Supplemental Material

DNA Ligases: Progress and Prospects

Stewart Shuman

Molecular Biology Program, Sloan Kettering Institute, New York, NY 10066

Supplemental Figures S1 – S4

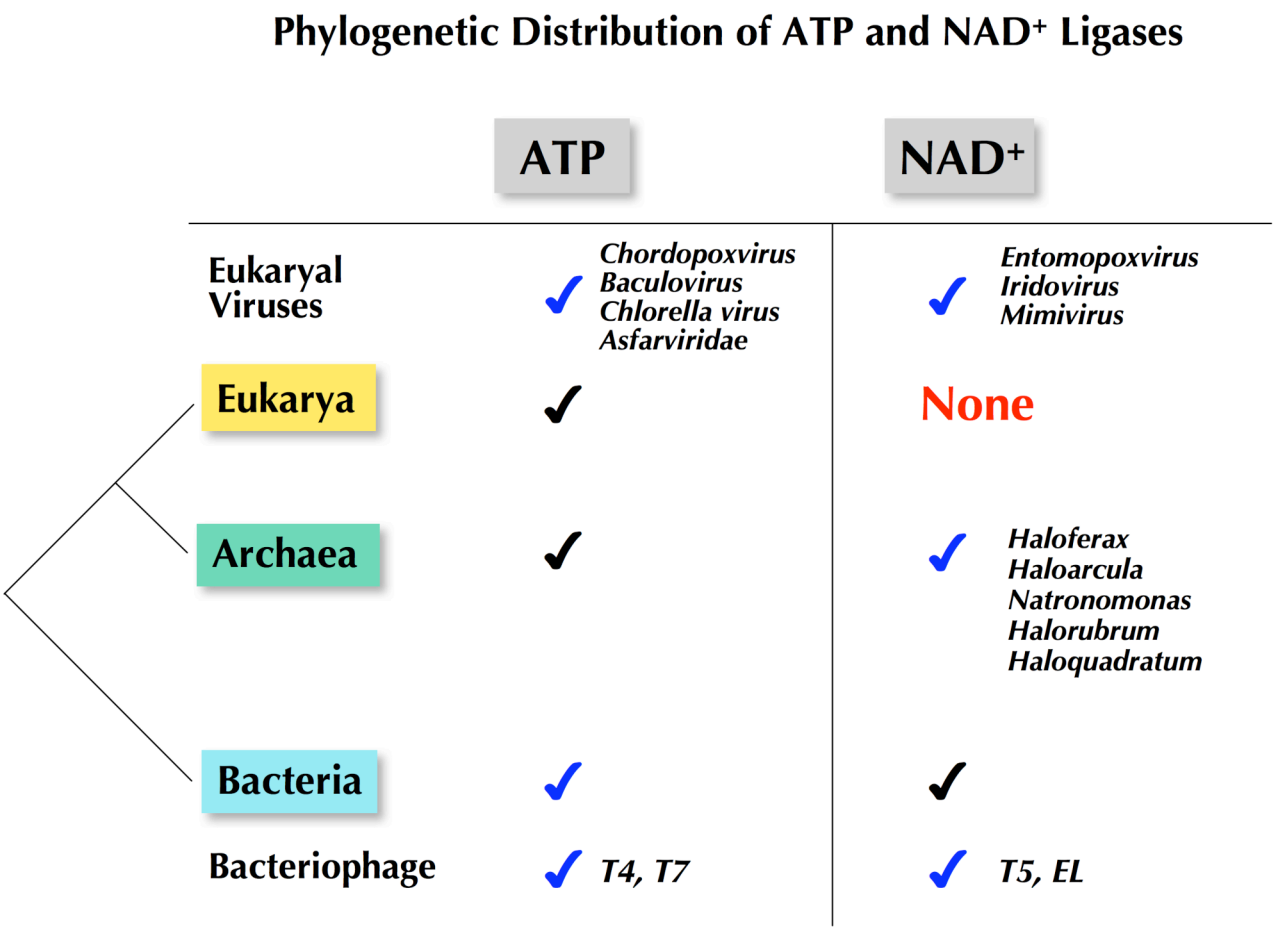
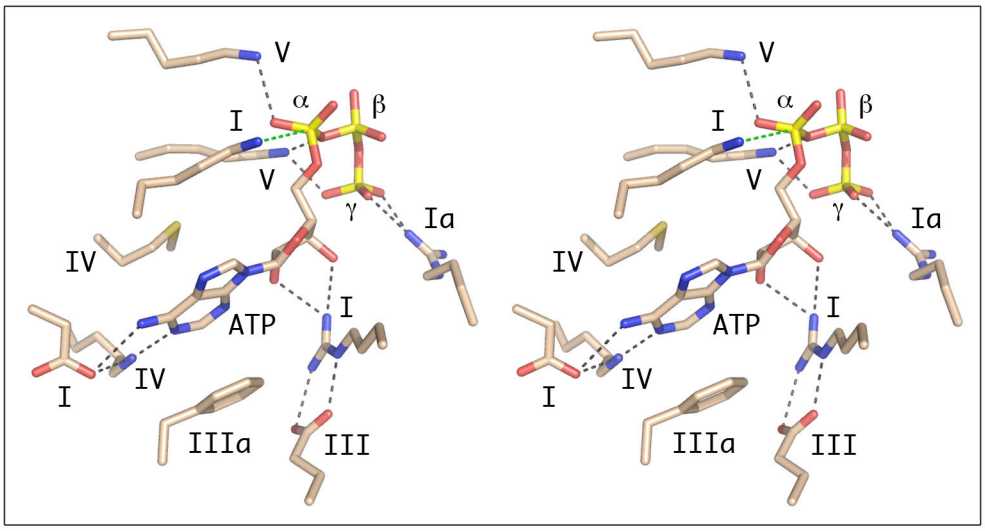


Figure S1. Distribution of ATP-dependent and NAD$^+$-dependent DNA ligases among cellular and viral taxa. Check marks denote the presence of biochemically characterized DNA ligases (or confidently assigned structural homologs) in the proteomes of species within the three domains of cellular life or the proteomes of DNA viruses that replicate in eukaryal or bacterial hosts. Black check marks indicate that ligases of a particular clade are found in all organisms within the domain. Blue check marks specify that a particular type of ligase is found in only some members of the organismal/viral taxon. To date, NAD$^+$-dependent ligases have not been found in any eukaryal organism. Biochemical evidence that NAD$^+$-dependent ligases are encoded by entomopoxviruses, mimivirus, and *Haloferax volcanii i*s reported in the publications cited below.

1. Sriskanda, V., Moyer, R.W., and Shuman, S. (2001) NAD$^+$-dependent DNA ligase encoded by a eukaryotic virus. *J. Biol. Chem.* **276**, 36100-36109.
2. Tong, J., Feng, H., Huang, J., Afonso, C.L., Rock, D.L., Barany, F., and Cao, W. (2004) *Biochim. Biophys. Acta* **1701**, 37-48.
3. Benarroch, D., and Shuman, S. (2006) Characterization of mimivirus NAD$^+$-dependent DNA ligase. *Virology* **353**, 133-143.
4. Zhao, A, Gray, F.C. and MacNeill, S.A. (2006) ATP- and NAD$^+$-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Mol. Microbiol.* **59**, 743-752.
5. Poivedin, L., and McNeill, S.A. (2006) Biochemical characterization of LigN, and NAD$^+$-dependent DNA ligase from the halophilic euryarchaeon *Haloferax volcanii* that displays maximal *in vitro* activity at high salt concentrations. *BMC Mol. Biol*. **7**, 44.

Sulfolobus solfataricus DNA ligase



Covalent Nucleotidyltransferse Superfamily Motifs

	I	Ia	III	IIIa	IV	V
EcoLigA	CELKLDGLA	TRG	LEVRGEVF	FCYGV	DGVVIK	AVAFK
EfaLigA	CELKIDGLA	TRG	VEVRGECY	FLYTV	DGIVIK	AIAYK
TfiLigA	VEHKVDGLS	TRG	LEVRGEVY	TFYAL	DGVVVK	ALAYK
MtuLigA	CELKIDGVA	TRG	LEVRGEVF	ICHGL	DGVVVK	AIAYK
HinLigA	CEPKLDGLA	TRG	LEVRGEVF	NAYGI	DGTVLK	AIAYK
BstLigA	CELKIDGLA	TRG	LEARGEAF	FVYGL	DGIVIK	AIAYK
T7Lig	AEIKYDGVR	SRT	FMLDGELM	KLYAI	EGLIVK	WWKMK
ChVLig	ATPKIDGIR	SRT	EGSDGEIS	YWFDY	EGVMIR	LLKMK
HuLig1	CEYKYDGQR	SRN	FILDTEAV	YAFDI	EGLMVK	WLKLK
SsoLig	VDYKYDGER	SRR	FIIEGEIV	FLFDL	EGVMVK	WIKLK
PfuLig	FEIKYDGAR	SRR	AIVEGELV	NLFDV	EGLMAK	WLKIK
MtuLigD	FEGKWDGYR	SRS	VVLDGEAV	WAFDL	EGVIAK	WVKDK
T4Rnl2	ARERIHGTN	KRT	YQVFGEFA	YVFDI	EGYVLK	AIKCK
TbREL1	ACEKVHGTN	KRS	LVLNGELF	FAFDI	EGVVIR	IIKLR
T4Rnl1	ILTKEDGSL	SKG	FTANFEFV	ILLNV	EGYVAV	HFKIK
PabRnl3	VEEKVDGYN	TRG	LILVGEMA	FLFDV	EGIIMK	IVKYV
ChVCE	VSEKTDGIR	DRA	SIFDGELC	VLFDA	DGLIIM	LFKLK
CalCE	VCEKTDGLR	TRE	TLLDGELV	VIFDA	DGLIYT	LLKWK

Figure S2. **Defining features of the covalent nucleotidyltransferase superfamily**. The superfamily comprises NAD^+-dependent DNA ligases, ATP-dependent DNA and RNA ligases, and GTP-dependent RNA guanylyltransferases (capping enzymes), all of which catalyze NMP transfer from NTP to polynucleotide 5'-monophosphate (ligases) or 5' diphosphate (capping enzymes) termini through a covalent enzyme-(lysine-Nζ)-NMP intermediate. The active site, depicted in stereo in the *top panel* for *Sulfolobus solfataricus* DNA ligase (pdb 2HIX), is composed of amino acids located in the six NTase motifs. Motifs I, Ia, III, IIIa, IV and VI are aligned in the *bottom panel* for all superfamily members for which crystal structures have been determined. These comprise subfamilies of NAD^+-dependent DNA ligases (LigA, from *Escherichia coli*, *Enterococcus faecalis*, *Thermus filiformis*, *Mycobacterium tuberculosis*, *Haemophilus influenzae*, *Bacillus stearothermophilus*), ATP-dependent DNA ligases (phage T7, *Chlorella* virus, human Lig1, *S. solfataricus*, *Pyrococcus furiosus*, *M. tuberculosis* LigD), ATP-dependent RNA ligases (T4 Rnl2, *Trypanosoma brucei* REL1, T4 Rnl1, *Pyrococcus abyssi* Rnl3) and GTP-dependent RNA capping enzymes (from *Chlorella* virus and *Candida albicans*). The motif I lysine is the step 1 nucleophile that forms the covalent lysyl-NMP adduct. In the *Sso*Lig structure, the motif I lysine Nζ is 3.7 Å from the α-phosphorus in an apical orientation (154°) with respect to the bridging oxygen of the PP_i leaving group. The purine base is characteristically sandwiched between the motif IIIa aromatic side chain and a motif IV aliphatic residue (both in green in the bottom panel) Adenine specificity is typically conferred by two charged amino acids: an acidic motif I residue that accepts a H-bond from adenine N7 and a motif IV lysine that donates a H-bond to adenine-N1. The adenine-specificity residues of ATP-dependent DNA/RNA ligases are highlighted in blue in the bottom panel (except in the few cases where these contacts do not apply).

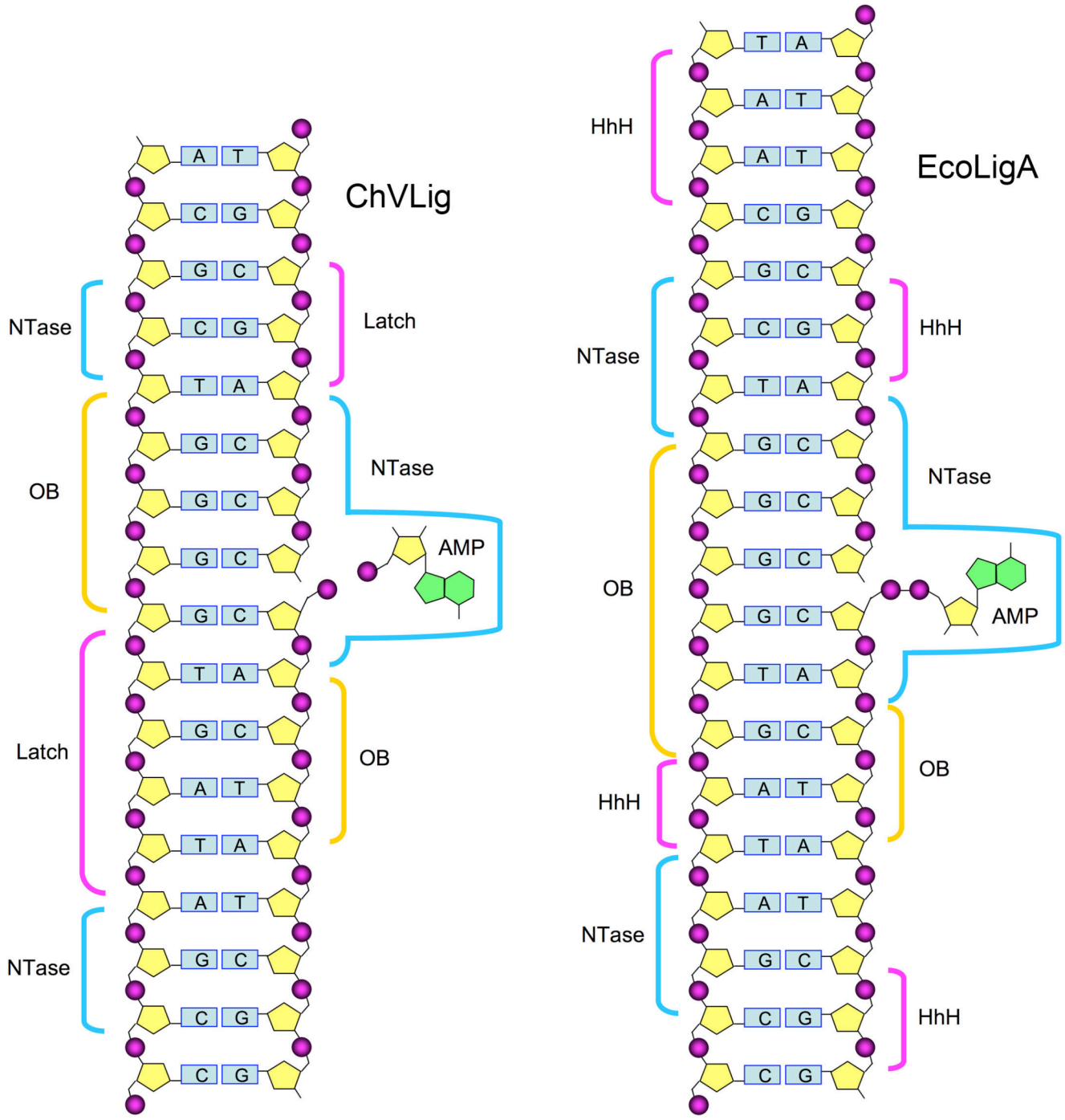


Figure S3. **Domain "footprint" of ChVLig and *E. coli* LigA contacts to DNA.** The nicked duplex DNA is depicted as a two-dimensional cartoon, with the continuous template strand on the left and the nicked strands on the right. The ChVLig structure (left) captures the ligase-adenylate intermediate bound to a 5'-PO_4 nick. The EcoLigA structure (right) captures the ligase bound to nicked DNA-adenylate. In both cases, the extrahelical 5'-adenylate is bound within a conserved AMP-binding pocket of the NTase domain. The DNA strand segments contacted by the individual ligase domain modules are indicated in brackets. The NTase and OB domains have similar DNA footprints in both ligase structures.

Archaeon	Nucleotide specificity
Acidithiobacillus ferrooxidans[1]	ATP
Aeropyrum pernix[2]	ATP ADP
Ferroplasma acidarmanus[3]	ATP
Ferroplasma acidophilum[1]	ATP NAD^+
Methanobacterium thermoautotrophicum[4]	ATP
Picrophilus torridus[1]	ATP NAD^+
Pyrococcus horikoshii[5]	ATP
Staphylothermus marinus[6]	ATP ADP
Sulfophobococcus zilligii[7]	ATP ADP GTP
Sulfolobus acidocaldarius[1]	ATP
Sulfolobus shibatae[8]	ATP
Thermococcus fumicolans[9]	ATP NAD^+
Thermococcus kodakarensis[10]	ATP NAD^+
Thermococccus sp.[11]	ATP NAD^+
Thermoplasma acidophilum[1]	ATP NAD^+

Figure S4. **Diverse nucleotide substrate specificities of archaeal ATP-dependent DNA ligases**. All known archaea have an ATP-dependent DNA ligase. The sources of those archaeal ligase that have been characterized biochemically are listed above; the relevant literature is cited below. The archaeal ligases fall into two broad classes: (i) monospecific ligases that utilize ATP only as the nucleotide substrate; and (ii) dual-specificity ligases that utilize ATP plus another nucleotide. The second nucleotide substrate can be NAD^+ or ADP, as specified. The *S. zilligii* DNA ligase is exceptional in that it is reported to utilize three nucleotide substrates: ATP, ADP, and GTP.

1. Ferrer, M. et al., (2008) *Proc. Natl Acad. Sci. USA* 105, 8878-8883.
2. Jeon, S.J., and Ishikawa, K. (2003) *FEBS Lett*. 550, 69-73.
3. Jackson, B.R. et al. (2007) *Extremophiles* 11, 315-327.
4. Sriskanda, V., et al. (2000) *Nucleic Acids Res*. 28, 2221-2228.
5. Keppetipola, N., and Shuman, S. (2005) *J. Bacteriol*. 187, 6902-6908.
6. Seo, M.S. et al. (2007) *J. Biotechnol*. 128, 519-530.
7. Sun, Y. et al. (2008) *Envir. Microbiol*. 10, 3212-3224.
8. Lai, X. et al. (2002) *Extremophiles* 6, 469-477.
9. Rolland, J. et al. (2004) *FEMS Microbiol Lett.* 236, 267-273.
10. Nakatani, M. et al. (2000) *J. Bacteriol*. 182, 6424-6433.
11. Kim Y.J. et al. (2006) *Biotechnol. Lett*. 28, 401-407.