Characterization of proteolytic fragments of bacteriophage T7 DNA ligase

Aidan J. Doherty, Stephen R. Ashford and Dale B. Wigley*

Laboratory of Molecular Biophysics, University of Oxford, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

ABSTRACT

Treatment of T7 DNA ligase with a range of proteases generates two major fragments which are resistant to further digestion. These fragments, of molecular weight 16 and 26 kDa, are derived from the N- and C-termini of the protein, respectively. The presence of ATP or a non-hydrolysable analogue, ADPNP, during limited proteolysis greatly reduces the level of digestion. The N-terminal 16 kDa region of the intact T7 ligase is labelled selectively in the presence of [α-^{32}P]ATP, confirming that it contains the active site lysine residue. In common with the intact enzyme, the C-terminal portion of the protein retains the ability to band shift DNA fragments of various lengths, implicating it in DNA binding. It can also inhibit ligation by the intact protein, apparently by competing for target sites on DNA. We conclude that the N-terminal region, which contains the putative active site lysine, plays a role in the transfer of AMP from the enzyme–adenylate complex to the 5′ phosphate at the nick site, while the C-terminal 26 kDa fragment appears to position the enzyme at the target site on DNA.

INTRODUCTION

DNA ligases catalyse the formation of phosphodiester bonds at single-strand breaks between adjacent 3′ hydroxyl and 5′ phosphate termini in double-stranded DNA (for reviews see [1,2,3]). DNA ligase is an essential enzyme for cells. It is required for a number of important cellular processes, including replication of DNA and the repair of damaged DNA, as evidenced by the number of viruses which have genes encoding their own DNA ligases. The enzymes appear to fall into two groups; those requiring NAD+ for activity and those requiring ATP. The NAD+-requiring DNA ligases have only been found in prokaryotic organisms [1,2], but the DNA ligase from an archeon bacteria *Desulfurolobus ambivalens* requires ATP [4]. The eukaryotic enzymes all utilize ATP, as do the virally encoded enzymes. They show a wide variety of molecular weights, from 103 kDa for the human type I enzyme to 41 kDa for the enzyme from bacteriophage T7. The discrepancy between these figures is due to the presence, in the mammalian enzymes, of a domain in the protein which, when phosphorylated, is the signal for translocation of the protein from the cytoplasm of the cell to the nucleus [3]. Even in the proteins that are extended at the N-terminus, the level of sequence homology is much greater in the C-terminal region (Fig. 1). The active site lysine has been identified and is in a region that is homologous across all known sequences of both DNA and RNA ligases [5,6]. The T bacteriophages also encode their own ATP-dependent DNA ligases. The T4 enzyme has found widespread use in recombinant DNA techniques. Probably because of this and the ready availability of the enzyme, it has been the subject of considerable study, though it has never been crystallized. The enzyme from bacteriophage T7 is smaller than its T4 counterpart, but shows very similar properties. The T7 enzyme appears to represent the smallest functional unit for a DNA ligase and sequence homology with other ligases shows that it corresponds to the highly conserved C-terminal region of other ATP-dependent ligases.

In 1967, DNA ligase activity was purified from T7-infected *Escherichia coli* [7]. Studies with mutant bacteriophage deficient in polynucleotide ligase showed that while the enzyme was not essential for phage growth, since the host ligase can substitute for the phage-induced enzyme, the amount of phage produced upon infection is drastically reduced and there is an accumulation of small DNA fragments in the host [8]. Assays with crude extracts containing the ligase showed that it has the ability to join oligo(dT)·poly(A) and oligo(A)·poly(dT) [9]. The T7 ligase was later co-purified with T7 gene 4 protein [10]. The enzyme can ligate DNA substrates in the presence of ATP, but also at a reduced rate with dATP [2,10]. In common with other DNA ligases, the enzyme requires divalent cations for activity. This requirement appears to be fulfilled by Mg^{2+} *in vivo*, although other ions, such as Ca^{2+}, can substitute to give reduced activity [2]. The position of the *lig* gene (*lig* gene 1.3) was identified [11] and shown to encode a protein with an approximate molecular weight of 41 kDa. The complete sequencing of the T7 genome [12] confirmed the position and size of the gene and predicted an enzyme with a molecular weight of 41.133 kDa, composed of 359 amino acids, making it the smallest known DNA ligase. Despite the availability of the gene sequence for T7 ligase it has not been cloned or characterized in detail. Since the majority of the biochemical characterization of ligases has been performed on the larger T4 and *E.coli* enzymes, little is known about the structure and mechanism of the smaller enzymes such as T7, T3 and african swine fever virus ligase.

It is not known whether DNA ligases have functional domains which are involved in either catalysis or DNA binding and, if so, how these interact together to perform a ligation event. Lehman and co-workers isolated a large fragment of *E.coli* DNA ligase, using limited proteolysis, which retained the ability to form the

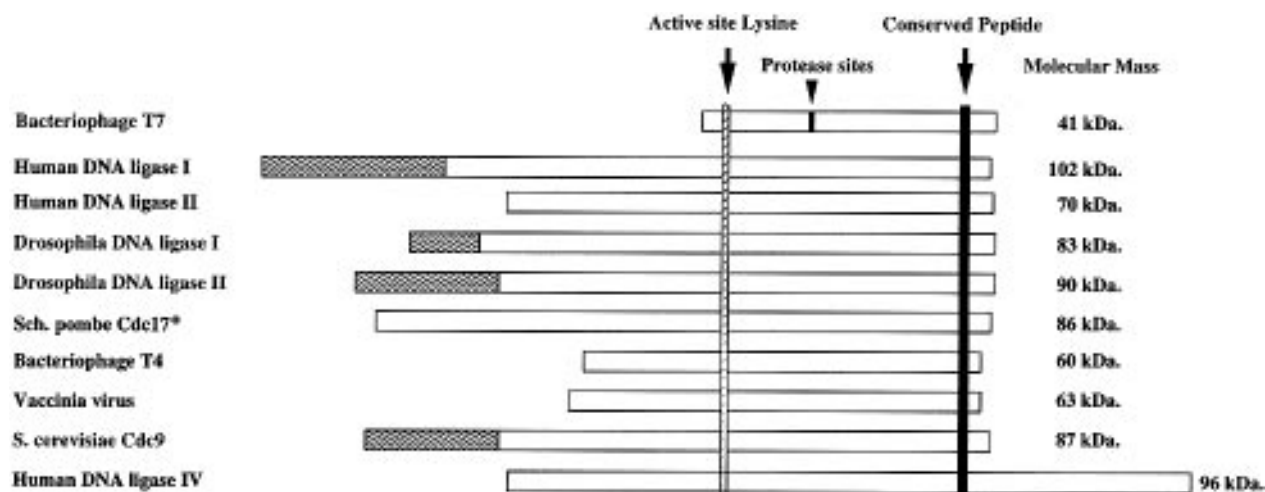---

* To whom correspondence should be addressed

Figure 1. Alignment of ATP-dependent DNA ligases. Schematic diagram of a representative sample of ATP-dependent DNA ligases aligned at their active sites and a conserved peptide. The areas filled in with wavy lines indicate regions of the enzymes that have been shown to be dispensable for catalytic activity. The enzymes shown are bacteriophage T7 (12), human DNA ligase I (5,14,36), human DNA ligase II (37,38),*Drosophila* DNA ligase I (39), *Drosophila* DNA ligase II (40), *S.pombe* Cdc17[+] DNA ligase (41), bacteriophage T4 DNA ligase (32),*Vaccinia* virus DNA ligase (42), *S.cerevisiae* Cdc9 DNA ligase (43,44) and human DNA ligase IV (34). The protease-sensitive site of T7 ligase (see text) is indicated with an arrow.

enzyme–adenylate intermediate but had no DNA ligation activity (13). These results suggested that the fragment had lost the ability to bind DNA, but this was not proven. More recently it was shown that endogenous proteolytic removal of the N-terminal 216 amino acids of human DNA ligase I resulted in a fully active fragment of 703 amino acids (3,14).

Previously we reported the over-expression, crystallization and biochemical characterization of T7 DNA ligase (15). In this report, we extend this work by employing classical limited proteolysis techniques to begin to elucidate the structure and mechanism of action of T7 ligase. Two major proteolytic fragments were isolated and N-terminally sequenced. The fragments were subsequently cloned and over-expressed. We present evidence that the smaller proteolytic fragment (16 kDa), containing the putative active site motif, is the site of AMP attachment. The C-terminal fragment (26 kDa) is able to inhibit the ligation activity of the intact enzyme by binding competitively to ligation sites on DNA.

MATERIALS AND METHODS

DNA manipulations

All restriction endonucleases and modification enzymes were obtained from Boehringer Mannheim UK Ltd (Lewes, UK) or Gibco BRL. Protein purification resins and columns were obtained from Pharmacia (Uppsala, Sweden). All other chemicals were obtained from Sigma (Poole, Dorset, UK) unless stated otherwise. [α-^{35}S]dATP and [α-^{32}P]ATP were obtained from Amersham International Ltd. Taq polymerase (AmpliTaq) was obtained from Perkin Elmer Cetus (Emoryville, CA). Sequenase II was obtained from United States Biochemicals (Cleveland, OH).

All restriction endonucleases and modification enzymes were used according to the manufacturers' protocols. All the transformations, DNA isolations and manipulations were performed essentially as described previously unless otherwise stated (16). Oligonucleotides were synthesized using an Applied Biosystems 381A DNA synthesizer. The oligonucleotides were desalted on

NAP G200 columns (Pharmacia) according to the manufacturer's instructions

The following strains of *E.coli* were used as hosts for pET21 (Novagen) constructs: XL1-Blue (*sup*E44, *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *th*i, *rel*A, [F′ *pro*AB, *lac*I^{q}ZΔM15, Tn*10*(*tet*^{R})]) (17) for the propagation of clones and the preparation of single-stranded DNA for sequencing; GR501 (*Hfr*, *thi*-1, *pts*I$^+$, *lig*-251) (18), a conditional lethal DNA ligase mutant, to test if the isolated genomic fragments had complementary ligase activity; B834(DE3) (*hsd*S *gal* (*lcI*ts857 *ind1 met*$^-$ *Sam* 7 *nin* 5*lac*UV5-T7 gene 1) (19,20) for the over-expression of T7 DNA ligase. *Escherichia coli* strains were grown aerobically in Luria broth or on Luria broth agar plates containing the appropriate antibiotics.

Limited proteolysis

Proteolytic cleavage of purified T7 ligase (2 mg/ml) was carried out at 30°C for 10 min in the appropriate buffer containing 5 mM magnesium chloride. A number of proteases and conditions were tried but the best results were obtained with 5 µl clostripain (2.5 U/ml) with a buffer containing 50 mM Tris–HCl, pH 7.5, 5 mM CaCl$_2$, 5 mM DTT, 10% glycerol in a total volume of 20 µl. Reactions were terminated either by the addition of protease inhibitors or EDTA followed by the addition of SDS–PAGE buffer, and heated at 95°C for 5 min. Protein samples were analysed by SDS–PAGE in 10, 12 and 15% gels with 4% stacking gels (21). Gels were stained with Coomassie Brilliant Blue and destained in 10% acetic acid and 25% methanol. N-Terminal sequencing of the ligase fragments was performed using an Applied Biosystems model 470A protein sequenator connected to an on-line 120A high pressure liquid chromatograph.

DNA ligase labelling

T7 DNA ligase (1 mg/ml) was incubated in 50 mM Tris–HCl, pH 7.5, 10 mM MgCl$_2$, 5 mM DTT and 25 µCi [α-^{32}P]ATP in a total volume of 20 µl for 15 min at 25°C. The reactions were stopped by boiling in SDS buffer for 5 min and analysed by electrophoresis

on a 12% SDS gel (21). The gels were dried before being autoradiographed on Fuji RX X-ray film.

Cloning and over-expression of the 16 and 26 kDa fragments

In order to clone and over-express the 16 and 26 kDa fragments, we amplified the respective gene fragments from T7 genomic DNA using PCR. DNA primers were designed based on the amino acid sequencing of the proteolytic fragments and the known gene sequence (GenBank accession no. G15572). Two oligonucleotides were designed for PCR with each fragment based on the N-terminal amino acid sequence of the proteolytic fragments: a 5′ primer (5′-GATATACCATGGTGAACATTAA-GACTAAC-3′) and a 3′ primer (5′-GGGTTTGAATTCCTAAC-GGATTGGTTCAACGAATAA-3′) for PCR of the 16 kDa fragment and a 5′ primer (5′-GATATACCATGGATAAAGTTC-CCTTTAAGCTGCAC-3′) and 3′ primer (5′-GGGTTTTAAGC-TTACATTTTCTCTTGAGGG-3′) for PCR of the 26 kDa fragment. The 5′ primers contained an *Nco*I site and the 3′ primers had a *Hin*dIII site after the stop codon, allowing the PCR products to be cloned into the T7 promoter-based expression vector pET21d. PCR was performed with these primers and bacteriophage T7 genomic DNA (Sigma) using Amplitaq in 100 µl of PCR buffer and cycled as described (22). The PCR reaction was electrophoresed in 1.0% agarose (Gibco BRL) and the product excised from the gel and purified using a Qiaex DNA extraction kit (Qiagen). The DNA was digested with *Nco*I and *Hin*dIII and further gel purified. The fragments were ligated to dephosphory-lated, *Nco*I/*Hin*dIII-cleaved pET21d, transformed into *E.coli* XL1-Blue and plated onto LB agar plates containing 100 µg/ml ampicillin.

Clones containing the fragments pT716 and pT726 were identified by restriction digestion. Clones were then screened for expression by transforming into B834(DE3)[pLysS], growing selected colonies at 37°C in 5 ml Luria broth containing ampicillin and chloramphenicol to an A_{600} of 0.6, induced with 1 mM IPTG and grown for a further 3 h. The level of induction was monitored by electrophoresis of cell extracts on 15% SDS–polyacrylamide gels.

The dideoxy method (23) was used to confirm the nucleotide sequence of the T7 *lig* fragments. Single-stranded template DNA was produced in *E.coli* XL1-Blue using helper phage M13KO7 (Pharmacia) according to the manufacturer's instructions.

Purification of the 26 kDa fragment

Two litre cultures of Luria broth containing 100 µg/ml ampicillin and 50 µg/ml chloramphenicol were inoculated with a 5 ml culture of B834(DE3)[pLysS][pT726] and grown at 37°C until the A_{600} reached 0.6–0.7. The cultures were induced by the addition of 0.5 mM IPTG. Growth was continued for a further 3 h before harvesting the cells by centrifugation at 5000 *g*. The cell pellets were stored at –20°C until required. The pellets were lysed by sonication of a 10% (w/v) cell suspension in buffer A (50 mM Tris–HCl, pH 7.5, 2 mM EDTA, 5 mM DTT and 100 µM PMSF). The cells were pelleted by centrifugation at 20 000 *g*. The pellet was washed with buffer A containing 1% Triton X-100 and recentrifuged at 20 000 *g*. This procedure was repeated several times with buffer A alone. The pellet was allowed to dissolve slowly in buffer C (5 M guanidine hydrochloride, 50 mM Tris–HCl, pH 7.5 and 25 mM NaCl). This solution was dialysed

for 16 h against buffer A containing 10% glycerol. The dialysed solution was applied to a heparin–Sepharose column (20 ml) equilibrated with 90% buffer A and 10% buffer B (buffer A containing 2 M sodium chloride). After washing the column with a further two column volumes of this mixture, the protein was step eluted by washing with 40% buffer B. SDS–PAGE confirmed that this peak contained the semi-purified enzyme at high concentrations. The conductivity of the peak fractions containing the 26 kDa fragment was reduced to that of 50 mM NaCl by dilution with buffer A. This was loaded onto a Q Sepharose column (40 ml) pre-equilibrated in buffer A. The column was then washed with 80 ml of buffer A before eluting the protein with 13% buffer B. Purity was again monitored by electrophoresising samples on a 12% SDS–polyacrylamide gel.

Purification of the 16 kDa fragment

The 16 kDa fragment was purified from SDS–polyacrylamide gels essentially as described previously (24). Clostripain-treated T7 ligase (20 µg/track) was applied to a 15% SDS–polyacrylamide gel. After electrophoresis, the gel was cut in two and one half was stained with Coomassie brilliant blue and the position of the 16 kDa fragment identified. This allowed identification of the position of the fragment on the unstained gel, which was then excised, 200 µl of elution buffer (50 mM Tris–HCl, pH 7.5, 0.1 mM EDTA, 5 mM DTT, 150 mM NaCl) was added and the gel was crushed thoroughly and left for 2 h. The eluted protein was purifed away from the gel with a Micropure microporous separator (Amicon) and the solution was buffer exchanged (50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 5 mM DTT) and concentrated using a Microcon concentrator (Amicon).

ATPase assays

ATPase assays were carried out by detection of inorganic phosphate produced using acidic ammonium molybdate and malachite green as described previously (25).

DNA ligase assays

DNA ligase assay substrate (22mer) was radiolabelled by incubating 20 µg of the oligonucleotide with 100 µCi of [γ-^{32}P]ATP (3000 Ci/mmol; Amersham) and 50 U T4 poly-nucleotide kinase for 45 min at 37°C followed by 10 min at 70°C. The unincorporated label was removed by centrifugation through a S-200 microspin column (Pharmacia). The DNA ligase assay was performed essentially as described previously (26). The complementary 18mer and 22mer oligonucleotides were annealed to single-stranded M13mp19 by incubation at 70°C for 2 min and allowed to cool for 1 h. The annealed DNA was incubated with ligase buffer (50 mM Tris–HCl, pH 7.5, 10 mM MgCl₂, 5 mM DTT), unless otherwise stated, in the presence of enzyme and nucleotide cofactors as indicated, in a total volume of 10 µl for 15 min at 25°C. The reactions were terminated by the addition of sequencing stopping buffer (Sequenase kit; US Biochemicals) followed by heating at 95°C for 5 min. The ligation products were subjected to electrophoresis on a 15% polyacrylamide–urea gel and to autoradiography with Fugi RX X-ray film.

DNA mobility shift assays

These were performed essentially as described previously (27). An aliquot of 20 µg of oligonucleotide was labelled by incubation
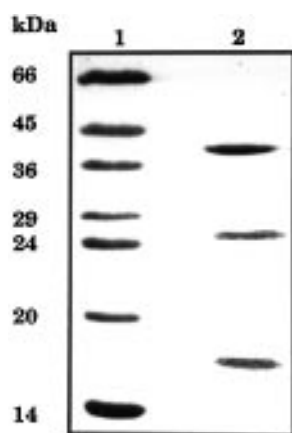
Figure 2. Limited proteolysis of T7 ligase. A 15% SDS–polyacrylamide gel of the limited proteolysis products of T7 ligase produced by treatment with clostripain. Lane 1 contains standards with molecular weights as indicated, while lane 2 is a partial proteolytic digest showing the intact protein (42 kDa) and the two major digestion products at 26 and 16 kDa.

with 100 μCi [γ-^{32}P]ATP (3000 Ci/mmol) and 50 U T4 polynucleotide kinase for 45 min at 37°C followed by 10 min at 70°C. The unincorporated label was removed by centrifugation through a S-200 microspin column (Pharmacia). Binding reactions were carried out in a total volume of 10 μl, containing 50 mM Tris–HCl, pH 7.5, 5 mM DTT, 5 mM MgCl$_2$. Reaction mixtures were incubated at room temperature for 20 min and run on 6–8% polyacrylamide gels which contained 5% glycerol.

RESULTS

Protease cleavage analysis of T7 DNA ligase

T7 DNA ligase was digested under limiting conditions with a range of common proteases, including trypsin, chymotrypsin, clostripain, V8, elastase and thermolysin. Despite the different cleavage specificities of these enzymes, they all produced a major digestion product of ~24–26 kDa as determined by SDS–PAGE. Some proteases (clostripain and V8) also produced another smaller fragment of ~16–18 kDa (Fig. 2). The sizes of the fragments, the intensity of the bands and their dependence on proteolytic digestion time suggested that the two fragments resulted from a single cleavage at a protease hypersensitive site on the ligase. The 26 and 16 kDa fragments were relatively resistant to further digestion by the enzymes, suggesting that they might represent discrete domains. Pre-incubation of T7 ligase with ATP made the enzyme more resistant to proteolysis (data not shown), suggesting that binding of nucleotide may induce a conformational change in the protein. To establish the relationship between each of these clostripain fragments and their associated activities, purification of the two fragments was attempted under native conditions. The digestion products were loaded onto a Superdex-S75 column, but it was not possible to separate the fragments from the full-length enzyme. A single protein peak eluted from the column and had an apparent molecular weight of 80–85 kDa. It should be noted that this anomalous elution profile on gel filtration has been observed previously for the undigested enzyme (15) and is thought to be caused by the extended shape of the molecule in solution.

SDS–PAGE analyses revealed that the 26 and 16 kDa fragments co-eluted in this single peak. Under native conditions the fragments are very tightly associated and can only be separated under denaturing conditions. Limited separation of the two fragments was achieved in gel filtration experiments in the presence of 8 M urea or 6 M guanidine hydrochloride, but this method proved impractical for isolation of the fragments.

T7 DNA ligase digested with clostripain was separated by electrophoresis on a 15% SDS–polyacrylamide gel and the fragments were excised and N-terminally sequenced. Clostripain, also known as endoproteinase Arg-C, cleaves at the carboxylic side of arginine residues (28). The protease-sensitive site for the larger fragment was shown to be located between Arg130 and Lys131 by N-terminal sequencing (Fig. 1). The smaller 16 kDa fragment starts at the extreme N-terminal methionine, which is not removed, while the larger 26 kDa fragment commences C-terminal to the protease site. N-Terminal sequencing of equivalent 26 kDa fragments, produced by digestion with a diverse range of proteases (data not shown), confirmed that this region is prone to cleavage by these enzymes.

Cloning, over-expression and purification of T7 ligase fragments

Cloning of the 16 and 26 kDa proteolysis fragments was performed as described in Materials and Methods. A large number of positives clones containing the correct gene inserts (pT716 and pT726) were identified by digestion of the mini-preparations of DNA with *Nco*I and *Hin*dIII. DNA sequencing of both constructs confirmed that no mutations had been introduced during PCR. In order to determine whether these clones had DNA ligase activity *in vivo*, the two constructs (pT726 and pT716) were transformed into *E.coli* GR501, a conditional lethal DNA ligase mutant which is not viable at 42°C. Transformants were grown overnight at either 30°C (permissive temperature) or 42°C (non-permissive temperature) to assay for functional complementation of the replication-defective phenotype. Previously we have observed that expression of T7 ligase, despite being under the control of a T7 promoter in this plasmid, was sufficient to complement *E.coli* GR501 (18). The fragments were unable to complement this ligase-deficient strain, indicating either that they have no ligation activity or that it is too low to rescue growth.

The expression constructs (pT716 and pT726) were transformed into B834(DE3)[pLysS] and these transformants screened for over-expression of the T7 ligase fragments by inoculating single colonies into small cultures of LB medium and growing for several hours. The cells were induced at mid log phase, by the addition of IPTG, and grown for several more hours. Samples were taken and analysed by SDS–PAGE. In all cases there was large production of proteins which had apparent molecular weights of ~26 and 16 kDa respectively, but unfortunately both of these proteins were totally insoluble. Various methods were employed to try and resolubilize the proteins, with mixed success. It was possible to resolubilize >80% of the 26 kDa fragment using the method described in Materials and Methods. This protein was purified to homogeneity using the three column procedure outlined above, giving a final yield of >60 mg/l. Despite trying a wide range of expression conditions and solubilization methods, attempts to re-fold the 16 kDa fragment proved unsuccessful.

It has been shown previously that it is possible to recover proteolytic fragments from polyacrylamide gels following electro-
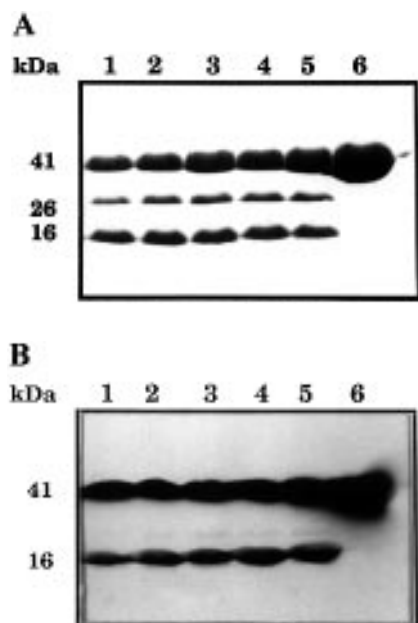
Figure 3. Limited-proteolysis of [α-^{32}P]-labelled T7 ligase. (**A**) A 15% SDS–polyacrylamide gel, stained with Coomassie blue, of limited proteolyis products of [α-^{32}P]-labelled T7 ligase digested with clostripain (0.01 U) for varying lengths of time. Lanes 1–5, 25 µg T7 ligase incubated with clostripain for 30, 20, 15, 10 and 5 min. Lane 6, undigested T7 ligase. (**B**) Autoradiogram of the same gel showing that the intact T7 ligase and the 16 kDa fragment are adenylated selectively.

Figure 4. Inhibition of ligation by the 26 kDa fragment. The synthetic nicked substrate consists of a complementary and 5′-^{32}P-labelled 22mer and an adjacent 18mer annealed with single-stranded M13 DNA. Reactions containing DNA substrate and 1 µM T7 DNA ligase were incubated for 15 min at room temperature in ligation buffer (see Materials and Methods) with the 26 kDa fragment as indicated. The panel shows the autoradiograph of electrophoresis products separated on a 15% denaturing polyacrylamide gel. Lanes 1–5 contain 0, 0.1, 1.0, 10 and 100 µM 26 kDa fragment, respectively. Lane 6 contains no intact T7 DNA ligase or 26 kDa fragment.

phoresis in the presence of SDS (24,29). We resorted to purifying the proteolytically produced 16 kDa fragment using this method and reconstituting the protein activity in essentially the same manner as described previously (24). In order to test the feasibility of this approach for T7 ligase, we used intact enzyme as a control. The proteins were separated by SDS–PAGE, eluted from the gel as described in Materials and Methods and tested for activity. We found that it was possible to re-nature the intact T7 ligase and restore ATPase and ligase activities (data not shown).

Activity of the 16 and 26 kDa fragments

Digestion of [α-^{32}P]ATP-labelled T7 ligase with clostripain revealed that the 16 kDa fragment contains the site of AMP attachment (Fig. 3), presumably at the lysine of the conserved active site motif (<u>K</u>YDGVR). To ascertain whether this N-terminal region had intrinsic adenylation activity, we incubated the renatured 16 kDa fragment with [α-^{32}P]ATP. The 16 kDa fragment could not be adenylated (data not shown). We also looked to see whether the adenylation reaction had become uncoupled from the ATPase activity in the 16 kDa fragment. No ATPase activity could be detected under our assay conditions. The lower level of detection in this system corresponds to a k_{cat} of 0.2/s.

The 26 kDa fragment had no DNA ligation activity even at very high concentrations of the protein (data not shown). However, a surprising phenomenon was observed when both T7 ligase and the 26 kDa fragment were added together in the DNA ligation assays. The 26 kDa fragment had an inherent ability to inhibit the ligation of nicked or cohesive-end DNA substrates catalysed by
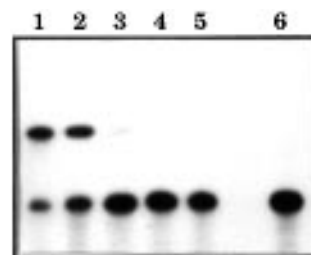
the pre-adenylated form of T7 ligase in the absence of ATP (Fig. 4). Similar results were obtained when T7 ligase was replaced with the T4 enzyme (data not shown). Furthermore, this inhibition was competitive in nature. At equimolar concentrations of the intact enzyme and 26 kDa fragment, ligation activity was completely abolished. In all cases, the DNA was pre-incubated with the fragment before addition of the intact enzyme. This type of inhibition can be explained in either of two ways. The first explanation is that the fragment has ATPase activity or a high affinity for ATP. This does not appear to be the case, since we were unable to demonstrate either binding or hydrolysis of [α-^{32}P]ATP (data not shown) and the C-terminal fragment actually inhibits the activity of pre-adenylated enzyme in the complete absence of ATP. An alternative, more feasible, explanation is that the 26 kDa fragment has a high affinity for the ligase binding site on the DNA substrate and can compete directly with the intact enzyme for this site.

In order to investigate the substrate specificity of T7 DNA ligase, DNA gel shift assays were performed with a number of short double-stranded DNA fragments (40–100 bp). The gel shift method is based on DNA–protein complexes having different electrophoretic mobilities compared to the free DNA. It can be used to separate the free DNA from the protein–DNA complexes, but complexes with different protein stoichiometries on the same DNA can also be separated from each other (30,31). Consequently, if the ligase were to bind to DNA at a specific site, gel shift analysis would show a single retarded band of DNA. However, if the protein were to bind to more than one site on the DNA, additional bands of retarded DNA should be observed. We show that T7 DNA ligase can shift short DNA fragments (Fig. 5) and this interaction appears to be specific, since only a single band shift is observed. The 26 kDa domain can also retard the mobility of these fragments in a similar manner (Fig. 5), suggesting that the C-terminal region of the enzyme is responsible for mediating this site-specific DNA binding. We observed that an ~10-fold higher concentration of the intact enzyme, compared to the 26 kDa fragment, was required to shift a comparable proportion of the DNA, suggesting that the fragment has a higher affinity for DNA. The simplest explanation for a single DNA band shift by both T7 ligase and the 26 kDa domain results from the binding of one protein molecule to a specific site on one molecule of duplex
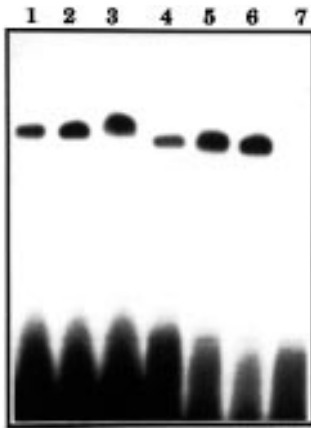
Figure 5. DNA gel shift assays. Autoradiogram of a 5% polyacrylamide gel showing the shifting of a labelled double-stranded 40mer DNA fragment by T7 ligase and the 26 kDa fragment. Lanes 1–3 contain 2, 4 and 6 µg 26 kDa fragment, respectively; lanes 4–6 contain 2, 4 and 6 µg intact T7 ligase; lane 7, 40mer alone.

DNA. However, a closer examination of the gels suggests that the situation may be a bit more complicated. The most obvious discrepancy is that the mobility of the 26 kDa:DNA band is less than the T7 ligase–DNA complex, which should be the reverse if a 1:1 complex exists between the protein and DNA. Whether this is due to multiple binding sites for proteins on the DNA or to the differing charge on the fragment compared to the full-length protein has not been analysed. Further investigation will be required to understand the mode of DNA binding of these two proteins.

DISCUSSION

T7 DNA ligase appears to represent a minimal size 'core' DNA ligase and is therefore an ideal model to study the essential features, both structural and catalytic, of this class of enzyme. Alignment of the amino acid sequences of eukaryotic ligases using a pairwise alignment method for the determination of sequence identities (4) reveals that these enzymes contain sequence homology with other ATP-dependent DNA ligases, but little similarity to the NAD⁺-dependent DNA ligases. Despite the different sizes of the eukaryotic ligases it is possible to align their sequences because of a high degree of conservation in the C-terminal region (Fig. 1). It has been noted that on alignment of homologus sequences of T4 and T7 DNA ligases, the C-termini coincided almost exactly, with the majority of the differences being located at the N-termini (32). It has also been shown that the distance between the presumed active site lysine and the C-terminus is very similar (~330 residues) across a diverse range of ATP-dependent ligases (4,5; Fig. 1) and the sequence analysis suggests that there is probably a high degree of structural homology between the ATP-dependent DNA ligases.

Based on the data we present here and complementary data from other studies on DNA ligases, we can propose a model for DNA binding which is consistent with the biological and catalytic roles of these enzymes. ATP-dependent DNA ligases consist of two functional regions, an N-terminal catalytic region and a C-terminal DNA binding region. The variable N-terminal region contains the active site motif (KYDGxR) (5) and has been proposed to be responsible for AMP attachment and transfer to

the target nick site on DNA. In the larger eukaryotic enzymes, such as mammalian I and II, *Drosophila* II and *S.cerevisiae* Cdc9, it has been demonstrated that the extreme N-terminal region of the enzyme, which is the least conserved, is not required for catalytic activity (Fig. 1; 3,33). It has been suggested that this region may play a role in nuclear translocation or be employed in specific protein–protein interactions during DNA replication and repair/recombination (3). Phosphorylation of the N-terminus of human DNA ligase I is required for activity of the catalytic domain, suggesting that this end region of the larger eukaryotic enzyme has important regulatory functions which are not required in the smaller enzymes. In T7 ligase, the N-terminus is much shorter, presumably because it does not require a modulation domain to regulate activity *in vivo*. Limited proteolysis of T7 ligase has provided data about the possible roles of different regions of the enzyme. Although ATPase activity could not be demonstrated for the 16 kDa N-terminal fragment, we have confirmed that this region is the site of covalent attachment of AMP, as shown previously for human ligase I (5). Mutagenesis studies of human ligase I showed that deletion of 23 amino acids at the C-terminus caused complete loss of enzyme activity and also prevented enzyme–AMP formation. Pre-incubation of T7 ligase with ATP prior to proteolysis greatly reduces cleavage of the protein into the two major cleavage fragments (16 and 26 kDa), suggesting that binding of the nucleotide causes a conformational change making the protein resistant to proteolysis. One explanation is that there is an association of the N- and C-terminal domains which could play a role in ATP binding and hydrolysis.

After completion of this work, we determined the crystal structure of T7 DNA ligase at 2.6 Å resolution (34), allowing us to compare the proteolysis results directly with the structure of the enzyme. The protein comprises two distinct domains (Fig. 6), a larger N-terminal Domain 1 (residues 2–240) which contains the ATP binding site and a smaller Domain 2 (residues 241–349). The two domains have a distinct groove running between them, which is proposed to be the DNA binding site. Although we expected the proteolytically sensitive site to be situated between the two domains, it was actually located in an exposed surface loop (Fig. 6). The accessibility of this loop region explains the hypersensitivity of this region to protease digestion. The 26 kDa fragment has lost 130 amino acids (16 kDa) at the N-terminus, which accounts for over half of Domain 1. This N-terminal region is involved in extensive contacts with the remainder of the protein, which may explain why the 16 kDa fragment was apparently unable to fold correctly on its own. The truncated protein (26 kDa) retains half of Domain 1, the cleft between the two domains and the whole of Domain 2. The crystal structure strongly suggests that DNA binds in the cleft between the two domains of the protein. Data from the limited proteolysis experiments described in this work revealed that removal of 130 amino acids at the N-terminus does not affect the ability of the truncated protein to bind to DNA, but destroys its catalytic activity. The loss of catalytic activity of the resulting 26 kDa fragment is presumably due to removal of the active site lysine residue (Lys34). However, the fragment retains the ability to bind to double-stranded DNA, with a binding affinity that is approximately one order of magnitude higher than the intact protein. It is also a very effective competitive inhibitor of intact T4 and T7 DNA ligases and again ~10-fold less of the fragment is sufficient to abolish the ligation activity. The apparent correlation of these data suggests that inhibition is probably due to a tighter binding of the fragment to ligation sites. The 26 kDa

Figure 6. Crystal structure of T7 DNA ligase. Ribbon diagram of T7 DNA ligase showing the domain structure of the enzyme. Domain 1 comprises the brown and yellow portions of the protein, while Domain 2 is coloured green. The 16 kDa proteolytic fragment is coloured brown and the 26 kDa fragment consists of the yellow and green parts of the structure. The protease-hypersensitive site (indicated by an arrow) is located in the extended loop between the brown and yellow parts of the protein. Part of this loop is disordered in the crystal structure and is shown as a short gap in the structure between the brown and yellow regions. The N- and C-termini of the protein are also indicated.

fragment retains the conserved cleft between the domains and hence these data provide further evidence that this cleft is the DNA binding site. This would also explain the ability of the 26 kDa fragment to act as a competitive inhibitor of the intact enzyme. We are over-expressing Domains 1 and 2 separately in order to characterize the biochemical properties of these regions of the enzyme.

Recently, Lindahl and co-workers have isolated a fourth human DNA ligase isoenzyme (35), which, in contrast with other ligases, contains an extended C-terminal region of 323 amino acids with no homology to any other protein (Fig. 1). Despite this difference, it retains the central core region, supporting the view that certain essential features are present in this portion of all ATP-dependent DNA ligases. Bacteriophage T7 and T3 have evolved the smallest DNA ligases and by discovering how these enzymes work we will begin to gain a better understanding of the larger eukaryotic enzymes.

ACKNOWLEDGEMENTS

REFERENCES

1 Lehman,I.R. (1974) *Science*, **186**, 790.
2 Engler,M.J. and Richardson,C.C. (1982) In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, New York, NY, Vol. XV, pp. 3–29.
3 Lindahl,T. and Barnes,D.E. (1992) *Annu. Rev. Biochem.*, **61**, 251–281.
4 Kletzin,A. (1992) *Nucleic Acids Res.*, **20**, 5389–539.
5 Tomkinson,A.E., Totty,N.F., Ginsburg,M. and Lindahl,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 400–404.
6 Heaphy,S., Singh,M. and Gait,M.J. (1987) *Biochemistry*, **18**, 2980–2984.
7 Becker,A., Glenda,L., Gefter,M. and Hurwitz,J. (1967) *Proc. Natl. Acad. Sci. USA*, **58**, 1996.
8 Masamune,Y., Frenkel,G.D. and Richardson,C.C. (1971) *J. Biol. Chem.*, **246**, 6874–6879.
9 Fareed,G.C., Wilt,E.M. and Richardson,C.C. (1971) *J. Biol. Chem.*, **246**, 925–932.
10 Hinkle,D.C. and Richardson,C.C. (1975) *J. Biol. Chem.*, **250**, 5523–5529.
11 Studier,F.W. (1972) *Science*, **176**, 367–376.
12 Dunn,J.J. and Studier,F.W. (1981) *J. Mol. Biol.*, **148**, 303–330.
13 Panasenko,S.M., Modrich,P. and Lehman,I.R. (1976) *J. Biol. Chem.*, **251**, 3432–3435.
14 Tomkinson,A.E., Lasko,D.D., Daly,G. and Lindahl,T. (1990) *J. Biol. Chem.*, **265**, 12611–12617.
15 Doherty,A.J., Ashford,S.R., Subramanya,H.S. and Wigley,D.B. (1996) *J. Biol. Chem.*, in press.
16 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
17 Bullock,W.O., Fernandez,J.M. and Short,J.M. (1987) *BioTechniques*, **5**, 376–379.
18 Dermody,J.J., Robinson,G.T. and Sternglanz,R. (1979) *J. Bacteriol.*, **139**, 701–704.
19 Wood.W.B. (1966) *J. Mol. Biol.*, **16**, 118–133.
20 Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
21 Laemmli,U.K (1970) *Nature*, **227**, 680–685.
22 Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
23 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
24 Hager,D.A. and Burgess,R.R. (1980) *Anal. Biochem.*, **109**, 76–86.
25 Lanzetta,P.A., Alvarez,L.J., Reinach,P.S. and Candia,O.A. (1979) *Anal. Biochem.*, **100**, 95–97.
26 Yang,S.-W. and Chan,Y.H. (1992) *J. Biol. Chem.*, **267**, 8117–8122.
27 Fried,M. and Crothers,D.M. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
28 Giles,A.-M., Imhoff,I.M. and Keil,B. (1979) *J. Biol. Chem.*, **254**, 1462.
29 Reece,R.J. and Maxwell,A. (1989) *J. Biol. Chem.*, **264**, 19648–19653.
30 Fried,M.G. (1989) *Electrophoresis*, **10**, 366–376.
31 Taylor,J., Badcoe,I.G., Clarke,A.R. and Halford,S.E. (1991) *Biochemistry*, **30**, 8743–8753.
32 Armstrong,J., Brown,R.S. and Suigata,A. (1983) *Nucleic Acids Res.*, **11**, 7145–7156.
33 Husain,I., Tomkinson,A.E., Burkhart,W.A., Moyer,M.B., Ramos,W., Mackey,Z.B., Besterman,J.M. and Chen,J. (1995) *J. Biol. Chem.*, **270**, 9683–9690.
34 Subramanya,H.S., Doherty,A.J., Ashford,S.R. and Wigley,D.B. (1996) *Cell*, in press.
35 Wei,Y.-F., Robins,P., Carter,K., Caldecott,K., Pappin,D.J.C., Yu,G.-L., Wang,R.-P., Shell,B.K., Nash,R.A., Schar,P., Barnes,D.E., Haseltine,W.A. and Lindahl,T. (1995) *Mol. Cell. Biol.*, **15**, 3206–3216.
36 Barnes,D.E., Johnston,L.H., Kodama,K., Tomkinson,A.E., Lasko,D.D. and Lindahl,T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6679–6683.
37 Wang,Y.C.J., Burkhart,W.A., Mackey,Z.B., Moyer,M.B., Ramos,W., Husain,I., Chen,J., Besterman,J.M. and Tomkinson,A.E. (1994) *J. Biol. Chem.*, **269**, 31923–31928.
38 Roberts,E., Nash,R.A., Robins,P. and Lindahl,T. (1994) *J. Biol. Chem.*, **269**, 3789–3792.
39 Rabin,B.A., Hawley,R.S. and Chase,J.W. (1986) *J. Biol. Chem.*, **261**, 10637–10645.
40 Takahashi,M. and Senshu,M. (1987) *FEBS Lett.*, **213**, 345–352.
41 Barker,D.G., White,J.H.M. and Johnson,L.H. (1987) *Eur. J. Biochem.*, **162**, 658–667.
42 Smith,G.L., Chan,Y.S. and Kerr,S.M. (1989) *Nucleic Acids Res.*, **17**, 9051–9062.
43 Barker,D.G., White,J.H.M. and Johnson,L.H. (1987) *Nucleic Acids Res.*, **13**, 8323–8337.
44 Tomkinson,A.E., Tappe,N.J. and Friedberg,E.C. (1992) *Biochemistry*, **31**, 11762–11771.