CORE
Metadata, citation and similar papers at core.ac.uk
Provided by University of East Anglia digital repository

PROTEINS: Structure, Function, and Genetics 51:321–326 (2003)

SHORT COMMUNICATION

NAD$^+$-dependent DNA Ligases of *Mycobacterium tuberculosis* and *Streptomyces coelicolor*

Adam Wilkinson,[1] Heather Sayer,[2] Desmond Bullard,[2] Andrew Smith,[2] Jonathan Day,[3] Tobias Kieser,[4] and Richard Bowater[2]*

[1]*Phico Therapeutics, Ltd., Babraham Hall, Babraham, Cambridge, United Kingdom*
[2]*Molecular Biology Sector, School of Biological Sciences, University of East Anglia, Norwich, United Kingdom*
[3]*Division of Molecular Genetics, Anderson College, University of Glasgow, Glasgow, United Kingdom*
[4]*Department of Genetics, John Innes Centre, Norwich, United Kingdom*

ABSTRACT **Sequencing of the genomes of *Mycobacterium tuberculosis* H37Rv and *Streptomyces coelicolor* A3(2) identified putative genes for an NAD$^+$-dependent DNA ligase. We have cloned both open reading frames and overexpressed the protein products in *Escherichia coli*. In vitro biochemical assays confirm that each of these proteins encodes a functional DNA ligase that uses NAD$^+$ as its cofactor. Expression of either protein is able to complement *E. coli* GR501, which carries a temperature-sensitive mutation in *ligA*. Thus, in vitro and in vivo analyses confirm predictions that *ligA* genes from *M. tuberculosis* and *S. coelicolor* are NAD$^+$-dependent DNA ligases. Proteins 2003;51:321–326.**
© 2003 Wiley-Liss, Inc.

Key words: DNA end-joining; DNA repair; ligation; mycobacteria; streptomycetes

INTRODUCTION

DNA ligases are essential enzymes required for joining of DNA strand breaks formed during DNA replication, recombination, and repair. Biochemical studies of DNA ligases purified from a variety of organisms and viruses have identified that the reaction mechanism can be broken down into three steps.[1–4] In the first step of the reaction, a covalent enzyme-adenylate intermediate is formed via a specific lysine in the catalytic motif of the enzyme (KXDG in motif I—Fig. 1). The adenylate group can be provided by either adenosine triphosphate (ATP) or nicotinamide-adenine dinucleotide (NAD$^+$), leading to the classification of the enzymes as ATP- or NAD$^+$-ligases.

The paradigm for studies of NAD$^+$-ligases is the *Escherichia coli* K12 enzyme encoded by the essential gene *ligA*.[1] NAD$^+$-ligases are also essential for *Bacillus subtilis*[5] and *Staphylococcus aureus*,[6] and are likely to be indispensable in all eubacteria. Several *E. coli* strains carrying temperature-sensitive mutations in *ligA* have been isolated, such as *E. coli* GR501, which is unable to grow at temperatures above 42°C.[7] NAD$^+$-ligases have been purified from other eubacteria and have similar biochemical characteristics.[2,8] Recently, an NAD$^+$-ligase was identified within the genome of a virus,[9] but they have not been found in eucaryotic genomes. Thus, NAD$^+$-ligases have been suggested as possible targets for novel antibacterial compounds.[6,9–12]

Within every eubacterial genome sequenced so far are open reading frames predicted to encode NAD$^+$-ligases of similar size and extensive amino acid sequence homology.[8] A series of six conserved motifs occur in sequence alignments of DNA ligases and mRNA capping enzymes[3,13,14] (e.g., Fig. 1). Motifs I–V lie in the core domains of these enzymes and make important contacts to the bound nucleotide, indicating that the active site has been conserved among all DNA ligases.[2,3] X-ray crystallographic studies have provided high-resolution structures of NAD$^+$-ligases from *Thermus filiformis*[11] and *Bacillus stearothermophilus*.[10] The good agreement between the two X-ray crystallographic structures and the high degree of homology of protein sequence suggests that all NAD$^+$-ligases are likely to have similar overall structures.[3,11]

Recent studies identified active ATP-ligases in *Haemophilus influenzae*[15] and *B. subtilis*.[5] Analysis of genome sequences predicts that several eubacteria encode distinct ATP-ligases in addition to their NAD$^+$-ligase.[2,8,16] For example, genomic analysis of *Mycobacteria tuberculosis* H37Rv[17] and *Streptomyces coelicolor* A3(2)[18] identified putative genes for three ATP-ligases and one NAD$^+$-ligase. These actinomycetes are complex eubacteria and an ancient synteny exists between their genomes,[18] but

Grant sponsors: The Biotechnology and Biological Sciences Research Council (BBSRC), Guildford Bench Methodology Fund, John and Pamela Salter Charitable Trust, Nuffield Foundation, Society for General Microbiology, and the University of East Anglia.

*Correspondence to: Richard Bowater, Molecular Biology Sector, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK. E-mail: r.bowater@uea.ac.uk

Received 9 September 2002; Accepted 25 November 2002

© 2003 WILEY-LISS, INC.

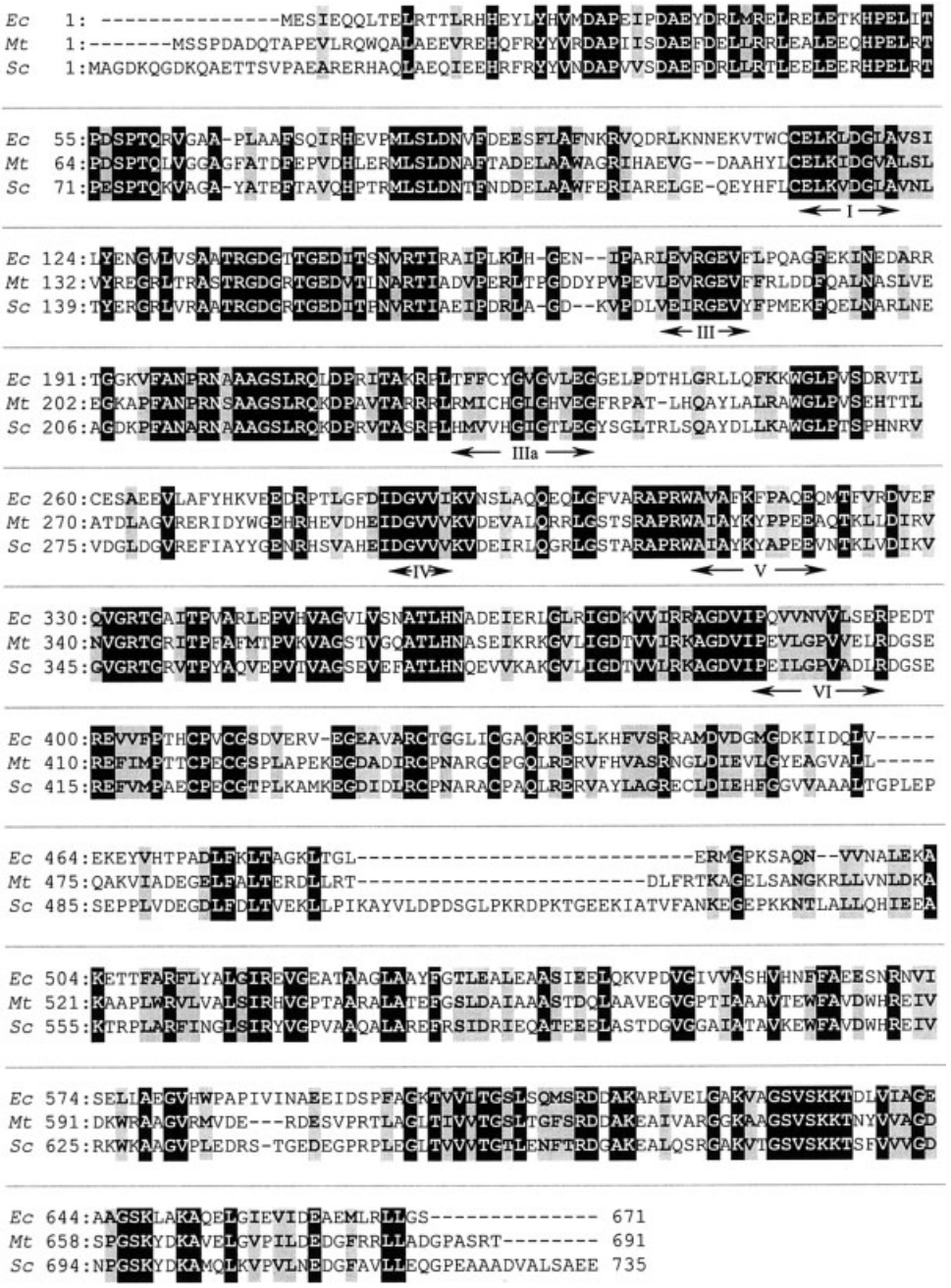

Fig. 1. Sequence alignment of *LigA* from *E. coli* K12, *M. tuberculosis* H37Rv, and *S. coelicolor* A3(2). Amino acid sequence of *E. coli* K12 *LigA* was aligned with the use of ClustalW with the predicted open reading frames for *M. tuberculosis* H37Rv *LigA* (gene = *Rv3014c*) and *S. coelicolor* A3(2) *LigA* (gene = *SC8D9.06*). Identical residues are shown in white on a black background. Amino acids highlighted in bold on a gray background are conserved within standard functional groups as follows: acidic = DE; basic = HKR; hydrophobic = A, F, I, L, M, P, V, W; polar = C, G, N, Q, S, T, Y. Arrows indicate the six sequence motifs (Motifs I, III, IIIa, IV, V, and VI) conserved in DNA ligases and mRNA capping enzymes.

they are phylogenetically distinct to enterobacteria such as *E. coli*.

DNA ligases from *E. coli* have been well studied, but it remains possible that the physiological roles of these enzymes may be altered in different eubacteria. Furthermore, although the good homology between the amino acid sequences of DNA ligases identified within various eubacterial genomes suggest that all are likely to be functional NAD^+-ligases, the validity of these predictions must be established by experimental analysis. To begin to address whether the NAD^+-ligases from actinomycetes have different biochemical roles, or whether they may be influenced by the potential ATP-ligases, we wished to confirm that the predicted *ligA* genes act as functional NAD^+-ligases. In this

respect, here we report the cloning and overexpression of the *ligA* homologues from *M. tuberculosis* and *S. coelicolor*.

MATERIALS AND METHODS

Cloning of DNA Ligases

Various NAD^+-ligases were amplified by polymerase chain reaction (PCR) with a proof-reading DNA polymerase from genomic DNA as follows:

E. coli K12 NAD^+-ligase (EcLigA, 671 amino acids) was amplified using 5′ primer (5′-CAT ATG GAA TCA ATC GAA CAA C-3′) and 3′ primer (5′-GGA TCC TCA GCT ACC CAG CAA ACG CA-3′).

M. tuberculosis H37Rv NAD^+-ligase (MtLigA, 691 amino acids) was amplified using 5′ primer (5′-CAT ATG GTG AGC TCC CCA GAC GCC GA-3′) and 3′ primer (5′-GGA TCC TCA CGT TCG TGA CGC GGG TCC GT-3′).

S. coelicolor A3(2) NAD^+-ligase (ScLigA, 735 amino acids) was amplified using 5′ primer (5′-CAT ATG GCC GGC GAC AAG CAG GG-3′) and 3′ primer (5′-GGA TCC CTA TTC CTC AGC CGA AAG CGC-3′).

For each PCR amplification, the 5′ primers contained an *Nde*I site, and the 3′ primers contained a *BamH*I site. PCR products were cloned using the Zero Blunt TOPO® Cloning kit (Invitrogen) and sequenced to confirm that no mutations were introduced during PCR. Fragments were excised from the TOPO vectors with the use of the *Nde*I and *BamH*I sites and cloned into pET16b (Novagen). Proteins overexpressed from this vector contain a 10-His tag within an extra 21 amino acids (2.5 kDa) at the N-terminus of the protein. To allow overexpression of proteins in *E. coli* GR501, full-length ligases plus the His-tag were excised from pET-16b vectors with the use of the *Nco*I and *BamH*I sites, and cloned into p*Trc99*A (Amersham Pharmacia).

Protein Purification

For protein expression, all *E. coli* cultures were grown at 37°C in Luria–Bartani (LB) medium containing ampicillin (100 μg/mL) and chloramphenicol (50 μg/mL). The pET16b clones were transformed into *E. coli* BL21 (DE3) pLysS, plated on LB-agar containing antibiotics, and grown overnight. Single colonies were inoculated into 5 mL liquid media, grown overnight, and diluted 100-fold into 50 mL fresh media. After growth to midlog phase (OD_{600} = 0.4–0.6), protein expression was induced by addition of IPTG to 0.4 m*M*. Cells were harvested after 4 h further growth, and proteins were purified from the soluble fraction on 2 mL Ni^+-agarose columns (Novagen His-bind resin). The proteins were dialyzed into 20 m*M* Tris-HCl pH 7.5, 50 m*M* NaCl, 20% glycerol.

The purified *E. coli* LigA (including the His-tag) was used to raise rabbit polyclonal antibodies (Davids Biotechnologie, Germany).

Preparation of DNA Substrate

A double-stranded DNA 40-bp substrate carrying a single-strand nick between bases 18 and 19 was used as the standard substrate for ligation assays.[19] This substrate was created in Tris-borate-EDTA buffer (TBE) buffer by annealing an 18-mer (5′-GTA AAA CGA CGG CCA GTG-3′) and a 22-mer (5′-AAT TCG AGC TCG GTA CCC GGG G-3′) to a complementary 40-mer (5′-CCC CGG GTA CCG AGC TCG AAT TCA CTG GCC GTC GTT TTA C-3′). At their 5′ ends, the 18-mer contained a fluorescein molecule and the 22-mer was phosphorylated.

Analysis of Ligation Activity

The nicked 40-bp substrate (50 pmoles) was used to assay in vitro ligation activity of each enzyme (35 pmoles). Ligations were performed in 10 μL reactions and incubated at 25°C for up to 1 h. Reactions were conducted in the presence of 1 m*M* ATP or 26 μ*M* NAD^+ in 50 m*M* Tris (pH 7.8), 10 m*M* $MgCl_2$, 10 m*M* dithiothreitol, 25 μg/mL bovine serum albumen (BSA). At the end of the incubation, the samples were mixed with an equal volume of formamide-loading buffer, heated to 95°C, loaded on to a 15% polyacrylamide-urea gel (10 × 10 cm), and run at 300 V for 1 h in IX TBE. Reaction products on the gel were visualized using a Molecular Dynamics Storm phosphorimager.

To assay for ligation activity in vivo, we used *E. coli* GR501, which has a temperature-sensitive mutation in *ligA*.[20] Cells were transformed with p*Trc99A* containing the relevant DNA ligase and grown overnight at 30°C on LB-agar containing ampicillin (100 μg/mL). Single colonies were streaked on to fresh LB-agar plates and grown at 30°C or 44°C.

RESULTS AND DISCUSSION

Predicted NAD^+-Ligases of *M. tuberculosis* and *S. coelicolor*

The genome sequences of *M. tuberculosis* H37Rv[17] and *S. coelicolor* A3(2)[18] contain one putative NAD^+-ligase, named *ligA*. We refer to the protein products of the *M. tuberculosis* and *S. coelicolor ligA* genes as MtLigA and ScLigA, respectively. Alignment of these sequences with other NAD^+-ligases confirmed that they contained the six conserved motifs that are typical of DNA ligases (Fig. 1). Note that the levels of G+C-bases in each of these *ligA* genes is dramatically different: *E. coli* K12 is 54%, *M. tuberculosis* H37Rv is 69% and *S. coelicolor* is 70% G+C. Despite these differences at the gene level, there is very good homology between the amino acid sequences of MtLigA, ScLigA, and EcLigA (protein product of *E. coli* K12 *ligA*). As expected from phylogenetic relationships, MtLigA and ScLigA are most similar, with 57% identity and 68% similarity. Comparison to EcLigA gives 42% identity and 56% similarity for MtLigA, and 40% identity and 55% similarity for ScLigA. The most distinct differences between the proteins occur in ScLigA, with two insertions of 5 and 24 amino acids that begin at amino acids 480 and 506, respectively. These extensions are within two of the conserved helix–hairpin–helix motifs that have been implicated in non-sequence-specific binding of DNA by NAD^+-ligases.[3,21] High-resolution analysis of ScLigA will be required to determine whether these additional amino acids have a dramatic influence on the structure of the protein.

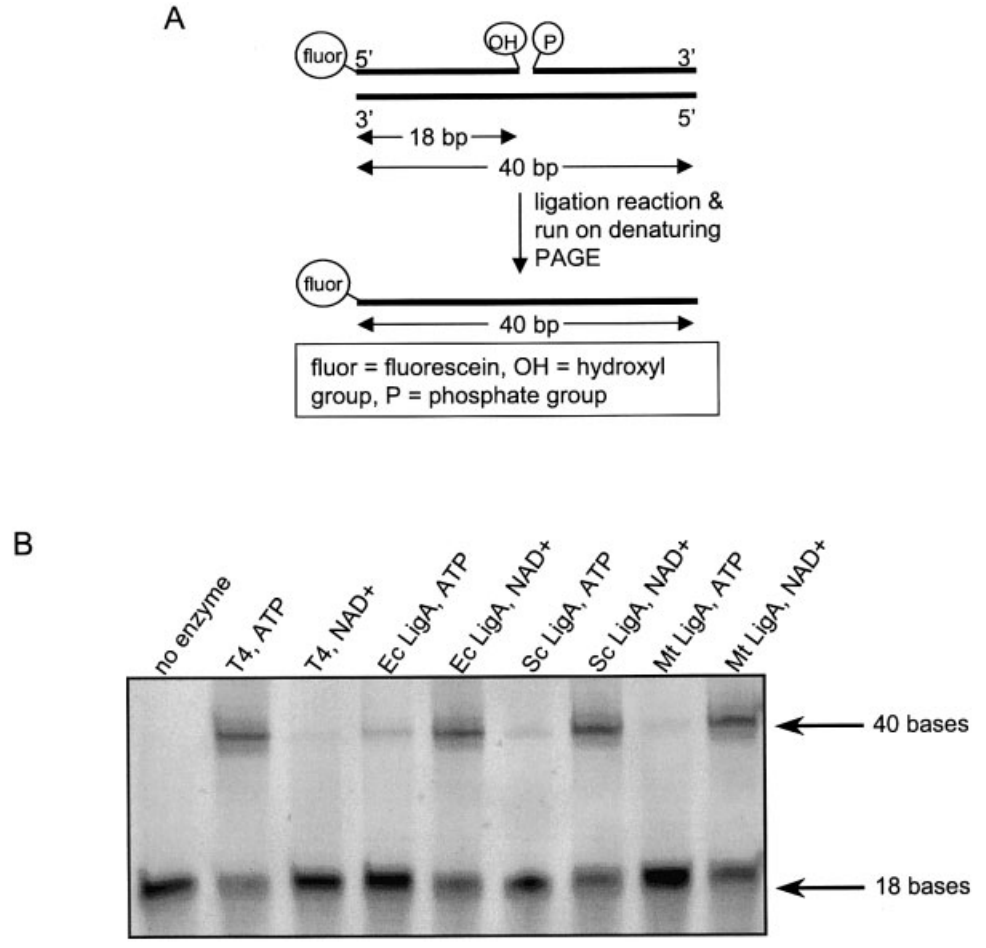

Fig. 2. *LigA* genes from *E. coli, M. tuberculosis*, and *S. coelicolor* encode functional NAD^+-ligases. **(A)** Schematic details of Fluorescein-oligo assay of in vitro DNA ligation activity. If ligation occurs, the size of the DNA product increases from 18 to 40 bases. **(B)** In vitro ligation assays use indicated protein in the presence of ATP or NAD^+. LigA from *E. coli, M. tuberculosis*, and *S. coelicolor* all had DNA end-joining activity in the presence of NAD^+. Note that all eubacterial *LigA*'s had low ligation activity in the presence of ATP, because preadenylated enzyme allowed one round of ligation. T4 was used as a control for ATP-dependent ligation and was thus more active in the presence of ATP.

Functional NAD^+-ligases of *M. tuberculosis* and *S. coelicolor*

PCR amplification of the relevant genomic DNA generated full-length genes for MtLigA and ScLigA, which were cloned into an *E. coli* expression system. To compare our results with previous studies, we also cloned full-length EcLigA into the same expression system. The NAD^+-ligases were overexpressed and purified by affinity chromatography to an in-frame N-terminal His-tag. This method of purification has been used in a number of studies of different DNA ligases,[9,22–24] and there is no evidence to suggest that the His-tag influences activity of the enzyme. Analysis on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) showed that the His-tagged proteins were >90% pure, and Western blots with a polyclonal antibody to EcLigA detected no contamination with chromosomally expressed *E. coli* LigA (data not shown).

To examine whether each of the purified proteins had DNA ligase activity, we used a standard in vitro assay for DNA ligation[19] that uses a 40-bp double-stranded DNA containing a gap between bases 18 and 19 of one strand [Fig. 2(A)]. The eubacterial proteins purified in this study all joined the DNA "nick" in the presence of NAD^+ and were thus, confirmed to have DNA ligase activity [Fig. 2(B)]. Commercially available bacteriophage T4 (T4) DNA ligase was used as a control for ATP-ligases and, as expected, DNA end-joining was observed in the presence of ATP. Note that there was a low level of activity of all proteins (including T4) in the absence of their appropriate cofactor [Fig. 2(B)], which was proportional to the amount of added enzyme (data not shown). This has been observed with other DNA ligases[6,23] and occurs because the overexpressed enzymes are preadenylated when purified, allowing a single turnover of each enzyme.

The essential nature of LigA is illustrated by *E. coli* strain GR501, which contains a temperature-sensitive mutation in *ligA* that renders the strain nonviable at or above temperatures of 42°C.[7] This strain has been used previously to confirm ligation activity in *E. coli* of human

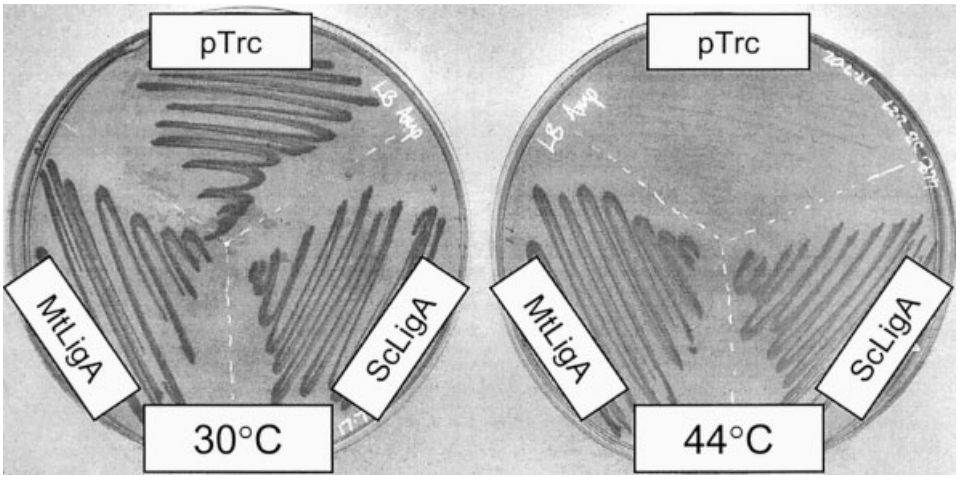

Fig. 3. *M. tuberculosis* and *S. coelicolor ligA* genes function as NAD^+-ligases in *E. coli*. *E. coli* GR501, which contains a temperature-sensitive mutation in chromosomal *ligA*, was transformed with p*Trc99A* overexpressing MtLigA or ScLigA and grown overnight at 30°C on LB-agar containing ampicillin (100 μg/mL). *E. coli* GR501 transformed with p*Trc99*A alone was used as a control for cells containing only temperature-sensitive *ligA*. Single colonies from both transformations were streaked on to fresh LB-agar plates containing ampicillin and grown at 30°C (left plate) or 44°C (right plate). Note that over expression of MtLigA or ScLigA complemented the temperature-sensitive mutation and allowed *E. coli* GR501 to grow at 44°C.

DNA ligase I.[25] We used this strain to test whether MtLigA and ScLigA can act as functional NAD^+-ligases in vivo. EcLigA, MtLigA, and ScLigA were cloned into the *E. coli* expression vector p*Trc99*A. Control experiments showed that *E. coli* GR501 was nonviable at 44°C when transformed with p*Trc99*A alone (Fig. 3). As expected, expression of full-length EcLigA from p*Trc99*A allowed growth at 44°C (data not shown). Thus, the presence of wild-type EcLigA complements the mutation and allows GR501 to grow at elevated temperature. Similarly, expression of MtLigA or ScLigA from p*Trc99*A permitted growth of *E. coli* GR501 at 44°C (Fig. 3). Thus, MtLigA and ScLigA can function as NAD^+-ligases in *E. coli*.

CONCLUSIONS

The presence of putative genes for both ATP- and NAD^+-ligases in some eubacteria is intriguing and indicates a complex evolutionary relationship for DNA ligases.[8] Current evidence suggests that NAD^+-ligases cannot be replaced by eubacterial ATP-ligases[5]; thus, the NAD^+-dependent enzymes are likely to be essential in all eubacteria. Therefore, NAD^+-ligases are still potential targets for antibacterial compounds. To identify whether this will be a valuable area of study, it is important to perform biochemical characterization of NAD^+-ligases from a wide range of eubacteria. In this regard, we have cloned and overexpressed the *ligA* homologs from *M. tuberculosis* and *S. coelicolor*. In vitro and in vivo analyses confirmed that these genes encode functional NAD^+-ligases.

ACKNOWLEDGMENTS

We thank Drs. Laura Bowater and Manuel Lavesa-Curto for their comments on this manuscript.

REFERENCES

1. Lehman IR. DNA ligase: Structure, mechanism, and function. Science 1974;186:790–797.
2. Timson DJ, Singleton MR, Wigley DB. DNA ligases in the repair and replication of DNA. Mutat Res 2000;460:301–318.
3. Doherty AJ, Suh SW. Structural and mechanistic conservation in DNA ligases. Nucleic Acids Res 2000;28:4051–4058.
4. Tomkinson AE, Levin DS. Mammalian DNA ligases. BioEssays 1997;19:893–901.
5. Petit MA, Ehrlich SD. The NAD-dependent ligase encoded by yerG is an essential gene of Bacillus subtilis. Nuclic Acids Res 2000;28: 4642–4648.
6. Kaczmarek FS, Zaniewski RP, Gootz TD, Danley DE, Mansour MN, Griffor M, Kamath AV, Cronan M, Mueller J, Sun D, Martin PK, Benton B, McDowell L, Biek K, Schmid MB. Cloning and functional characterization of an NAD (+)-dependent DNA ligase from *Staphylococcus aureus*. J Bacteriol 2001;183:3016–3024.
7. Dermody JJ, Robinson GT, Sternglanz R. Conditional-lethal deoxyribonucleic acid ligase mutant of *Escherichia coli*. J Bacteriol 1979;139:701–704.
8. Wilkinson A, Day J, Bowater R. Bacterial DNA ligases. Molec Microbiol 2001;40:1241–1248.
9. Sriskanda V, Moyer RW, Shuman S. NAD^+-dependent DNA ligase encoded by a eukaryotic virus. J Biol Chem 2001;276:36100–36109.
10. Singleton MR, Hakansson K, Timson DJ, Wigley DB. Structure of the adenylation domain of an NAD^+-dependent DNA ligase. Structure 1999;7:35–42.
11. Lee JY, Chang C, Song HK, Moon J, Yang JK, Kim HK, Kwon ST, Suh SW. Crystal structure of NAD(+)-dependent DNA ligase: Modular architecture and functional implications. EMBO J 2000; 19:1119–1129.
12. Sriskanda V, Shuman S. Conserved residues in domain Ia are required for the reaction of *Escherichia coli* DNA ligase with NAD^+. J Biol Chem 2002;277:9695–9700.
13. Aravind L, Koonin EV. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol 1999;287:1023–1040.
14. Shuman S, Schwer B. RNA capping enzyme and DNA ligase: A superfamily of covalent nucleotidyl transferases. Mol Microbiol 1995;17:405–410.
15. Cheng C, Shuman S. Characterisation of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*. Nuclic Acids Res 1997;25:1369–1374.
16. Weller GR, Doherty AJ. A family of DNA repair ligases in bacteria? FEBS Lett 2001;505:340–342.
17. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R,

Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream M-A, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537–544.

18. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Niel S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 2002;417:141–147.
19. Timson DJ, Wigley DB. Functional domains of an NAD^+-dependent DNA ligase. J Mol Biol 1999;285:73–83.
20. Dermody JJ, Bourguignon GJ, Foglesong PD, Sternglanz R. Nalidixic acid-sensitive and resistant modes of DNA replication in *Escherichia coli*. Biochem Biophys Res Commun 1974;61:1340–1347.
21. Doherty AJ, Serpell LC, Ponting CP. The helix–hairpin–helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA. Nuclic Acids Res 1996;24:2488–2497.
22. Shuman S, Ru XM. Mutational analysis of vaccinia DNA ligase defines residues essential for covalent catalysis. Virology 1995;211:73–83.
23. Sriskanda V, Shuman S. A second NAD^+-dependent DNA ligase (LigB) in *Escherichia coli*. Nucleic Acids Res 2001;29:4930–4934.
24. Sriskanda W, Kelman Z, Hurwitz J, Shuman S. Characterisation of an ATP-dependent DNA ligase from the thermophilic archaeon *Methanobacterium thermoautotrophicum*. Nucleic Acids Res 2000;28:2221–2228.
25. Kodama K-I, Barnes DE, Lindahl T. In vitro mutagenesis and functional expression in *Escherichia coli* of a cDNA encoding the catalytic domain of human DNA ligase I. Nucleic Acids Res 1991;19:6093–6099.