THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 22, Issue of June 1, pp. 19172–19181, 2001

Biochemical Analysis of Point Mutations in the 5′-3′ Exonuclease of DNA Polymerase I of *Streptococcus pneumoniae*

FUNCTIONAL AND STRUCTURAL IMPLICATIONS*[S]

Received for publication, September 22, 2000, and in revised form, February 19, 2001
Published, JBC Papers in Press, March 7, 2001, DOI 10.1074/jbc.M008678200

Mónica Amblar‡, Mario García de Lacoba, Maria A. Corrales, and Paloma López

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Velázquez 144, 28006 Madrid, Spain

To define the active site of the 5′-3′ exonucleolytic domain of the *Streptococcus pneumoniae* DNA polymerase I (Spn pol I), we have constructed His-tagged Spn pol I fusion protein and introduced mutations at residues Asp^{10}, Glu^{88}, and Glu^{114}, which are conserved among all prokaryotic and eukaryotic 5′ nucleases. The mutations, but not the fusion to the C-terminal end of the wild-type, reduced the exonuclease activity. The residual exonuclease activity of the mutant proteins has been kinetically studied, together with potential alterations in metal binding at the active site. Comparison of the catalytic rate and dissociation constant of the D10G, E114G, and E88K mutants and the control fusion protein support: (i) a critical function of Asp^{10} in the catalytic event, (ii) a role of Glu^{114} in the exonucleolytic reaction, being secondarily involved in both catalysis and DNA binding, and (iii) a nonessential function of Glu^{88} for the exonuclease activity of Spn pol I. Moreover, the pattern of metal activation of the mutant proteins indicates that none of the three residues is a metal-ligand at the active site. These findings and those previously obtained with D190A mutant of Spn pol I are discussed in relation to structural and mutational data for related 5′ nucleases.

The polymerase function of type-I-like DNA polymerases has been studied in considerable detail, with biochemical and mutagenesis analysis proceeding in parallel with the structural studies (1–6). Such detailed study have provided a prototypical molecular model of DNA-dependent DNA polymerization and important insights into the architecture of the primer and nucleotide binding sites. By contrast, despite the mutational and structural analysis of several 5′-3′ exonucleolytic domains of eubacterial polymerases (7–9) and related bacteriophage 5′ nucleases (10–12), the molecular mechanism of the exonucleolytic reaction still remains obscure.

Sequence comparisons and enzymatic studies indicate that the eubacterial pol[1] I-associated 5′ nucleases share significant sequence homology with the polymerase-independent 5′ nucleases from several bacteriophages (13). The prokaryotic 5′ nucleases are also related to mammalian FEN-1 proteins and several yeast proteins of the RAD2 family, having two large blocks of sequence similarity that bear some resemblance to the bacterial and bacteriophage nuclease sequences (14, 15). Some clues to identify important residues in the bacterial 5′-3′ exonuclease family derive from the multiple sequence alignment of 10 bacterial and bacteriophage nucleases (13). Six conserved sequence motifs containing 14 invariant amino acids were identified, 9 of which were carboxylate residues. The presence of highly conserved carboxylate residues led to the proposal that some of these amino acids could be involved in metal binding at the active site of the 5′-3′ exonucleases, as occurs in other enzymes catalyzing phosphoryl transfer reactions (1). This hypothesis has been further supported by structural data from the 5′ nucleases from bacteriophage T5 (10) and *Taq* pol (9), and from T4 RNase H (11). In these three proteins, the 5′ nuclease active site consists of a set of carboxylate residues, which coordinate metal ligands that are essential for the nuclease activity. However, several intriguing differences among the three active sites exist, which pose some important questions.

DNA polymerase I of *Streptococcus pneumoniae* (Spn pol I) is a bifunctional protein having two enzymatic activities: DNA polymerase and 5′-3′ exonuclease (16). These activities are located on different domains of the protein that are arranged in the same order in all pol I-like DNA polymerases (17, 18). Like other DNA polymerases of the family, both enzymatic activities of Spn pol I are involved in DNA-repair processes (19, 20). Unlike that of *Escherichia coli* (21), the exonucleolytic domain has proved to be essential for pneumococcal cell viability (20). Previous studies on the exonuclease activity of Spn pol I showed an essential role for Asp^{10} and Asp^{190} in the exonucleolytic reaction, since the substitution of these carboxylate residues by Ala led to an almost total inactivation of the nuclease domain in Spn polID10A protein (22) and to a drastic reduction of the catalytic efficiency as well as to an altered metal binding in the Spn polID190A mutant protein (23).

In this paper, we describe the overproduction and purification of a His-tagged fusion form of Spn pol I and the introduc-

* This work was performed under the auspices of the Consejo Superior de Investigaciones Científicas and was supported by European Union Grant QLK2-CT-2000-00543 and by the Program of Strategic Groups of the Comunidad de Madrid. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains Figs. S1–S3.

‡ To whom correspondence should be addressed. Instituto de tecnologie Química e Biológica. Universidade Nova de Lisboa, Apart. 127, 2781-901 Deiras, Portugal. Tel.: 351-214469548; Fax: 351-214411277.

[1] The abbreviations used are: pol, polymerase; Spn pol I, *S. pneumoniae* DNA polymerase; Spn polID10A, Spn pol I with the amino acid substitution D10A; Spn polID190A, Spn pol I with the amino acid substitution D190A; FEN, flap endonuclease; Taq pol, *Thermus aquaticus* DNA polymerase I; Pfu pol, *Pyrococcus furiosus* DNA polymerase II; IPTG, thio-β-D-galactoside; Spn pol I-(His), Spn pol I fused to an His tag at the C-terminal end; Eco pol I, *E. coli* DNA polymerase I; Mtb pol I, *Mycobacterium tuberculosis* DNA polymerase I; *Mj*FEN-1, *Methanococcus jannaschii* FEN-1; *h*FEN-1, human FEN-1; bp, base pair(s); PCR, polymerase chain reaction; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.

tion of mutations at three residues of the exonucleolytic domain of the protein: Asp^{10} and Glu^{114} (proposed metal ligands in other nucleases) and Glu^{88} (highly conserved among prokaryotic 5′ nucleases). We also report the use of these enzymes in a kinetic study to explore the roles of the conserved carboxylate residues in the exonucleolytic reaction. Finally, we present a three-dimensional model of the putative 5′-3′ exonucleolytic domain of Spn pol I, built by homology modeling, in order to provide us with a framework to support some structural explanations of the Spn pol I mutant activities.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes, T4 DNA ligase, and T4 polynucleotide kinase were purchased from New England Biolabs. T7 DNA polymerase and inorganic pyrophosphatase were obtained from Amersham Pharmacia Biotech, Taq pol from Roche Molecular Biochemicals, and Pfu pol from Stratagene. Unlabeled oligonucleotide primers were synthesized in a Gene Assembler (Amersham Pharmacia Biotech) at the Centro de Investigaciones Biológicas (Madrid, Spain). The *E. coli* strains used were JM109 (*endA*1 *recA*1 *gyr*A96 *thi hsdR*17 (r_K^- m_K^+) *rel*A1 *supE*44 I^- Δ(*lac-proAB*) F′ (*traD*36 $proA^+B^+$ $lacI^qZ\Delta M15$) (24) for cloning experiments and BL21(DE3) (F^- r_B^- m_B^- *gal omp*T (*int*::P_{lacUV5} T7 *gen*1 *imm*21 *nin*5) (25) for expression of enzymes to be purified.

Construction of Plasmids Expressing Spn pol I Derivatives—The pMA6 plasmid was constructed by cloning a 5.3-kilobase pair *Eco*RI fragment from plasmid pSM29 (26), containing the pneumococcal *polA* gene, into the *Eco*RI site of the phagemid pAlter-1 vector (Promega). By site-directed mutagenesis, two new restriction sites were introduced at both sides of the *polA* gene on pMA6 plasmid: *Eco*RI (between the −10 and −35 boxes) and *Xho*I (at the stop codon of the gene). Then, the mutated *polA* gene was excised from the pMA6 derivative by digestion at the newly created *Eco*RI and *Xho*I restriction sites and ligated to the pET21b (fusion vector derived from the *E. coli* pET5 expression vector; Ref. 27). The resulting plasmid, named pMA9, contains the *polA* gene fused at its 3′ end to an His tag coding sequence under the control of T7 gene ϕ10 promoter (27). As a consequence, the stop codon of the structural gene was replaced by a DNA fragment encoding the amino acid sequence $LE(H)_6$ obtaining the fusion derivative named Spn pol I-(His), which can be selectively bounded to chelating Sepharose resin through the histidine residues. Plasmid pMA10 was derived from pMA9 by removal of the 80-bp *Xba*I-*Eco*RI fragment containing the ribosome binding site of ϕ10 gene and the T7 tag coding region. Expression plasmids for pol I derivatives having mutations at the 5′-3′ exonucleolytic domain of the protein, were obtained by swapping the 399-bp *Eco*RI-*Nhe*I restriction fragment from pMA10 for the corresponding PCR mutagenic products. The sequence of the mutated *polA* genes was determined by thermal cycle sequencing of the resulting recombinant plasmids using Taq pol, the Abitrin 377 automatic sequencer and the corresponding kit supplied by Applied Biosystems Inc., at the Centro de Investigaciones Biológicas.

Mutagenesis of the polA Gene—Introduction of the restriction sites *Eco*RI and *Xho*I at the pneumococcal *polA* gene was carried out using ECR (5′-CAATGGTATTTTTTGAATTCTTTCCTTTATA-3′) and XHO (5′-CTGGTACGAGGCTAAACTCGAGGGGGGCTAGTCCTC-3′) oligonucleotides, and the Altered Site *in vitro* mutagenesis system from Promega.

Mutations of the 5′-3′ exonucleolytic domain coding region of *polA* gene were obtained by using the high yield method for site-directed mutagenesis by PCR and three primers as described by Steingberg *et al.* (28). The mutagenic primer used was 5′-AAATTATTATTGATTNNNGGGTCTTCTGTAGCT-3′, where N corresponds to degenerated position within the 10th codon, at which the four nucleotides can be introduced. ECR and A-PCR (5′-AGGGATATTATCCGACTTATCACCCA-3′) oligonucleotides were used as primers for the amplification reaction. PCR products were digested with *Eco*RI and *Nhe*I and subcloned into the expression plasmid pMA10. The resulting reaction mixture was introduced in *E. coli* JM109 strain by electroporation, and a set of the clones obtained were sequenced to determine the mutation efficiency. The PCR mutagenic procedure was first carried out using Taq pol during the amplification reaction. The analysis of 36 clones showed that 21 of them contained mutations at non-targeted positions of the *polA* gene, such as insertions or deletions (resulting in frameshift mutations), or multiple base substitutions. Only 4 of the 36 clones contained different mutations (GCG, CAG, CGU, and UAG) at the 10th codon (GAU), giving the amino acid substitutions D10A, D10Q, D10R, and D10stop, respectively. One of the non-targeted mutations obtained during PCR reactions caused a change at codon 88 (GAG by AAG), resulting in the E88K amino acid substitution in Spn pol I. Then, a second round of PCR mutagenesis was carried out using Pfu pol (enzyme with high fidelity of polymerization). Of the 37 clones analyzed, 17 were found to carry plasmids with mutations at the 10th codon. Moreover, this analysis revealed five new mutations, in which codon number 10 was replaced by AGU, GUG, ACU, AAA, or GGC. Such changes correspond, respectively, to the amino acid substitutions D10S, D10V, D10T, D10K, and D10G. Only one clone contained, in addition to D10A mutation, the amino acid substitution E114G (change of GAG by GGG at codon 114). Since Glu^{114} is highly conserved in the family of prokaryotic 5′-3′ exonucleases (13), the mutation at this position was subcloned as a single amino acid substitution of Spn pol I into the expression plasmid pMA10, making use of *Nde*I and *Xho*I sites of *polA* gene that are bracketing the 114th codon.

Induction Protocol of Spn pol I and Its Derivatives—Plasmid pMA10 and its derivatives were transferred to the *E. coli* BL21(DE3) overproducer strain. The expression of the wild-type pol I enzyme and its fusion derivatives was achieved by IPTG induction of the host strain BL21(DE3) (27) containing the corresponding expression plasmid (pSM23 for wild-type enzyme (26) and pMA10 set plasmids for fusion proteins).

Cells containing the corresponding plasmid were grown in M9 medium, supplemented with 200 μg/ml ampicillin at 37 °C to an absorbance of 0.45 at 600 nm, and then induced by addition of 0.75 mM IPTG. Samples were withdrawn at different induction times and crude extracts prepared to test enzymatic activities. The crude extracts were also examined by 0.1% SDS, 8% PAGE. Quantification of the proportion of Spn pol I and mutant forms in crude extracts was performed by scanning the gels with the Molecular Analyst system (Bio-Rad).

Preparation of Small Scale Crude Extracts—Cell pastes obtained from 1.5-ml IPTG-induced cultures of the appropriate strain were washed by suspension in 1 ml of buffer I (10 mM Tris-HCl, pH 7.6, 3 mM β-mercaptoethanol), centrifuged, and suspended in 58 μl of lysis buffer II (323 μg/ml lysozyme in buffer I). The resulting suspensions were incubated 5 min at 37 °C, supplemented with 92 μl of buffer I, and incubated for another 5 min at 37 °C. Then, samples were treated with 0.1% Triton X-100 for 5 min at 37 °C and subjected to three cycles of freezing and thawing at −70 °C and 37 °C, respectively. The viscosity of the extracts was reduced by passage through a 0.36-mm inner diameter needle; extracts were then centrifuged and the supernatant stored at −70 °C.

Purification of His Tag Fusion Proteins—Cells from 100-ml IPTG-induced cultures were harvested by centrifugation, washed with 50 ml of buffer A (0.5 M NaCl, 20 mM HPO_4Na_2, pH 7.6), and suspended in 10 ml of the same buffer. Cell lysis was achieved by incubation with 0.3 mg/ml lysozyme for 30 min at 0 °C and three cycles of freezing and thawing. The crude extracts were ultracentrifuged at 81,000 × *g*, 50 min, and the soluble fraction recovered. After addition of 10 mM imidazole, the clarified extracts were added to 4 ml of Chelating Sepharose Fast Flow (Amersham Pharmacia Biotech), previously charged and equilibrated in buffer A plus 10 mM imidazole. The mixture was incubated by end-over-end rotation with gentle agitation for 30 min at 4 °C. The resin was sedimented by centrifugation at 2000 × *g*, 3 min, and washed three times by addition of five volumes of 10 mM imidazole/buffer A and incubation for 5 min at 4 °C as described above. After washing, the resin was again incubated with two volumes of 40 mM imidazole/buffer A, and the protein was eluted by incubation with an equal volume of 100 mM imidazole/buffer A. The imidazole was removed from the sample by dialysis against buffer A plus 50% glycerol, allowing at the same time the concentration and equilibration in the storage buffer. Quantification of the proteins was carried out by gel fractionation of the sample and scanning of the gels with a Molecular Analyst system (Bio-Rad).

Polymerase Assay—Polymerase activity was determined on activated calf thymus DNA following the method described previously (18). One unit of polymerase activity is defined as the amount of enzyme catalyzing the incorporation of 10 nmol of dNTP into DNA in 30 min at 37 °C.

5′-3′ Exonuclease Assays—The nuclease activity in DNA-containing 0.1% SDS, 10% polyacrylamide gels after electrophoresis and removal of SDS was assayed as previously described by Rosenthal and Lacks (29). Exonuclease activity, assayed using salmon sperm DNA, was determined in the presence of 0.1 mM $MnCl_2$ after 30 min at 37 °C, as described previously (16). The salmon sperm substrate, (previously nicked with pancreatic deoxyribonuclease I) was labeled with [^{3}H]dTTP using Spn polIc269 (an Spn pol I derivative that only contains the polymerase domain; Ref. 18). Exonuclease activity was also tested using a 5′-^{32}P-labeled (5′-CCAGTCACGACGTTGT-3′) or 3′-^{32}P-labeled (5′-CCAGTCACGACGTTGTA-3′) oligonucleotide annealed to M13mp2

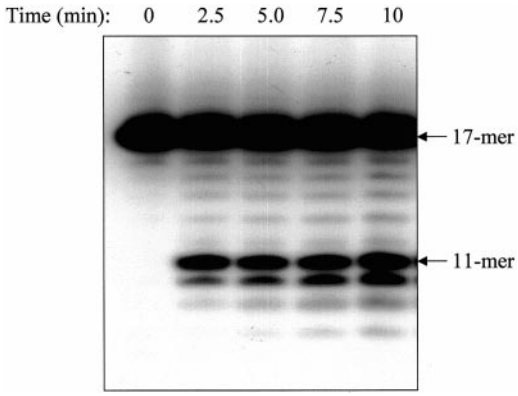

FIG. 1. **Exonucleolytic reaction catalyzed by Spn pol I-(His) on 3′ end-labeled oligonucleotide annealed to ssDNA from M13mp2.** An autoradiogram is shown of the gel fractionation of the reaction products obtained from the exonucleolytic reaction performed with 0.042 unit of exonuclease activity from Spn pol I-(His) and 100 nM DNA substrate. The experiment was carried out as described under "Experimental Procedures" withdrawing samples at the times indicated.

ssDNA as described (23). The concentration of DNA substrate, $MnCl_2$ or $MgCl_2$ and enzymes as well as reaction time are indicated under "Results." In the case of reactions with 5′ end-labeled substrate, the concentration of mutant proteins used was as high as possible without exceeding the inhibitory concentration of 8% glycerol in the reaction (enzymes were stored in 50% glycerol). One unit of exonuclease activity is defined as the amount of enzyme required for the release of 10 nmol of nucleotide from DNA in 30 min at 37 °C.

To measure the 5′-3′ exonuclease rates, experiments were performed using the 3′-^{32}P-labeled 17-mer oligonucleotide annealed to M13mp2 ssDNA as substrate, as described previously (23). The assays were carried out with the preferred divalent metal ion at the optimal concentration for the nuclease activity (0.1 mM $MnCl_2$) and increasing substrate concentrations. The amount of each enzyme was adjusted to obtain linear conditions, and samples (2 μl) were removed at appropriate times during incubation at 37 °C. The products were fractionated in a 20% polyacrylamide gel, where the degradation products are distinguished as discrete bands. A typical experiment is depicted in Fig. 1. The experimental objective was to determine the velocity of exonucleolysis ($\nu_{n,n-1}$) by measuring the proportion of reaction products of n and $n-1$ nucleotides in the gel. Since every primer that reached position $n-1$ also reached position n, the velocity at the site can be expressed by the following equation.

$$\nu_{n,n-1} = \left(\frac{\sum_{i=m}^{n} I_i}{t}\right)\left(\frac{\sum_{i=m}^{n-1} I_i}{I_n}\right) \qquad \text{(Eq. 1)}$$

t is the reaction time, I_i is the integrated intensity at site i expressed as a percentage of total substrate, and m is the length of the final product of the reaction. The velocity of the conversion reaction $\nu_{15,14}$ and $\nu_{14,13}$ was measured in order to minimize the contribution of both initial DNA-protein complex formation and the high dissociation rate at 11-mer product observed for all proteins analyzed (Fig. 1 and results not shown). The velocities remained essentially constant for t up to 30 min for all fusion proteins, indicating that the DNA-enzyme complexes at the point of analysis were in steady state (data not shown). In addition, the concentration of the 15-mer and 14-mer substrates remained lower than 5% of the total DNA, indicating that only one turnover was measured. The relationship of velocity and concentration of primed M13mp2 DNA substrate conformed to the Michaelis-Menten equation, as indicated by linearity in the Lineweaver-Burk plots (30) of [1/v] *versus* [1/primed M13mp2] (Fig. S1). The double-reciprocal plots depicted in Fig. S1 (available as supplemental material in the on-line version of this article) were fitted by a linear least-squares regression analysis and used to determine, from the intercepts, V_{max} (corresponding to the maximum value of I_{n-1}/I_n) and K_m (corresponding to the value of [primed M13mp2] when I_{n-1}/I_n is at half-maximum). The catalytic rate of the exonuclease reaction (k_{cat}) was calculated as $V_{max}/[E_{total}]$, ($[E_{total}]$ being the concentration of enzyme used in the assay).

Filter-binding Assays—Formation of DNA-Spn pol I derivative complexes was measured by using alkali-treated nitrocellulose filters (Millipore, type HAWP 45 μm) as described by McEntee *et al.* (31). The DNA substrate was obtained by PCR amplification of the *Bgl*II-*Xba*I 2239-bp *polA* gene fragment using ECR and A-PCR oligonucleotides and [α-^{32}P]dCTP and Taq pol enzyme, which generate dsDNA with 3′ protruding ends. The resulting 616-bp dsDNA was treated with *Nhe*I restriction enzyme, and the 414-bp product, which contains only one 5′ protruding end, was purified by gel electrophoresis. As the presence of metal ion in the binding buffer was required to detect specific DNA retention on the filters (data not shown), it was necessary to standardize conditions of the binding assay to minimize 5′-3′ exonuclease activity. The standard binding reaction was carried out in 30 μl of buffer B (10 mM Tris-HCl, pH 7.6, 1 mM dithiothreitol, 50 mM KCl, 2.6% glycerol, 0.1 mM $MnCl_2$) with 0.6 nM of the α-^{32}P-labeled 414-bp PCR product and increasing amounts of protein. The mixtures were incubated for 10 min at 15 °C in order to reach DNA-protein equilibrium and the reactions stopped by addition of 150 μl of ice-cold buffer B. Samples were filtered, washed with 9 ml of the same buffer, and dried, and their radioactivity measured by scintillation counting. The quantification of the amount of Spn pol I derivative-DNA complexes retained on the filters was corrected by subtracting the nonspecific retention of labeled DNA in the absence of pol I derivatives and in the presence of an equal bovine serum albumin concentration.

Molecular Modeling—A three-dimensional model of the putative 5′-3′ exonucleolytic domain of Spn pol I was built from its amino acid sequence, the 2.4-Å resolution x-ray structure of the 5′-3′ exonucleolytic domain of Taq pol (9), and 2.5-Å resolution x-ray structure of the T5 5′ nuclease (10) by using knowledge-based protein modeling methods. Their cartesian coordinates were from the Brookhaven Protein Data Bank, with identification codes 1TAQ and 1EXN, respectively. The structural conserved regions and the definition of the corresponding connecting loops were identified from the multiple alignment of the amino acid sequences of Spn pol I, Taq pol, and T5 5′ nuclease (Fig. 6). This alignment was initially identified from the multiple alignment of the nuclease domains of 38 prokaryotic and eukaryotic proteins (Fig. S2, available as supplemental material in the on-line version of this article). The overall conformation of the 5′-3′ exonucleolytic domain of Spn pol I was subjected to energy minimization until convergence, using a combination of steepest descent and conjugate gradients algorithms. The energy calculations were carried out under the AMBER force field (33). Computations were performed on a Power Challenge R10000 by using the BIOSYM software package, release 95.0 (Molecular Simulations, Inc., San Diego, CA).

RESULTS

Production, Expression, and Purification of Spn pol I-(His) and Its Derivatives

Spn pol I-(His)—After a 3.5-h IPTG induction of *E. coli* BL21(DE3) containing pMA10, Spn pol I-(His) was the major protein product in cell extract, corresponding to about 9% of the total protein content (Fig. 2*A*, *lane 4*). This yield was very similar to the 10% obtained with cells harboring pSM23 (26), which encodes Spn pol I (Fig. 2*A*, *lane 3*). The nuclease activity in cell extracts from both BL21(DE3)[pMA10] and BL21(DE3)[pSM23] cultures, was detected *in situ* using a DNA containing polyacrylamide gel (Fig. 2*B*). In both extracts, a degradation band was detected corresponding to a polypeptide of ~100 kDa, the predicted size for Spn pol I and its fusion derivative (Fig. 2*B*, *lanes 2* and *3*). In addition, other bands of activity were observed that presumably corresponded to proteolytic fragments of the pneumococcal enzymes, since they were not detected in extracts carrying the pET5 vector (Fig. 2*B*, *lane 1*). Quantification of the amount of Spn pol I and Spn pol I-(His) (Fig. 2*A*, *lanes 3* and *4*) *versus* their nuclease activity (Fig. 2*B*, *lanes 2* and *3*) revealed that the fusion of the His tag did not affect the 5′-3′ exonuclease specific activity. However, this fusion resulted in a 5-fold reduction of the polymerase activity of the pneumococcal enzyme (producing 60 units of polymerase activity/mg of protein after a 3.5-h induction, compared with 318 units for the vector coding for the wild-type Spn pol I, after correction for the endogenous polymerase activity (3 units) of the host organism). Thus, the Spn pol I-(His) construct

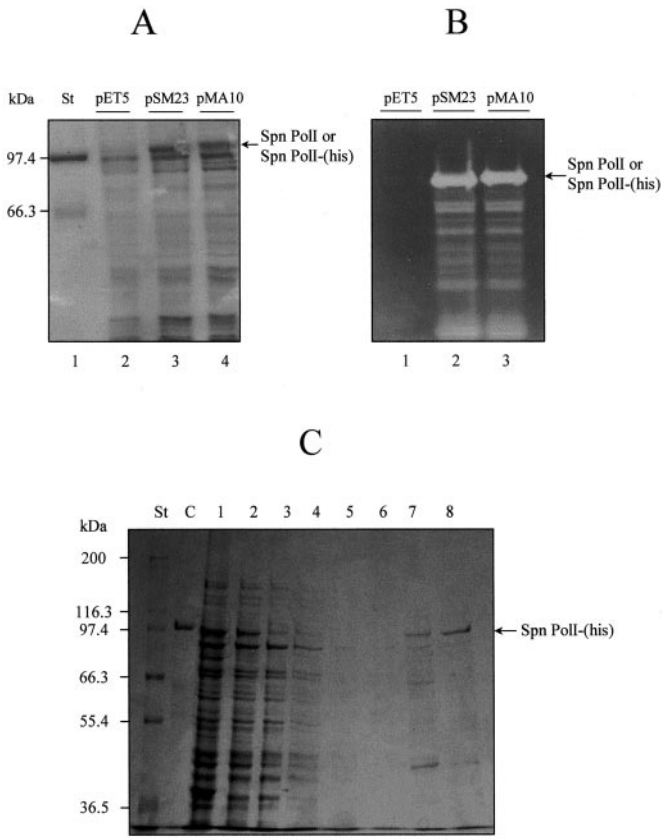

FIG. 2. **Overexpression and purification of Spn pol I-(His).** 15 μg of crude extracts from 3.5-h IPTG-induced cultures of *E. coli* BL21(DE3) carrying the indicated plasmids were fractionated in a 8% polyacrylamide gel. The protein content was analyzed by Coomassie Brilliant Blue stain (*A*), and the nuclease activity of the extract was assayed by DNase gel assay (*B*). The purity of Spn pol I-(His) at each purification step was evaluated by 0.1% SDS, 8% PAGE and staining the gel with Coomassie Brilliant Blue (*C*). *Lane C*, 225 ng of purified Spn pol I; *lane 1*, 1/1400 volume of crude extract; *lane 2*, 1/1400 volume of the sample applied to chelating Sepharose resin; *lane 3*, 1/1400 volume of the sample recovered after application to the resin; *lanes 4–6*, 1/600 volume of each of the three wash steps with 10 mM imidazole; *lane 7*, 1/500 volume of the wash step with 40 mM imidazole; *lane 8*, 1/250 volume of the eluted fraction with 100 mM imidazole. *St*, mixture of polypeptides of known molecular weight.

is appropriate for mutational analysis of the 5′-3′ exonuclease activity of the pneumococcal enzyme.

Spn pol I-(His) was purified from IPTG-induced cultures by binding to a chelating Sepharose resin (Fig. 2*C*). By testing different concentrations of imidazole-eluting agent, we developed a purification procedure yielding 0.37 mg of Spn pol I-(His)/liter of induced culture. The purity of the protein sample was greater than 60% (Fig. 2*C*, *lane 8*), with polymerase and exonuclease specific activities of 614 and 1049 units/mg of protein, respectively.

5′-3′ Exonuclease Mutant Forms of Spn pol I-(His)—The Asp^{10} of Spn pol I is highly conserved among the 5′ nucleases (7), and structural data support the involvement of this residue in the active site (9–11). Substitution of Asp^{10} by Ala in Spn pol I (22) or its equivalent residues in Eco pol I (7) or Mtb pol I (8) by Asn, yielded proteins with so little 5′ nuclease activity that they could not be further analyzed. To obtain a mutant protein that could be kinetically characterized, we mutated the pneumococcal *polA* gene at the 10th codon. Ten different mutant forms of Spn pol I-(His) were obtained, eight of them with different amino acid changes at the Asp^{10} (D10G, D10V, D10T, D10K, D10A, D10Q, D10S, and D10R) and two of them containing either E88K or E114G amino acid substitutions. These carboxylate residues, Glu^{88} and Glu^{114}, are also conserved among the prokaryotic 5′ nucleases (7). All mutant forms were overproduced in *E. coli* BL21(DE3) and purified as for Spn pol I-(His). Analysis of the protein production and polymerase activity after IPTG induction at 30 or 37 °C of two representative mutant derivatives (Spn polID10A-(His) and Spn polID10G-(His)) revealed that incubation for 3 h at 37 °C was optimal induction, without significant insolubilization of the proteins (data not shown). This was surprising, since we have observed previously that 91% of Spn polID10A was insoluble upon induction at 37 °C (22). This procedure allowed us to purify the mutant proteins with a typical yield of 0.4 mg of protein/liter of induced culture at greater than 70% purity and 200–600 units of polymerase activity/mg of protein.

TABLE I
Specific enzymatic activities of His tag fusion proteins

Polymerase activity was assayed on activated calf thymus DNA. The 5′-3′ exonuclease activity was determined from the experiments performed with 5′-^{32}P-end-labeled oligonucleotide annealed to M13mp2 ssDNA. The enzymatic activities are expressed as units/μg of protein (units · μg^{-1}). The figures are the average of at least three independent experiments. Standard deviations are indicated.

Enzyme	Polymerase activity	5′ Nuclease activity	5′ Nuclease/ polymerase
	units · μg^{-1}	*units* · μg^{-1}	%
Spn polI-(his)	0.43 ± 0.06	$(7.2 \pm 0.8) \cdot 10^{-3}$	100
D10Q	0.42 ± 0.03	$<1.0 \cdot 10^{-5}$	<0.14
D10S	0.46 ± 0.04	$<1.4 \cdot 10^{-5}$	<0.08
D10V	0.23 ± 0.02	$<3.7 \cdot 10^{-5}$	<0.96
D10T	0.42 ± 0.03	$<2.9 \cdot 10^{-5}$	<0.41
D10K	0.30 ± 0.07	$<1.1 \cdot 10^{-5}$	<0.22
D10R	0.29 ± 0.03	$<1.7 \cdot 10^{-5}$	<0.35
D10G	0.32 ± 0.07	$(1.5 \pm 0.4) \cdot 10^{-4}$	4.69
D10A	0.25 ± 0.04	$(5.8 \pm 0.9) \cdot 10^{-6}$	0.14
E88K	0.66 ± 0.11	$(1.9 \pm 0.3) \cdot 10^{-3}$	17.2
E114G	0.38 ± 0.03	$(6.1 \pm 1.0) \cdot 10^{-5}$	0.95

Enzymatic Activities of Spn pol I-(His) and Its Derivatives—All the purified proteins were assayed for polymerase activity on activated DNA (Table I). Similar specific activities were obtained for all mutant and control fusion proteins, indicating that the polymerase domain of Spn pol I was functionally unaffected by the single amino acid changes in the 5′-3′ exonucleolytic domain. Next, we determined the exonuclease activity of the enzymes on salmon sperm DNA substrate. Only E88K and the control protein showed detectable activity: 890 and 1000 units/mg of protein, respectively. Due to the low exonuclease activity of the mutant proteins, it was necessary to use a more sensitive assay to determine the effect of the mutations introduced at the exonucleolytic domain. Therefore, a 5′-^{32}P-labeled oligonucleotide annealed to M13mp2 ssDNA was used as substrate to compare the 5′-3′ exonuclease activity of Spn pol I-(His) and all the mutant proteins (Fig. 3). With this substrate we observed the first nucleolytic event in the exonucleolytic reaction. Spn pol I-(His) generated the same products as those previously obtained with Spn pol I (23), derived from both the 5′-3′ exonuclease (mononucleotides) and the 5′ end-dependent endonuclease activities (dinucleotides), which are present in all 5′ nucleases of the family (34). Of the 10 mutant proteins, only those carrying the mutations E114G, E88K, D10G, and D10A gave detectable activities, with that of the Spn polID10A-(His) mutant only being visible after 16 h of incubation, as observed previously with the Spn polID10A mutant (22). In addition, none of these four mutations altered the ratio of mononucleotide to dinucleotide products (~15:1), in contrast to the previously described behavior of Spn polID190A, which showed a prevalence of endonucleolytic over

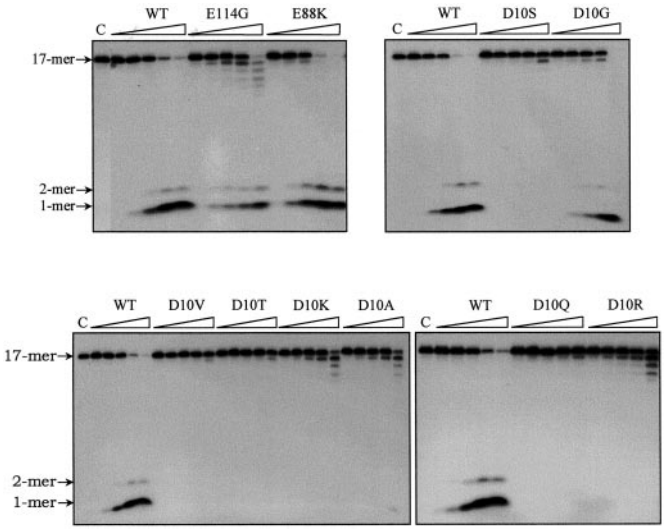

FIG. 3. **Exonuclease activity of His tag fusion proteins on 5′ end-labeled oligonucleotide annealed to M13mp2 ssDNA.** Reaction mixtures contained 1.5 pmol of DNA and 1 mM $MnCl_2$. Samples were removed at 10 min, 30 min, 1 h, 2 h, and 16 h. The concentration of the enzymes in the reaction were: 0.05 pmol of Spn pol I-(His) 2.2 pmol of Spn polIE114G-(His), 0.18 pmol of Spn polIE88K-(His), 0.26 pmol of Spn polID10S-(His), 0.62 pmol of Spn polID10G-(His), 0.096 pmol of Spn polID10V-(His), 0.12 pmol of Spn polID10T-(His), 0.33 pmol of Spn polID10K-(His), 0.91 pmol of Spn polID10A-(His), 0.35 pmol of Spn polID10Q-(His), and 0.21 pmol of Spn polID10R-(His). *C*, control; *WT*, Spn polI-(His); *E114*, polIE114G; *E88*, polIE88K; *D10*, polID10V.

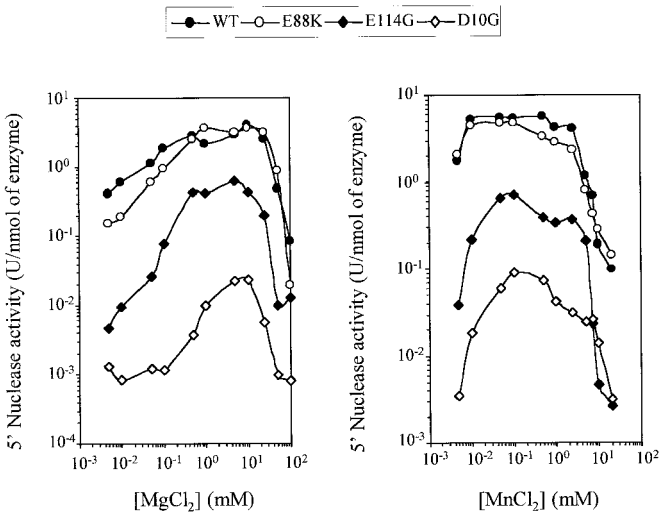

FIG. 4. **Metal dependence of the exonuclease activity of fusion proteins.** Reactions were performed with 3 pmol of 3′-^{32}P-end-labeled oligonucleotide annealed to M13mp2 ssDNA and increasing concentrations of $MgCl_2$ or $MnCl_2$. To determine the exonuclease activity in the presence of Mg^{2+}, the incubation time was 30 min for reactions catalyzed by Spn pol I-(His) and Spn polIE88K-(His), or 2 h for those catalyzed by the Spn polID10G-(His) and Spn polIE114G-(His). Reaction time in the presence of Mn^{2+} was 10 min for Spn pol I-(His) and Spn polIE88K-(His), 30 min for Spn polIE114G-(His), and 2 h for Spn polID10G-(His).

exonucleolytic cleavages (23). In some reactions, products most likely generated by removal of mononucleotides at the 3′ end of the oligonucleotide substrate were also observed. These products were presumably produced by residual contamination of the samples with 3′-5′ exonucleases. Quantification of the experiments, shown in Fig. 3, allowed us to categorize the mutant proteins in terms of their exonuclease activity (Table I). The E88K mutant possessed 26% of the exonuclease specific activity present in Spn pol I-(His), indicating that Glu^{88} is not an essential residue for this enzymatic activity. By contrast, the mutations introduced at Glu^{114} or Asp^{10} resulted in a decrease of more than 98% of the exonuclease activity compared with that of the control enzyme. In the case of the amino acid substitutions of Asp^{10}, 5′-3′ exonuclease activity was only measurable in D10G and D10A mutants. These results argue in favor of an essential role of the Asp^{10} and Glu^{114} residues in the exonuclease activity. Therefore, we proceeded to determine the role played by Asp^{10}, Glu^{88}, and Glu^{114} residues by analyzing the mutant forms Spn polID10G-(His) (the only mutant at Asp^{10} that retains significant exonuclease activity), Spn polIE88K-(His), and Spn polIE114G-(His).

Metal Dependence of the 5′-3′ Exonuclease Activity of His Tag Spn pol I Derivatives—The divalent metal ion requirements for the nuclease activity of the His tag Spn pol I fusion proteins were determined using a 3′-^{32}P-end-labeled M13-primed substrate and measuring the 5′-3′ exonuclease activity at different Mn^{2+} or Mg^{2+} concentrations. The reaction products were fractionated in a denaturing 20% polyacrylamide gel and the degradation bands quantified as described previously (23).

The results obtained revealed that the Mn^{2+} and Mg^{2+} dependence of the exonuclease activity was similar for the four fusion proteins analyzed (Fig. 4), and correlated with that previously obtained for Spn pol I (23). Thus, neither the fusion of the His tag at the C-terminal end of the protein nor the amino acid changes introduced altered the metal dependence of the 5′-3′ exonuclease activity of the *S. pneumoniae* enzyme. The His tag proteins showed maximum exonuclease activity over a wide range of Mn^{2+} concentration, ranging between 10 μM and 1 mM (or even 5 mM) $MnCl_2$, but diminished dramatically at higher concentrations. In the case of Mg^{2+}, the exonuclease activity of the four fusion proteins increased gradually from the lowest concentration tested (5 μM $MgCl_2$), reaching its maximum value at 10 mM, above which the activity drastically decreased. The diminution of exonuclease activity observed at high metal ion concentration for all proteins was probably due to the metal binding to the substrate, rather than to an inhibitory effect on the protein.

Comparative Kinetic Analysis of His Tag-Spn pol I Fusion Proteins—To investigate the role of the Asp^{10}, Glu^{88}, and Glu^{114} residues in the 5′-3′ exonuclease activity of Spn pol I, we determined the kinetic parameters of the D10G, E88K, and E114G mutant enzymes and of Spn pol I-(His) control protein (Table II). The apparent catalytic rate (k_{cat}) and K_m values were calculated from experiments performed using 3′-^{32}P-end-labeled 17-mer oligonucleotide annealed to M13mp2 ssDNA as substrate for the exonucleolytic reaction. The scheme of the reaction is as follows.

$$E + D_n \underset{k_{-0}}{\overset{k_0}{\rightleftharpoons}} E\cdot D_n \xrightarrow{dNMP} E\cdot D_{n-1} \xrightarrow{dNMP} E\cdot D_{n-2} \xrightarrow{dNMP} E\cdot D_{n-3} \xrightarrow{dNMP} E\cdot D_{n-4} \longrightarrow \ldots \longrightarrow E\cdot D_m \longrightarrow E + D_m$$

$$\begin{array}{cccccccccccc} & & & & k_{cat} & & k_{cat} & & k_{cat} & & k_{cat} & & k_{cat} & & k_{dis} \\ & & & & k_{-a} \Uparrow k_a & & k_{-b} \Uparrow k_b & & k_{-c} \Uparrow k_c & & k_{-d} \Uparrow k_d & & & & \\ & & & & E + D_{n-1} & & E + D_{n-2} & & E + D_{n-3} & & E + D_{n-4} & & & & \end{array}$$

SCHEME 1

TABLE II
5′-3′ exonuclease kinetic constants of His tag fusion proteins

Exonuclease rates were measured at different substrate concentrations as indicated under "Experimental Procedures." The K_m and k_{cat} values are the average of those obtained for the two catalytic events analyzed in Fig. S1. The K_D values were determined from filter binding assays as described in the text. ND, not determined.

Enzyme	K_m	k_{cat}	K_D
	nM	s^{-1}	nM
Spn polI-(His)	721 ± 35	5.1 ± 0.6	8.9 ± 3.4
Spn polIE88K-(His)	777 ± 65	0.9 ± 0.3	6.6 ± 2.9
Spn polIE114G-(His)	328 ± 50	$(5.0 \pm 1.0) \cdot 10^{-2}$	113 ± 46
Spn polID10G-(His)	206 ± 8	$(8.7 \pm 1.4) \cdot 10^{-4}$	11 ± 5.2
Spn polID10K-(His)	ND	ND	14 ± 6.9

E represents the enzyme, D_n is the original 17-mer primer bound to DNA, D_{n-1}, D_{n-2}, D_{n-3}, and D_{n-4} are the products of excision of one by one mononucleotide (*dNMP*), and *Dm* is the final product of the reaction. The side pathways (downward pointing arrows) lead to dissociation of complexes $E{\cdot}D_{n-1}$, $E{\cdot}D_{n-2}$, $E{\cdot}D_{n-4}$, with rate constants k_{-a}, k_{-b}, k_{-d}. The reverse association of the enzyme and the reaction products (D_{n-1}, $D_{n-2,}$ and D_{n-4}) can occur via k_a, k_b, k_d (upward pointing arrows).

In order to minimize the contribution of both initial DNA-protein complex formation (k_{-0}/k_0) and the high dissociation rate at 11-mer product (k_{dis}) (see Fig. 1), we measured the velocity of the conversion of 15-mer substrate to 14-mer product ($v_{15,14}$) and of 14-mer substrate to 13-mer product ($v_{14,13}$) (see details under "Experimental Procedures"). In addition, association of the enzyme and the substrates of the reactions analyzed (D_{15} and D_{14}) was minimized, since at the time periods assayed the concentration of D_{15} and D_{14} was very small in relation to original primer (D_{17}) (see Fig. 1). Moreover, correction for further conversion of both substrate and product to shorter products (see details under "Experimental Procedures") allow us to neglect association and dissociation rates downstream of the target reaction. The Lineweaver-Burk plots of [1/*v*] *versus* [1/primed M13mp2] are depicted in Fig. S1 (available on-line as supplemental material to this article), and they were used to calculate the kinetic constants shown in Table II. The apparent K_m and k_{cat} values were similar for the two catalytic events analyzed for each protein except for the E88K mutant, where the rate of conversion of D_{14} to D_{13} (1.25 s^{-1}) was slightly higher than that of D_{15} to D_{14} (0.53 s^{-1}). This indicated that the kinetic constants were not significantly affected by the nature of the nucleotide present at the 5′ end of each substrate (dAMP or dGMP). The apparent K_m of the E88K mutant was similar to that of Spn pol I-(His), and this mutation resulted in only a 5-fold reduction of the apparent k_{cat}. These results indicate that Glu^{88} is not essential for the exonucleolytic reaction. By contrast, the D10G and E114G mutants possessed k_{cat} values ~6000-fold and ~100-fold lower than that of Spn pol I-(His), respectively. This sharp reduction of the apparent k_{cat} values suggested an important role for Asp^{10} and Glu^{114} residues in the exonuclease catalytic event. The apparent K_m value of D10G and E114G were lower than that of the control protein, implying that these mutations resulted in a higher DNA binding affinity. However, it is important to keep in mind that K_m is not an equilibrium dissociation constant (K_D) because it is also affected by k_{cat}. Moreover, in our kinetic analysis of a catalytic even *x*, we can assume that the apparent kinetic constant measured is $K_m = (k_{-x} + k_{cat})/k_x$. Therefore, the fast catalytic rates of Spn pol I-(His) and E88K mutant could mask the actual affinity for the substrate yielding the high apparent K_m values observed. Supporting this hypothesis, the K_m previously obtained for Spn pol I (K_m = 100 nM; Ref. 23) was also lower than that for Spn pol I-(His), probably due to the low k_{cat} of wild-type enzyme (k_{cat} = 0.11 s^{-1}; Ref. 23). The differences in k_{cat} between both wild-type and Spn pol I-(His) enzymes can be explained because of the different purification procedures used. The His tag fusion proteins were obtained using a 1-day purification step, whereas Spn pol I had been previously purified over 1 week using a more complex protocol consisting of three chromatographic steps (23). Therefore, the inactivation of the wild-type enzyme was probably higher than that of the fusion proteins. This was borne out by the fact that the specific exonuclease activity of the enzymes prepared for the kinetic experiments were 14 units/nmol for Spn pol I-(His) and 1 unit/nmol for Spn pol I. In conclusion, the kinetic analyses allowed the determination of the catalytic rate of the enzymes, but the apparent K_m values were not a measure of the DNA binding capability of these His tag fusion proteins. Therefore, we proceeded to determine the dissociation constant (K_D) to establish whether the substrate affinity was affected by the mutations.

DNA Binding Affinity of His Tag Fusion Proteins—The effect of the mutations in Spn pol I-(His) on DNA-exonucleolytic domain dissociation equilibrium constant (K_D) was measured by filter binding assays using homogeneously ^{32}P-labeled dsDNA substrate. The retention of DNA in the filters requires the interaction of the DNA with the protein. Since Spn pol I possesses two enzymatic domains, for which DNA is a substrate, either domain could bind DNA. To minimize interference of DNA interactions with the polymerase domain on binding of the substrate to the exonucleolytic domain of Spn pol I, we used assay conditions (see "Experimental Procedures") in which no DNA binding was detected with Spn polIc269 (data not shown). This protein only contains the polymerase domain of Spn pol I and possesses the same K_m for DNA as the wild-type protein (18).

Complex formation between each protein and DNA was measured as a function of protein concentration, and the saturation curves are depicted in Fig. 5. The data were fitted by direct nonlinear least-squares regression program to the equilibrium binding Equation 2 for independent sites.

$$[C] = ([D][P]^n)/(K + [P]^n) \quad \text{(Eq. 2)}$$

[C], [D] and [P] represent concentrations of formed complex, DNA, and free protein, respectively, *K* is the apparent dissociation rate constant (K_D), and *n* is the Hill coefficient. Transformation of Equation 2 into a linear form yielded Equation 3.

$$\log\frac{[C]}{[D] - [C]} = n \log[P] - \log[K] \quad \text{(Eq. 3)}$$

The Hill plots of the data log([C]/([D]-[C])) *versus* log[P] are depicted in Fig. S3 (see supplemental material, available online). The Hill coefficient *n* for the protein tested ranged from 0.8 to 1.2, indicating a stoichiometry of 1:1 for the protein-DNA complexes (as expected since the DNA substrate contains only one 5′ protruding end). Therefore, the apparent K_D of Spn pol I-(His) as well as of D10G, E88K, E114G, and D10K (which displayed a non-measurable 5′-3′ exonuclease activity) mutant enzymes was calculated from Equation 4 at each protein concentration tested, and the averages are shown in Table II.

$$\log[K] = n \log[P] - \log\frac{[C]}{[D] - [C]} \quad \text{(Eq. 4)}$$

Spn pol I-(His) and the D10G, D10K and E88K mutant enzymes showed a similar K_D value, ranging between 7 and 14 nM. These results show that differences of more than 6000-fold on the catalytic rate of the proteins are not reflected in the

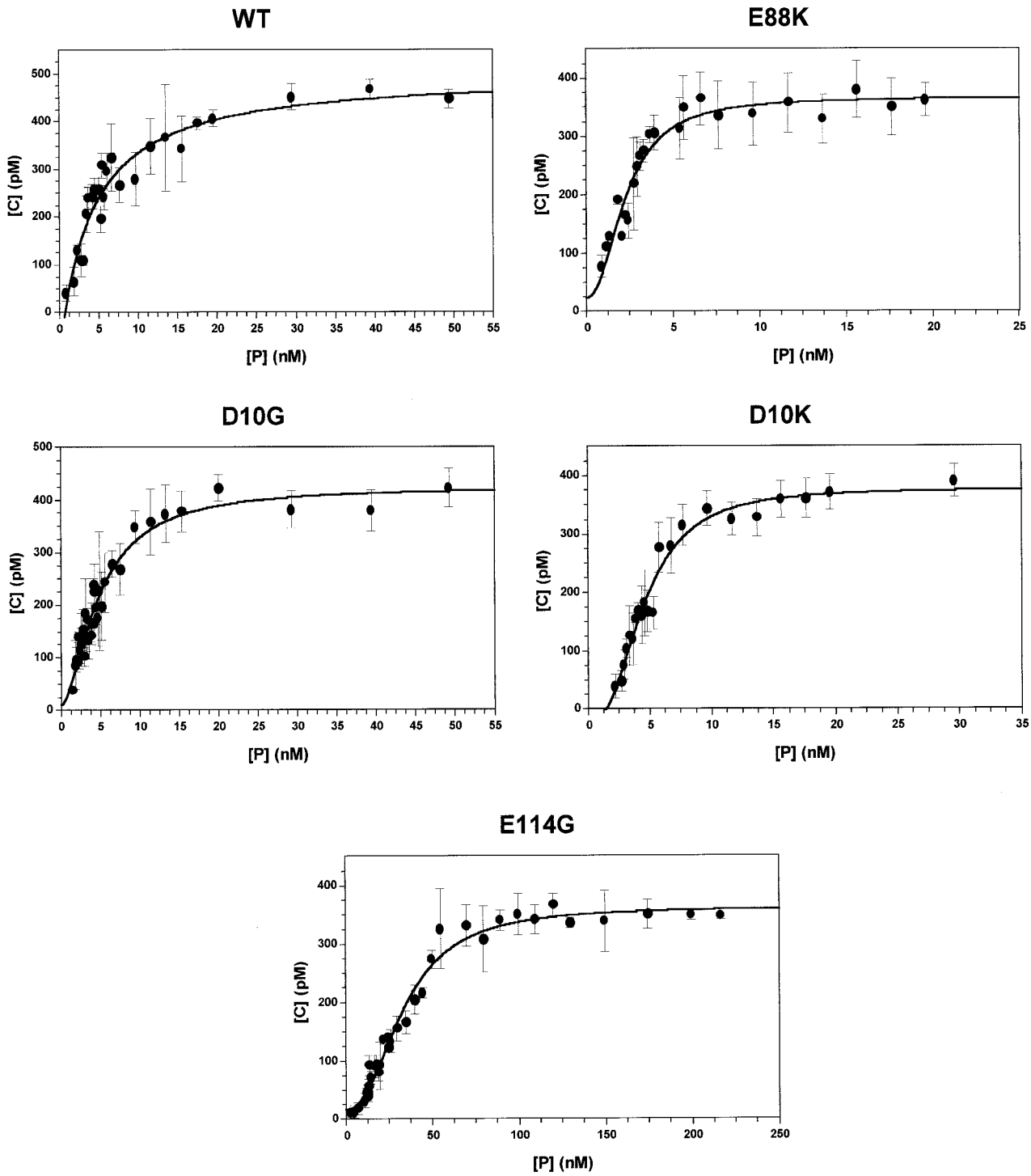

FIG. 5. **Detection of DNA-protein complex formation by filter assay.** The binding reactions with the indicated proteins were performed as described under "Experimental Procedures." Concentrations of retained DNA-protein complexes [C], [P] obtained upon increasing protein concentration are depicted. Experimental data were treated through a non-linear regression analysis program (Curve Expert 1.3, Microsoft Corp.).

binding assay. Moreover, the results indicate that neither Asp^{10} nor Glu^{88} are involved in DNA binding. However, the E114G mutation produced a 13-fold increase of the K_D, suggesting a direct involvement of Glu^{114} in DNA binding of Spn pol I through its exonucleolytic domain.

DISCUSSION

In recent years many structural and mutational data have emerged from studies on different 5′ nucleases, showing that catalysis is supported by divalent metal ions whose ligands are carboxylate residues that are highly conserved in all 5′ nucleases. A sequence alignment from 10 bacterial DNA polymerases and related bacteriophage 5′ nucleases (13), revealed the presence of 10 invariant or highly conserved carboxylate residues, 9 of which appear to be important for the exonucleolytic reaction (reviewed in Ref. 1). The subsequent inclusion in the analysis of eight sequences of polymerase-dependent and independent 5′ nucleases from different bacteria, plus the comparison of the 5′ nuclease domain of Eco pol I and the eukaryotic FEN-1 enzyme, revealed that only six invariant residues are present in all prokaryotic and eukaryotic nucleases (7). Moreover, the multiple alignment of the 5′ nuclease domain of

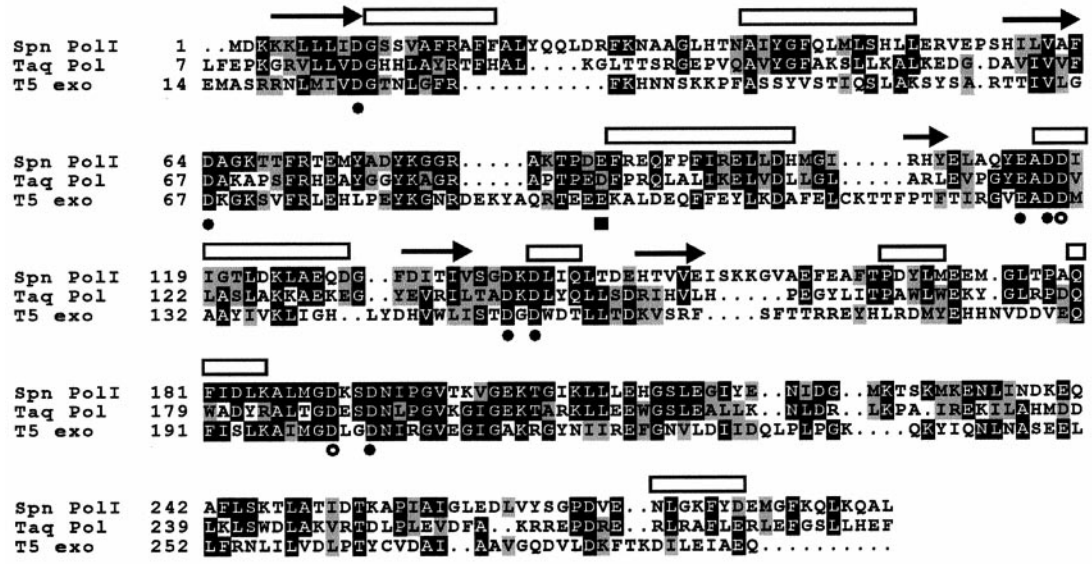

FIG. 6. **Amino acid sequence alignment of the 5′ nuclease from T5 and the 5′-3′ exonucleolytic domains of Spn pol I and Taq pol.** This alignment was derived from the multiple sequence alignment of 38 exonucleolytic domains of the family A of DNA polymerases, and prokaryotic and eukaryotic nucleases (see Fig. S2, available as supplementary material on-line). The *numbers* indicate the amino acid position relative to the N terminus of each sequence. Invariant residues are shown as *white letters* over *black background*, whereas similarity is indicated by *gray background*. Conserved secondary structural motifs (α-helices as hollow rectangles and β-strands as *arrows*) are indicated at the *top* of the aligned sequences. Invariant acidic residues at the active site are indicated at the *bottom* of the aligned sequences. ●, residues present in both prokaryotic and eukaryotic nucleases; ○, residues exclusively present in the prokaryotic enzymes; ■, acidic residue conserved in prokaryotic enzymes.

38 prokaryotic and eukaryotic proteins (Fig. S2 (see on-line supplemental material)) allowed us to confirm and further support the previously observed conservation patterns of acidic residues.

We therefore carried out a mutational analysis of the 5′ nuclease domain of Spn pol I-(His), to gain insights into the structure-function relationships in 5′ nucleases. Three carboxylate residues of the 5′ nuclease domain of Spn pol I were analyzed: Asp^{10}, Glu^{88}, and Glu^{114}. Although Glu^{88} is not an invariant amino acid, in 82% of the sequences aligned, this position is occupied by an acidic residue (Fig. S2). In addition, some of the amino acids surrounding Glu^{88} are invariant and have been shown to be important for the 5′ nuclease activity of Eco pol I (35) and 5′ nuclease from T5 (36, 37). Therefore, it has been proposed that the Glu^{88} region could be involved in DNA binding (7). However, our results revealed that substitution of Glu^{88} by Lys did not drastically alter either the catalytic constants or the DNA binding ability of Spn pol I. Thus, it seems that Glu^{88} specifically does not play an essential role in the exonucleolytic reaction.

The crystal structures of Taq pol, T5 nuclease, and T4 RNase H, and the related eukaryotic flap endonuclease *Mj*FEN-1 have been reported recently (9–11, 38). Although they differ in detail, all four share an active site made up of conserved residues that coordinate, at least, two divalent metal ions. Despite the high identity of the residues that constitute the active site, differences exist in the number of metals bound and the distances between them. The 5′ nuclease from T5, T4 RNase H, and *Mj*FEN-1, contain two divalent metal ions separated by 5–8 Å, whereas in the exonucleolytic domain of Taq pol, three different metal binding sites were detected: site I (crystals soaked in Zn^{2+}), and sites II and III (crystals soaked in Mn^{2+}). Sites I and II are separated by about 5 Å and are each about 10 Å from site III. The 5-Å distance between sites I and II in Taq pol is similar to the 3.9 Å that separates the two metal ions bound to the 3′-5′ exonucleolytic domain of the Klenow fragment (39). Accordingly, a two-divalent metal ion mechanism, analogous to that used by the 3′-5′ proofreading exonuclease of Klenow fragment, was proposed for the 5′-3′ exonuclease of Taq pol (7, 9). Such a mechanism implies that two divalent metal ions (I and II) promote the formation of a hydroxyl ion that attacks the scissile phosphodiester bond (MgA) and stabilize the oxyanion leaving group (MgB) and the pentavalent transition state formed during the reaction (MgA and MgB).

The two divalent metal ions bound at the active site of T4 RNase H and 5′ nuclease from T5 are further apart (7 and 8 Å, respectively). This greater distance makes it difficult to propose a mechanism similar to that of the 3′-5′ exonucleases. It is more likely that the divalent metal ions act independently in the exonucleolytic reaction, one of them (located at Me1 metal binding site) being essential for catalysis and the other (at the Me2 metal binding site) playing a more indirect role, probably in substrate binding (11).

The alignment of the 5′-3′ exonuclease domains of Spn pol I and Taq pol, and the 5′ nuclease from bacteriophage T5 (derived from the multiple alignment of the 38 nucleases depicted in Fig. S2 (see on-line supplemental materials)) revealed a high sequence homology, Spn pol I possessing a 62% and a 51% similarity with Taq pol and T5, respectively (Fig. 6). This homology allowed us to build a three-dimensional model of the exonucleolytic domain of Spn pol I (residues 1–290) based on the crystal structures of the *Taq* pol and the T5 enzyme. The model predicts that the 5′-3′ exonucleolytic domain of Spn pol I should adopt a overall conformation similar to that of those nucleases, consisting of a central β-sheet, surrounded by clusters of helices on either side, and organized in two subdomains. The superimposition of the three-dimensional model on the structure of the 5′-3′ exonucleolytic domain of Taq pol (Fig. 7*A*) and T5 5′ nuclease (Fig. 7*B*) clearly revealed that most of the secondary structural elements are conserved in all three proteins (particularly between Spn pol I and Taq pol) and that they are largely present in the same relative positions. Another interesting feature of the Spn pol I model is the presence of a helical arch similar to that of the T5 nuclease, which has been proposed as being important for the threading mechanism of the structure-specific endonuclease activity of these types of

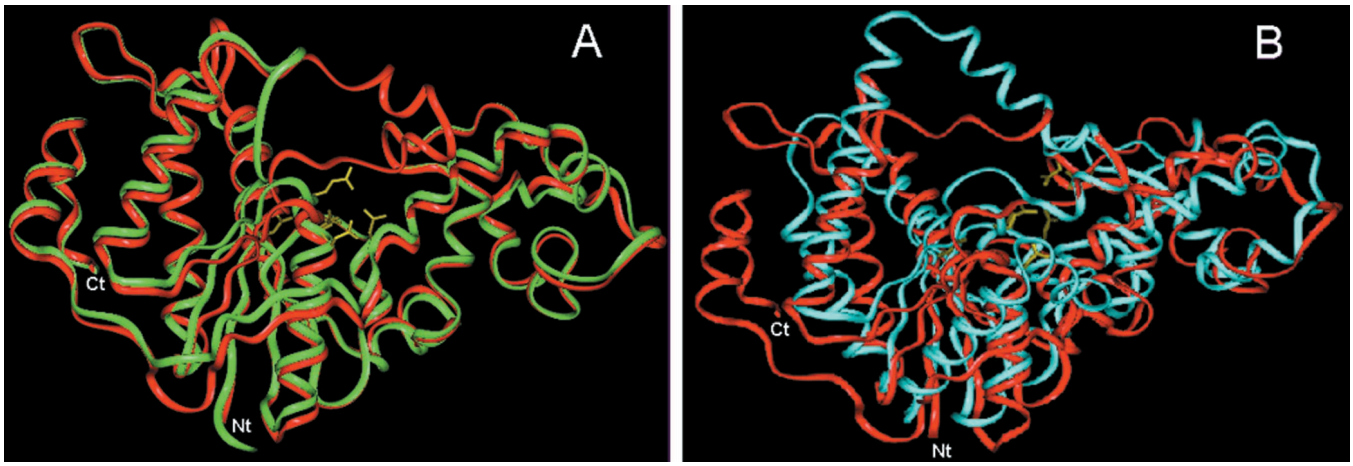

FIG. 7. **Structural modeling of the putative 5′-3′ exonucleolytic domain of Spn pol I.** *Solid ribbon* represents the optimally superimposed polypeptide backbone of the energy-minimized model of the 5′-3′ exonucleolytic domains of Spn pol I (*red*) and *Taq* pol (*green*) (*A*) or 5′ nuclease from T5 (*cyan*) (*B*). The amino acid side chains bound to metal ions at the active site of *Taq* pol (*A*) or T5 5′ nuclease (*B*) are drawn with *solid sticks* (*yellow*). *Nt* and *Ct* indicate the N- and C-terminal ends, respectively.

enzymes (10). In our model, the helical structure is composed of two helices made up of 36 residues, smaller than the three-helix arch of 44 residues present in the bacteriophage nuclease (the corresponding region in Taq pol is crystallographically disordered; Ref. 9).

The predicted active site of the nuclease domain of Spn pol I (Fig. 8) is located in a similar position to those of the crystallized enzymes. The nine acidic residues conserved in the prokaryotic nucleases (Fig. 6; see also Fig. S2 in on-line supplementary material), including the Asp^{10} and Glu^{114}, cluster in a sphere with 11-Å radius that should accommodate the divalent metal ions. Similar radii are observed in the spheres containing the analogous residues in 5′ nuclease from T5 (9 Å) and *Taq* pol (10.5 Å), indicating that, in an analogous manner, two or three metal ions could be bound at the active site of Spn pol I exonuclease.

From our model of Spn pol I, we can also infer that Asp^{190} (residue conserved among all prokaryotic 5′ nucleases and analyzed previously by the mutation D190A; Ref. 23) is located at the active site of the exonucleolytic domain. Analysis of this mutant showed that the replacement of Asp^{190} by Ala produced a sharp reduction of the k_{cat} for the exonuclease activity, which was accompanied by a deficiency in catalytic DNA-protein complex formation as well as a marked preference of the protein for Mn^{2+} over Mg^{2+} and an alteration of the optimal concentration of Mn^{2+} for catalysis (23). These data suggest that Asp^{190} is an essential residue, being involved simultaneously in DNA binding and catalysis, and that its function is mediated by the metal coordination at the active site. The crystal structure of T5 nuclease showed that Asp^{201} (counterpart of Asp^{190}) together with Asp^{153}, Asp^{155}, and Asp^{204} (counterparts of Asp^{139}, Asp^{141}, and Asp^{193} in Spn pol I) are within coordination distance of the metal ion located at the Me2 site, which has been implicated in substrate binding. Superimposition of the active site of the Spn pol I model on the structure of the T5 enzyme (data not shown),revealed a RMS deviation between the Asp^{190}-Spn pol I residue and the bacteriophage nuclease Asp^{201} of only 2.6 Å, supporting the involvement of Asp^{190} in the coordination of the metal ion bound at the putative Me2 site in Spn pol I.

It can be also inferred from our model that Asp^{10} and Glu^{114} could coordinate the metal ions bound at the active site, so that these residues, which are present in all prokaryotic and eukaryotic 5′ nucleases, could be essential for its nuclease activity. The available structural data indicate that these residues are involved in binding a divalent metal ion at the active site

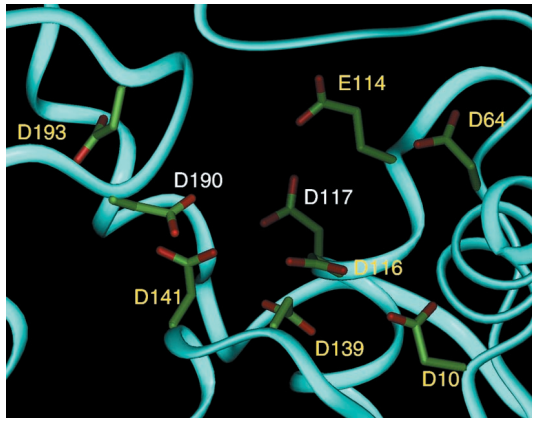

FIG. 8. **Close-up view of the active site of the structural model of Spn pol I 5′-3′ exonucleolytic domain.** The polypeptide backbone of Spn pol I is represented as a *solid ribbon* (*cyan*). The conserved-amino acid side chains are represented as *solid sticks* with the carbon atoms in *green* and the oxygen atoms in *red*. The seven residues present in both prokaryotic and eukaryotic nucleases are labeled in *yellow*, and those exclusively present in the prokaryotic enzymes are labeled in *white*.

(Asp^{10} in Me1 of T5 nuclease and site I of Taq pol, and Glu^{114} in Me1 of T5 and site III of Taq pol). However, our results show that substitution of Asp^{10} or Glu^{114} by Ala drastically reduces exonuclease activity of Spn pol I without changing its Mn^{2+} and Mg^{2+} dependence. Thus, they indicate that these residues do not coordinate metal ions. Nevertheless, they do not rule out that, upon removal of one of these putative ligands, metal ions can remain bound to the active site of the exonucleolytic domain of Spn pol I through interactions with other acidic residues, although an alteration of affinity for the metals should be expected, as was the case for the D190A mutant of Spn pol I (23). Moreover, in our structural model of Spn pol I, Asp^{64}, Asp^{116}, and Asp^{117} (metal ligands in other 5′ nucleases) cluster within a sphere of ~7.5-Å radius (Fig. 8) and the conformational freedom of their side chains could allow an arrangement to bind a divalent metal ion at the putative Me1 site. Therefore, at least, we can infer that Asp^{10} and Glu^{114} do not have a crucial role in metal coordination at the nuclease active site of Spn pol I.

The essential role of Asp^{10} in exonuclease activity is unquestionable, since mutations at this position in all nucleases studied led to an almost total inactivation of the protein, as reported for D13N mutation in Eco pol I (7), D21N in Mtb pol I (8), D34A

in *h*FEN-1 (40), and D19N in T4 RNase H (12). We have demonstrated that, in Spn pol I, the Asp^{10} is involved exclusively in the catalytic event of the exonucleolytic reaction, because the substitution D10G produced a ~6000-fold reduction of the catalytic rate without affecting the DNA binding capability. A similar role has been proposed for the analogous Asp^{34} residue in *h*FEN-1, since its D34A mutant is inactive in cleavage but not in DNA binding (40). Our experimental results suggest that the Asp^{10} in Spn pol I is not critical for metal coordination. Therefore, this residue could be essential by acting like a general base during the reaction, activating a water molecule and generating the hydroxyl group required for nucleophilic attack. For this purpose, Asp^{10} must be in close proximity to the metal ion bound to the putative Me1 site, as shown by our nuclease domain model of the pneumococal enzyme. This metal should stabilize the pentavalent transition state formed during the reaction and therefore would be essential for catalysis.

The role played by Glu^{114} in exonucleolytic reaction is less clear. This amino acid of Spn pol I, as well as the equivalent residue of *h*FEN-1 (40), seems to be involved in substrate binding, and has been considered less essential than other residues analyzed for the exonuclease activity of Eco pol I (7) and Mtb pol I (8). The structural data are highly contradictory, since the counterpart of Glu^{114} seems to be involved in binding of the catalytic metal ion (Me1) in T5 nuclease (10) and *Mj*FEN-1 (38), does not interact with metal ions at the T4 RNase H active site (11), and is part of the metal binding site III in Taq pol (for which a nonessential role has been proposed (9)). Thus, it is likely that Glu^{114} of Spn pol I will be at the exonuclease active site, as shown by the Spn pol I exonuclease conformational model, contributing to the extensive hydrogen-bonding/electrostatic network surrounding both metal ions, rather than interacting in a direct or indirect manner with any one of them. If this hypothesis is correct, substitution of Glu^{114} by Gly should not produce a loss of a coordination with one or two metal ions (as indicated by our results) but a reorientation of the metal(s) at the active site, making the catalytic center less accessible to the DNA substrate and reducing the catalytic efficiency of the reaction.

Summarizing the above results and the available data from other 5′ nucleases, it seems that the involvement of two metal ions playing different roles is the more plausible mechanism for exonucleolytic reaction in 5′ nucleases. A water molecule (probably activated by Asp^{10} or an equivalent residue) would be responsible for the nucleophilic attack of the phosphodiester bond. The divalent metal ion bound to the Me1 site would be essential for the reaction, stabilizing the generated transition state, whereas the metal ion bound to the Me2 site positions the DNA correctly at the active site for cleavage (probably by interacting with the 5′ end). Thus, the role played by the Me2 site would be also important for the reaction, since an incorrect orientation of the DNA at the active site would lead to a drastic reduction of the exonuclease activity. Therefore, its function may not be as indirect as proposed previously (8, 12). This mechanistic model does not necessary exclude the possibility of a reaction similar to that of the 3′-5′ exonuclease of the Klenow fragment proposed for Taq pol. The 5′ nucleases are able to recognize a wide variety of DNA substrates with different structures and to catalyze endo- or exonucleolysis; thus, they must possess a very flexible active site capable of adopting different conformations depending on the substrate and the reaction taking place. It is possible that the differences of active site architecture seen in distinct 5′ nucleases reflect the different functional conformations that these active sites can adopt. For example, in the case of Taq pol, one metal binding site was detected in presence of Zn^{2+} whereas two different sites appeared in presence of Mn^{2+}. This could be taken as an indication of the flexibility of its active site, being capable of adopting *in vivo* two different conformations depending of the requirements of the enzyme. By contrast, it is also possible that the differences between distinct nucleases are a consequence of the individual substrate specificities of the enzymes.

Acknowledgments—We thank Dr. M. Espinosa, Dr. L. Blanco, and Dr. S. W. Elson for helpful discussions and critical reading of the manuscript. We are grateful to Drs. T. A. Steitz and S. H. Eom for providing the refinement of the tertiary structure of Taq pol and to Dr. A. Ceska for the initial model of the active site of Spn pol I, based in the three-dimensional structure of T5 nuclease, and for helpful discussion.

REFERENCES

1. Joyce, C. M., and Steitz, T. A. (1994) *Annu. Rev. Biochem.* **63,** 777–822
2. Doublie, S., and Ellenberg, T. (1998) *Curr. Opin. Struct. Biol.* **8,** 704–712
3. Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17,** 7514–7525
4. Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature* **391,** 304–307
5. Doublie, S., Sawaya, M. R., and Ellenberger, T. (1999) *Struct. Fold Des.* **7,** R31–R35
6. Steitz, T. A. (1999) *J. Biol. Chem.* **274,** 17395–17398
7. Xu, Y., Derbyshire, V., Ng, K., Sun, X. C., Grindley, N. D. F., and Joyce, C. M. (1997) *J. Mol. Biol.* **268,** 284–302
8. Mizrahi, V., and Huberts, P. (1996) *Nuc. Acids Res.* **24,** 4845–4852
9. Kim, Y., Eom, S. H., Wang, J., Lee, D., Suh, S. W., and Steitz, T. A. (1995) *Nature* **376,** 612–616
10. Ceska, T. A., Sayers, J. R., Stier, G., and Suck, D. (1996) *Nature* **382,** 90–93
11. Mueser, T. C., Nossal, N. G., and Hyde, C. C. (1996) *Cell* **85,** 1101–1112
12. Bhagwat, M., Meara, D., and Nossal, N. G. (1997) *J. Biol. Chem.* **272,** 28531–28538
13. Gutman, P. D., and Minton, K. W. (1993) *Nucleic Acids Res.* **21,** 4406–4407
14. Harrington, J. J., and Lieber, M. R. (1994) *Genes Dev.* **8,** 1344–1355
15. Robins, P., Pappin, D. J., Wood, R. D., and Lindahl, T. (1994) *J. Biol. Chem.* **269,** 28535–28538
16. López, P., Martínez, S., Díaz, A., Espinosa, M., and Lacks, S. A. (1989) *J. Biol. Chem.* **264,** 4255–4263
17. Díaz, A., Pons, M. E., Lacks, S. A., and López, P. (1992) *J. Bacteriol.* **174,** 2014–2024
18. Pons, M. E., Díaz, A., Lacks, S. A., and López, P. (1991) *Eur. J. Biochem.* **201,** 147–155
19. López, P., Martínez, S., Díaz, A., and Espinosa, M. (1987) *J. Bacteriol.* **169,** 4869–4871
20. Díaz, A., Lacks, S. A., and López, P. (1992) *Mol. Microbiol.* **6,** 3009–3019
21. Joyce, C. M., and Grindley, N. D. F. (1984) *J. Bacteriol.* **158,** 636–643
22. Amblar, M., Sagner, G., and López, P. (1998) *J. Biotechnol.* **63,** 17–27
23. Amblar, M., and López, P. (1998) *Eur. J. Biochem.* **252,** 124–132
24. Yanish-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* (*Amst.*) **33,** 103–119
25. Studier, F. W., and Moffatt, B. A. (1986) *J. Mol. Biol.* **189,** 113–130
26. Martínez, S., López, P., Espinosa, M., and Lacks, S. A. (1986) *Gene* (*Amst.*) **44,** 79–88
27. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89
28. Steinberg, R. A., and Gorman, K. B. (1994) *Anal. Biochem.* **219,** 155–157
29. Rosenthal, A. L., and Lacks, S. A. (1977) *Anal. Biochem.* **80,** 76–90
30. Cornish-Bowden, A. (1995) *Fundamentals of Enzyme Kinetics*, 2nd Ed., pp. 30–32, Portland Press, London
31. McEntee, K., Weinstock, G., and Lehman, I. R. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 857–861
32. Deleted in Proof
33. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1996*) J. Comp. Chem.* **7,** 230–246
34. Ceska, T. A., and Sayers, J. R. (1998) *Trends Biochem. Sci.* **23,** 331–336
35. Joyce, C. M., Fujii, D. M., Laks, H. S., Hughes, C. M., and Grindley, N. D. F. (1985) *J. Mol. Biol.* **186,** 283–293
36. Pickering, T. J., Garforth, S. J., Thorpe, S. J., Sayers, J. R., and Grasby, J. A. (1999) *Nucleic Acids Res.* **27,** 730–735
37. Garforth, S. J., Ceska, T. A., Suck, D., and Sayers, J. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 38–43
38. Hwang, K. Y., Baek, K., Kim, H.-Y., and Cho, Y. (1998) *Nat. Struct. Biol.* **5,** 707–713
39. Beese, L. S., and Steitz, T. A. (1991) *EMBO J.* **10,** 25–33
40. Shen, B., Nolan, J. P., Sklar, L. A., and Park, M. S. (1997) *Nucleic Acids Res.* **25,** 3332–3338