Footprinting of *Chlorella* Virus DNA Ligase Bound at a Nick in Duplex DNA*

(Received for publication, January 8, 1999, and in revised form, March 3, 1999)

Mark Odell and Stewart Shuman‡

*From the Molecular Biology Program, Sloan-Kettering Institute, New York, New York 10021*

The 298-amino acid ATP-dependent DNA ligase of *Chlorella* virus PBCV-1 is the smallest eukaryotic DNA ligase known. The enzyme has intrinsic specificity for binding to nicked duplex DNA. To delineate the ligase-DNA interface, we have footprinted the enzyme binding site on DNA and the DNA binding site on ligase. The size of the exonuclease III footprint of ligase bound a single nick in duplex DNA is 19–21 nucleotides. The footprint is asymmetric, extending 8–9 nucleotides on the 3′-OH side of the nick and 11–12 nucleotides on the 5′-phosphate side. The 5′-phosphate moiety is essential for the binding of *Chlorella* virus ligase to nicked DNA. Here we show that the 3′-OH moiety is not required for nick recognition. The *Chlorella* virus ligase binds to a nicked ligand containing 2′,3′-dideoxy and 5′-phosphate termini, but cannot catalyze adenylation of the 5′-end. Hence, the 3′-OH is important for step 2 chemistry even though it is not itself chemically transformed during DNA-adenylate formation. A 2′-OH cannot substitute for the essential 3′-OH in adenylation at a nick or even in strand closure at a preadenylated nick. The protein side of the ligase-DNA interface was probed by limited proteolysis of ligase with trypsin and chymotrypsin in the presence and absence of nicked DNA. Protease accessible sites are clustered within a short segment from amino acids 210–225 located distal to conserved motif V. The ligase is protected from proteolysis by nicked DNA. Protease cleavage of the native enzyme prior to DNA addition results in loss of DNA binding. These results suggest a bipartite domain structure in which the interdomain segment either comprises part of the DNA binding site or undergoes a conformational change upon DNA binding. The domain structure of *Chlorella* virus ligase inferred from the solution experiments is consistent with the structure of T7 DNA ligase determined by x-ray crystallography.

ATP-dependent DNA ligases catalyze the joining of a 5′-phosphate-terminated strand to a 3′-hydroxyl-terminated strand via three sequential nucleotidyl transfer reactions (1, 2). In the first step, nucleophilic attack on the α-phosphorus of ATP by ligase results in release of pyrophosphate and formation of a covalent intermediate (ligase-adenylate) in which AMP is linked via a phosphoamide (P–N) bond to the e-amino group of a lysine. In the second step, the AMP is transferred to the 5′-end of the 5′-phosphate-terminated DNA strand to form a DNA-adenylate intermediate, A(5′)pp(5′)N. In this reaction, a 5′-phosphate oxygen of the DNA strand attacks the phosphorus of ligase-adenylate; the active site lysine side chain is the leaving group. In the third step, ligase catalyzes attack by the 3′-OH of the nick on the DNA-adenylate to join the two polynucleotides and release AMP.

ATP-dependent ligases are ubiquitous in eukaryotes. They are also encoded by certain eubacteria, bacteriophages, and eukaryotic DNA viruses. Sequence comparisons suggest that a catalytic domain common to all ATP-dependent ligases is embellished by additional isozyme-specific protein segments at the N or C termini (3). The catalytic domain includes a set of six collinear motifs (I, III, IIIa, IV, V, and VI) that define a superfamily of covalent nucleotidyl transferases encompassing the ATP-dependent polynucleotide ligases and GTP-dependent mRNA capping enzymes (4, 5). The crystal structure of T7 DNA ligase with bound ATP shows that the nucleotide binding pocket is composed of motifs I, III, IIIa, IV, and V (6). The lysine in motif I (KXDXGXR) is the site of covalent attachment of AMP to the enzyme (7).

Crystallography and mutagenesis have illuminated enzymic functional groups that are involved in nucleotide binding and nucleotidyl transfer (6, 8–10). What remains obscure is the structural basis for DNA nick recognition by DNA ligases. Conversion of nicks into phosphodiester bonds is the common final step in the DNA repair and replication pathways. Nicks are potentially deleterious DNA lesions that, if not corrected, may give rise to lethal double-strand breaks. One imagines a mechanism to ensure that ligases are directed to sites where their action is required, either via the interaction of ligase with other replication or repair proteins assembled at a nick (3) or via an intrinsic capacity to discriminate nicks from other DNA structures.

We are examining the interaction of eukaryotic ligases with defined DNAs using virus-encoded enzymes as models. Vaccinia virus DNA ligase and *Chlorella* virus DNA ligase each forms a discrete complex with a singly nicked DNA ligand in the absence of magnesium that can be resolved from free DNA by native polyacrylamide gel electrophoresis (11, 12). The viral ligases do not form stable complexes with the following ligands: (i) DNA containing a 1- or 2-nucleotide gap, (ii) the sealed duplex DNA product of the ligation reaction, (iii) a singly nicked duplex containing a 5′-OH terminus at the nick instead of a 5′-phosphate, or (iv) a singly nicked duplex containing an RNA strand on the 5′-phosphate side of the nick (10–15). These DNA ligases apparently do have an intrinsic nick sensing function. Discrimination at the substrate binding step may account for the feeble activity of the eukaryotic viral DNA ligases in sealing across gaps and in joining nicked molecules containing a 5′-phosphate-terminated RNA strand (14, 15). Nick recognition by vaccinia virus DNA ligase and *Chlorella* virus DNA ligase also depends on occupancy of the AMP binding pocket on the enzyme, i.e. mutations of the ligase active site that abolish

*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.

‡ To whom correspondence should be addressed: Molecular Biology Program, Sloan-Kettering Institute, 1275 York Ave., New York, NY 10021. Tel.: 212-639-7145; Fax: 212-717-3623; E-mail: s-shuman@ski.mskcc.org.
the capacity to form the ligase-adenylate intermediate also eliminate nick recognition, whereas a mutation that preserves ligase-adenylate formation, but inactivates downstream steps of the strand joining reaction, has no effect on binding to nicked DNA (10, 13).

To physically map the ligase-DNA interface, we sought to obtain footprints of the ligase binding site on DNA and the DNA binding site on ligase. We conducted this analysis using the 298-amino acid Chlorella virus PBCV-1 DNA ligase, which is the smallest eukaryotic DNA ligase known and likely constitutes the minimal catalytic unit (12). A confounding technical factor in nuclease-footprinting the DNA side of the ligase-DNA interface was the requirement for a divalent cation cofactor at the nuclease digestion step, the problem being that the ligase-adenylate, which binds to nicked DNA in the absence of divalent cations, immediately catalyzes strand joining upon addition of a divalent cation and then dissociates from the closed duplex ligation product. The enabling factor in obtaining a nuclease footprint of the ligase-DNA complex was the finding that an alanine substitution for Asp-29 in motif I of Chlorella virus DNA ligase (298-KXTGXR298) blocks step 2 of the ligation pathway without affecting enzyme-adenylate formation and nick recognition (10). This “step-arrest” mutant ligase remains bound at the nick in the presence of magnesium; consequently, the D29A-DNA complexes could be footprinted with exonuclease III. We report that the ligase footprint extends from 8–9 nucleotides on the 3'-OH side of the nick to 11–12 nucleotides on the 5'-phosphate side of the nick. To examine the protein component of the ligase-DNA interface, we used the classical approach of limited proteolysis to probe the structure of Chlorella virus DNA ligase in the free and DNA-bound states. The Chlorella virus ligase consists of two protease-resistant domains connected by a protease-sensitive interdomain segment located between motifs V and VI. Protection of the interdomain segment from proteolysis upon binding to nicked DNA implies this region as a component of the DNA binding surface of the enzyme.

**EXPERIMENTAL PROCEDURES**

**Enzyme Purification**—Recombinant wild type Chlorella virus DNA ligase (12) was a gift of Dr. C. K. Ho. Mutant enzyme D29A (10) was expressed in bacteria without an affinity tag and purified from soluble cell lysates by S-Sepharose and blue-Sepharose column chromatography. The chromatographic behavior of D29A was monitored by SDS-PAGE analysis of the column fractions. The lysate was applied to S-Sepharose that had been equilibrated with 0.2 M NaCl in Buffer A (50 mM Tris-HCl (pH 7.5), 2 mM EDTA, 5 mM DTT, 10% glycerol). The column was step-eluted with Buffer A containing 0.4, 0.6, 0.8, and 1 M NaCl concentrations (not shown). Analysis by denaturing gel electrophoresis confirmed that the 32P-labeled 18-mer strand was bound at the nick when magnesium was included in the reaction mixtures (not shown). Analysis by denaturing gel electrophoresis confirmed that the 32P-labeled 18-mer strand was bound at the nick when magnesium was included in the reaction mixtures (not shown).

**Preparation of Labeled DNA Ligands**—DNA ligands used in DNA binding and footprinting experiments were prepared by annealing a 5' 32P-labeled oligonucleotide to unlabeled complementary strands to form the nicked and gapped structures depicted in the figures. End-labelling reactions and gel purification of the 32P-labeled strands were performed as described (9). Labeled oligonucleotides were annealed to an 4-fold molar excess of unlabeled strands in a solution of 0.2 mM NaCl by heating the mixture for 5 min at 65 °C and then cooling slowly to 22 °C.

**Preparation of 2',3'-Dideoxy and 2'-OH DNA Strands**—A 2',3'-dideoxy-terminated 19-mer strand was prepared in two stages. First, an 18-mer strand (dATCCGATTTCGCTCCCT) was 5' 32P-labeled and gel-purified. Then, the labeled 18-mer was reacted for 30 min at 37 °C with 1 mM dATP (cordycepin triphosphate) and calf thymus terminal transferase. The 19-mer 2'-OH product was separated from residual 18-mer substrate by preparative electrophoresis through a 20% polyacrylamide gel and recovered as described above.

**RESULTS**

**Exonuclease III Footprint of DNA Ligase Bound at a Nick**—To footprint the ligase-DNA interface on the 5'-phosphate side of a DNA nick, we incubated the purified D29A enzyme with a singly nicked ligand consisting of a 42-nucleotide 3'-OH-terminated hairpin DNA and a 5' 32P-labeled 18-mer strand annealed to the 5'-tail of the hairpin strand (Fig. 1A). The binding reaction mixtures contained 50 nM nicked DNA and up to 0.8 μM ligase. At these concentrations of DNA and protein, it is the adenylated form of Chlorella virus ligase that binds to the nick to form a discrete complex (10). Analysis of the binding reaction mixtures by native gel electrophoresis showed that nearly all of the 32P-labeled nicked DNA was bound at 0.2 μM D29A, which suggests that about one-fourth of the enzyme molecules in the preparation were adenylated (Fig. 1B). The D29A-adenylate is defective in the transfer of AMP to the 5'-phosphate of the nick (10). Hence, the enzyme remained bound at the nick when magnesium was included in the reaction mixtures (not shown). Analysis by denaturing gel electrophoresis confirmed that the 32P-labeled 18-mer strand was unreactive with a molar excess of D29A in the presence of magnesium during a 10-min incubation at 22 °C (Fig. 1C, lane + and data not shown).

The ligase-DNA mixtures and a control sample containing DNA alone were treated with exonuclease III. The 5' labeled digestion products were resolved by denaturing gel electrophoresis and detected by autoradiography. The amount of exonuclease III included in the reactions was sufficient to convert all of the 32P-labeled 18-mer into a rapidly migrating species of less than 5 nucleotides (Fig. 1C). Ligase binding to the DNA resulted in the appearance of a nuclease resistant doublet at 38–39 nucleotides (denoted by the asterisk in chromosome XDG). The labeled 18-mer was incubated with a molar excess of D29A ligase and calf thymus terminal transferase (57 units; Promega) in primer extension buffer containing 0.1 mM potassium cacodylate (pH 7.2), 10 mM CoCl2, and 1 mM DTT. The labeled strands were labeled by autoradiography, and the 19-mer was eluted from an excised gel slice and recovered by ethanol precipitation.

A 3' -deoxy, 2'-hydroxyl-terminated 19-mer strand was prepared in similar fashion. First, an 18-mer strand (dATCCGATTTCGCTCCCT) was 5' 32P-labeled and gel-purified. The labeled 18-mer was annealed to form the structure shown in Fig. 2. This DNA was reacted for 30 min at 37 °C with 1 mM 3' dATP (cordycepin triphosphate) and calf thymus terminal transferase. The 19-mer 2'-OH product was separated from residual 18-mer substrate by preparative electrophoresis through a 20% polyacrylamide gel and recovered as described above.

Footprint of DNA Ligase Bound at a Nick—To footprint the ligase-DNA interface on the 5'-phosphate side of a DNA nick, we incubated the purified D29A enzyme with a singly nicked ligand consisting of a 42-nucleotide 3'-OH-terminated hairpin DNA and a 5' 32P-labeled 18-mer strand annealed to the 5'-tail of the hairpin strand (Fig. 1A). The binding reaction mixtures contained 50 nM nicked DNA and up to 0.8 μM ligase. At these concentrations of DNA and protein, it is the adenylated form of Chlorella virus ligase that binds to the nick to form a discrete complex (10). Analysis of the binding reaction mixtures by native gel electrophoresis showed that nearly all of the 32P-labeled nicked DNA was bound at 0.2 μM D29A, which suggests that about one-fourth of the enzyme molecules in the preparation were adenylated (Fig. 1B). The D29A-adenylate is defective in the transfer of AMP to the 5'-phosphate of the nick (10). Hence, the enzyme remained bound at the nick when magnesium was included in the reaction mixtures (not shown). Analysis by denaturing gel electrophoresis confirmed that the 32P-labeled 18-mer strand was unreactive with a molar excess of D29A in the presence of magnesium during a 10-min incubation at 22 °C (Fig. 1C, lane + and data not shown).

The ligase-DNA mixtures and a control sample containing DNA alone were treated with exonuclease III. The 5' labeled digestion products were resolved by denaturing gel electrophoresis and detected by autoradiography. The amount of exonuclease III included in the reactions was sufficient to convert all of the 32P-labeled 18-mer into a rapidly migrating species of less than 5 nucleotides (Fig. 1C). Ligase binding to the DNA resulted in the appearance of a nuclease resistant doublet at the expense of the complete digestion product. The upper band comigrated with a 5' 32P-labeled 12-mer pATTCCGATAGTG identical in sequence to the first 12 nucleotides of the input 18-mer strand (lane M). The ligase concentration dependence of exonuclease III protection was similar to the ligase dependence of the protein-DNA complex formation (Fig. 1 and data not shown). We conclude that the ligase footprint on the 5'-phosphate strand extends 11–12 nucleotides downstream of the nick.

The ligase footprint on the 3'-OH side of the nick was determined using a different DNA ligand composed of three strands annealed to form the structure shown in Fig. 2. This DNA was 32P-labeled at the 5'-end of the 60-mer template strand. Incubation of the nicked DNA with a molar excess of D29A ligase did not alter the size of the labeled strand. Exonuclease III digestion of the DNA in the absence of ligase converted all of the 60-mer into shorter 5' labeled oligonucleotides, principally a cluster of three species migrating at ~11–13 nucleotides and a minor product of ~23 nucleotides. Binding of D29A ligase resulted in the appearance of an exonuclease III-resistant doublet migrating at 38–39 nucleotides (denoted by the asterisk in chromosome XDG). The labeled 18-mer was incubated with a molar excess of D29A ligase and calf thymus terminal transferase (57 units; Promega) in primer extension buffer containing 0.1 mM potassium cacodylate (pH 7.2), 10 mM CoCl2, and 1 mM DTT. The labeled strands were labeled by autoradiography, and the 19-mer was eluted from an excised gel slice and recovered by ethanol precipitation.
Footprint of DNA Ligase Bound at a Nick

**Fig. 1.** Exonuclease III footprint of Chlorella virus DNA ligase bound at a nick. A, the structure of the nicked duplex ligand used in the DNA binding and footprinting reactions is shown. The 32P-labeled 5'-phosphate at the nick is indicated by the asterisk. The vertical arrows indicate the margins of the exonuclease III footprint of ligase on the 5'-phosphate strand. B, DNA binding. Reaction mixtures (10 μl) containing 50 mM Tris-HCl (pH 8.0), 5 mM DTT, 0.5 mM of nicked ligand, and 0.25, 0.5, 1, or 2 pmol of D29A ligase (proceeding left to right) were incubated for 5 min at 22 °C. A control mixture (lane −) contained no ligase. The samples were adjusted to 8% glycerol and then analyzed by electrophoresis through a native 8% polyacrylamide gel containing TBE (90 mM Tris borate, 2.5 mM EDTA). An autoradiogram of the gel is shown. C, exonuclease III footprint. Reaction mixtures (10 μl) containing 50 mM Tris-HCl (pH 8.0), 5 mM DTT, 0.66 mM MgCl₂, 0.5 pmol of nicked ligand, and D29A ligase as specified. Control reactions containing 8 pmol of ligase (lane +) or no ligase (lane −) were incubated for 10 min at 22 °C and then quenched in 45 mM EDTA and 20% formamide. Footprinting reaction mixtures (lanes Exo III) containing no ligase (lane −) or 1, 2, 4, or 8 pmol of ligase (proceeding left to right) were incubated for 5 min at 22 °C and then supplemented with 10 units of exonuclease III (New England Biolabs). After digestion for 10 min at 22 °C, the reactions were quenched in EDTA and formamide. The reaction products were analyzed by electrophoresis through a 20% polyacrylamide gel containing 7 M urea. An autoradiogram of the gel is shown. A 5'-32P-labeled 12-mer oligonucleotide (pATTCCGATAGTG) was analyzed in parallel (lane M).

Fig. 2. Footprinting the ligase-DNA interface on the 3'-OH side of the nick. The structure of the nicked duplex ligand used in the footprinting reactions is shown. The 32P-labeled 5'-phosphate on the 60-mer template strand is indicated by the asterisk. The vertical arrows indicate the margins of the exonuclease III footprint of ligase on the 60-mer strand. The 5'-phosphate at the nick is unlabeled. Footprinting was performed as described in Fig. 1. Control reactions containing 8 pmol of ligase (lane +) or no ligase (lane −) were not subjected to nuclease treatment. Reaction mixtures containing no ligase (lane −) or 1, 2, 4, or 5 pmol of ligase (proceeding left to right) were digested with exonuclease III. The reaction products were analyzed by electrophoresis through a denaturing 12% polyacrylamide gel. An autoradiogram of the gel is shown. A mixture of 5'-32P-labeled oligonucleotides (41-, 30-, 18-, and 12-mer) was run in parallel in lane M.

The 5'-phosphate group and the 3'-nucleotide at the nick are critical for DNA recognition (10, 12). The role of the 3'-OH moiety in DNA binding has not been evaluated. In the experiment shown in Fig. 4, we analyzed by native gel electrophoresis the binding of wild type ligase to a DNA ligand containing a standard 3'-OH/5'-phosphate nick versus ligands containing variant structures at the termini of the discontinuous top strands. The ligase bound to nicked DNA, but it did not form a complex with nicked DNA lacking a 5'-phosphate or with DNA containing a 1-nucleotide gap between the 5'-phosphate and the 3'-OH groups (Fig. 4B). To test specifically the role of the 3'-OH group, we prepared a ligand containing a single 3'-deoxynucleotide at the nick (Fig. 4A). The instructive finding was that ligase did bind to a 3'-deoxy/5'-phosphate nick (Fig. 4B). Similar experiments performed with purified recombinant vaccinia virus DNA ligase showed that it was also capable of binding to a 3'-deoxy/5'-phosphate nick (not shown). Thus, the 3'-OH moiety is not required for nick recognition by the two virus-encoded enzymes.

The 3'-OH of the Nick Is Required for DNA-Adenylate Formation...
The second step of the ligase pathway is the reaction of DNA-bound ligase-adenylate with the 5′-phosphate at the nick to form the nicked DNA-adenylate intermediate. DNA-adenylate is normally undetectable during reactions catalyzed by eukaryotic viral ligases (11, 12); this is because the enzyme remains bound to the nicked DNA-adenylate after step 2 and rapidly catalyzes the attack of the 3′-OH on DNA-adenylate to form a phosphodiester. By definition, the 3′-OH is required for step 3. But does the 3′-OH participate in step 2? We have shown here that a 3′-OH is not required for the binding of ligase to the nicked substrate. If the 3′-OH plays no role in step 2 chemistry, we would expect to detect the accumulation of DNA-adenylate when ligase reacts with the 3′-deoxy/5′-phosphate nicked DNA substrate.

3′-OH/5′-Phosphate-nicked DNA and 3′-deoxy/5′-phosphate-nicked DNA substrates (both 5′-32P-labeled at the nick) were incubated with an 8-fold molar excess of wild type Chlorella virus DNA ligase in the presence of magnesium and ATP. Reaction with the standard nick resulted in near quantitative joining of the 5′-32P-labeled 18-mer substrate and a preadenylated nicked duplex in which the 3′-OH moiety is critical for DNA-adenylate formation by the two virus-encoded DNA ligases.

A 2′-OH Is Unable to Substitute for the 3′-OH in the Second and Third Steps of the Ligation Reaction—We performed an additional experiment in which Chlorella virus DNA ligase was reacted in stoichiometric amounts with a nicked substrate consisting of a 5′-32P-labeled 18-mer (pTCCGATAGTGACCTA-CAG) and a 2′-OH-terminated 19-mer strand (pCATACTCGGTGCCTTA) annealed to a complementary 36-mer template strand. (The 2′-OH-terminated 19-mer was prepared by 3′ extension of an 18-mer with cordycepin triphosphate.) The ligase catalyzed no detectable joining of the two strands to form an expected 37-mer ligation product and there was only a trace conversion of the 5′-32P-labeled 18-mer to DNA-adenylate (data not shown). Thus, we conclude that the essential role of the 3′-OH at the nick in DNA adenylate formation cannot be fulfilled by a 2′-OH moiety. Note that the presence of a 2′-OH has no inhibitory effects on the action of Chlorella virus DNA ligase, which is perfectly capable of binding and sealing a nicked duplex in which the 3′-OH strand is all RNA (15).

Can a 2′-OH serve as the nucleophile in step 3 of the ligation reaction? To answer this question, we tested the ability of Chlorella virus ligase to seal a preadenylated nicked duplex DNA containing a 2′-OH at the nick (Fig. 6). The use of a preadenylated nicked molecule bypasses the requirement for the 3′-OH during step 2 and allows us to focus exclusively on the reactivity of a 2′-OH as the nucleophile in the strand closure step. The preadenylated 2′-OH substrate and a preadenylated 3′-OH control nicked DNA substrate were reacted with...
The sites of $^{32}$P-labeling are denoted by dots. The substrates used in the ligation assays are shown.

late intermediate (confirmed that the N-terminal sequence (AITKPL) corresponds after transfer to a polyvinylidene difluoride membrane

Sequencing of the 40-kDa species by automated Edman chemistry (predicted size of 34 kDa) is anomalously slow (12).

The electrophoretic mobility of the 298-amino acid PBCV-1 products: a set of 3 polypeptides migrating at

concentrations of trypsin yielded two clusters of digestion products. A set of 3 polypeptides migrating at

an 8-fold excess of ligase for 60 min. The 3'-OH substrate was efficiently joined to the adenylated strand to yield the expected ligation product, and the reaction was complete in 15 min (Fig. 6). In contrast, the 2'-OH strand was completely unreactive with the DNA-adenylate (Fig. 6). We conclude that Chlorella virus ligase displays a strict requirement for a 3'-OH during the step of phosphodiester bond formation.

Probing the Structure of DNA Ligase by Limited Proteolysis with Trypsin—Purified D29A ligase was subjected to proteolysis with increasing amounts of trypsin. SDS-PAGE analysis of the undigested enzyme preparation revealed an ~40 kDa polypeptide corresponding to the Chlorella virus ligase (Fig. 7). The electrophoretic mobility of the 298-amino acid PBCV-1 ligase (predicted size of 34 kDa) is anomalously slow (12). Sequencing of the 40-kDa species by automated Edman chemistry after transfer to a polyvinylidene difluoride membrane confirmed that the N-terminal sequence (A1T1KPL) corresponded to that of the Chlorella virus enzyme beginning at residue Ala-2 of the ligase polypeptide. Apparently, the ligase suffered removal of the initiating methionine during expression in Escherichia coli. Initial scission of the ligase at low concentrations of trypsin yielded two clusters of digestion products: a set of 3 polypeptides migrating at ~26–29 kDa and a smaller set of polypeptides migrating at ~10–12 kDa (Fig. 7). Sequencing of these cleavage products revealed that each of the polypeptides in the larger cluster derived from the N terminus of the enzyme. The lower molecular mass cluster was a mixture of at least 4 polypeptides. The N-terminal sequences of the two major species, RSTHKSGK and RSTHKSGK, indicated that these products arose by tryptic cleavage at Lys-219/Arg-220 and Arg-220/Asp-221. Two less abundant peptides within the lower molecular mass cluster derived from tryptic cleavages at Lys-210/Thr-211 and Lys-224/Ser-225. The trypsin-accessible sites fall within a 16-amino acid segment between nucleotidyl

transerase motifs V and VI (see Fig. 10). The finding that the two clusters of tryptic products persisted at trypsin concentrations in excess of the amount sufficient to cleave all of the input

Fig. 6. A 2'-OH cannot replace the 3'-OH in strand closure at a preadenylated nick. The adenylated 18-mer oligonucleotide used to form the preadenylated nicked substrate was synthesized by ligase-mediated AMP transfer to the 5'-32P-labeled 18-mer strand of a DNA molecule containing a 1-nucleotide gap (13). The radiolabeled DNA-adenylate strand was gel-purified and annealed to an unlabeled 36-mer template in the presence of a 2'-OH 19-mer strand to form the 2'-OH structure shown in the figure. The molar ratio of 18-mer DNA-adenylate to 36-mer to 19-mer strands in the annealing reaction was 1:2:3. The 19-mer 2'-OH strand was 5'-32P-labeled at 1/400 the specific activity of the 18-mer DNA-adenylate strand. A control 3'-OH substrate was prepared by using an unlabeled 3'-OH 18-mer strand in place of the 19-mer 2'-OH strand. The structure of the 3'-OH substrate is shown. Ligation reaction mixtures containing (per 10 µl) 50 mM Tris-HCl (pH 8.0), 5 mM DTT, 10 mM MgCl2, 80 fmol of preadenylated nicked DNA, and 640 fmol of ligase were incubated at 22 °C. Aliquots (10 µl) were removed after 15, 30, and 60 min and quenched immediately with EDTA and formamide. The products were resolved by denaturing polyacrylamide gel electrophoresis. An autoradiogram of the gel is shown. The positions of the input 32P-labeled DNA-adenylate strand (AppDNA), the nonadenylated 5'-32P-labeled 18-mer (pDNA, analyzed in lane M), and the 36-mer ligation product formed with the 3'-OH substrate are indicated on the right.

Fig. 7. Limited proteolysis of Chlorella virus DNA ligase with trypsin. Reaction mixtures (10 µl) containing 50 mM Tris-HCl (pH 8.0), 5 mM DTT, 1 µg of D29A ligase, and 1, 2, 4, or 8 ng of trypsin (proceeding from left to right) were incubated for 10 min at 22 °C. Trypsin was omitted from a control reaction (lane –). The reactions were quenched by the addition of SDS sample buffer, and the digests were analyzed by electrophoresis through a 16% polyacrylamide gel containing 0.1% SDS. A photograph of the Coomassie Blue-stained gel is shown. The positions and sizes (kDa) of marker proteins are indicated on the right. A parallel set of reaction mixtures was resolved by SDS-PAGE, and the indicated polypeptide bands were subjected to automated N-terminal sequencing as described (17). The N-terminal sequences are denoted in single-letter code.
Footprint of DNA Ligase Bound at a Nick

**FIG. 8. DNA protects Chlorella virus ligase from digestion by trypsin.** Reaction mixtures (10 μl) containing 50 mM Tris-HCl (pH 8.0), 5 mM DTT, and 5 μg (150 pmol) of D29A ligase were preincubated for 10 min at 22 °C in the presence (lanes 6–10) or absence (lanes 1–5) of 150 pmol of singly nicked DNA. The structure of the DNA ligand is shown above the right panel. The samples were then digested for 10 min at 22 °C with 1, 2, 4, or 8 ng of trypsin (proceeding from left to right within each titration series). Trypsin was omitted from control reactions (lanes 1 and 6). The digestion products were analyzed by SDS-PAGE. A Coomassie Blue-stained gel is shown. Polypeptides corresponding to the full size ligase and N-terminal tryptic fragments are denoted on the right.

protein suggested that these species corresponded to folded protein domains. The sensitivity of wild type ligase to tryptic proteolysis was identical to that observed with the D29A enzyme (not shown). These results are consistent with a bipartite ligase structure consisting of a larger N-terminal domain and a smaller C-terminal domain connected by a linker segment distal to motif V.

**Nicked DNA Affords Protection of Ligase from Tryptic Proteolysis**—The tryptic sensitivity of free ligase was compared with that of ligase preincubated with an equimolar concentration (15 μM) of DNA ligand containing a single 3’-OH/5’-phosphate nick. The trypsin concentration dependence of cleavage of the free enzyme to yield the N-terminal cluster (Fig. 8) was similar to that seen in Fig. 7. Nicked DNA afforded substantial (~8-fold) protection of the ligase from trypsic digestion (Fig. 8). We did not observe protection of ligase from tryptic digestion when the ligase was preincubated with the 42-mer hairpin strand alone (not shown). These results suggest that the trypsin-sensitive segment of the free ligase that is protected in the presence of DNA may comprise part of the DNA binding site on the enzyme. Alternatively, protection from proteolysis may occur because of a ligand-induced conformational change involving a site on the enzyme other than the DNA-binding surface. A ligand-induced conformational change would be supported by a finding of DNA-induced acquisition of tryptic sensitivity at novel sites not cleaved in the free enzyme. However, we found that the electrophoretic mobility of the digestion products generated in the presence of nicked DNA did not differ from those produced in the absence of DNA.

**Limited Proteolysis of DNA Ligase with Chymotrypsin**—To further probe the structure of Chlorella virus DNA ligase and its binding site for DNA, we carried out limited proteolysis with chymotrypsin. Digestion of the enzyme in the absence of DNA yielded two sets of cleavage products: a set of 3 polypeptides migrating at ~29 kDa and a smaller, lower mass set of polypeptides migrating at ~10–12 kDa (Fig. 9). The polypeptides in the lower cluster derived from the N terminus of the enzyme. The lower molecular mass cluster was of a mixture of at least 5 polypeptides. The N-terminal sequences of two of these small polypeptides, KSGKVG and SKRST, suggested that initial scission by chymotrypsin occurred at His-223/Lys-224 and Tyr-217/Ser-218. These two chymotryptic cleavage sites fall within the same segment of the ligase between motifs V and VI that is accessible to trypsin (Fig. 10). Other peptides within the low molecular mass cluster of the chymotryptic digest derived from the N terminus or from proteolytic cleavages at Tyr-89/Asn-90 and Met-83/Thr-84 (Fig. 7). The latter chymotryptic sites are located between motifs III and IIIa (Fig. 10).

Prior incubation with nicked DNA protected the ligase from initial scission by chymotrypsin and altered the distribution of cleavage products within the high molecular mass cluster (Fig. 9). The major species of ~29 kDa generated by chymotrypsin in the presence of DNA comigrated with the least abundant of the 3 polypeptides in the high molecular mass cluster seen in the absence of DNA. Sequencing of the 29-kDa polypeptide showed that it derived from the N terminus of the ligase. This N-terminal polypeptide was relatively refractory to further proteolysis. The lower molecular mass cluster of chymotryptic products formed in the presence of nicked DNA consisted of the same set of peptides detected in the digest of the free enzyme, except that the NAKFSY product of cleavage at Tyr-89/Asn-90 was relatively more abundant in the DNA-containing digests than in the digest of the free ligase. Although the C termini of the chymotryptic fragments are not known, the findings suggest that the 29-kDa N-terminal fragment is the initial chymotryptic product (a consequence of scission at His-223/Lys-224) and that nearby secondary cleavages (e.g., at Tyr-217/Ser-218 and at least one other site) generate the slightly smaller products in the high molecular mass cluster. In some chymo-tryptic digests, we detected among the low molecular mass products a peptide with the sequence KNTNT, which may be among the smaller products in the high molecular mass cluster. In some chymo-tryptic digests, we detected among the low molecular mass products a peptide with the sequence KNTNT, which indicated cleavage at Phe-204/Lys-205. Nicked DNA apparently protects the ligase from primary and secondary cleavages at these sites.

**Effect of Limited Proteolysis on DNA Binding**—The D29A ligase was subjected to limited proteolysis with increasing amounts of trypsin or chymotrypsin in the absence of DNA. Aliquots of the proteolyzed enzyme were then incubated with 32P-labeled nicked duplex DNA, and DNA-protein complex formation was assayed by native gel electrophoresis (Fig. 11A). The remainder of the digest was analyzed by SDS-PAGE to ascertain the extent of scission of the ligase polypeptide (Fig.
consistent with the third scenario. We conclude that the integrity of the protease-sensitive linker segment is critical for nick recognition.

DISCUSSION

The experiments presented in this study enhance our understanding of nick recognition and catalysis by ATP-dependent DNA ligases in four respects: (i) we provide the first physical characterization of the enzyme binding site on a singly nicked duplex DNA; (ii) we identify a protease-sensitive interdomain linker that may comprise part of the substrate binding site on the ligase; (iii) we show that the 3’-OH group at the nick is not required for nick recognition but is essential for DNA-adenylate formation by Chlorella virus ligase; and (iv) we find that a 2’-OH cannot substitute for the 3’-OH during steps 2 or 3 of the ligase reaction.

DNA Footprint of Ligase Bound at a Nick—We used exonuclease III to footprint Chlorella virus D29A ligase bound at a single nick in duplex DNA. The size of the binding site was 19–21 nucleotides. The ligase footprint was slightly asymmetric, extending 8–9 nucleotides on the 3’-OH side of the nick and 11–12 nucleotides on the 5’-phosphate side. The margins of the footprint were remarkably discrete. Also, there was nearly complete exonuclease arrest at these margins when ligase was present at concentrations sufficient to bind nearly all of the input ligand. We surmise that the Chlorella virus ligase does not dissociate from the DNA or diffuse linearly away from the nick during the time frame of the in vitro assays. Given that Chlorella virus ligase is a monomer in solution and that ligase binding to nicked DNA yields a single discrete complex when analyzed by native gel electrophoresis, we attribute the observed footprint to the binding of a single enzyme molecule at the nick. The use of the D29A enzyme for the nuclease footprint analysis was dictated by the need to prevent ligase-adenylate from immediately catalyzing strand joining once bound to the nicked DNA in the presence of magnesium. The footprint observed for the D29A enzyme is likely to mimic that of the wild type Chlorella virus ligase, insofar as: (i) the affinity of D29A for the nick is equivalent to that of wild type ligase, (ii) the mobility of the D29A-nicked DNA complex during native gel electrophoresis is identical to that of the wild type ligase-nicked DNA complex, and (iii) the tertiary structures of the D29A and wild type ligase are grossly similar as gauged by partial proteolysis.

The 298-amino-acid Chlorella virus ligase represents a minimal catalytic domain of an ATP-dependent ligase. As such, its footprint on nicked DNA may represent a minimal interface required for intrinsic nick sensing. Other eukaryotic DNA ligases, which are much greater in size (552–922 amino acids), may well cover a larger segment of DNA than does the Chlorella virus enzyme. To examine this point, we attempted to footprint the 552-amino-acid vaccinia virus DNA ligase bound at a nick. We engineered and purified a mutated version of vaccinia ligase (D233A) in which the motif I Asp was replaced by alanine with the intent of recapitulating the step-arrrest phenotype of the Chlorella virus D29A mutant. Vaccinia ligase D233A was active in enzyme-adenylate formation, and the mutant protein bound to nicked DNA with normal affinity. Unfortunately, a low level of residual strand joining activity (1% of that of the wild type vaccinia ligase) was sufficient to allow single-turnover ligation within the time frame of the footprinting assays once the D233A ligase-DNA complexes resolved footprint to the binding of a single enzyme molecule at the nick. The experiments presented in this study enhance our understanding of nick recognition and catalysis by ATP-dependent DNA ligases in four respects: (i) we provide the first physical characterization of the enzyme binding site on a singly nicked duplex DNA; (ii) we identify a protease-sensitive interdomain linker that may comprise part of the substrate binding site on the ligase; (iii) we show that the 3’-OH group at the nick is not required for nick recognition but is essential for DNA-adenylate formation by Chlorella virus ligase; and (iv) we find that a 2’-OH cannot substitute for the 3’-OH during steps 2 or 3 of the ligase reaction.
ligase provides a low resolution view of the enzyme’s domain structure and a clue to the location of the DNA binding site on the enzyme. The peptide bonds in the native enzyme that were most accessible to cleavage by trypsin and chymotrypsin cluster within a short segment located distal to motif V. This suggests that Chlorella virus ligase consists of an N-terminal domain that includes nucleotidyl transferase motifs I, III, IIIa, IV, and V, plus a smaller C-terminal domain. The domain structure inferred from the proteolysis experiment is consistent with the actual domain structure of T7 DNA ligase determined by x-ray crystallography (6).

It is worth noting that the principal sites of protease accessibility in the T7 and vaccinia ligases are situated between motifs III and IIIa (16, 17) rather than at the interdomain boundary deduced from the crystal structure. The protease-accessible region between motifs III and IIIa corresponds to an exposed surface loop that is disordered in the T7 ligase crystal (6). The Chlorella virus ligase is cleaved by chymotrypsin between motifs III and IIIa, but this site is less accessible than the sites located distal to motif V. The segment between motifs III and IIIa is shorter in Chlorella virus ligase (24 amino acids) than in T7 ligase (49 amino acids) and is less rich in potential tryptic sites than the corresponding segments of the T7 and vaccinia enzymes. Thus, the relative protease-resistance of the putative loop within domain 1 of Chlorella virus ligase fortuitously permitted the delineation of a protease-sensitive interdomain linker elsewhere in the protein. The amino acid sequence of the linker segment of the Chlorella virus ligase (FGYSKRSTHKSG) is loosely conserved at an analogous location distal to motif V in the DNA ligases of Schizosaccharomyces pombe (YGGKGRSSTVYGA), Saccharomyces cerevisiae (YGRGKRTGTYGG), Desulfurolobus ambivalens (HGGKRGK-GKSS), human ligase I (LQGRKARGYGG), and bacteriophage T4 (YPHRKDPKAGG) (18).

The interdomain linker of the Chlorella virus ligase was less accessible to proteolysis when the enzyme was bound to nicked duplex DNA. The simple interpretation of this result is that the linker composes part of the DNA binding site and protection from proteolysis reflects steric hindrance by the bound substrate. The proposition that the interdomain segment might be directly involved in DNA binding is consonant with the suggestion that duplex DNA might bind within a deep, positively charged cleft separating domains 1 and 2 of the T7 ligase (6). Because there is no obvious sequence conservation between the Chlorella virus ligase linker and domain 2 of T7 ligase, we are unable to model it on the T7 structure. We do not exclude an alternative interpretation of the data whereby DNA binding to ligase elicits a conformational change in the linker segment. Conformational flexibility of a related nucleotidyl transferase is revealed in the crystal structure of Chlorella virus RNA-capping enzyme, which can adopt an open conformation with a wide interdomain groove and a closed form with a narrow groove (19). The conformational transition in the capping enzyme entails local changes in the secondary structure of motif V and distal elements within domain 2.

Catalytic Role of the 3′-OH of the Nick—The 5′-phosphate and 3′-OH groups at the nick act sequentially as nucleophiles during steps 2 and 3 of the strand joining pathway. The 5′-phosphate is also essential for the binding of ligase-AMP to nicked DNA, implying that functional groups on the enzyme make direct contact with the 5′-phosphate prior to step 2 chemistry. Here, we have shown that the 3′-OH moiety is not required for nick recognition. The Chlorella virus ligase and the vaccinia enzyme bind to a nicked ligand containing 3′-deoxy and 5′-phosphate termini. An instructive finding was that such nicks are not adenylated by the viral ligases, \textit{i.e.} that the 3′-OH is important for step 2 chemistry even though it is not itself chemically transformed during DNA-adenylate formation.

Our results confirm and extend the early findings of Modrich and Lehman (20) in their classic paper describing the steady-state kinetic analysis of the composite ligation reaction and the partial reactions catalyzed by the NAD-dependent \textit{E. coli} DNA ligase. Using homopolymeric substrates consisting of reactive poly(dT) strands annealed to poly(dA), they were able to detect the addition of AMP to the 5′-phosphate ends of poly(dT) that was modified by the addition of a single 3′-dideoxythymidylate residue. However, the rate of synthesis of DNA-adenylate on the homopolymeric 3′-deoxy substrate was at least three orders of magnitude slower than the rate of strand joining with a 3′-OH substrate. Modrich and Lehman (20) proposed, as we do, that the 3′-OH is essential for step 2. A more recent study by Yang and Chan (21) demonstrated that 3′-dideoxyadenylate substitution at the nick completely blocked DNA-adenylate formation and strand joining by human DNA ligases I and II, but the issue of whether the human ligases could bind to the 3′-deoxy nicks was not addressed. The present study of Chlorella virus ligase clearly excludes a requirement for the 3′-OH in nick recognition as an explanation for the step 2 block. We suggest that the 3′-OH at the nick interacts with an essential step 2 catalyst, either an amino acid on the Chlorella virus enzyme or the divalent cation cofactor. The key role of the 3′-OH during step 2 of the ligation reaction of Chlorella virus DNA ligase cannot be fulfilled by a 2′-OH group. A need for the 3′-OH during step 2 may not apply to every DNA ligase. Tomkinson et al. (22) have shown that the \textit{S. cerevisiae} DNA ligase Cdc9p is able to catalyze DNA-adenylate formation at a 3′-dideoxy/5′-phosphate nick. This is remarkable, given that yeast Cdc9p is regarded as the structural and functional counterpart of mammalian DNA ligase I (22), as noted above, stringently requires the 3′-OH for catalysis of DNA-adenylate formation.

The next milestone in understanding ligase action in molecular detail will be the determination of the structure of the enzyme bound to DNA. The present findings inform such efforts, insofar as (i) knowledge of the ligase footprint can direct the design of nicked substrates for cocrysalization trials and (ii) stable nick binding by step-arsenic mutations makes such mutants good candidates for crystallization.

REFERENCES

1. Lehman, I. R. (1974) \textit{Science} \textbf{186}, 790–797
2. Engler, M. J., and Richardson, C. C. (1982) \textit{Enzymes} \textbf{15}, 3–29
3. Tomkinson, A. E., and Mackey, Z. B. (1998) \textit{Mutat. Res.} \textbf{407}, 1–9
4. Shuman, S., and Schwer, B. (1995) \textit{Mol. Microbiol.} \textbf{17}, 405–410
5. Shuman, S. (1996) \textit{Structure} \textbf{4}, 653–656
6. Subramanya, H. S., Doherty, A. J., Ashford, S. R., and Wigley, D. B. (1996) \textit{Cell} \textbf{85}, 667–675
7. Tomkinson, A. E., Totty, N. F., Gisnaburg, M., and Lindahl, T. (1991) \textit{Proc. Natl. Acad. Sci. U. S. A.} \textbf{88}, 400–404
8. Kodama, K., Barnes, D. E., and Lindahl, T. (1991) \textit{Nucleic Acids Res.} \textbf{19}, 6093–6099
9. Shuman, S., and Ru, X. (1995) \textit{Virology} \textbf{211}, 73–83
10. Sriskanda, V., and Shuman, S. (1998) \textit{Nucleic Acids Res.} \textbf{26}, 525–531
11. Shuman, S. (1995) \textit{Biochemistry} \textbf{34}, 16138–16147
12. Ho, C. K., Van Etten, J. L., and Shuman, S. (1997) \textit{J. Virol.} \textbf{71}, 1931–1937
13. Sekiguchi, J., and Shuman, S. (1997) \textit{J. Virol.} \textbf{71}, 9679–9684
14. Sekiguchi, J., and Shuman, S. (1997) \textit{Biochemistry} \textbf{36}, 9073–9079
15. Sriskanda, V., and Shuman, S. (1998) \textit{Nucleic Acids Res.} \textbf{26}, 3536–3541
16. Doherty, A. J., Ashford, S. R., and Wigley, D. B. (1996) \textit{Nucleic Acids Res.} \textbf{24}, 2282–2287
17. Sekiguchi, J., and Shuman, S. (1997) \textit{Nucleic Acids Res.} \textbf{25}, 727–734
18. Kletzin, A. (1992) \textit{Nucleic Acids Res.} \textbf{20}, 5389–5396
19. Hakansson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997) \textit{Cell} \textbf{89}, 545–553
20. Modrich, P., and Lehman, I. R. (1973) \textit{J. Biol. Chem.} \textbf{248}, 7502–7511
21. Yang, S., and Chan, J. Y. H. (1992) \textit{J. Biol. Chem.} \textbf{267}, 8117–8122
22. Tomkinson, A. E., Tappe, N. J., and Friedman, E. C. (1992) \textit{Biochemistry} \textbf{31}, 11762–11771