). Mutants R627S (lanes 14 and 15), I810S (lanes 18 and 19) and G640A (lanes 21 and 22) display most clearly this progressively decreasing processivity during processive transcription; mutant G640A, in particular, displays a clear intermediate pattern between the least active mutants that progress through abortive transcription but then largely terminate in the GC-rich +19 to +34 stretch (lanes 24–35) and the more active mutants

in lanes 1–13. For every mutant enzyme it is found that decreased processivity during abortive transcription is associated with decreased processivity during the processive stage of transcription and, in general, the correlation between processivity during these two stages of transcription appears similar for all of the mutants. However, exceptions can be noted; relative to other mutants, the D552S mutant shows a proportionally greater decrease in processivity during processive transcription relative to its decrease in processivity during abortive transcription (lanes 6 and 7). If the poorly active Y639 (lanes 24–27) and D812 (lanes 30–33) series of mutants are ordered in terms of their apparent processivities then this order is the same for both abortive and processive transcription. The similarly poorly active H811A (lanes 28 and 29) and H811S (lanes 34 and 35) mutants, on the other hand, are less processive than the Y639 or D812 mutants during abortive transcription, but are more processive than these mutants during processive transcription: the H811 mutants progress through the GC rich stretch more often as revealed by the presence of the longer transcripts in lanes 29 and 35 (asterisked). The 40 h exposures of the Y639 and D812 mutant reactions (not shown) show much less progression through the CG rich stretch.

Discussion

Table I presents values for the relative activity of a collection of mutant enzymes as determined from their rates of synthesis of long (DE81 retainable) transcripts on a supercoiled plasmid containing a single T7 promoter. The differences in the relative activity of these mutants are not attributable to changes in the K_m for NTP or promoter utilization, since these assays were carried out at NTP and promoter concentrations well above their respective K_m values. Under these assay conditions the differential activity of these mutants could therefore be attributed to differences in K_i (the rate of initiation of transcription, where initiation is defined as synthesis of a dinucleotide or larger mRNA molecule), the processivity parameter P (defined as the probability that a polymerase will extend a transcript or terminate at each step in NTP addition), K_{cat} (the rate of NTP addition to the growing transcript) and a poorly defined parameter (specific for DNA dependent RNAPs), which would express the efficiency with which a polymerase progresses from abortive transcription to processive transcription. As defined by McClure and Chow (1980) the processivity parameter itself is a function of two kinetic parameters, K_{cat} and K_{diss} (the rate of ternary complex dissociation) according to the equation $P = K_{cat}/(K_{cat} + K_{diss})$. From the characterization presented in this study we cannot draw rigorous conclusions regarding how each of these kinetic parameters have been affected by the described mutations. However, we will argue that the simplest and most consistent interpretation of the observations contained in Table I and Figure 3 is that the effects of these mutations are mediated largely through reductions in K_{cat}, the turnover rate for addition of an NTP to the nascent RNA.

Examination of Figure 3 eliminates two possible interpretations for the effects of these mutations. First, it is clear that the reduced activity of these mutants cannot be explained simply as a consequence of a reduced rate of transcription initiation. If this were the case, the less active mutants would not display the reduced processivity seen in Figure 3; the less active mutants would show reduced levels of total transcription, but there would not be any redistribution of transcript lengths towards shorter transcripts. The transcript patterns observed in Figure 3 are also inconsistent with the interpretation that the reduced activity of the mutant enzymes is due to a specific block or reduction in the efficiency of the transition from abortive to processive transcription. Such specific deficiencies in the abortive to processive transition have been previously described in mutant or otherwise modified T7 RNAPs. Muller et al. (1988) described an N-terminally proteolyzed enzyme that could carry out abortive transcription but could not enter processive transcription. Patra et al. (1992) described two N-terminally localized T7 RNAP mutants that displayed greatly increased levels of abortive transcript synthesis relative to long transcript synthesis. In the latter case there were no dramatic decreases in processivity during the processive stage of transcription as we observe for the mutants characterized here. In both the Muller et al. (1988) and the Patra et al. (1992) studies, the modified enzymes displayed deficits specific to the abortive stage of transcription or the point of transition from abortive to processive transcription. However, it is clear in Figure 3 that reduced enzyme activity, decreased processivity during abortive transcription and decreased processivity during processive transcription consistently correlate. Decreased processivity during abortive transcription manifests itself as an increase in the levels of the very short transcripts produced during this stage of transcription relative to the longer transcripts produced during the processive stage of transcription. It also results in a skewing towards shorter transcript lengths within the abortive transcript size range. Decreased processivity during processive transcription manifests itself as a ladder of longer terminated transcripts extending up to the length of the full-length product or, in the case of the least active mutants, extending beyond the abortive transcription range, but terminating predominantly within a CG rich region of the template.

Processivity is a function of the turnover rate for NTP addition (K_{cat}) and the dissociation rate of the ternary complex (defined as the dissociation of any one of the three components of the ternary complex). The decreased processivities of the mutant enzymes could therefore be due to decreased ternary complex stability or decreased catalytic efficiency (reduced K_{cat}). Ternary complex stability during processive transcription is mediated partly through an interaction, between the nascent RNA and the RNA polymerase, that becomes established once the nascent RNA reaches a length of more than ~9 bases. Evidence suggests that the binding site for nascent RNA on the polymerase accounts for most of the non-specific nucleic acid binding activity of T7 RNAP (Ikeda and Richardson, 1987; Muller et al., 1988). A reduction in ternary complex stability caused by a polymerase mutation might therefore be expected to be reflected in reduced non-specific nucleic acid binding. The stability of the ternary complex during abortive transcription is believed to depend partly on maintenance of the promoter–polymerase interaction throughout this phase of transcription (Ikeda and Richardson, 1986; Krummel and Chamberlin, 1989; Sousa et al., 1992). A reduction in ternary complex stability during this stage of transcription might therefore be expected to be reflected in

reduced affinity of the polymerase for the promoter. Our gel retardation experiments with promoter and non-promoter DNA (Table I) did not reveal any changes in promoter or non-promoter DNA binding that might account for the differences in ternary complex stability necessary to generate the dramatic differences in mutant enzyme processivities seen in Figure 3. The transition from abortive to processive transcription has been shown to depend, at least in part, on a reduction in the rate of ternary complex dissociation in the processive versus abortive ternary complex (Muller *et al.*, 1988). It also involves the loss of the promoter–polymerase interaction and the establishment of an interaction between the nascent RNA and a nascent RNA binding site located at least partly on the N-terminal domain of the enzyme (Muller *et al.*, 1988; Patra *et al.*, 1992). The kinetic parameter and the nature of the interactions governing ternary complex stability are therefore quite different in the processive and abortive ternary complexes. On the other hand, the catalytic mechanism of NTP addition is expected to be similar or identical during both stages of transcription. Indeed there is no evidence for a change in K_{cat} in the transition from abortive to processive transcription. Given all these considerations, it is easier to account for the observations contained in Table I and Figure 3 if we suppose that the effects of these mutations are mediated largely through reductions in K_{cat}, the turnover rate for NTP addition to the nascent chain. This is particularly the case if we are to account for the high level of correlation between reduced processivity during abortive and processive transcription that is demonstrated in Figure 3. Since ternary complex stability is governed by a different set of interactions in the abortive complexes compared with the processive ternary complexes, one would expect that some mutations could effect ternary complex stability during abortive transcription, but not processive transcription, and vice versa. In those cases, a correlation between processivity during abortive and processive transcription would not be observed. Effects on the turnover rate for NTP addition to the nascent RNA would not, however, be expected to be confined to either abortive or processive transcription given the reasonable expectation of utilization of common catalytic machinery for NTP addition to the nascent RNA during both stages of transcription. Effects on K_{cat} would therefore be expected to give rise to exactly what we observe in Figure 3: a high degree of correlation between reduced processivity during abortive and processive transcription. We therefore conclude that the simplest interpretation, consistent with our observations, is that the effects of the mutations described here are mediated largely through reductions in K_{cat}. We realize, of course, that our arguments are not unequivocal and that this conclusion cannot be rigorous. A more complete characterization of these mutants in which processivity and other kinetic parameters are being characterized with a variety of templates is underway.

A further point that must be made is that the measured relative activity parameters (the relative V_{max} values of Table I) are not quantitatively equivalent to the reductions in K_{cat} that underlie these reductions in activity. This is primarily because a mutation that results in a relatively small decrease in K_{cat} can give rise to much larger reduction in relative activity as assessed by the rate of incorporation of NTP into long mRNA molecules. This is because a modest decrease in processivity during abortive transcription can greatly decrease the fraction of elongation complexes that proceed through abortive transcription to the processive stage of transcription, and processive transcription normally accounts for most (>99%) of the incorporation of NTP into mRNA. It is useful to present a numerical example of this here. For T7 RNAP K_{cat} has been measured to be ~230/s (Golomb and Chamberlin, 1974). Based on this rate constant and the measured average transcript length of T7 RNAP, Martin *et al.* (1988) estimated an upper limit for the dissociation rate during processive transcription of 0.003/s. This value is in good agreement with the measured half-life of a paused T7 RNAP elongation complex of ~5 min (Shi *et al.*, 1988). This would imply a processivity for T7 RNAP during processive transcription close to 1 ($Pp > 0.999987$). During abortive transcription the rate of termination at each NTP addition step is 5–10% (Martin *et al.*, 1988) resulting in a processivity value, $Pa = \sim 0.90-0.95$. These values, in conjunction with a measured initiation rate (K_i) of 50/min (Martin and Coleman, 1987), allowed Martin *et al.* (1988) to conclude that the switch from abortive to processive transcription must involve a decrease in the dissociation rate of the processive elongation complex compared with that of the abortive elongation complex. If we assume that the increased processivity of the elongation complex depends exclusively on a decrease in the complex dissociation rate then we can estimate that the dissociation rate during abortive transcription would be 10–20/s. Using these numbers we can determine the fraction of polymerases that achieve synthesis of a more than ~9 base mRNA during abortive transcription, at which point they would enter processive transcription. For the wild type enzyme this is: $\{(230\ s)/[(230\ s) + (20\ s)]\}^8 = \sim 50\%$. If a mutation reduces K_{cat} by a factor of 10, only $(23/43)^8 = 0.6\%$ of all transcripts would be extended to the length necessary for entrance into processive transcription. A 10-fold decrease in the turnover rate for NTP addition would then give rise to an ~100-fold decrease in relative activity. It is important to keep this in mind when comparing the magnitudes of the effects on relative V_{max} characterized here with the K_{cat} values measured for mutations in DNAP I or other polymerases which do not display two distinct phases of polymerization with very different processivities.

Our primary observation in this study is therefore that mutation of T7 RNAP residues within motifs A, B and C as defined by Delarue *et al.* (1990) usually results in large decreases in the rate of incorporation of NTP into RNA, probably as a result of a reduction in the catalytic efficiency of the polymerase. These results support the proposal that the three motifs identified in the Delarue *et al.* alignment do, in fact, identify catalytically critical regions in these polymerases. The data presented in Table II provide further support for the significance of this alignment: there is a good correlation between the degree of conservation of a particular amino acid within the alignment and the degree of reduction of transcriptional activity which results when this amino acid is mutated. The exception to this is the poorly conserved isoleucine at position 810 which, when mutated to serine, lowers activity ~70-fold. Table II shows, however, that a large hydrophobic residue occurs at this position in 88% of the polymerase sequences analyzed by Delarue *et al.* (1990) and this isoleucine is close to the most catalytically critical residue identified in this study, D812.

Because in this study we have 'duplicated' in T7 RNAP

a set of mutations previously generated in DNAP I, we can ask whether the putatively corresponding mutant phenotypes are similar in both enzymes. For the D882 series Polesky *et al.* (1990) found the following effects on DNAP I activity: D882N and D882S, K_{cat} decreased by >3000-fold; D882A, K_{cat} decreased by ~500-fold; D882E, K_{cat} decreased by ~400-fold. D882 in DNAP I aligns with D812 in T7 RNAP for which we found the following effects on activity: D812S, activity decreased by ~50 000-fold; D812N, activity decreased by ~12 000-fold; D812A, activity decreased by ~6700-fold; D812E, activity decreased by ~3400-fold. None of these mutations had a significant effect on NTP K_m in either enzyme. Significantly perhaps, the order of effect of these mutations in T7 RNAP parallels that in DNAP I: the mutations to serine and asparagine are the most disruptive in both enzymes, while the structurally conservative glutamic acid to aspartic acid mutation is the least disruptive and the mutation to alanine has an intermediate effect. While the magnitude of these effects does not appear to correlate it must be considered, as pointed out previously, that modest changes in K_{cat} can give rise to much larger changes in the relative activity of an RNAP than a DNAP. The H881A mutation in DNAP I shows a 10-fold reduction in K_{cat} but little effect on NTP K_m. The corresponding 811A mutation in T7 RNAP shows an ~4000-fold reduction in activity with little effect on NTP K_m. Lastly, the Y766S mutation in DNAP I revealed a 3- to 4-fold decrease in activity and a 2- to 3-fold increase in NTP K_m, while a Y766F mutation showed wild type NTP K_m and K_{cat} values (Carrol *et al.*, 1991) Similarly, the corresponding Y639F mutation in T7 RNAP shows wild type activity but the Y639S mutation shows a more extreme effect: an ~2000-fold reduction in activity. This comparison shows that genetic analysis identifies these three residues as catalytically critical in both enzymes and pinpoints D812/882 as the most catalytically critical of these; it also implies that it is the phenolic ring of Tyr766/659 that is the significant functional group of this residue in both enzymes. However, there are also clear differences between 'corresponding' mutant phenotypes: the Y639 and H811 mutations in T7 RNAP have larger effects on activity than would be expected given the modest reduction in K_{cat} observed when the corresponding residues were mutated in DNAP I. Therefore, while the similarities between corresponding mutant phenotypes are consistent with a gross similarity in the active site structures of these enzymes, the observed differences indicate that there must be at least modest differences in active site structure and consequent larger differences in catalytic mechanism/kinetic parameters. In and of itself this comparison offers only limited support to the argument of similarity in active site structure, not simply because some differences between corresponding mutant phenotypes are found, but primarily because it is limited to only three residues that have been mutated in both enzymes and because the mutant phenotypes are not particularly distinctive. After these three residues, the most catalytically critical residue identified in this study was D537 (corresponds to D705 in DNAP I), mutation of which reduced the activity by >300-fold. It has also been reported that mutation of D705 in DNAP I is strongly disruptive of DNAP I activity, though the fold reduction in K_{cat} or activity obtained upon mutation of this residue has not yet been reported (Joyce, 1991).

We might also ask if the phenotypes presented in Table I can be made sense of in terms of the three-dimensional structure of motifs A, B and C, as seen in the KF structure: i.e. if we assume that the regions corresponding to motifs A, B and C in T7 RNAP and KF have similar three-dimensional structures, then will the KF structure provide a reasonable basis for interpreting the differential effects of mutating different residues within these motifs? Figure 4 shows the α-carbon backbone of the KF of DNAP I centered on the structure formed by motifs A (in red), B (in green)

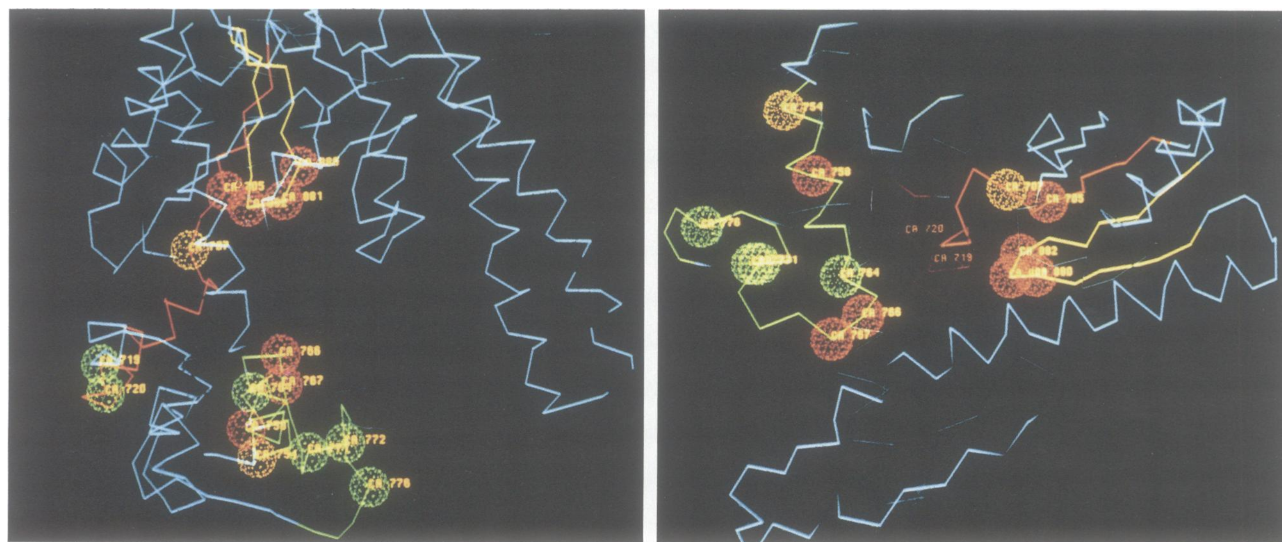

Fig. 4. The α-carbon structure of the KF is shown in blue with the segments corresponding to motifs A (in red), B (in green) and C (in yellow). The panel on the left shows the view looking down the long axis of the putative DNA binding cleft, the panel on the right shows a view looking into the structure formed by the motif A, B and C segments and perpendicular to the long axis of the DNA binding cleft. Indicated on the structure are the α-carbon positions of residues that are conserved between DNAP I and T7 RNAP. In T7 RNAP, mutation of the conserved residues, which would corresond to the residues highlighted here by van der Waals radii in red, results in large (~100-fold or more) reductions in activity, mutation of residues highlighted by yellow radii reduces activity a moderate amount (~25-fold), while mutation of residues highlighted in green only modestly affects activity (~10-fold or less). For reference, the following secondary structure designations are according to Ollis *et al.* (1985): motif A includes β-strand 9, α-helix L and the four residue connecting element between these two structures; motif B includes α-helix O and the 12 residue irregularly structured segment C-terminal to this helix; motif C includes β-strands 12 and 13, and the segment connecting these two strands.

and C (in yellow). Indicated on this structure are the α-carbon positions of residues in DNAP I that are identical in T7 RNAP and were mutated in this study. These α-carbon positions are highlighted by van der Waals radii in red if they correspond to T7 RNAP residues which show drastic (~100-fold or more) reductions in activity when mutated, in yellow if they correspond to residues that show moderate reductions in activity (~25-fold), and in green if they show small effects on activity (~10-fold or less). Figure 4 shows that a consistent interpretation of these mutant phenotypes can be made based on the KF structure: it appears that the degree to which the mutation of a particular residue affects the activity correlates with the proximity of that residue to a catalytic site that would be centered on the C-terminal half of KF helix O, the turn segment immediately C-terminal to the end of helix O and the turn connecting β-strands 12 and 13. The exception to this is the L637A mutation, which occurs at the C-terminal end of helix O, but has no effect on activity. We note, however, that this leucine is the least well conserved residue mutated in this study and that alanine occurs at this position in motif B 15% of the time in the Delarue *et al.* alignment. Alanine is therefore probably an allowed substitution at this position.

Materials and methods

Enzymes and chemicals

Restriction and nucleic acid modifying enzymes were from New England Biolabs and Bethesda Research Laboratories. Nucleotides were from United States Biochemicals.

Site-directed mutagenesis of T7 gene 1

pAR1219 (Davanloo *et al.*, 1984) was digested with *Bam*HI to release a 4.3 kb fragment containing T7 gene 1. This fragment was cloned into the *Bam*HI site of phagemid pUC118 to create pDPT7. Site-directed mutagenesis of pDPT7 was carried out using the method of Kunkel (1985) and the Bio-Rad Muta-Gene kit with the following modifications: (i) the growth period of the *dut-*, *ung-* bacterium (CJ236) containing the phagemid for preparation of single-stranded uracil-containing DNA for mutagenesis was extended from the recommended 4−6 h to 20 h in TBG medium (per l: 12 g Bactotryptone, 24 g yeast extract, 4 ml glycerol, 2.3 g KH_2PO_4 and 12.5 g K_2HPO_4 and 20 mM glucose) rather than the 2 × YT medium recommended by the manufacturer; (ii) the primer annealing step involved a 20 min incubation at 37°C; (iii) the second strand synthesis buffer (using T7 DNA polymerase, from Bio-Rad) was 5 mM Tris, pH 8.0, 0.8 mM $MgCl_2$, 0.1 mM ATP, 0.2 mM DTT, 0.1 mg/ml BSA, 0.02 mM dNTP; (iv) incubation conditions for second strand synthesis−ligation (after addition of all buffer and enzyme components) were 1 min on ice followed by 15 min at 37°C, immediate cooling to −20°C and storage at −20°C until used for transformation. Mutations were identified by sequencing using modified T7 DNA polymerase (Tabor and Richardson, 1987). Only the region in the immediate neighbourhood of the targeted mutation was sequenced. In order to ensure that the observed phenotypes were in fact due to the targeted mutation the following steps were taken: (i) each mutant was isolated at least twice and each of these independent isolates was characterized to confirm that they would display identical *in vitro* phenotypes; (ii) 10 clones that were not mutated in the targeted region but were isolated during the course of the mutagenesis were characterized *in vitro*. All 10 of these clones displayed wild type activity. HPLC-purified oligonucleotides were prepared by the University of Pittsburgh DNA synthesis facility on an Applied Biosystems DNA synthesizer. DNA sequencing, purification, analyses and manipulation of nucleic acid and protein preparations were carried out according to standard protocols (Ausubel *et al.*, 1987).

Polymerase expression and purification

Wild type and mutant T7 RNAPs were expressed from pDPT7 by IPTG induction according to the procedure of Davanloo *et al.* (1984). Analysis of *in vivo* polymerase expression and proteolysis susceptibility was carried out by resuspending the cell paste from 1 ml of culture in 100 μl of SDS protein sample buffer (Ausubel *et al.*, 1987) and lysing by boiling for 60 s. Lysates were loaded immediately on 7.5% Pharmacia polyacrylamide phast gels and proteins were visualized by silver staining according to the manufacturer's instructions. Wild type and mutant T7 RNAPs were purified as described by Sousa *et al.* (1989) except that purification was stopped after the first column purification, phosphocellulose chromatography. Selected fractions from the phosphocellulose column were then pooled and stored at −20°C in 20 mM Na_2PO_4, pH 8.0, 0.5 M NaCl, 1 mM EDTA, 1 mM DTT and 50% glycerol for use in transcription assays.

In vitro characterization of mutant enzymes

Determination of mutant enzyme kinetic parameters and mutant enzyme processivity (as evaluated by denaturing gel analysis of radiolabeled transcripts) was carried out as described in Patra *et al.* (1992) using mutant and wild type enzymes purified as described above. Analysis of promoter and non-promoter binding activity was carried out as described by Muller *et al.* (1988) using synthetic, double-stranded 24 bp promoter and non-promoter DNAs of identical sequence to those described by Muller *et al.* (1988).

Acknowledgements

This work was supported by ACS grant NP-737 (to E.M.L.). The use of an AMBIS Radioanalytic imaging system was supported by NIH small instrumentation award GM 43454.

References

Ausubel,M.F., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (1987) *Current Protocols in Molecular Biology*, Volume I. Greene Publishing and Wiley-Interscience, New York.

Carrol,S.S., Cowat,M. and Benkovic,S.J. (1991) *Biochemistry*, **30**, 804−813.

Catalano,C.E., Allen,D.J. and Benkovic,S.J. (1990) *Biochemistry*, **29**, 3612−3621.

Chung,Y.J., Sousa,R., Rose,J.P., Lafer,E.M. and Wang,B.C. (1990a) In Wu,F.Y.-H. and Wu,C.-W. (eds), *Structure and Function of Nucleic Acids and Proteins*, Raven Press, New York, pp. 55−59.

Chung,Y.J., Sousa,R.S., Rose,J.P., Lafer,E.M. and Wang,B.C. (1990b) In Papas,T.S. (ed.), *Gene Regulation and AIDS: Transcriptional Activation, Retroviruses, and Pathogenesis (Advances in Applied Biotechnology* series, Volume 7), Gulf Publishing Co., Houston, TX, pp. 3−7.

Davanloo,P., Rosenberg,A.H., Dunn,J.J. and Studier,F.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2035−2039.

Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) *J. Protein Engng*, **10**, 461−467.

Gunderson,S.I., Chapman,K.A. and Burgess,R.R. (1987) *Biochemistry*, **26**, 1539−1546.

Golomb,M. and Chamberlin,M. (1974) *J. Biol. Chem.*, **249**, 2858−2863.

Ikeda,R.A. and Richardson,C.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3614−3618.

Ikeda,R.A. and Richardson,C.C. (1987) *J. Biol. Chem.*, **262**, 3790−3799.

Joyce,C.M. (1991) *Curr. Opin. Struct. Biol.*, **1**, 123−129.

Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **84**, 4767−4771.

Martin,C.T. and Coleman,J.E. (1987) *Biochemistry*, **26**, 2690−2696.

Martin,C.T., Miller,D.K. and Coleman,J.E. (1988) *Biochemistry*, **27**, 3966−3974.

McClure,W.R. and Chow,Y. (1980) *Methods Enzymol.*, **64**, 277−297.

Mullen,G.P., Shenbagamurti,P. and Mildvan,A.S. (1989) *J. Biol. Chem.*, **264**, 19637−19647.

Muller,D.K., Martin,C.T. and Coleman,J.E. (1988) *Biochemistry*, **27**, 5763−5771.

Ollis,D.L., Brick,P. and Steitz,T.A. (1985) *Nature*, **313**, 765−769.

Patra,D., Lafer,E.M. and Sousa,R. (1992) *J. Mol. Biol.*, **224**, in press.

Polesky,A.H., Steitz,T.A., Grindley,N.D.F. and Joyce,C.M. (1990) *J. Biol. Chem.*, **265**, 14579−14591.

Rush,J. and Konigsberg,W.H. (1990) *J. Biol. Chem.*, **265**, 4821−4827.

Sankar,S. and Porter,A.G. (1992) *J. Biol. Chem.*, **267**, 527−534.

Shi,Y., Gamper,H. and Hearst,J.E. (1988) *J. Biol. Chem.*, **263**, 527−534.

Sousa,R. (1991) PhD Dissertation, University of Pittsburgh.

Sousa,R., Rose,J.P., Chung,Y.J., Lafer.E.M. and Wang,B.C. (1989) *Proteins*, **5**, 226−270.

Sousa,R., Rose,J.P., Chung,Y.J., Lafer,E.M. and Wang,B.C. (1991) *Proceedings of the Annual Meeting of the American Crystallographic Association* (Toledo, OH). ACA Publishers, p. 46.

Sousa,R., Patra,D. and Lafer,E.M. (1992) *J. Mol. Biol.*, **224**, in press.

Tabor,R. and Richardson,C.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4767−4771.

Received on April 1, 1991; revised on June 1, 1992