Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining

Jiafeng Gu and Michael R. Lieber

DNA double-strand breaks (DSB) represent the most deleterious form of DNA damage. Two mechanistically distinctive repair pathways have evolved to mend these breaks in eukaryotes: homologous recombination (HR) and nonhomologous DNA end-joining (NHEJ) (Lieber et al. 2003). HR is restricted to late S or G2 of the cell cycle, whereas NHEJ can function throughout the cell cycle and is the primary repair pathway for DSBs. NHEJ is also distinctive for the mechanistic flexibility of the nucleases, polymerases, and ligases in eukaryotes (Lieber 2007; Lieber et al. 2007).

Initially, NHEJ was thought to be restricted to eukaryotes because the best-studied prokaryote, Escherichia coli, has no ability to join DNA ends (if it did, molecular cloning would have had a different history) (Table 1). Only when bioinformaticists discovered a distantly diverged Ku-like gene in diverse prokaryotic genomes did researchers begin to realize the existence of a similar NHEJ pathway in bacteria (Aravind and Koonin 2001; Doherty et al. 2001; d’Adda di Fagagna et al. 2003). The bacterial Ku homolog is suspected to form a homodimer with a conjectured structure similar to the ring-shaped eukaryotic Ku heterodimer (Weller et al. 2002). The gene for an ATP-dependent ligase named LigD was often found to be adjacent to the Ku gene on the bacterial chromosome (Aravind and Koonin 2001; Doherty et al. 2001; Weller and Doherty 2001). This linkage between Ku and an ATP-dependent ligase prompted more extensive studies and later defined a bacterial NHEJ pathway.

So why do some but not all bacteria retain this DNA repair pathway? Bacterial NHEJ is nonessential under rapid proliferation conditions because HR is active and a duplicate genome is present to provide homology donors (Pitcher et al. 2007a; Shuman and Glickman 2007). However, those bacteria that retain the NHEJ pathway spend much of their life cycle in stationary phase, at which point HR is not available for DSB repair for lack of homology donors. In addition, in nature, desiccation and dry heat are two naturally occurring physical processes that produce substantial numbers of DSBs in bacteria. Hence, bacterial Ku and LigD are present in bacterial species that often form endospores, because during sporulation, NHEJ is efficient for the DSB repair (Pitcher et al. 2007a).

In many bacterial species, unlike the eukaryotic NHEJ ligase IV, LigD is a large, multidomain protein that contains three components within a single polypeptide: a polymerase (POL) domain, a phosphoesterase (PE) domain, and a ligase (LIG) domain (Shuman and Glickman 2007). Therefore, it is tempting to suggest that bacterial Ku and LigD are necessary and sufficient to repair all the DSBs generated in vivo (Della et al. 2004). However, in this issue of Genes & Development, the study by Aniukwu et al. (2008) from the Shuman and Glickman laboratories describes extensive genetic studies combined with in vivo DSB repair analysis in Mycobacterium smegmatis and clearly demonstrates the complexity and flexibility of the NHEJ pathway in bacteria. They uncover a faithful NHEJ pathway specifically for 3’ overhang DSB repair, which is Ku- and LigD-independent. Their results also show that the structure of the broken ends determines the pathway and outcome of the DSB repair.

Complexity within the LigD POL domain

Genetic and biochemical studies have demonstrated that POL X polymerases are involved in the eukaryotic NHEJ pathway (Ma et al. 2004; Bertocci et al. 2006). In mammalian cells, the POL X family consists of pol β, pol μ, pol λ, and TdT. The latter three all function in NHEJ, whereas pol β does not. Pol μ and pol λ are expressed in mammalian somatic cells, but TdT is expressed only in pre-B and pre-T lymphoid cells in the vertebrate immune system.

A crystal structure of the bacterial LigD POL domain has revealed features that are distinct from those of the POL X family polymerases (Zhu et al. 2006; Pitcher et al. 2007b). The LigD POL domain fold more closely resembles the members of the archaeo-eukaryotic primase (AEP) superfamily (Iyer et al. 2005). Although structurally unrelated, the LigD POL domain still shares many
Table 1. Possible corresponding components between NHEJ in prokaryotes and eukaryotes

| Functional component       | Prokaryotes                  | Eukaryotes                                      |
|----------------------------|------------------------------|-------------------------------------------------|
| Toolbelt protein           | Ku (30–40 kDa)               | Saccharomyces cerevisiae                        |
| Polymerase                 | POL domain of LigD           | Multicellular eukaryotes                       |
| Nuclease                   | ?                            | Ku 70/80                                        |
| Kinase/phosphatase         | PE domain of LigD            | Pol4                                            |
| Ligase                     | LIG domain of LigD           | Rad50:Mre11:Xrs2 [& FEN-1]                     |
|                            |                              | Artemis:DNA-PKcs                                |
|                            |                              | Ttp1 and others                                 |
|                            |                              | PKNP and others                                 |
|                            |                              | XLF:XRCC4:Ligase IV                             |

Bacterial Ku shows some distant homology with eukaryotic Ku, and structurally they are suspected to be similar. Eukaryotic Ku is a toroidal and slides onto DNA ends (Walker et al. 2001), and bacterial Ku may do the same. Once bound to the DNA end, Ku can function to recruit other proteins. In this sense, Ku functions like a toolbelt onto which other enzymes (nucleases, polymerases, kinases/phosphatases, and ligases) can bind or debind and do so repeatedly, if necessary. For bacterial NHEJ, multiple activities reside in one polypeptide. In eukaryotic NHEJ, the activities are independent of one another. Many of the end-modifying activities (polymerases or ligases) have similar mechanistic flexibilities, and some of these may be examples of convergent evolution, rather than the divergent evolution of Ku. The nucleases for prokaryotic NHEJ are still the subject of continuing investigation.

Gu and Lieber

functional features with POL X family polymerases, especially polymerase μ. One notable property of the LigD POL domain is that it prefers ribonucleoside triphosphates (rNTPs) over deoxyribonucleoside triphosphates (dNTPs) as substrates in the NHEJ processing step, although templated ribonucleotide addition is restricted to a short RNA tract [Zhu and Shuman 2005b; Yakovleva and Shuman 2006]. This distinctive feature of polymerase activity may be an example of convergent evolution because it is also found in polymerase μ [Nick McElhinny and Ramsden 2003]. This reduced ability to discriminate rNTPs from dNTPs may be advantageous for the repair of chromosomal breaks by the NHEJ pathway in stationary phase when cells are depleted of dNTPs, but still have substantial rNTP pools.

Another unique feature of the LigD POL domain is that it has both template-dependent and template-independent polymerase activities (Della et al. 2004; Gong et al. 2005; Zhu and Shuman 2005b). Aniukwu et al. (2008) demonstrated that the LigD POL domain preferentially adds a single nontemplated nucleotide to the blunt-ended duplex DNA substrate in vivo, which largely accounts for the infidelity of blunt NHEJ. When presented with a 5’ overhang substrate, LigD POL domain performs the fill-in synthesis before sealing. The template-independent activity of the LigD POL domain is also implicated in the ligation between a blunt end and a 3’ overhang end (Della et al. 2004). This theme is a second example of convergent evolution between prokaryotic NHEJ and eukaryotic NHEJ. Human pol μ can add nucleotides in a template-independent manner under physiologic conditions [Gu et al. 2007a]. This strategy in both prokaryotic NHEJ and eukaryotic NHEJ allows random addition at a DNA end to provide “candidate” terminal microhomology sites for annealing between the two DNA ends in instances in which no suitable microhomology exists.

The study by Aniukwu et al. (2008) shows that inactivating the LigD POL domain by mutagenesis is quite different from the deletion of the entire domain. They mutated the metal-binding residues in the LigD POL active site, which abolishes the activities of both template-dependent and template-independent addition. After inactivating the polymerase activity, a slight decrease in the efficiency of blunt NHEJ was observed, whereas the efficiency of 5’ overhang and 3’ overhang NHEJ was not affected. However, the fidelity of both blunt and 5’ overhang is increased, which is partially due to the template-independent activity within the LigD POL domain. Template-dependent fill-ins are still observed in the 5’ overhang NHEJ, which clearly indicates that another mycobacterial polymerase is able to compensate for the loss of the LigD POL activity in the NHEJ pathway. A corresponding functional redundancy is also seen in eukaryotes with both polymerase μ and λ functioning in NHEJ.

However, Aniukwu et al. (2008) observed that the effect of deletion of this LigD POL domain more closely resembles the effect of deleting the entire LigD. These results lead to the conclusion that the LigD POL domain provides an important structural role in addition to its polymerase activity in bacterial NHEJ. Recently, the mycobacterial LigD POL domain has been proposed to mediate the approximation of two incompatible DNA ends [termed synapsis in the recombination field] (Brissett et al. 2007), which may account for the structural role of the LigD POL domain in NHEJ.

The backup ligases are less efficient and less mechanistically flexible than LigD

A dedicated ATP-dependent DNA ligase is a distinctive feature of the NHEJ pathway (Shuman and Glickman 2007). In eukaryotes, ligase IV has been confirmed both genetically and biochemically as the NHEJ ligase [Chotikawong et al. 1997; Grawunder et al. 1997, 1998; Schar et al. 1997]. It forms a complex with XRCC4. XRCC4 stabilizes the ligase IV protein in cells and improves its joining activity by increasing the efficiency of the adenylation of ligase IV [Grawunder et al. 1997; Modesti et al. 1999]. The recently identified XLF factor stimulates the ability of the XRCC4–Ligase IV complex to join incompatible ends [Gu et al. 2007b; Tsai et al. 2007]. The mechanistic flexibility of XRCC4–ligase IV is exempli-
fied by its ability to ligate both incompatible DNA ends and across gaps (Gu et al. 2007a).

Ligation occurs through three consecutive steps (Tomkinson et al. 2006): (1) The ATP-dependent ligase reacts with ATP to form an enzyme–adenylate complex; (2) there is transfer of the activated AMP group from the adenylation to the 5' phosphate group of the broken end; and (3) a 3’ hydroxyl group from the other broken DNA end serves as a nucleophile to attack the phosphoanhydride, and the AMP serves as the leaving group to permit completion of ligation.

Unlike DNA ligase IV, which is the obligatory NHEJ ligase in eukaryotes, the LigD LIG domain is not essential for bacterial NHEJ (Aniukwu et al. 2008). When the entire LIG domain is deleted (∆LIG), 12% and 16% of the joining persists for NHEJ of blunt-end pairs and for compatible 5’ overhang-end pairs, respectively, whereas nearly half of 3’ overhang NHEJ is retained. This result indicates that another mycobacterial ligase is able to compensate for the lack of the LIG domain of LigD. In addition to deleting the entire LIG domain, Aniukwu et al. (2008) further tested another two strains with LIG mutations. In one strain, they replaced the active lysine residue of the LIG domain with an alanine (K484A), which abolishes the ability of the LIG domain to form a LIG-AMP intermediate, but this mutant LIG is still able to seal the preadenylated DNA intermediate. In order to ablate both of the activities, they introduced a new mutation within the active site in the second strain (E533A). By comparison of the NHEJ activities in those three different LIG domain mutant strains, they found that NHEJ activities in ∆LIG more closely resembled those in E533A, and the joining efficiency in K484A is higher than that in the other two. These results indicate that K484A is able to seal the preadenylated DNA intermediate provided by the backup ligase, whose action is slower compared with the LigD LIG domain.

Mycobacteria have another two ATP-dependent DNA ligases: LigB and LigC (Shuman and Glickman 2007). Since LigC is able to accumulate the preadenylated DNA intermediate in vitro (Gong et al. 2004) and bacterial Ku can stimulate the DSB repair by LigC in vitro (Zhu and Shuman 2005a), this pathway joins about one-third of the 3’ overhangs. It would be interesting to determine whether 3’ overhangs are also repaired by the alternative NHEJ pathway.

Bacterial NHEJ nuclease

In eukaryotes, DNA-PKcs and Artemis exist as a complex within cells (Ma et al. 2002). During NHEJ, DNA-PKcs can activate the 5’- and 3’-endonuclease activities of Artemis that are likely responsible for overhang processing in NHEJ (Ma et al. 2002). Ionizing radiation studies suggest that Artemis is required for survival of 10%–20% of mammalian cells, but not the remaining cells (Riballo et al. 2004). This may mean that there are additional nucleases that could contribute; but it could also mean that some of the joining events do not require a nuclease at all. The latter possibility is supported by the fact that some incompatible DNA ends can be ligated simply with Ku and XLF:XRCC4:DNA ligase IV in biochemically defined systems (Gu et al. 2007a,b).

Within the LigD gene, there is a PE domain. Initially, this PE domain was regarded as the nuclease for bacterial NHEJ. Doherty and colleagues (Della et al. 2004) reconstituted mycobacterial NHEJ in vitro and in yeast and reported that the nucleotide resections observed in the joining are due to the LigD PE domain. Another study ectopically expressed mycobacterial Ku and LigD in E. coli and observed large amounts of nucleotide deletion in the plasmid DSB repair assay, which was also attributable to the LigD PE domain (Malyarchuk et al. 2007). When Doherty and colleagues (Pitcher et al. 2005) purified the PE domain separately, they found that it did not have 3’-exonuclease activity. Shuman and colleagues (Zhu and Shuman 2005a) found that the LigD PE domain has 3’-ribonuclease and 3’-phosphatase activities, but they did not detect 3’-exonuclease activity in this domain. In vivo, Aniukwu et al. (2008) reported that nucleotide resections observed in the joining are not decreased when comparing NHEJ activity in LigD PE mutant strains with wild-type strains of M. smegmatis. It is possible that 3’-phosphatase activity is involved in those in vivo DSBs with 3’-phosphate or 3’-phosphoglycolate termini, which must be removed before ligation. This activity has not yet been systematically tested because the linearized plasmids generated in vitro have no 3’ phosphates on the overhangs. Further biochemical and in vivo studies will be of substantial interest for determining the full range of activities of the PE domain and the role of any additional nucleases.

Alternative NHEJ pathway

An additional instructive aspect of the study by Aniukwu et al. (2008) is that they identified a Ku- and LigD-independent NHEJ pathway specifically for 3’ overhangs that are fully compatible with 4 base pairs of microhomology. In wild-type cells, this pathway joins about one-third of the 3’ overhangs with 100% fidelity. Since this special joining activity is retained in the triple knockout of all the ATP-dependent ligases, the only remaining ligase in the mutant strain, LigA, is considered to be responsible for this special faithful 3’ overhang joining. LigA is an NAD+-dependent ligase, which is conserved in all bacteria and is essential for bacterial growth (Shuman and Glickman 2007). E. coli has no NHEJ activity and cannot repair any DSBs, even compatible 3’ overhangs. It would be interesting to determine whether some other mycobacterial-specific factors are involved.
in this LigA-facilitated 3’ overhang-joining. In addition, what is the evolutionary advantage of retaining this faithful 3’ overhang-joining in mycobacteria? Although the alternative role of LigA is intriguing, substantial lengths of microhomology would be rare at naturally occurring DSBs in vivo, and hence this pathway might turn out to be more of experimental interest.

Future directions

In addition to many interesting mechanistic questions, this line of work opens up other important avenues. Mycobacteria, in infections such as tuberculosis, are notorious for persisting in stationary phase for long periods. Host defense against mycobacterial infections include the usual oxidative attack on the bacteria, which would cause some DSBs. Inhibition of bacterial NHEJ potentially provides promising new avenues in a key area of infectious disease research. Hence, insight into bacterial NHEJ has interesting evolutionary relevance, but also important public health implications.

Acknowledgments

Research in M.R.L.’s laboratory is supported by NIH grants. The Perspective format limits the number of citations, and we apologize to authors who were not directly cited.

References

Aniukwu, J., Glickman, M.S., and Shuman, S. 2008. The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. Genes & Dev [this issue], doi: 10.1101/gad.1631908.
Aravind, L. and Koonin, E.V. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res. 11: 1365–1374.
Bertocci, B., De Smet, A., Weill, J.C., and Reynaud, C.A. 2006. Mechanism of non-homologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku ligase D and ligase C. Nat. Struct. Mol. Biol. 12: 304–312.
Gong, C., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S., and Glickman, M.S. 2005. Mechanism of non-homologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku ligase D and ligase C. Nat. Struct. Mol. Biol. 12: 304–312.
Grawunder, U., Wilm, M., Wu, X., Kuleszsa, P., Wilson, T.E., Mann, M., and Lieber, M.R. 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature 388: 492–495.
Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M.R. 1998. DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. Mol. Cell 2: 477–484.
Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M.F., and Lieber, M.R. 2007a. XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. EMBO J. 26: 1010–1023.
Gu, J., Lu, H., Tsai, A.G., Schwarz, K., and Lieber, M.R. 2007b. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4–DNA ligase IV complex: Influence of terminal DNA sequence. Nucleic Acids Res. 35: 5755–5762.
Iyer, L.M., Koonin, E.V., Leipe, D.D., and Aravind, L. 2005. Origin and evolution of the archea-eukaryotic primase superfamily and related palm-domain proteins: Structural insights and new members. Nucleic Acids Res. 33: 3875–3896.
Lieber, M.R. 2007. The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. 283: 1–5.
Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. Nat. Rev. Mol. Cell Biol. 4: 712–720.
Lieber, M.R., Lu, H., Gu, J., and Schwarz, K. 2007. Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: Relevance to cancer, aging, and the immune system. Cell Res. 18: 125–133.
McElhinney, S.A. and Ramsden, D.A. 2003. Polymerase μ is a DNA-directed DNA/RNA polymerase. Mol. Cell. Biol. 23: 2309–2315.
Millar, S., Wright, D., Castore, R., Klepper, E., Weiss, B., Doherty, A.J., and Harrison, L. 2007. Expression of Mycobacterium tuberculosis Ku and Ligase D in Escherichia coli results in RecA and RecB-independent DNA end-joining at regions of microhomology. DNA Repair (Amst.) 6: 1413–1424.
Malyarchuk, S., Wright, D., Castore, R., Klepper, E., Weiss, B., Doherty, A.J., and Harrison, L. 2007. Expression of Mycobacterium tuberculosis Ku and Ligase D in Escherichia coli results in RecA and RecB-independent DNA end-joining at regions of microhomology. DNA Repair (Amst.) 6: 1413–1424.
Modesti, M., Hesse, J.E., and Gellert, M. 1999. DNA binding of Xrc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. EMBO J. 18: 2008–2018.
Nick McElhinney, S.A. and Ramsden, D.A. 2003. Polymerase μ is a DNA-directed DNA/RNA polymerase. Mol. Cell. Biol. 23: 2309–2315.
Pitcher, R.S., Tenkin, L.M., Green, A.J., and Doherty, A.J. 2005. Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. J. Mol. Biol. 351: 531–544.
Pitcher, R.S., Brissett, N.C., and Doherty, A.J. 2007a. Non-homologous end-joining in bacteria: A microbial perspective. Annu. Rev. Microbiol. 61: 259–282.
Pitcher, R.S., Brissett, N.C., Picher, A.J., Andrade, P., Juarez, R.,
Thompson, D., Fox, G.C., Blanco, L., and Doherty, A.J. 2007b. Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J. Mol. Biol.* 366: 391–405.

Riballo, E., Kuhne, M., Rief, N., Doherty, A., Smith, G.C., Recio, M.J., Reis, C., Dahm, K., Fricke, A., Krempler, A., et al. 2004. A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to γ-H2AX foci. *Mol. Cell* 16: 715–724.

Schar, P., Herrmann, G., Daly, G., and Lindahl, T. 1997. A newly identified DNA ligase of *S. cerevisiae* involved in RAD52-independent repair of DNA double-strand breaks. *Genes & Dev.* 11: 1912–1924.

Shuman, S. and Glickman, M.S. 2007. Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.* 5: 852–861.

Tomkinson, A.E., Vijayakumar, S., Pascal, J.M., and Ellenberger, T. 2006. DNA ligases: Structure, reaction mechanism, and function. *Chem. Rev.* 106: 687–699.

Tsai, C.J., Kim, S.A., and Chu, G. 2007. Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc. Natl. Acad. Sci.* 104: 7851–7856.

Walker, J.R., Corpina, R.A., and Goldberg, J. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412: 607–614.

Weller, G.R. and Doherty, A.J. 2001. A family of DNA repair ligases in bacteria? *FEBS Lett.* 505: 340–342.

Zhu, H. and Shuman, S. 2005b. A primer-dependent polymerase function of *Pseudomonas aeruginosa* ATP-dependent DNA ligase (LigD). *J. Biol. Chem.* 280: 418–427.

Zhu, H. and Shuman, S. 2007. Characterization of *Agrobacterium tumefaciens* DNA ligases C and D. *Nucleic Acids Res.* 35: 3631–3645.

Zhu, H., Nandakumar, J., Aniukwu, J., Wang, L.K., Glickman, M.S., Lima, C.D., and Shuman, S. 2006. Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. *Proc. Natl. Acad. Sci.* 103: 1711–1716.