Nucleic Acids Research, 1997, Vol. 25, No. 7 **1369–1374**

Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*

Chonghui Cheng and Stewart Shuman*

Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10021 USA

Received December 10, 1996; Revised and Accepted February 10, 1997

ABSTRACT

We report that *Haemophilus influenzae* encodes a 268 amino acid ATP-dependent DNA ligase. The specificity of *Haemophilus* DNA ligase was investigated using recombinant protein produced in *Escherichia coli*. The enzyme catalyzed efficient strand joining on a singly nicked DNA in the presence of magnesium and ATP (K_m = 0.2 μM). Other nucleoside triphosphates or deoxynucleoside triphosphates could not substitute for ATP. *Haemophilus* ligase reacted with ATP in the absence of DNA substrate to form a covalent ligase–adenylate intermediate. This nucleotidyl transferase reaction required a divalent cation and was specific for ATP. The *Haemophilus* enzyme is the first example of an ATP-dependent DNA ligase encoded by a eubacterial genome. It is also the smallest member of the covalent nucleotidyl transferase superfamily, which includes the bacteriophage and eukaryotic ATP-dependent polynucleotide ligases and the GTP-dependent RNA capping enzymes.

INTRODUCTION

DNA ligases catalyze the joining of 5′ phosphate-terminated donor strands to 3′-hydroxyl-terminated acceptor strands via three sequential nucleotidyl transfer reactions (1,2). In the first step, nucleophilic attack by the ligase on ATP or NAD results in formation of a covalent intermediate in which AMP is linked to the ε amino group of a lysine. The nucleotide is then transferred to the 5′-end of the donor polynucleotide to form DNA–adenylate, an inverted (5′)-(5′) pyrophosphate bridge structure, AppN. Attack by the 3′-OH of the acceptor strand on the DNA–adenylate joins the two polynucleotides and liberates AMP.

The NAD-dependent ligases are found exclusively in eubacteria. Genes encoding NAD-dependent ligases have been identified in *Escherichia coli* (GenBank accession no. M30255), *Haemophilus influenzae* (U32789), *Thermus thermophilus* (M74792), *Zymomonas mobilis* (Z11910), *Rhodothermus marinus* (U10483), *Mycoplasma gentialium* (U39703) and *Synechocystis* (D90899). These proteins are of similar size (659–731 amino acids) and display extensive amino acid sequence conservation.

The ATP-dependent DNA ligases are found ubiquitously in eukaryotes and archaea (2,3). Animal cells contain multiple ATP-dependent DNA ligase isozymes encoded by at least three separate genes (4–6). ATP-dependent DNA ligases are also encoded by eukaryotic DNA viruses, e.g. the poxviruses (7), African swine fever virus (8) and *Chlorella* virus PBCV-1 (9). The only ATP-dependent DNA ligases described in eubacteria are those encoded by the T-odd and T-even bacteriophages (T4, T6, T3 and T7).

The ATP-dependent DNA ligases belong to a superfamily of covalent nucleotidyl transferases that includes the GTP-dependent eukaryotic mRNA capping enzymes (10). The ligase/capping enzyme superfamily is defined by a set of six short motifs (Fig. 1). The lysine within motif I (KxDGxR) is the active site of AMP transfer by the eukaryotic and bacteriophage ligases (11–13) and GMP transfer by the capping enzymes (14–17). Conserved residues within motifs I, III, IV and V are critical for covalent nucleotidyl transfer, as shown by mutational analyses (18–21). The recently reported crystal structure of T7 DNA ligase shows that the ATP binding site is made up of conserved motifs I, III, IIIa, IV and V (13).

The *Chlorella* virus PBCV-1 ligase, a 298 amino acid polypeptide, is the smallest ATP-dependent ligase described to date (9). Cellular DNA ligases are much larger; for example, human ligases I, III and IV are 919, 922 and 844 amino acid polypeptides respectively (5). The bacteriophage T7 and T3 enzymes (359 and 346 amino acid polypeptides respectively), the T-even phage enzymes (487 amino acids), the vaccinia virus and African swine fever virus ligases (552 and 419 amino acids respectively) and the achaebacterial ligases (542–600 amino acids) are of intermediate size (3). Sequence comparisons of cellular and virus-encoded proteins suggest that a catalytic domain common to all ATP-dependent ligases is embellished by additional isozyme-specific protein segments situated at their N- or C-terminus.

Here we report a new ATP-dependent DNA ligase encoded by the genome of a eubacterium, *H.influenzae*. The *H.influenzae* open reading frame encoding an ATP-dependent ligase-like protein was encountered by searching the GenBank database for gene products related to the DNA ligase of *Chlorella* virus PBCV-1. The predicted *H.influenzae* gene product includes the six motifs shared among the cellular and DNA virus-encoded ATP-dependent DNA ligases (Fig. 2). The order and spacing of these motifs in the *H.influenzae* ligase-like protein is similar to that seen in other ligase family members (Fig. 1). The *H.influenzae* polypeptide, at 268 amino acids, is smaller than any known ligase or capping enzyme. Potentially interesting questions about the evolution of the ligases are raised that can be addressed once the

*To whom correspondence should be addressed. Tel: +1 212 639 7145; Fax: +1 212 717 3623; Email: s-shuman@ski.mskcc.org

	I		III		IIIa		IV		V		VI
Hin	KLDGVR	-28-	FAIDGELF	-24-	KLYVFDVPDAEG	-47-	EGVVV	-14-	ILKLKTARGEE	-62-	PRFATYWREKK
ChV	KIDGIR *	-29-	EGSDGEIS *	-24-	SYYWFDYVTDDP	-55-	EGVMI *	-18-	LLKMKQFKDAE *	-92-	PVF-IGIRHEE
VAC	KYDGER	-41-	IVLDSEIV	-27-	CLFVFDCLYFDG	-52-	EGLVL	-13-	WLKIKRDYLNE	-120-	PRF-TRIREDK
SFV	KYDGER	-41-	FILDAELV	-28-	CLFVFDCLFYND	-52-	EGLML	-13-	WLKIKKDHLKT	-120-	PRC-TKIRDDK
FPV	KYDGER	-41-	MILDGEII	-28-	CIFIFDCLYFND	-52-	EGFVL	-13-	WLKIKKDYLDG	-119-	PRC-SRIREDK
Sc	KYDGER	-42-	LILDCEAV	-32-	CLFAFDILCYND	-52-	EGLMV	-18-	WLKLKKDYLEG	-118-	PRF-LRIREDK
Sp	KYDGER	-42-	FILDCEAV	-32-	CLFAFDILYLNG	-52-	EGLMV	-18-	WLKVKKDYLSG	-122-	PRF-IRIREDK
At	KYDGER	-42-	FILDCEVV	-32-	CIFAFDMLYLNG	-52-	EGLII	-17-	WLKLKKDYMDS	-120-	PRL-LRVREDK
Hu1	KYDGQR	-42-	FILDTEAV	-32-	CLYAFDLIYLNG	-52-	EGLMV	-17-	WLKLKKDYLDG	-120-	PRF-IRVREDK
Hu3	KYDGER	-41-	MILDSEVL	-28-	CLFVFDCIYFND	-52-	EGLVI	-13-	WLKVKKDYLNE	-121-	PRC-TRIRDDK
Hu4	KLDGER	-47-	CILDGEMM	-29-	CLCVFDVLMVNN	-52-	EGIMV	-14-	WLKIKPEYVSG	-121-	PRI-EKIRDDK
Xe	KYDGER	-42-	CILDTEAV	-32-	CVYAFDMLYLNG	-52-	EGLMV	-17-	WLKLKKDYLEG	-120-	PRF-LRIRDDK
Ce	KYDGER	-42-	FIVDAEVV	-30-	VVFLFDLLYFNG	-52-	EGLMI	-17-	WLKMKKDYVDG	-120-	PRF-LRIRDDK
Ca	KMDGDR	-47-	VILDGEMV	-40-	FFLVFDILFLNG	-52-	EGLVL	-16-	WIKVKPEYLEK	-113-	NHC-RKIREDK
ASF	KRNGVR	-41-	VYLDGELY	-22-	HFYVFDCFWSDQ	-51-	EGAIV	-18-	LAKLKPLLDAE		
Dam	KYDGER	-39-	FIVEGEII	-33-	NVFLFDLMYYEG	-51-	EGVMV	-17-	WIKFKRDYQSE	-117-	PRF-IRWRPDK
Mja	KYDGAR	-38-	LIVEGECV	-34-	RVYLFDILYKDG	-68-	EGVMI	-15-	MYKFKPTLESL	-95-	PRV-VRFRFDK
T4	KADGAR	-47-	VLIDGELV	-58-	KFQVWDYVPLVE	-54-	EGIIL	-14-	LYKFKEVIDVD	-85-	PI-AIRLREDK
T3	KYDGVR	-48-	FMLDGELM	-55-	SVRLYAVMPIHI	-60-	EGLIV	-14-	WWKLKPECEAD	-77-	PSF-EKFRGTE
T7	KYDGVR	-48-	FMLDGELM	-49-	HIKLYAILPLHI	-60-	EGLIV	-14-	WWKMKPENEAD	-96-	PSF-VMFRGTE

Figure 1. Conserved sequence elements that define a superfamily of covalent nucleotidyl transferases are conserved in *Haemophilus* ligase. Six co-linear sequence elements, designated motifs I, III, IIIa, IV, V and VI, are conserved in polynucleotide ligases and mRNA capping enzymes (10). The amino acid sequences are aligned for the DNA ligases encoded by *H.influenzae* (Hin) *Chlorella* virus PBCV-1 (ChV), vaccinia virus (VAC), Shope fibroma virus (SFV), fowlpox virus (FPV), *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Arabidopsis thaliana* (At), human ligase I (Hu1), human ligase 3 (Hu3), human ligase 4 (Hu4), *Xenopus laevis* (Xe), *Caenorrhabditis elegans* (Ce), *Candida albicans* (Ca), African swine fever virus (ASF), *Desulfurolobus ambivalens* (Dam), *Methanococcus jannaschii* (Mja) and bacteriophages T4 and T3 and T7. The number of amino acid residues separating the motifs is indicated by (-*n*-). Residues in the vaccinia virus DNA ligase that were found by mutational analysis to be essential for activity (21) are indicated by asterisks.

```
MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWN   50
GKQLLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEISSITKSFKG   100
DGWEKLKLYVFDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDK   150
THLYQFLAQVENLQGEGVVVRNPNAPYERKRSSQILKLKTARGEECTVIA   200
HHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNERENPPPIGSVITYKYR   250
GITNSGKPRFATYWREKK                                   268
```

Figure 2. Amino acid sequence of a putative *H.influenzae* DNA ligase. The sequence of the 268 amino acid polypeptide encoded by expression plasmid pET-Hin-ligase is shown. Conserved motifs I, III, IIIa, IV, V and VI are highlighted in boxes.

biochemical properties of the *H.influenzae* protein have been defined. This was accomplished by expressing the 268 amino acid polypeptide in *E.coli*. We found that the recombinant protein is indeed an ATP-dependent DNA ligase. A biochemical characterization of the *H.influenzae* ligase is presented.

MATERIALS AND METHODS

T7-based vector for expression of *H.influenzae* DNA ligase

Oligonucleotide primers complementary to the 5′- and 3′-ends of the putative *H.influenzae* ligase gene were used to amplify the 268 amino acid open reading frame. Plasmid p700 DNA (22; a generous gift of Dr Andrew Preston, University of Iowa) was used as the template for a polymerase chain reaction (PCR) catalyzed by Pfu DNA polymerase (Stratagene). The sequence of the 5′ flanking primer was 5′-GGGCCCCATATGAAATTTTAC-CGCACTTTG; the 3′ flanking primer was 5′-CAACCGCACT-TGGATCCACATACTGATGGA. These primers were designed to introduce *Nde*I and *Bam*HI restriction sites at the 5′- and 3′-ends of the ligase gene. The 0.9 kbp PCR product was digested with *Nde*I and *Bam*HI, then cloned into the *Nde*I and *Bam*HI sites of the T7-based expression plasmid pET3c to generate plasmid pET-Hin-ligase. The entire insert of pET-Hin-ligase was sequenced by the dideoxy chain termination method. The primary structure of the polypeptide encoded by pET-Hin-ligase is shown in Figure 2. pET-Hin-ligase was transformed into *E.coli* BL21.

Expression and purification of recombinant *H.influenzae* ligase

A 500 ml culture of *E.coli* BL21/pET-Hin-ligase was grown at 37°C in Luria–Bertani medium containing 0.1 mg/ml ampicillin until the A_{600} reached 0.8. The culture was infected with bacteriophage λCE6 as described (23) and incubation was continued at 37°C for 4 h. Cells were harvested by centrifugation and the pellet was stored at –80°C. All subsequent procedures were performed at 4°C. Thawed bacteria were resuspended in 50 ml lysis buffer [50 mM Tris–HCl, pH 7.5, 1 mM dithiothreitol (DTT), 10 mM EDTA, 10% sucrose] containing 0.15 M NaCl. Lysozyme and Triton X-100 were added to final concentrations of 60 μg/ml and 0.1% respectively and the sample was sonicated for 30 s. Insoluble material was removed by centrifugation for 30 min at 18 000 r.p.m. in a Sorvall SS34 rotor. SDS–PAGE analysis indicated that the expressed *Haemophilus* ligase protein was recovered in the pellet, not in the soluble lysate. The insoluble material was resuspended with a Dounce homogenizer in 50 ml lysis buffer. The centrifugation step was repeated and the pelleted material was resuspended in 10 ml lysis buffer containing 0.5 M NaCl. This suspension was centrifuged again. The recombinant *Haemophilus* ligase was partially soluble in 0.5 M NaCl. The 0.5 M NaCl pellet, which still contained substantial *Haemophilus* ligase, was resuspended in 10 ml lysis buffer containing 1 M NaCl. Additional *Haemophilus* ligase activity was recovered in the 1 M NaCl supernatant fraction after centrifugation. The 1 M NaCl pellet was then extracted with 10 ml lysis buffer containing 2 M NaCl. The 2 M NaCl pellet was resuspended in 5 ml lysis buffer. The 0.5, 1 and 2 M NaCl supernatants were combined

(30 ml, containing 4 mg protein) and dialyzed against 0.1 M NaCl in buffer A (50 mM Tris–HCl, pH 8.0, 1 mM EDTA, 2.5 mM DTT, 10% glycerol, 0.1% Triton X-100). This material was applied to a 1.5 ml column of phosphocellulose that had been equilibrated with 0.1 M NaCl in buffer A. The column was washed with the same buffer and then eluted step-wise with buffer A containing 0.2, 0.25, 0.5 and 1.0 M NaCl. The activity of the column fractions was monitored by label transfer from [α-^{32}P]ATP to protein. *Haemophilus* ligase adenylyltransferase activity was recovered in the 0.5 M fraction (0.6 mg protein). An aliquot of the 0.5 M phosphocellulose fraction was applied to a 4.8 ml 15–30% glycerol gradient containing 50 mM Tris–HCl, pH 8.0, 2 mM DTT, 0.5 M NaCl and 0.1% Triton X-100. The gradient was centrifuged at 50 000 r.p.m. for 24 h at 4°C in a Beckman SW50 rotor. Fractions were collected from the bottom of the tube. Marker proteins (bovine serum albumin and cytochrome c) were sedimented in a parallel gradient. Enzyme fractions were stored at –80°C and thawed on ice just prior to use. Protein concentrations were determined using the BioRad dye binding assay, with bovine serum albumin as the standard.

Enzyme–AMP complex formation

Reaction mixtures (10 µl) containing 60 mM Tris–HCl, pH 8.0, 0.5 mM DTT, 10 mM $MgCl_2$, 0.16 µM [α-^{32}P]ATP and enzyme were incubated for 10 min at 22°C, then halted by adding SDS to 1% final concentration. The samples were electrophoresed through a 12% polyacrylamide gel containing 0.1% SDS. Label transfer to the 31 kDa *Haemophilus* ligase polypeptide was visualized by autoradiographic exposure of the dried gel and was quantitated by scanning the gel with a Fujix BAS1000 Bio-Imaging Analyzer.

Ligase substrate

The standard substrate used in ligase assays was a 36 bp DNA duplex containing a centrally placed nick (24). This DNA was formed by annealing two 18mer oligonucleotides to a complementary 36mer strand. The 18mer constituting the donor strand was 5′-^{32}P-labeled and gel purified as described (24). The labeled donor was annealed to the complementary 36mer in the presence of a 3′-OH-terminated acceptor strand in 0.2 M NaCl by heating at 70°C for 10 min, followed by slow cooling to room temperature. The molar ratio of the 18mer donor to 36mer complement to 18mer acceptor strands in the hybridization mixture was 1:4:4.

DNA ligation

Reaction mixtures (20 µl) containing 50 mM Tris–HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$, 1 mM ATP, 0.2 pmol 5′-^{32}P-labeled DNA substrate and enzyme were incubated at 22°C. Reactions were initiated by addition of enzyme and halted by the addition of 1 µl 0.5 M EDTA and 5 µl formamide. The samples were heated at 95°C for 5 min and then electrophoresed through a 15% polyacrylamide gel containing 7 M urea in TBE (90 mM Tris-borate, 2 mM EDTA). The labeled 36mer ligation product was well resolved from the 5′-labeled 18mer donor strand. The extent of ligation [36mer/(18mer + 36mer)] was determined by scanning the gel using a Fujix BAS1000 phosphorimager.

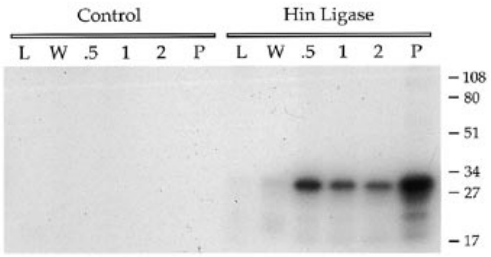

Figure 3. Adenylyltransferase activity of recombinant *Haemophilus* ligase. Whole cell lysates were prepared from λCE6-infected BL21 cells bearing either the pET vector (Control) or the pET-Hin-ligase expression plasmid (Hin Ligase) as described in Materials and Methods. Adenylyltransferase reaction mixtures (10 µl) contained 60 mM Tris–HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$, 0.16 µM [α-^{32}P]ATP and 1 µl of the protein fraction indicated: soluble bacterial lysate (lane L); lysis buffer wash fraction (lane W); 0.5 M NaCl extract of the insoluble pellet (lane 5); 1 M NaCl extract (lane 1); 2 M NaCl extract (lane 2); resuspended 2 M NaCl pellet (lane P). Incubation was for 10 min at 22°C. The reaction products were resolved by SDS–PAGE. An autoradiograph of the dried gel is shown. The positions and sizes (in kDa) of prestained marker polypeptides are indicated on the right.

RESULTS

Expression of the *H.influenzae* ligase-like protein in bacteria

The *H.influenzae* open reading frame URF1 encoding a ligase-like protein was amplified by PCR and cloned into a T7 RNA polymerase-based bacterial expression vector. The pET-Hin-ligase plasmid was introduced into *E.coli* BL21. Expression of the target gene was induced by infection with bacteriophage λCE6. This phage contains the gene encoding T7 RNA polymerase. A prominent 31 kDa polypeptide was detectable by SDS–PAGE in whole cell extracts of infected bacteria (not shown). This polypeptide was not present when bacteria containing the pET vector alone were infected with λCE6. After centrifugal separation of the crude lysate, the 31 kDa protein was recovered in the insoluble pellet fraction. This protein was partially solubilized by sequential extraction of the pellet fraction with buffers containing 0.5, 1 and 2 M NaCl.

Recombinant *H.influenzae* protein forms a covalent enzyme–adenylate complex

The initial step in DNA ligation involves formation of a covalent enzyme–adenylate intermediate, EpA. The formation of EpA by ATP-dependent DNA ligases can be detected with high sensitivity and specificity by label transfer from [α-^{32}P]ATP to the enzyme. In order to assay adenylyltransferase activity of the expressed *H.influenzae* protein, we incubated the soluble lysate and the salt-extracted material from the insoluble fraction of λCE6-infected BL21/pET-Hin-ligase cells in the presence of [α-^{32}P]ATP and a divalent cation. The salt extracts formed an SDS-stable nucleotide–protein adduct that migrated as a 31 kDa species during SDS–PAGE (Fig. 3, Hin Ligase). No labeling of this protein was detected when the enzyme fraction was incubated with [α-^{32}P]GTP (not shown). Labeling of the 31 kDa polypeptide was not detected using salt extracts prepared from λCE6-infected bacteria that lacked the *Haemophilus* ligase plasmid (Fig. 3, Control).

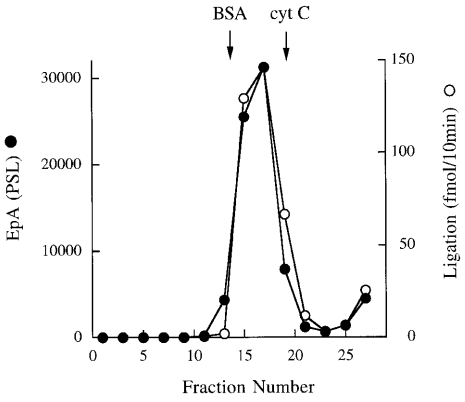

Figure 4. Glycerol gradient sedimentation. The phosphocellulose ligase preparation was sedimented in a 15–30% glycerol gradient as described in Materials and Methods. Aliquots of the glycerol gradient fractions were assayed for enzyme–adenylate formation (filled circles) and DNA strand joining activities (open circles). Adenylyltransferase reaction mixtures contained 0.16 μM [α-^{32}P]ATP and 1 μl of the indicated fractions. Incubation was for 10 min at 22°C. Adenylyltransferase activity was gauged by the signal intensity of the radiolabeled 31 kDa ligase polypeptide (PSL; photostimulatable luminescence). DNA ligation reactions are described in Figure 5. Marker proteins [bovine serum albumin (BSA) and cytochrome c (cyt C)] were centrifuged in a parallel gradient; their sedimentation peaks are indicated by arrowheads.

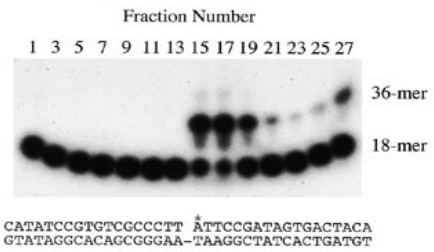

Figure 5. DNA ligation. The structure of the nicked duplex substrate is shown with the position of the 5′-labeled nucleotide indicated by the asterisk. The DNA ligation reaction mixtures contained 50 mM Tris–HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$, 1 mM ATP, 200 fmol nicked duplex DNA and 0.5 μl of the indicated gradient fractions. Incubation was for 10 min at 22°C. The reaction products were resolved by denaturing polyacrylamide gel electrophoresis. An autoradiogram of the gel is shown.

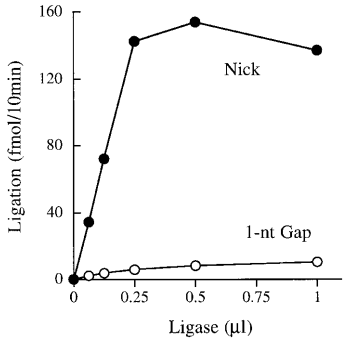

Figure 6. DNA substrate specificity. Reaction mixtures (20 μl) containing 50 mM Tris–HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$, 1 mM ATP, 200 fmol nicked DNA substrate or 1 nt gap DNA substrate and the indicated amounts of *Haemophilus* ligase (μl of glycerol gradient fraction) were incubated for 10 min at 22°C. The extent of ligation (fmol ends joined) is plotted as a function of input ligase.

The adenylyltransferase remained soluble after dialysis of the NaCl extract against buffer containing 0.1 M NaCl. The enzyme was applied to phosphocellulose and was recovered by step elution with 0.5 M NaCl. When the phosphocellulose preparation fraction was centrifuged through a 15–30% glycerol gradient in 0.5 M NaCl, a single peak of adenylyltransferase activity was detected (Fig. 4). We estimated a sedimentation coefficient of 3 S relative to marker proteins sedimented in a parallel gradient. This suggested that the *H.influenzae* adenylyltransferase is a monomer of the 31 kDa polypeptide.

DNA ligation

We assayed the ability of the recombinant *H.influenzae* protein to seal a 36mer synthetic duplex DNA substrate containing a single nick (24). Ligase activity was evinced by conversion of the 5′-^{32}P-labeled 18mer donor strand into an internally labeled 36mer product (Fig. 5). The DNA ligase activity profile across the glycerol gradient paralleled that of enzyme–adenylate formation (Figs 4 and 5). These results demonstrate that the *H.influenzae* protein is indeed a DNA ligase. Further characterization of the ligase was performed using the glycerol gradient preparation (peak fraction 17).

The extent of ligation of the nicked duplex during a 10 min incubation at 22°C in the presence of 1 mM ATP increased linearly with added enzyme (Fig. 6). The reaction saturated with 77% of the labeled donor strand converted to 36mer in 10 min. This upper limit of ligation probably reflected incomplete annealing of all three component strands to form the nicked substrate. Ligation by 0.25 μl enzyme in the presence of 1 mM ATP and 10 mM $MgCl_2$ was linear with time up to 10 min (not shown). Ligation depended on a divalent cation in excess of the input 1 mM ATP; activity was enhanced as Mg was increased from 2 to 20 mM (Fig. 7A). The divalent cation requirement was satisfied by 10 mM Mn, but not by 10 mM Co, Ca, Cu or Zn (Fig. 7B).

The rate of ligation was dependent on the concentration of ATP included in the reaction (Fig. 8A). A K_m of 0.2 μM ATP was calculated from a double-reciprocal plot of the data. Other rNTPs at 100 μM concentration could not substitute for ATP. dATP was also incapable of supporting strand joining (Fig. 8B).

DNA substrate specificity

The structure of the ligation substrate was altered such that the 3′-hydroxyl-terminated acceptor strand was separated from the 5′-phosphate donor terminus by a 1 nt gap (24). The strand joining activity of *H.influenzae* ligase on a 1 nt gap substrate was 4% of the activity of a nicked duplex DNA (Fig. 6). The implication is that the 3′-OH must be positioned fairly precisely relative to the 5′-phosphate donor terminus for ligation to occur.

Analysis of enzyme–AMP formation

The *H.influenzae* ligase reacted specifically with [α-^{32}P]ATP. The amount of enzyme–AMP complex formed during a 10 min

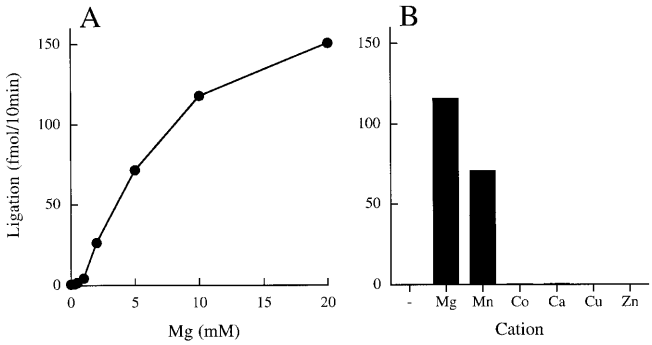

Figure 7. Magnesium concentration dependence and divalent cation specificity of strand joining. (**A**) Magnesium dependence. Reaction mixtures (20 µl) containing 50 mM Tris–HCl, pH 8.0, 5 mM DTT, 1 mM ATP, 200 fmol nicked DNA, 0.25 µl ligase and $MgCl_2$ as indicated were incubated for 10 min at 22°C. The extent of ligation (fmol) is plotted as a function of magnesium concentration. (**B**) Divalent cation specificity. Reaction mixtures (20 µl) containing 50 mM Tris–HCl, pH 8.0, 1 mM ATP, 200 fmol nicked DNA, 0.25 µl ligase and 10 mM of the indicated divalent cation were incubated for 10 min at 22°C. Divalent cation was omitted from a control reaction (–). Mg, Mn, Ca and Co were added as the chloride salt, Cu and Zn as the sulfate salt.

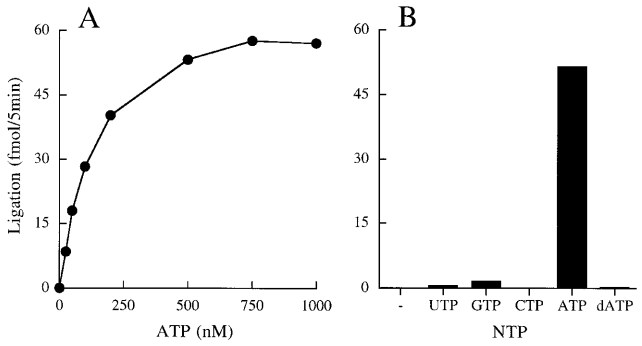

Figure 8. ATP concentration dependence and nucleotide specificity of strand joining. (**A**) ATP dependence. Reaction mixtures (20 µl) containing 50 mM Tris–HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$, 200 fmol nicked DNA, 0.25 µl ligase and ATP as indicated were incubated for 5 min at 22°C. The extent of ligation (fmol) is plotted as a function of ATP concentration. (**B**) Nucleotide specificity. Reaction mixtures (20 µl) containing 50 mM Tris–HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$, 200 fmol nicked DNA, 0.25 µl ligase and 100 µM NTP or dNTP as indicated were incubated for 5 min at 22°C. Nucleotide was omitted from a control reaction (–).

incubation at 22°C in the presence of 0.16 µM [α-^{32}P]ATP and 10 mM $MgCl_2$ was proportional to the amount of added enzyme with 0.063–2 µl of the glycerol gradient preparation (data not shown). The extent of EpA formation by 0.25 µl enzyme did not vary as a function of [α-^{32}P]ATP concentration from 25 to 2000 nM (data not shown). The concentration of available adenylation sites in the glycerol gradient enzyme preparation was 64 nM. This value is a minimal estimate of the concentration of active ligase molecules, as it does not take into account any ligase molecules that are already in the AMP-bound state. Using this value, we calculated from the data in Figure 6 that 10 fmol DNA substrate were ligated per fmol input ligase (as EpA units).

EpA formation depended on a divalent cation cofactor. This requirement was satisfied by either 5 mM magnesium or 5 mM manganese and to a lesser extent by 5 mM cobalt (Fig. 9).

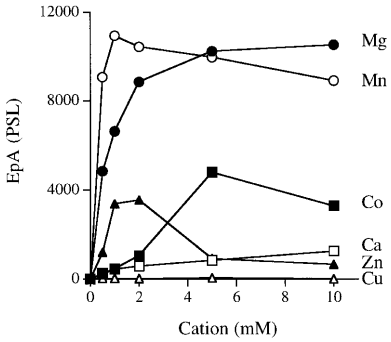

Figure 9. Divalent cation dependence and specificity of enzyme–AMP formation. Reaction mixtures (10 µl) containing 60 mM Tris–HCl, pH 8.0, 0.16 µM [α-^{32}P]ATP, 0.25 µl ligase and the divalent cation as indicated were incubated for 10 min at 22°C. Mg, Mn, Ca and Co were added as the chloride salt, Cu and Zn as the sulfate salt. The yield of labeled 31 kDa enzyme–adenylate complex (PSL) is plotted as a function of divalent cation concentration.

Manganese was a more effective cofactor than magnesium at 0.5–1 mM. Zinc supported EpA formation in a narrow concentration range from 1 to 2 mM, but activity decreased sharply at 5–10 mM. Calcium was a poor effector at all concentrations tested between 1 and 10 mM; the yield of EpA at 10 mM calcium was ~10% of the optimal value formed with magnesium or manganese. Copper was inactive at all concentrations examined (Fig. 9).

DISCUSSION

A *H.influenzae* gene encoding a putative ATP-dependent DNA ligase was identified on the basis of sequence similarity to members of the ligase/capping enzyme superfamily. We show that the 268 amino acid gene product is an ATP-dependent DNA ligase. This was achieved by expressing the *H.influenzae* protein in bacteria and characterizing its enzymatic properties. *Haemophilus influenzae* ligase, like other polynucleotide ligases, catalyzes strand joining via an enzyme–AMP intermediate. The *H.influenzae* enzyme displays strict specificity for ATP as the nucleotide cofactor. dATP is inactive. *Haemophilus* ligase resembles the T4, vaccinia virus, *Chlorella* virus and eukaryotic cellular enzymes in its discrimination of the NTP sugar moiety (9,21,24,25). The observed K_m of *Haemophilus* DNA ligase for ATP (0.2 µM) is comparable with values reported for mammalian DNA ligase I (0.5–1 µM) (2). The reported K_m values of *Chlorella* virus ligase (75 µM), vaccinia virus ligase (95 µM) and mammalian ligase II (40 µM) are significantly higher (9,24,26). The high efficiency of *H.influenzae* DNA ligase in strand joining across a nick in duplex DNA contrasts sharply with the low efficiency of ligation across a 1 nt gap. Vaccinia ligase, *Chlorella* virus ligase and yeast CDC9 ligase display similar properties (9,24,27).

The *H.influenzae* ligase is the smallest DNA ligase described to date. Insofar as *H.influenzae* ligase is also smaller than any known mRNA capping enzyme (18,28), it may constitute the catalytic core of the nucleotidyl transferase superfamily. *Haemophilus influenzae* ligase includes the six conserved motifs that define the family (10), but contains no additional amino acids at the C-terminus downstream of motif VI. It contains only 40 amino acids N-terminal of the presumptive active site, Lys41.

Note that the active site of the second smallest ligase, the 298 amino acid *Chlorella* virus enzyme, is at residue Lys27. The compaction of the *H.influenzae* ligase relative to the *Chlorella* virus protein is achieved by shortening the spacing between motif V and motif VI. This intervening segment is 62 amino acids in *H.influenzae* ligase versus 92 residues for *Chlorella* virus ligase. The biochemical function of conserved motif VI is not clear at present. It is surmised from the crystal structure of the T7 DNA ligase that motif VI is not a component of the ATP binding site (13). Conceivably, motif VI may be involved in binding of ligase to the nucleic acid substrate.

Haemophilus influenzae ligase is the first example of an ATP-dependent DNA ligase encoded by a eubacterial genome. The *H.influenzae* gene encoding the ATP-dependent ligase resides within a 9.4 kbp *Eco*RI genomic restriction fragment that was cloned and sequenced by Preston *et al.* (22). Eleven open reading frames are contained within this DNA fragment. Ten of these are homologous to previously identified bacterial genes. The one reading frame that had no apparent eubacterial homolog is the one that we have shown encodes the DNA ligase. This gene was previously named URF1 (unknown reading frame 1) (22). Preston and Moxon (unpublished results) have documented the presence and integrity of the full-length URF1 open reading frame in four different strains of *H.influenzae*: these include the pathogenic b serotype strains RM7004 and Eagan, the non-pathogenic Rd strain and the KW20b strain (a multiply passaged derivative of Rd that was transformed to capsule type b with DNA from strain Eagan). We propose that the *H.influenzae* gene encoding the ATP-dependent ligase should henceforth be named *ligA* and further suggest that the *H.influenzae* gene encoding a homolog of *E.coli* NAD-dependent DNA ligase be named *ligN*.

Previous efforts to disrupt the *H.influenzae ligA* gene were unsuccessful, i.e. no viable bacteria containing a *ligA* deletion could be recovered (22). This cannot be attributed to a locus-specific failure to engage in genetic recombination, because Preston *et al.* (22) were able to disrupt the gene immediately flanking *ligA*. A reasonable inference from these observations is that the *ligA* gene product is essential for growth of *H.influenzae*. A role for the ATP-dependent DNA ligase during DNA replication, repair and/or recombination is plausible; delineating these roles will require the isolation of temperature-sensitive *ligA* mutants.

The occurrence of a eukaryotic-type ligase in a eubacterial genome raises interesting evolutionary questions. *Haemophilus influenzae* is an obligate parasite that colonizes human respiratory mucosa and grows in culture only on rich medium. Its 1.8 Mb genome encodes many fewer genes than does the 4.7 Mb *E.coli* genome (29). The fastidious growth requirements stem from the lack of key enzymes involved in the metabolism of carbohydrates, pyrimidines, amino acids and coenzymes (30). Thus, *H.influenzae* depends on the human host to provide various nutrients. The parasitic relationship may also entail rare occurrences of horizontal gene transfer from the eukaryotic host to the bacterium. This was suggested by the finding that two *H.influenzae* proteins are homologous to eukaryotic amino acid transporters, but unrelated to any bacterial transporters (30). This raises the possibility that the *ligA* gene may have been acquired by *Haemophilus* from a eukaryote or a eukaryotic virus.

REFERENCES

1 Lehman,I.R. (1974) *Science*, **186**, 790–797.
2 Lindahl,T. and Barnes,D.E. (1992) *Annu. Rev. Biochem.*, **61**, 251–281.
3 Kletzin,A. (1992) *Nucleic Acids Res.*, **20**, 5389–5396.
4 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.
5 Wei,Y., Robins,P., Carter,K., Caldecott,K., Pappin,D., Yu,G., Wang,R., Shell,B., Nash,R., Schar,P., Barnes,D., Haseltine,W. and Lindahl,T. (1995) *Mol. Cell. Biol.*, **15**, 3206–3216.
6 Chen,J., Tomkinson,A.E., Ramos,W., Mackey,Z.B., Danehower,S., Walter,C.A., Schultz,R.A., Besterman,J.M. and Husain,I. (1995) *Mol. Cell. Biol.*, **15**, 5412–5422.
7 Smith,G.L., Chan,Y.S. and Kerr,S.M. (1989) *Nucleic Acids. Res.*, **17**, 9051–9062.
8 Hammond,J.M., Kerr,S.M., Smith,G.L. and Dixon,L.K. (1992) *Nucleic Acids Res.*, **20**, 2667–2671.
9 Ho,C.K., Van Etten,J.L. and Shuman,S. (1997) *J. Virol.*, **71**, 1931–1937.
10 Shuman,S. and Schwer,B. (1995) *Mol. Microbiol.*, **17**, 405–410.
11 Tomkinson,A.E., Totty,N.F., Ginsberg,M. and Lindahl,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 400–404.
12 Wang,Y.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.
13 Subramanya,H.S., Doherty,A.J., Ashford,S.R. and Wigley,D.B. (1996) *Cell*, **85**, 607–615.
14 Cong,P. and Shuman,S. (1993) *J. Biol. Chem.*, **268**, 7256–7260.
15 Niles,E.G. and Christen,L. (1993) *J. Biol. Chem.*, **268**, 24986–24989.
16 Schwer,B. and Shuman,S. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4328–4332.
17 Fresco,L.D. and Buratowski,S. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 6624–6628.
18 Shuman,S., Liu,Y. and Schwer,B. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 12046–12050.
19 Cong.P. and Shuman,S. (1995) *Mol. Cell. Biol.*, **15**, 6222–6231.
20 Kodama,K., Barnes,D.E. and Lindahl,T. (1991) *Nucleic Acids Res.*, **19**, 6093–6099.
21 Shuman,S. and Ru.X. (1995) *Virology*, **211**, 73–83.
22 Preston,A., Maskell,D., Johnson,A and Moxon,E.R. (1996) *J. Bacteriol.*, **178**, 396–402.
23 Shuman,S., Golder,M. and Moss,B. (1988) *J. Biol. Chem.*, **263**, 16401–16407.
24 Shuman,S. (1995) *Biochemistry*, **34**, 16138–16147.
25 Montecucco,A., Lestingi,M., Pedrali-Noy,G., Spadari,S. and Ciarrocchi,G. (1990) *Biochem. J.*, **271**, 265–268.
26 Teraoka,H., Sumikawa,T. and Tsukuda,K. (1986) *J. Biol. Chem.*, **261**, 6888–6892.
27 Tomkinson,A.E., Tappe,N.J. and Friedberg,E.C. (1992) *Biochemistry*, **31**, 11762–11771.
28 Ho,C.K., Van Etten,J.L. and Shuman,S. (1996) *J. Virol.*, **70**, 6658–6664.
29 Fleischmann,R.D. *et al.* (1995) *Science*, **269**, 496–512.
30 Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V. (1996) *Curr. Biol.*, **6**, 279–291.