Nucleic Acids Research, Vol. 19, No. 22 6093–6099

In vitro mutagenesis and functional expression in *Escherichia coli* of a cDNA encoding the catalytic domain of human DNA ligase I

Ken-ichi Kodama, Deborah E.Barnes and Tomas Lindahl*
Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK

Received September 23, 1991; Accepted October 24, 1991

ABSTRACT

Human cDNAs encoding fragments of DNA ligase I, the major replicative DNA ligase in mammalian cells, have been expressed as *lacZ* fusion proteins in *Escherichia coli*. A cDNA encoding the carboxyl-terminal catalytic domain of human DNA ligase I was able to complement a conditional-lethal DNA ligase mutation in *E.coli* as measured by growth of the mutant strain at the non-permissive temperature. Targeted deletions of the amino and carboxyl termini of the catalytic domain identified a minimum size necessary for catalytic function and a maximum size for optimal complementing activity in *E.coli*. The human cDNA was subjected to systematic site-directed mutagenesis *in vitro* and mutant polypeptides assayed for functional expression in the *E.coli* DNA ligase mutant. Such functional analysis of the active site of DNA ligase I identified specific residues required for the formation of an enzyme-adenylate reaction intermediate.

INTRODUCTION

DNA ligases catalyse the formation of phosphodiester bonds at single strand breaks in double helical DNA and are required in DNA replication, repair and recombination. DNA ligase I is the major DNA ligase activity in mammalian cells and is induced upon cell proliferation (1). This enzyme is apparently co-localised with other replication enzymes in the cell nucleus (2, 3). Human cDNAs encoding DNA ligase I have been isolated and expressed in a conditional-lethal *cdc9* DNA ligase mutant of the yeast, *Saccharomyces cerevisiae* (4). Functional complementation of the replication defective *cdc9* mutant confirms that DNA ligase I can act in DNA replication and, as the CDC9 gene product is required for joining of Okazaki fragments in *S.cerevisiae* (5), indicates a similar role for the human enzyme.

In the first step of the ligation reaction, DNA ligase interacts with a nucleotide derivative, ATP in the case of the virus-encoded and eukaryotic DNA ligases or NAD^+ for the bacterial enzymes, to form a covalent enzyme-adenylate reaction intermediate with the release of pyrophosphate or NMN (6). The AMP moiety is then transferred to the 5′-terminus of the DNA strand break prior to closure of the strand interruption and release of AMP. The AMP residue of the first reaction intermediate is covalently bound via a phosphoramidate bond to the ϵ-amino group of a specific lysine residue in the *E.coli*, bacteriophage T4 and mammalian enzymes (7,8). An active site peptide containing the reactive lysine residue of mammalian DNA ligase I has been isolated and sequenced (9).

Comparison of the active site peptide with the predicted coding sequence of human DNA ligase I and other ATP-dependent DNA ligases identified a conserved active site motif, Lys-Tyr/Ala-Asp-Gly-X-Arg, where the first residue is the reactive lysine and the residue at position 'X' is not highly conserved (9). Furthermore, the distance between the postulated adenylylation site and the carboxyl terminus of the polypeptide is very similar in the ATP-dependent DNA ligases. This region shows significant sequence homology between the *S.cerevisiae* and *Schizosaccharomyces pombe* enzymes, and human DNA ligase I (4, 10), and presumably defines essential enzyme functions. The NAD-dependent *E.coli* enzyme contains a degenerate but recognisable active site motif and has an extended carboxyl-terminal region. The size of the amino-terminal region of the different DNA ligases is highly variable and of unknown function, although it may be involved in interactions with chromatin or with other replication proteins, as well as being a possible target for post-translational modification of the enzyme by, for example, phosphorylation or ADP-ribosylation.

Heterologous functional expression of ATP-dependent DNA ligases within eukaryotic systems has been described (4, 11). In addition, the bacteriophage T4 ATP-dependent DNA ligase was able to substitute for the NAD-dependent enzyme of *Salmonella typhimurium in vivo* (12). Here we describe functional expression of the catalytic domain of the ATP-dependent human DNA ligase I in a conditional-lethal NAD-dependent DNA ligase (*lig*) mutant of *E.coli*. We have employed a simple system in which the same vector is used during site-specific *in vitro* mutagenesis of the human DNA ligase I cDNA and subsequent functional expression in the *E.coli lig* mutant host. Targeted deletion of the human cDNA produced a polypeptide with optimal activity in *E.coli* and

* To whom correspondence should be addressed

defined a minimum domain necessary for catalytic function. The resultant polypeptide was subjected to systematic site-specific mutagenesis and functional analysis of the active site region.

MATERIALS AND METHODS

Cloning vectors and bacterial strains

Human DNA ligase I cDNA sequences were propagated in the phagemid vector pBluescript SK$^-$ (Stratagene).

The following strains of *E.coli* were used as hosts for pBluescript constructs: DH5α (*sup*E44,*hsd*R17,*rec*A1,*end*A1, *gyr*A96,*thi*-1,*rel*A1, Δ*lac*U169[φ80*lacZ*ΔM15]) (13) for phagemid propagation and site-directed mutagenesis of uracil-containing DNA; XL1-Blue (*sup*E44,*hsd*R17,*rec*A1,*end*A1,*gyr*A96,*thi*-1,*rel*A1,[F′ *pro*AB, *lac*I^{q}ZΔM15,Tn10(*tet*^{R})]) (14) for the preparation of single-stranded DNA for sequencing; CJ236 (*dut*-1,*ung*-1,*thi*-1,*rel*A1/pCL105 [*cam*^{R} F′]) (15) for the preparation of uracil-containing template DNA for site-directed mutagenesis; GR501 (*Hfr*,*thi*-1,*pts*I$^+$,*lig*-251) (16) as the host strain for expression of human DNA ligase I/pBluescript constructs. GR501 is a conditional-lethal DNA ligase (*lig*) mutant, originally isolated as a temperature-sensitive DNA replication mutant, with no detectable DNA ligase activity at the non-permissive temperature of 40°C (16).

Human DNA ligase I constructs

The full-length human DNA ligase I cDNA was subcloned from a yeast expression vector, which allowed functional expression of the human cDNA in an *S.cerevisiae cdc9* conditional-lethal DNA ligase mutant (4), and placed under control of the *lac*

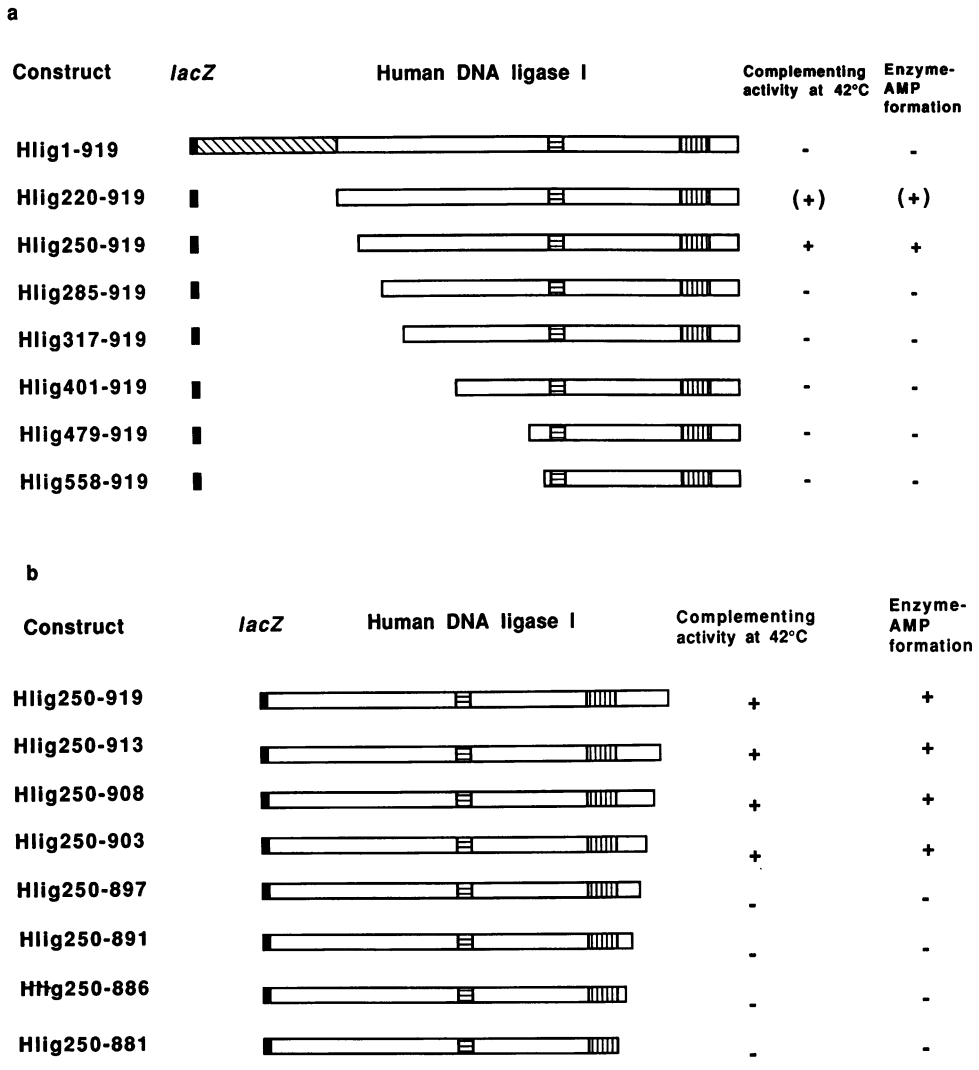

Figure 1. Schematic representation of human DNA ligase I cDNAs expressed in the *E.coli lig* mutant. The encoded residues of the intact DNA ligase I protein are given and the resultant polypeptide shown schematically for each construct by an open box. The black box represents the amino-terminal 23 amino acids of β-galactosidase encoded by the *E.coli lacZ* gene; the diagonally cross-hatched box represents the amino-terminal region of human DNA ligase I which is not required for catalytic activity *in vitro* or for complementation of an *S.cerevisiae cdc9* DNA ligase mutant (4, 20); the horizontally-hatched box represents the active site of DNA ligase I; the vertically-hatched box represents a peptide conserved in human DNA ligase I, and the yeast and vaccinia virus DNA ligases (20). The ability of the encoded polypeptide to complement the replication defect of the *E.coli lig* mutant leading to growth of the transformed host at 42°C, and to form an enzyme-adenylate complex in *E.coli* cell extracts (see figure 3), is indicated for each construct by a '+'. A negative result is denoted by a '−'; '(+)' denotes slow growth to microcolonies and trace quantities of enzyme-adenylate formed. **a.** Expression of the full-length protein (Hlig1−919) and amino-truncated forms. **b.** Expression of an active amino-terminal truncated protein (Hlig250−919) and carboxyl-terminal truncated forms.

promoter by insertion at the *Sal* I site of the pBluescript SK$^-$ polylinker. Since there are no *Sac* II sites in the DNA ligase I cDNA and a unique site for *Sac* II immediately downstream of the *lac* promoter in pBluescript SK$^-$, amino-terminal deletions of the human cDNA were introduced by; (a) engineering a *Sac* II restriction enzyme site in the coding region by site-directed mutagenesis, (b) removing unwanted human DNA ligase I sequences by digestion with *Sac* II, and (c) re-ligating the human DNA ligase I sequences in the pBluescript vector. The resultant constructs encoded amino-terminal truncated forms of human DNA ligase I in frame with 23 amino acids of β-galactosidase, with expression of the fusion proteins under control of the *lac* promoter. Constructs encoding the carboxyl terminal 700 (Hlig220−919), 670 (Hlig250−919), 635 (Hlig285−919), 603 (Hlig317−919), 519 (Hlig401−919), 441 (Hlig479−919) and 362 (Hlig558−919) amino acids together with a full length construct (919 amino acids) were produced in this way (figure 1a).

Deletions of the carboxyl terminus of DNA ligase I were achieved by introducing stop codons in the DNA ligase I cDNA by site-directed mutagenesis of Hlig250−919. In this way, a series of constructs was engineered encoding DNA ligase I from residue 250 of the full length polypeptide and lacking the carboxyl-terminal six residues (Hlig250−913), 11 residues (Hlig250−908), 16 residues (Hlig250−903), 22 residues (Hlig250−897), 28 residues (Hlig250−891), 33 residues (Hlig250−886) or 38 residues (Hlig250−881) (figure 1b).

Single or double base mutations were introduced in Hlig250−919 by site-directed mutagenesis to effect single amino acid substitutions within the active site of DNA ligase I. Resultant constructs (given in brackets) encoded the following amino acid changes; Glu566 to Lys (E566K), Tyr567 to Phe (Y567F), Tyr567 to Ile (Y567I), Lys568 to His (K568H), Lys568 to Arg (K568R), Tyr569 to Ala (Y569A),Tyr569 to Phe (Y569F), Asp570 to Glu (D570E), Asp570 to Asn (D570N), Asp570 to Gln (D570Q), Gly571 to Ala (G571A), Gly571 to Val (G571V), and Arg573 to Ala (R573A) (figure 2).

Production of single-stranded uracil-containing DNA and *in vitro* mutagenesis

Single-stranded DNA was prepared from pBluescript according to the manufacturer's protocol (Stratagene). Uracil incorporation was achieved by addition of uridine to the culture (to a final concentration of 10μM) at the same time as infection with the helper phage.

Site-directed mutagenesis of the uracil-containing DNA was performed by the single primer method (17) using mutagenic primers based on the human DNA ligase I cDNA sequence (4). All introduced mutations were confirmed by DNA sequencing (18) using Sequenase (United States Biochemical) and synthetic sequencing primers.

Functional expression of cDNA constructs in the *E.coli lig* mutant

E.coli strain GR501 was transformed with human DNA ligase I/ pBluescript constructs by the calcium chloride/rubidium chloride method (19). Transformants were grown overnight at either 30°C (permissive temperature) or, to assay for functional complementation of the replication-defective phenotype, colonies were scored after overnight incubation at the non-permissive temperature of 42°C.

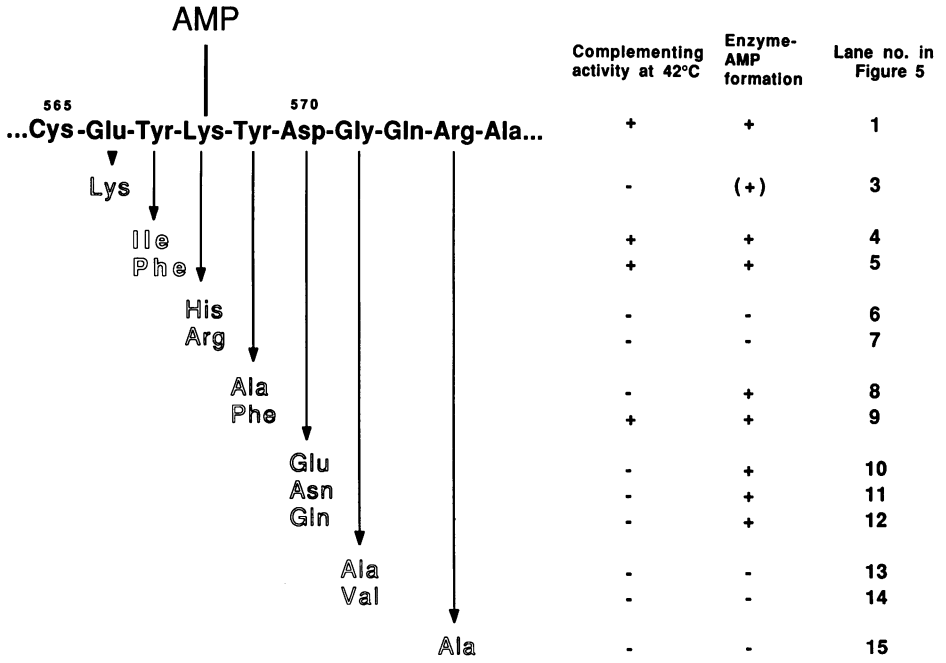

Figure 2. *In vitro* mutagenesis of the active site region of human DNA ligase I. The cDNA construct encoding amino acid residues 250 to 919 of the full-length protein (Hlig250−919) was subjected to *in vitro* mutagenesis as described in 'Materials and Methods' to effect single amino acid changes in the active site region of the protein. The amino acid sequence from residues 565 to 574 of the full-length DNA ligase I protein is shown in bold face and the active site Lys residue is denoted by attachment of the AMP group. Amino acid substitutions are shown in outline letters. Ability of a particular mutant protein to complement the *E.coli lig* mutant, leading to growth of the transformed host at 42°C is indicated by a '+'; a negative result is indicated by '−'. Ability to form an enzyme-adenylate complex in *E.coli* extracts is indicated; '+' : ability to form an enzyme-adenylate complex; '−' : no enzyme-adenylate complex detected even on long exposures of the autoradiograph; '(+)' : trace amounts of enzyme-adenylate formed.

Preparation of *E.coli* extracts and formation of DNA ligase-adenylate complex

One ml of an overnight GR501 culture grown at 30°C was collected by centrifugation. The bacterial pellet was resuspended in 0.7ml of reaction buffer (60mM Tris-HCl, pH8.0, 10mM $MgCl_2$, 5mM dithiothreitol, 50μg/ml bovine serum albumin) and disrupted by sonication. After centrifugation, cell extracts were adjusted to the same absorbance at 280nm by addition of reaction buffer. Aliquots of 18μl extract were incubated for 20 min at 20E^{0}C ıEwith 0.5μCi [α-^{32}P]ATP (3000Ci/mmol, Amersham Corp.) in a total volume of 20μl. Adenylylated polypeptides were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography.

Immunoblotting

Proteins in cell extracts were fractionated by SDS-PAGE and transferred to nitrocellulose filters (Scleicher & Schuell) in 25mM Tris-HCl, 192mM glycine, 20% methanol, pH8.3. After incubation for 2 hr at room temperature in 50mM Tris-HCl, pH7.5, 150mM NaCl, 0.05% Tween 20 (TBST) containing 5% w/v skimmed milk powder, filters were incubated for 12−16 hr at 4°C with a rabbit polyclonal antiserum diluted 1/500 in TBST containing 2.5% w/v skimmed milk powder. The rabbit antiserum was raised against a conjugated synthetic peptide corresponding to a peptide in the carboxyl-terminal region of the DNA ligases of vaccinia virus, *S.cerevisiae* and *S.pombe*, which is strongly conserved in human DNA ligase I (20). All subsequent steps were carried out at room temperature. Filters were washed four times with TBST to remove unbound antibody. After incubation for 5 min in TBST, ^{125}I-labelled donkey anti-rabbit (Ig) second antibody (Amersham Corp.) was added at 0.2μCi/ml and filters incubated for 2−3 hr. Filters were washed with TBST and analysed by autoradiography.

RESULTS AND DISCUSSION

Expression of full-length and amino-terminal truncated human DNA ligase I in the *E.coli lig* mutant

When *E.coli* strain GR501 was transformed with the full-length or amino-terminal truncated human DNA ligase I/pBluescript constructs, all transformants grew at 30°C (permissive temperature) but only transformants harbouring the Hlig220−919 or Hlig250−919 constructs formed colonies after overnight incubation at the non-permissive temperature of 42°C. Transformant colonies containing Hlig250−919 were of the same size as *E.coli lig*$^+$ control colonies, whereas those harbouring Hlig220−919 only formed microcolonies at the non-permissive temperature. Immunoblotting of crude *E.coli* cell extracts with an antibody against human DNA ligase I indicated that a polypeptide of the predicted molecular mass was expressed in soluble form in transformants containing each of the human DNA ligase I constructs (figure 3b). Thus, all mutant proteins were expressed in *E.coli* although, as seen in figure 3b, the amount of DNA ligase I protein varied with each construct, indicating that certain mutant proteins were somewhat less stable than others in *E.coli*. The Hlig250−919 was stably expressed in comparison to all the other constructs and, as judged by immunoblotting, the 80kDa fusion protein encoded by Hlig250−919 is the optimal size for expression of the human DNA ligase I protein in *E.coli*. Hlig1−919 (encoding the full-length protein) and Hlig317−919 were poorly expressed (see figure 3b, lanes 2 and 6). The apparent relative stabilities of the mutant proteins in *E.coli* may be due to differences in conformation or solubility, while the full-length protein (encoded by Hlig1−919; figure 3b, lane 2) appears to be susceptible to endogenous protease activity in *E.coli*.

An 85kDa (703 amino acids) carboxyl-terminal fragment of mammalian DNA ligase I is readily produced by endogenous proteolysis in crude extracts of mammalian cells or by subtilisin digestion *in vitro*, but is relatively resistant to further proteolysis. This fragment shows the same ability to form an enzyme-adenylate complex and join a DNA substrate in *in vitro* assays as the full-length protein, and apparently represents the catalytic domain of the enzyme (20). Furthermore, a truncated cDNA encoding this domain was able to complement a DNA ligase mutant of *S.cerevisiae* (4), indicating that the catalytic domain of DNA ligase I is sufficient for biological activity in the yeast and is able to fold into a functional protein without the presence of the amino-terminal region of the intact protein. Similarly, the protein encoded by the Hlig220−919 construct (which encodes a protein lacking just three amino acids of this previously identified catalytic domain) was able to complement the replication defect of the *E.coli lig* mutant leading to growth at

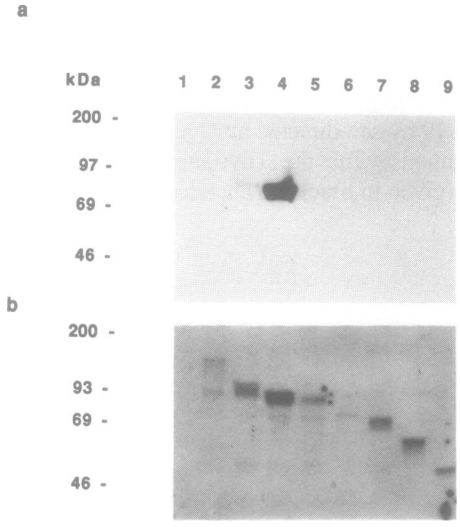

Figure 3. Expression of amino-terminal truncated human DNA ligase I cDNAs in the *E.coli lig* mutant. Amino-terminal truncations of the full-length DNA ligase I protein were achieved by *in vitro* mutagenesis of Hlig1−919 as described in 'Materials and Methods'. **a.** Enzyme-adenylate formation of *lac*Z/DNA ligase I fusion proteins expressed in *E.coli* . Extracts of *E.coli* transformants expressing *lac*Z/DNA ligase I fusion proteins were incubated with [α-^{32}P]ATP and adenylated peptides analysed by SDS-PAGE and autoradiography. [^{14}C] methylated molecular mass markers (Amersham) were; myosin, 200 kDa; phosphorylase b, 97 kDa; bovine serum albumin, 69kDa; and ovalbumin, 46kDa. **b.** Immunoblot of *lac*Z/DNA ligase I fusion proteins expressed in *E.coli*. Extracts of *E.coli* transformants expressing *lac*Z/DNA ligase I fusion proteins were electrophoresed through a 7.5% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. The membrane was incubated with a 1/500 dilution of a rabbit anti-peptide antiserum raised against a conjugated synthetic peptide corresponding to a strongly conserved peptide in the carboxyl-terminal region of the DNA ligases of vaccinia virus and yeasts, and human DNA ligase I (20). Polypeptides recognised by the antiserum were detected as described in 'Materials and Methods'. Unlabelled molecular mass markers (Amersham) were as [^{14}C] methylated markers (non-methylated phosphorylase b, 93 kDa). Extracts were prepared from *E.coli* strain GR501 transformed with the following constructs; lane 1: pBluescript vector; lane 2: Hlig1−919 (encoding full-length DNA ligase I protein); lane 3: Hlig220−919; lane 4: Hlig250−919; lane 5: Hlig285−919; lane 6: Hlig317−919; lane 7: Hlig401−919; lane 8: Hlig479−919; lane 9: Hlig558−919.

the non-permissive temperature, and was also able to form the enzyme-adenylate complex in *E.coli* cell extracts (figure 3a, lane 3). The 80kDa protein (670 amino acids) encoded by the Hlig250−919 construct was also able to form the enzyme-adenylate complex and, even allowing for the increased DNA ligase I protein level in transformants harbouring Hlig250−919 compared with the Hlig220−919 construct, this 80kDa protein appeared to be more readily adenylylated than the larger protein encoded by Hlig220−919. Thus, the Hlig250−919 construct defines a protein of optimal size for both catalytic and complementing activity in *E.coli*.

Proteins encoded by constructs Hlig285−919, Hlig317−919, Hlig401−919, Hlig479−919 and Hlig558−919 were not only unable to complement the replication defect of the *E.coli lig* mutant, but were also unable to form an enzyme-adenylate complex (figure 3a, lanes 5 to 9). No enzyme-adenylate complex was detected even on long exposures of the autoradiograph in figure 3a. Thus, reduction of the size of the catalytic domain of DNA ligase I from 670 (encoded by Hlig250−919) to 635 amino acids (encoded by Hlig285−919) by further amino-terminal deletion caused complete loss of enzyme activity *in vivo* and *in vitro*.

The full-length human DNA ligase I protein was not detectably able to form an enzyme-adenylate complex in *E.coli* cell extracts (figure 3a, lane 2), consistent with its inability to complement the replication defect of the host strain. This could be due to the proteolytic degradation of the full-length protein (figure 3b, lane 2) or may indicate that the extended hydrophilic amino terminus of the full-length protein exerts an inhibitory effect on the activity of the catalytic domain of the enzyme in this system. Thus, the enzyme might normally be activated by post-translational modification of the amino terminus and the requisite modifying process(es) may be absent or inefficient in *E.coli*. In agreement with the latter notion, investigations of the phosphorylation pattern of bovine DNA ligase I indicate that a non-phosphorylated, full-length form of the enzyme has little or no catalytic activity (C.Prigent, D.D. Lasko, K.K. and T.L., in preparation). Similarly to DNA ligase I, human and yeast DNA topoisomerase I have an extended amino-terminal region of presumed regulatory rather than catalytic function, and about 130 amino acids could be deleted from the amino terminus without affecting enzyme activity *in vitro* (21, 22). However, in contrast to the present results with DNA ligase I, this amino-terminal truncation of the yeast DNA topoisomerase I enzyme inactivated its function in *E.coli* (21). Human DNA topoisomerase I is also able to complement an *E.coli* mutant (23).

The results of expression of the full-length and amino-terminal truncated DNA ligase I proteins in the *E.coli lig* mutant are summarised in figure 1a.

Expression of carboxyl-terminal truncated human DNA ligase I in the *E.coli ig* mutant

Since the amino-terminal truncated protein encoded by the Hlig250−919 construct was found to be more efficiently expressed in *E.coli* than full-length human DNA ligase I, the effects of introducing carboxyl-terminal deletions in the DNA ligase I protein were determined by introducing stop codons in Hlig250−919 and assessing DNA ligase activity *in vitro* and *in vivo* following introduction of the mutant constructs into the *E.coli lig* host. The results of this analysis are summarised in figure 1b. All mutant proteins were stably expressed in *E.coli* (figure 4b), although removal of 22 or more amino acids from the carboxyl terminus resulted in a decreased level of expression (figure 4b, lanes 2 to 5). These shorter polypeptides were also unable to form an enzyme-adenylate complex in cell extracts of

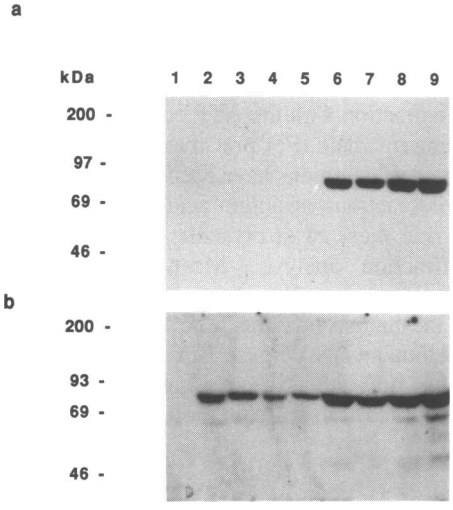

Figure 4. Expression of carboxyl-terminal truncated human DNA ligase I cDNAs in the *E.coli lig* mutant. Carboxyl-terminal truncations of DNA ligase I were introduced by *in vitro* mutagenesis of Hlig250−919, which encoded a polypeptide of optimal size for expression in *E.coli* (see figure 3). **a.** Enzyme-adenylate formation of *lacZ*/DNA ligase I fusion proteins expressed in *E.coli* (see legend for figure 3). **b.** Immunoblot of *lacZ*/DNA ligase I fusion proteins expressed in *E.coli* (see legend for figure 3). Molecular mass markers were as in figure 3. Extracts were prepared from *E.coli* strain GR501 transformed with the following constructs; lane1: pBluescript vector; lane 2: Hlig250−881; lane 3: Hlig250−886; lane 4: Hlig250−891; lane 5: Hlig250−897; lane 6: Hlig250−903; lane 7: Hlig250−908; lane 8: Hlig250−913; lane 9: Hlig250−919.

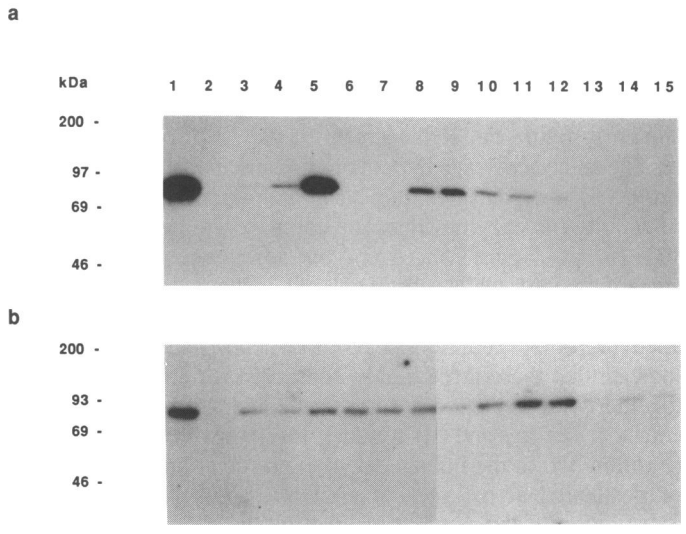

Figure 5. Site-directed mutagenesis of the active site region of the Hlig250−919 human DNA ligase I cDNA; expression of mutant proteins in *E.coli*. **a.** Enzyme-adenylate formation of mutant proteins expressed in *E.coli* (see legend for figure 3). Lanes 1 to 9 and 10 to 15 are from two different gels exposed on the same autoradiograph. **b.** Immunoblot of mutant *lacZ*/DNA ligase I fusion proteins expressed in *E.coli* (see legend for figure 3). Lanes 1 to 8 and 9 to 15 are from two different gels exposed to the same autoradiograph. Molecular mass markers were as in figure 3. Extracts were prepared from *E.coli* strain GR501 transformed with the following constructs; lane1: Hlig250−919; lane 2: pBluescript vector; lane 3: E566K; lane 4: Y567I; lane 5: Y567F; lane 6: K568H; lane 7: K568R; lane 8: Y569A; lane 9: Y569F; lane 10: D570E; lane 11: D570N; lane 12: D570Q; lane 13: G571A; lane 14: G571V; lane 15: R573A.

E.coli transformants (figure 4a, lanes 2 to 5). This loss of activity *in vitro* corresponded with loss of *in vivo* function as estimated by ability to complement the replication defect of the *E.coli lig* mutant at the non-permissive temperature. All proteins with carboxyl-terminal deletions of less than 22 amino acids retained enzyme activity both *in vitro* and *in vivo* (figure 1b; figure 4a, lanes 6 to 9).

The carboxyl-terminal region of human DNA ligase I shows homology with the DNA ligases of both *S.cerevisiae* and *S.pombe*, except that the human and *S.pombe* enzymes are, at optimal sequence alignment (4), 14 amino acids longer. There is little homology between the two enzymes in this 'tail' sequence and results here show that this sequence can be removed from the human enzyme (as in the protein encoded by Hlig250−903) without loss of function in *E.coli*. However, further removal of 8 amino acid residues (as in the protein encoded by Hlig250−897) caused complete loss of enzyme function. As immunoblotting analysis indicated that those proteins that retained enzyme function were also more stably expressed in *E.coli* (figure 4b, lanes 6 to 9), the carboxyl-terminal region of DNA ligase I may be required for correct protein folding. Alternatively, these residues may contain a site for post-translational modification which increases enzyme activity.

The results of expression of Hlig250−919 and carboxyl-terminal truncated DNA ligase I proteins in the *E.coli lig* mutant are summarised in figure 1b.

In vitro mutagenesis of the active site of human DNA ligase I

The active protein encoded by the Hlig250−919 construct was used to assess the effects of amino acid substitutions in the region of the active site. Single amino acid substitutions were effected by site-specific *in vitro* mutagenesis of Hlig250−919 and are shown in figure 2. All mutant proteins were expressed in *E.coli*, although at different levels, and were all less abundant than the 'parent' protein encoded by Hlig250−919 (figure 5b). Adenylylation of mutant proteins in crude *E.coli* cell extracts is shown in figure 5a. Replacement of the AMP-binding Lys568 residue at the active site by Arg or His caused total loss of enzyme activity. The replacement of Gly571 and Arg573 also caused loss of activity (no enzyme-adenylate complex was detected in lanes 13−15 of figure 5a even on long exposures of the autoradiograph), while substitution of Glu566 with a basic residue caused a very strong reduction in ability to form the enzyme-adenylate complex and loss of complementing activity. The data indicate that these three highly conserved residues (9) function together with the active site Lys in the interaction with the ATP cofactor. The Lys and Gly residues, together with the Asp residue (position 570 in the human enzyme) are invariant in the known or postulated active sites of all DNA and RNA ligases (9). Substitution of the Asp570 residue to Glu, Gln or Asn did not abolish ability to form the enzyme-adenylate complex but did lead to loss of ability to support growth of the *E.coli* host at the non-permissive temperature of 42°C. This indicates that the Asp residue is essential, not for formation of an enzyme-adenylate intermediate, but at a later step of the ligation reaction.

The amino acids immediately flanking the active site Lys on either side are hydrophobic residues in all DNA ligases (9). Substitution of Tyr567 or Tyr569 of human DNA ligase I by other hydrophobic residues caused either partial or no reduction in the ability of the enzyme to form an enzyme-adenylate complex. Phe or Ile residues were better tolerated than Ala in this regard. The proteins encoded by the constructs Y567F, Y567I and Y569F were also able to support growth of the *E.coli lig* mutant at the non-permissive temperature.

The bacteriophage T4 RNA ligase acts by a mechanism similar to DNA ligases in that an enzyme-adenylate reaction intermediate is initially generated, which then interacts with the nucleic acid substrate. Heaphy *et al.* (24) investigated the relative functional importance of several amino acid residues in the active site region of the enzyme by *in vitro* mutagenesis. T4 RNA ligase also contains the Lys-X-Asp-Gly residues of the active site motif common to DNA ligases, but shows little or no sequence homology to DNA ligases outside this short region. Replacement of the amino acid residue between the active site Lys and Asp residues with several alternatives was well tolerated. Substitution of the Asp residue allowed for the formation of an enzyme-adenylate intermediate, but a complete ligation reaction did not occur. These data (24) agree well with the present results for human DNA ligase I (summarised in figure 2) and point to mechanistic similarities at the active sites of these two enzymes.

General discussion

We have analysed functional aspects of human DNA ligase I by employing a single vector for both site-directed *in vitro* mutagenesis of the DNA ligase I cDNA and expression in an *E.coli lig* mutant. The use of uracil-containing template DNA in single primer-mediated site-directed *in vitro* mutagenesis was first described for the M13 filamentous phage vectors (17). We have used this method with the pBluescript phagemid vector and obtained similar mutation frequencies. The vector allowed efficient and stable expression of DNA ligase I cDNAs as *lacZ* fusion proteins in *E.coli*. A cDNA encoding the catalytic domain of the human enzyme was able to substitute for the defective NAD-dependent DNA ligase of a conditional-lethal *E.coli lig* mutant, allowing growth at the non-permissive temperature. By targeted *in vitro* mutagenesis, we were able to define an optimum domain size for the function of human DNA ligase I in *E.coli*, and identified several residues in the active site region of the human enzyme involved in the first step of the DNA ligation reaction, interaction with the ATP cofactor to form an enzyme-adenylate intermediate. For practical reasons, a limited number of amino acid substitutions have been analysed here. The precise roles and interactions of amino acid residues at the active site of the enzyme must await crystallization of DNA ligase I and X-ray diffraction analysis. Meanwhile, this mutagenesis/expression system is a direct means of investigating functional details of the human enzyme. The expression of the catalytic domain of human DNA ligase I in an *E.coli lig* mutant as an essential function at 42°C, but not 30°C, also provides a convenient approach in searching for specific chemical inhibitors of the human enzyme.

REFERENCES

1. Soderhall, S. and Lindahl, T. (1975) *J. Biol. Chem.*, **250**, 8438−8444.
2. Lasko, D. D., Tomkinson, A. E. and Lindahl, T. (1990) *J. Biol. Chem.*, **265**, 12618−12622.
3. Wilcock, D. and Lane, D. P. (1991) *Nature*, **349**, 429−431.
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. Johnston, L. H. and Nasmyth, K. A. (1978) *Nature*, **274**, 891−893.
6. Engler, M. J. and Richardson, C. C. (1982) In Boyer, P. D. (ed.), The Enzymes. Academic Press, New York, 3rd Ed., Vol. 15B, pp.3−29.
7. Gumport, R.I. and Lehman, I. R. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 2559−2563.

8. Soderhall, S. and Lindahl, T. (1973) *J. Biol. Chem.*, **248**, 672–675.
9. Tomkinson, A. E., Totty, N. F., Ginsburg, M. and Lindahl, T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 400–404.
10. Barker, D. G., White, J. H. M. and Johnston, L. H. (1987) *Eur. J. Biochem.*, **162**, 659 –667.
11. Barker, D. G. and Johnston, L. H. (1983) *Eur. J. Biochem.*, **134**, 315–319.
12. Park, U.E., Olivera, B. M., Hughes, K. T., Roth, J. R. and Hillyard, D.R. (1989) *J. Bacteriol.*, **171**, 2173–2180.
13. Hanahan, D. (1983) *J. Mol. Biol.*, **166,** 557–580.
14. Bullock, W. O., Fernandez, J. M. and Short, J. M. (1987) *BioTechniques*, **5**, 376–379.
15. Kunkel, T. A., Roberts, J. D. and Zakour, R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
16. Dermody, J. J., Robinson, G. T. and Sternglanz, R. (1979) *J. Bacteriol.*, **139,** 701–704.
17. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
18. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci USA.*, **74**, 5463–5467.
19. Kushner, S. R. (1978) In Boyer, H. B. and Nicosia, S. (ed.), Genetic engineering. Elsevier/North-Holland, Amsterdam, pp.17–23.
20. Tomkinson, A. E., Lasko, D. D., Daly, G. and Lindahl, T. (1990) *J. Biol Chem.*, **265**, 12611–12617.
21. Bjornsti, M.-A. and Wang, J. C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8971–8975.
22. D'Arpa, P., Machlin, P. S., Ratrie, III, H., Rothfield, N., Cleveland, D. W. and Earnshaw, W. C. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2543–2547.
23. Bjornsti, M.-A., Beneditti, P., Viglianti, G. A. and Wang, J. C. (1989) *Cancer Res.*, **49**, 6318–6323.
24. Heaphy, S., Singh, M. and Gait, M. J. (1987) *Biochemistry*, **26**, 1688–1696.