Bacterial NHEJ: A never ending story
Claire Bertrand, Annabelle Thibessard, Claude Bruand, François Lecointe, Pierre Leblond

To cite this version:
Claire Bertrand, Annabelle Thibessard, Claude Bruand, François Lecointe, Pierre Leblond. Bacterial NHEJ: A never ending story. Molecular Microbiology, 2019, 10.1111/mmi.14218. hal-02044826

HAL Id: hal-02044826
https://hal.univ-lorraine.fr/hal-02044826
Submitted on 21 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Bacterial NHEJ: A never ending story

Claire Bertrand¹, Annabelle Thibessard¹, Claude Bruand²*, François Lecointe³*, Pierre Leblond¹*

¹Université de Lorraine, Inra, DynAMic, F-54000 Nancy, France
²Laboratoire des Interactions Plantes-Microorganismes, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France.
³Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

*Co-corresponding authors

Running title: Growing complexity of bacterial NHEJ

Key words: Non-Homologous End Joining, Streptomyces, Sinorhizobium meliloti, Bacillus subtilis, Mycobacterium
Abstract

Double-strand breaks (DSBs) are the most detrimental DNA damage encountered by bacterial cells. DSBs can be repaired by homologous recombination thanks to the availability of an intact DNA template or by Non-Homologous End Joining (NHEJ) when no intact template is available. Bacterial NHEJ is performed by sets of proteins of growing complexity from *Bacillus subtilis* and *Mycobacterium tuberculosis* to *Streptomyces* and *Sinorhizobium meliloti*. Here we discuss the contribution of these models to the understanding of the bacterial NHEJ repair mechanism as well as the involvement of NHEJ partners in other DNA repair pathways. The importance of NHEJ and of its complexity is discussed in the perspective of regulation through the biological cycle of the bacteria and in response to environmental stimuli. Finally, we consider the role of NHEJ in genome evolution, notably in horizontal gene transfer.
Until *in silico* identification of putative NHEJ actors in bacterial genomes, it was thought that Homologous Recombination (HR) was the major double-strand break (DSB) repair mechanism in prokaryotes. However, HR needs an intact template to ensure repair of the break, a condition encountered by bacteria only during their dividing state and raising the question of the repair of DSB in other phases of their life cycle. The identification of a bacterial NHEJ answered, at least in part, this question since, in eukaryotes, NHEJ was known to ligate DNA ends together without any homologous template. This mechanism has been well documented in eukaryotes and involves a dozen of actors in human, including the Ku70/Ku80 heterodimer, and the ligase IV (Chang *et al.*, 2017).

The first hint for the existence of a bacterial NHEJ was provided by *in silico* approaches leading to the identification of Ku70/80 homologs and ATP-dependent DNA-ligase homologs (Doherty *et al.*, 2001; Aravind and Koonin, 2001). These Ku and DNA-ligase encoding genes are often colocalized, or in an operonic structure. The DNA-ligase genes encode a putative ligase domain (LigDom or LIG), fused or not with a polymerase domain (PolDom, POL or Prim-Pol) and/or a nuclease domain (NucDom or PE for Phosphoesterase). Whatever the domain combination, these multifunctional ligases were collectively named LigD, in reference to the main NHEJ ligase in Mycobacteria (Gong *et al.*, 2005).

The first experimental validation of a bacterial NHEJ came from the study of Weller *et al.* showing that *in vitro* the *Mycobacterium tuberculosis* LigD protein (LigD<sub>Mtub</sub>) is indeed an ATP-dependent DNA ligase that is stimulated by its cognate Ku<sub>Mtub</sub> partner, possibly via an interaction between Ku<sub>Mtub</sub> bound to DNA and LigD<sub>Mtub</sub> (Weller *et al.*, 2002). In parallel, authors described the
deleterious effect of the deletion of the *Bacillus subtilis* genes coding for LigD<sub>Bsub</sub> and Ku<sub>Bsub</sub> (ykoU-ykoV operon) on the survival of cells in stationary phase after ionizing radiation (IR). IR is a cause of multiple DNA lesions, including DSBs that are particularly hazardous. These two proteins from *B. subtilis* have been later shown to also perform NHEJ repair *in vitro* (de Vega, 2013). The Ku and/or LigD-dependent ability of *M. tuberculosis* and *Mycobacterium smegmatis* to recircularize linear plasmid DNA after transformation, and the IR sensitivity of *M. smegmatis* ku and ligD mutants in stationary phase, have given further evidence *in vivo* that several bacterial species are NHEJ proficient (Stephanou *et al.*, 2007).

LigD<sub>Mtub</sub> is not a common ligase. Its PolDom provides LigD<sub>Mtub</sub> with DNA-dependent DNA polymerase, DNA-dependent RNA polymerase and primase and non-templated nucleotide addition activities (Della *et al.*, 2004). Moreover, the Poldom interacts *in vitro* with Ku<sub>Mtub</sub>-DNA complex suggesting that it could be responsible for the recruitment of LigD<sub>Mtub</sub> on Ku-bound DSB *in vivo* (Pitcher *et al.*, 2005). LigD<sub>Mtub</sub> also shows a 3’-5’ exonuclease activity most probably carried by its NucDom (Della *et al.*, 2004). LigD from *Pseudomonas aeruginosa* (LigD<sub>Paer</sub>) and *Agrobacterium tumefaciens* (LigD2) display similar activities accordingly to their identical domain composition, though different domain organization, compared to LigD<sub>Mtub</sub> (Zhu and Shuman, 2005a; Zhu and Shuman, 2005b; Zhu *et al.*, 2005; Zhu and Shuman, 2007). Thus, LigD displays all activities allowing to process DNA ends, if necessary, and ligate them without the need for other proteins, except Ku. Accordingly, expression of Ku and LigD from *M. tuberculosis i)* partially restores NHEJ in yeast ku70 or dnl4 mutant strains (Della *et al.*, 2004) and ii) renders *Escherichia coli* NHEJ-proficient (Malyarchuk *et al.*, 2007).

Altogether, these studies have led to the following model for the bacterial NHEJ: Ku homodimer binds to DNA ends generated by a break and recruits LigD (one multidomain polypeptide, such
as LigD_{Mtub}, or a holoenzyme consisting of several subunits similar to the LigDom, PolDom and/or NucDom of LigD_{Mtub} probably via a direct physical interaction on the DNA between Ku and the PolDom of the LigD holoenzyme. LigD is able to process DNA ends, if broken ends are non-directly ligatable, thanks to its Poldom and/or Nucdom, before ligation by LigDom (Pitcher et al., 2007a).

Most of the past studies on bacterial NHEJ have already been extensively reviewed (Bowater and Doherty, 2006; Pitcher et al., 2007a; Shuman and Glickman, 2007; Gu and Lieber, 2008; Brissett and Doherty, 2009), with the exception of the most recent ones on structural data of different LigD domains and mechanistic details of DNA end processing by NHEJ actors before ligation (Nair et al., 2010; Zhu and Shuman, 2010; Brissett et al., 2011; Brissett et al., 2013). The archaea *Methanocella paludicola* contain genes coding for homologs of PolDom, NucDom, LigDom and Ku. Their characterization have also deepened our knowledge on the processing of DNA ends during NHEJ (Bartlett et al., 2013; Brissett et al., 2013; Bartlett et al., 2016). In the present report, we wished to focus on recent data obtained with different bacterial models highlighting the complexity of the prokaryotic NHEJ pathway in terms of involved actors, their interactions, their regulation, their links with other DNA repair pathways and finally their impact on bacterial evolution.

II- Multiple components NHEJ pathways

The above model of minimal NHEJ, with one Ku and one LigD holoenzyme performing bacterial NHEJ, does not take into account the existence of other proteins involved in this mechanism. These proteins can be orthologues of Ku or of domains of the LigD holoenzyme, and/or proteins interacting with them. Figure 1 illustrates diverse levels of NHEJ complexity in four bacterial models.
A recent and large scale search for \textit{ku}-like genes suggests the existence of a potential NHEJ repair system in about 25% of the prokaryotes for which a genome was available in public databases (McGovern \textit{et al.}, 2016), with a scattered distribution. Among the bacterial species predicted to possess a NHEJ repair system, some actually encode several putative Ku homologs, mainly actinobacteria of the streptomycetes family, and proteobacteria of the \(\alpha\) subdivision (McGovern \textit{et al.}, 2016). For instance, more than 70\% of \textit{Streptomyces} species encode more than one \textit{ku} gene (up to 4 in \textit{Streptomyces coelicolor}) (Hoff \textit{et al.}, 2016), and more than half of \(\alpha\)-proteobacteria possessing a \textit{ku} gene actually have two or more (up to 6 in \textit{Rhizobium leguminosarum bv trifolii}). Some of the \textit{ku} genes are located on plasmids, suggesting their putative acquisition via horizontal gene transfer. Importance of multiple \textit{ku} genes in some NHEJ proficient bacteria is discussed further below.

Putative NHEJ ligases are also often present in multiple copies. However, because of the more variable organization of LigD, it is more difficult in this case to predict whether the genes encode \textit{bona fide} NHEJ ligases. This difficulty is reinforced by the fact that several ATP-dependent DNA ligases can be used to perform NHEJ in the same bacterium. For instance, \textit{M. smegmatis} encodes four ATP-dependent DNA ligases (LigB\textsubscript{Msme}, LigC\textsubscript{1Msme}, LigC\textsubscript{2Msme} and LigD\textsubscript{Msme}) (Gong \textit{et al.}, 2005). Among them, LigD\textsubscript{Msme} is the main NHEJ ligase and contains the PolDom, LigDom and NucDom domains, as its homolog LigD\textsubscript{Mtub}. However, LigC\textsubscript{1Msme} (a LigDom-only enzyme) can act with Ku in an alternative NHEJ in \textit{M. smegmatis} (Gong \textit{et al.}, 2005; Akey \textit{et al.}, 2006; Aniukwu \textit{et al.}, 2008; Bhattarai \textit{et al.}, 2014). This LigC\textsubscript{1Msme}-dependent alternative NHEJ requires the PolDom of LigD\textsubscript{Msme}, the LigDom and NucDom being dispensable for this mechanism (Bhattarai \textit{et al.}, 2014). This result, supplemented by the fact that the PolDom of LigD\textsubscript{Msme} plays a structural role
independent of its polymerase activity in the efficiency of recircularization of linearized plasmid in vivo (Aniukwu et al., 2008), highlights that the PolDom could serve as an interaction platform for an ATP-dependent ligase protein containing only a LigDom. Other LigC proteins, LigC1 and LigC2 from *A. tumefaciens* are endowed with ligase activities that are stimulated by Ku (Zhu and Shuman, 2007), suggesting that alternative NHEJ pathways are conserved across bacterial species. It has to be noticed that *M. smegmatis* contains two *polD* genes expressing two PolDom-only polymerases that can suppress partially the sensitivity to IR of a mutant encoding a polymerase deficient LigD, suggesting that, as observed with LigC1, several PolDom proteins can be used in the NHEJ pathway in Mycobacteria (Bhattarai et al., 2014).

In *Streptomyces*, PolDom, LigDom and NucDom only exist as stand-alone domains and not as multifunctional proteins, making even more complex the identification of actual NHEJ partners. The analysis of 38 *Streptomyces* genomes revealed a conserved gene set, composed of two loci, *kuA* and its neighbor *polK* and *ligC* (encoding a putative ligase displaying an higher level of homology with LigC1 from *M. smegmatis* compared to LigD<sub>Mtub</sub>) and its neighbor *polR* (*polK* and *polR* being two PolDom encoding genes). Surprisingly, *ligD*, whose product is the closest homolog of LigD<sub>Mtub</sub>, is not ubiquitous in these genomes (present in 75% of the analyzed species), suggesting that the core *ligC* gene provides the main NHEJ ligase function. The variable set of *Streptomyces* NHEJ-like genes also includes *kuB* and *kuC*, the latter often being plasmid borne and often laying in the close vicinity of *ligD*. Although they are not conserved, these genes take part in DNA repair since their mutation in *Streptomyces ambofaciens* confers spore sensitivity to IR and increases the sensitivity of the core gene set mutants (Hoff et al., 2016). In *S. ambofaciens*, the involvement of KuA in a NHEJ system has been demonstrated together with the existence of
a secondary illegitimate DSB repair system that could imply these variable genes (Hoff et al., 2018).

The *Streptomyces* organization also questions the role of the two conserved PolDom encoding genes (*polK* and *polR*). In this regard, *Streptomyces* is not an exception since *M. smegmatis* (see above and below, PolD1, PolD2) and numerous eukaryotes, especially vertebrates, also possess several polymerases involved in NHEJ (Ramsden and Asagoshi, 2012). A first hypothesis to explain this observation is based on the difference in expression profile: indeed, *S. ambofaciens* *polR* seems constitutively expressed whereas *polK* transcription was not detected in the tested conditions (Hoff et al., 2016), but might be induced in response to some genotoxic stress. Alternatively or concomitantly, these PolDoms could display different template requirements, as for the four polymerases of the X family found in vertebrates (Polβ, Polλ, Polμ and TdT, for review see Ramsden and Asagoshi, 2012). Therefore, each PolDom could act specifically on different types of damaged DNA ends. In favor of this hypothesis, in *M. smegmatis*, a mutation of the PolDom of LigD abolishes the nucleotide addition but not the fill-in activity suggesting that at least two polymerase activities (probably driven by PolD1 and/or PolD2, Bhattarai et al., 2014) would be required for full DSB repair achievement (Aniukwu et al., 2008). Finally, due to its similarity with the mycobacterial locus encoding *ligC1* and *prim-polC* involved in a BER system (Płociński et al., 2017), we surmise that the *Streptomyces ligC-polR* locus could also be involved in this DNA repair pathway in addition to its role in NHEJ.

Another bacterium displaying a complex organization is *Sinorhizobium meliloti*, an α-proteobacterium which encodes four Ku and four LigD orthologues (Galibert et al., 2001); here, LigD proteins are defined as multidomain ATP-dependent ligases, with LigD3 and LigD1 lacking a Nuc and a Pol domain, respectively. The Ku proteins of α-proteobacteria can be classified in
several families based on sequence similarity, but a “representative” Ku protein present in all α-
proteobacteria could not be identified. Single and multiple *S. meliloti* *ku* mutations conferred IR
sensitivity to stationary phase bacteria, suggesting a role of these genes in DNA repair (Kobayashi
*et al.*, 2008; Dupuy *et al.*, 2017). Involvement of the various *ku* and *ligD* genes in NHEJ was
assessed more precisely using a linear plasmid recircularization assay, which showed the
existence of at least two distinct NHEJ pathways (Fig. 1) (Dupuy *et al.*, 2018). The first major
pathway is dependent on *ku2* and *ligD2*, two adjacent genes apparently organized in operon,
although they can be transcribed from independent promoters. A second, less efficient pathway,
dependent on *ku3*, *ku4* and *ligD4* was also detected in stress conditions (see further below). Both
*ku3* and *ku4* are organized in operon, whereas *ligD4* is located elsewhere in the genome, but all
three genes are co-regulated (see further below) and *ku3*, *ku4* and *ligD4* orthologues are often
found in close proximity, mostly in operon, in α-proteobacterial genomes. Interestingly, genetic
analyses showed that both *ku3* and *ku4* are needed for activity of this second pathway. This
observation, together with the quasi-systematic co-occurrence and co-regulation of *ku3*-
and *ku4*-like genes in NHEJ equipped α-proteobacteria, led to the suggestion that Ku3 and Ku4 might
function as heterodimers, like eukaryotic Ku proteins. Although LigD4 was systematically found
associated to the Ku3-Ku4-dependent pathway, LigD2, and to a much lesser extent LigD3, are
also able to function, probably aspecifically, with Ku3 and Ku4 on certain types of DNA ends and
in certain genetic contexts. Both *ku1* and *ligD1* are weakly expressed, and consistently, no NHEJ
activity could be associated with these genes in the tested conditions. PolD1 (also called LigD1p,
Kobayashi *et al.*, 2008), encoded by the gene located just downstream *ligD1*, is likely the vestigial
PolDom of a larger ancestral LigD1 protein. Deletions of *ligD1* analyzed by Dupuy *et al.* (2018)
encompassed *polD1*, showing that this gene is not involved in NHEJ in the tested conditions.
S. meliloti also encodes three ligC and one ligB genes, whose role in NHEJ or another DNA repair system has not been tested so far (Fig. 1).

Inactivation of NHEJ in M. smegmatis and S. meliloti unmasked a background level of Ku- and LigD-independent end-joining activity, called alternative End-Joining (A-EJ), observed during recircularization of linear transforming plasmids (Aniukwu et al., 2008; Dupuy et al., 2018). Depending on the bacterial species and the DNA end type, the observed efficiencies and fidelity of repair greatly varied. It is unclear whether A-EJ occurs by the same mechanism in all cases, but as it was observed in M. smegmatis ligBCD mutants, it is believed to depend on the replicative ligase LigA. Interestingly, LigA-dependent, LigB-independent A-EJ has been described in the NHEJ-deficient E. coli (Chayot et al., 2010). When not resulting in faithful ligation of compatible DNA ends (which occurred in some cases with 100% efficiency; Aniukwu et al., 2008; Dupuy et al., 2018), A-EJ repair events mostly involved deletions of variable lengths, with frequent occurrence of microhomologies at the junctions (Aniukwu et al., 2008), very similar to the observed products of A-EJ in E. coli (Chayot et al., 2010). These observations highlight the importance of DNA end resection for homology search, which in E. coli is in part due to RecBCD. The resection enzymes involved in the A-EJ of NHEJ-proficient bacteria have not been characterized yet.

**NHEJ complexity: are there other actors involved?**

Proteins other than Ku and LigD domains have been suggested to be involved in NHEJ, either via their physical and/or functional interaction with Ku and/or LigD or via the impact of their inactivation on DSB repair. This is the case for the helicase UvrD1 from Mycobacteria, which interacts with Ku in a yeast two-hybrid assay and which forms ternary complex with Ku and dsDNA (Sinha et al., 2007). The fact that the helicase activity of UvrD1 is stimulated in presence of KuMsme and that the deletion of uvrD1 in M. smegmatis sensitizes this bacterium to IR makes
sense with a role of UvrD1 in DSB repair via NHEJ. However, its role in this process remains to be elucidated.

Another protein interacting with the NHEJ actors in Mycobacteria is Sir2. This deacetylase interacts physically with both Ku and LigD and its inactivation decreased the NHEJ efficiency in vivo (Li et al., 2011). Its putative role in the NHEJ process is discussed below (see “Regulation of NHEJ”).

**NHEJ complexity: the multiple DNA binding abilities of Ku**

Bacterial Ku is composed of a N-terminal Ku core domain homologous to the central domain of the Ku70 and Ku80 eukaryotic homologs (Aravind and Koonin, 2001; Kushwaha and Grove, 2013a; McGovern et al., 2016; Pitcher et al., 2007a). Ku70/80 heterodimer has been shown to bind DSB and to form a ring (involving the Ku core domains of the two subunits) able to encircle a dsDNA molecule (Walker et al., 2001). This particular structure of the eukaryotic Ku allows its threading inward the DNA molecule when additional Ku proteins enter by the ends (de Vries et al., 1989; Paillard and Strauss, 1991; Downs and Jackson, 2004). This threading ability is conserved at least for Ku from *B. subtilis*, *M. tuberculosis* and most probably *M. smegmatis* (Weller et al., 2002; Kushwaha and Grove, 2013a; McGovern et al., 2016). The Ku core domain of bacterial Ku is associated to a variable C-terminal (C-ter) tail which contains the ubiquitous minimal C-ter domain and, in most cases, an extended C-ter tail (McGovern et al., 2016). The minimal C-ter is required for the interaction of Ku<sub>Subl</sub> with LigD<sub>Subl</sub>, is involved in the stimulation of the ligase activity of LigD<sub>Subl</sub> and LigD<sub>Paer</sub> and is contained in a segment of Ku<sub>Mtub</sub> interacting with LigD<sub>Mtub</sub> in a yeast two-hybrid assay (Gong et al., 2005; Zhu and Shuman, 2010; McGovern et al., 2016). These observations led to the proposal that the minimal C-ter is a part of the LigD-
interacting domain in bacterial Ku. The extended C-ter is predominantly a domain rich in basic residues (lysine and arginine) that is involved, but not required, in the Ku-dependent stimulation of the ligase activity of LigD in *P. aeruginosa* and *B. subtilis* (Zhu and Shuman, 2010; McGovern et al., 2016). This domain also confers to Ku the ability to bind DNA without free ends (Kushwaha and Grove, 2013b; McGovern et al., 2016) and limits the threading of Ku*sub* inward the DNA molecule. Its function could be to reduce putative negative effects of this threading inside the chromosome on other DNA repair processes. The regulation of the threading of Ku may also be important for the recruitment of LigD near, or at the DSB site, ensuring an efficient NHEJ repair since LigD alone is a poorly efficient ligase without its Ku partner *in vitro* (Weller et al., 2002; Della et al., 2004; Zhu and Shuman, 2010; de Ory et al., 2014; McGovern et al., 2016). The function of Ku in the NHEJ could be more proactive than previously anticipated by allowing the bridging of DNA ends, thanks to its extended C-ter (McGovern et al., 2016), an ability shared by the human Ku (Ramsden and Gellert, 1998). The different functions of this extended C-ter domain raise the question about the efficiency of the NHEJ in bacteria encoding a Ku without this domain and/or about the role of Ku and its extended C-ter in another DNA repair mechanism (for instance the BER process, see below). In *Streptomyces*, one of the accessory Ku (KuB), present in only 8 of 38 species (Hoff et al., 2016), possesses an extra C-ter domain weakly related to the RhoN domain (Pfam 07498) which role in Ku protein remains to be elucidated.

Finally, a very recent study demonstrated that the eukaryotic Ku orthologue encoded by the bacteriophage Mu (d’Adda di Fagagna et al., 2003) is able to promote NHEJ repair of plasmidic and chromosomal DSBs, most probably by interacting with and stimulating the *E. coli* replicative ligase LigA (Bhattacharyya et al., 2018). Surprisingly, MuGam also stimulates *in vitro* the LigB ligase of *E. coli*, although MuGam and LigB do not seem to interact on DNA. These results raise
the possibility that a bacterial Ku protein might also increase the ligase activity of a non-NHEJ ligase. In this line, Ku<sub>Msmeg</sub> (Kushwaha and Grove, 2013a), but not Ku<sub>Mtub</sub> (Weller <i>et al.</i>, 2002), has been shown to stimulate the T4 DNA ligase <i>in vitro</i>. To our knowledge, no study assessed the possibility that a bacterial Ku redirects the activity of a bacterial replicative ligase.

III- Regulation of NHEJ

In eukaryotes, NHEJ repairs DSBs throughout the whole cell cycle and is predominant in G1 phase while HR relies on the presence of an intact sister chromatid and therefore functions in S/G2 phase (Ira <i>et al.</i>, 2004; Aylon <i>et al.</i>, 2004; Clerici <i>et al.</i>, 2008). Similarly, during quiescent states such as sporulation (<i>Bacillus</i>, <i>Streptomyces</i>) or late-stationary phase (<i>Mycobacterium</i>, <i>Sinorhizobium</i>) in bacteria, HR would not be operational because stationary phase cells and spores are assumed to be single genome compartments. Note however that spores (Setlow and Setlow, 1996; Moeller <i>et al.</i>, 2007; Hoff <i>et al.</i>, 2016) or stationary phase bacteria (Pitcher <i>et al.</i>, 2007b; Dupuy <i>et al.</i>, 2017) of HR-deficient recA mutants display IR sensitivity. This is most likely because of the involvement of recA in the repair of other DNA damages, in particular via its role in the SOS response, since i) spores of <i>B. subtilis</i> addAB and recJ mutants, deficient in the resection step of HR (no other genes coding for end resection nucleases are known to be involved in HR in <i>B. subtilis</i>), displayed IR resistance similar to the one of wild type spores (Vlašić <i>et al.</i>, 2014), ii) the effects of recA and NHEJ mutations on IR sensitivity of <i>B. subtilis</i> and <i>S. ambofaciens</i> spores or <i>M. smegmatis</i> and <i>S. meliloti</i> stationary phase bacteria were found to be additive (Stephanou <i>et al.</i>, 2007; Pitcher <i>et al.</i>, 2007b; Moeller <i>et al.</i>, 2012; Hoff <i>et al.</i>, 2016; Dupuy <i>et al.</i>, 2017). Thus, DSB repair in quiescent cells would essentially rely on NHEJ using available free non-homologous DNA ends. Furthermore, DSB repair is essential during these quiescent states because cells are more exposed to harsh environmental conditions such as desiccation, that can
damage the DNA. How the balance between NHEJ and HR is regulated is far from understood but crucial regarding the consequences of the repair.

The best characterized bacterial NHEJ regulation comes from a study aiming at characterizing gene expression in the forespore of \textit{B. subtilis} (Wang et al., 2006). The operon containing \textit{ku}\textit{Bsub} and \textit{ligD}\textit{Bsub} is expressed inside the forespore under the control of the sigma factor SigG and the transcriptional factor SpoVT. Their role in the DSB repair during spore germination (LigD is an ATP-dependent ligase and ATP is present at very low levels in spores, Korza et al., 2016) has been demonstrated by the localization of a Ku-GFP fusion on the spore nucleoid after initiation of the germination, and by the high sensitivity of NHEJ deficient spores (mutated for \textit{ku}\textit{Bsub} or \textit{ku}\textit{Bsub} and \textit{ligD}\textit{Bsub}) to dry heat, a cause of DSB inside spores (Wang et al., 2006). Other sources of DNA lesions, such as vacuum desiccation, heavy ions bombardment and ultraviolet radiations have a higher toxic effect on NHEJ deficient compared to wild type spores of \textit{B. subtilis} (Moeller et al., 2007; Moeller et al., 2008; Moeller et al., 2014). According to the study of Wang et al., expression of \textit{ku}\textit{Bsub} and \textit{ligD}\textit{Bsub} was observed three to four hours after \textit{B. subtilis} cells initiate their sporulation program, concomitantly with the expression of SigG and SpoVT (Nicolas et al., 2012). However, it was found that the \textit{ku}\textit{Bsub} \textit{ligD}\textit{Bsub} operon was slightly up-regulated (~2 times) compared to its basal level (i.e. in non sporulating conditions) when bacteria encounter a cold stress. In this condition neither SpoVT nor SigG were expressed suggesting that the NHEJ genes in \textit{B. subtilis} could be regulated by other transcriptional factors (Nicolas et al., 2012). The NHEJ proficiency of these cold stressed bacteria remains to be assessed. Besides the overexpression of the \textit{ku}\textit{Bsub} and \textit{ligD}\textit{Bsub} genes during the sporulation, lack of impact of inactivation of the HR-end resection genes on the X-ray radiation sensitivity of \textit{B. subtilis} spores led to the proposal for a mechanism that specifically inhibits the action of AddAB and RecJ nucleases inside germinating spores.
spores, rendering the NHEJ pathway the only efficient DSB repair mechanism in this condition (Vlašić et al., 2014; Raguse et al., 2017). Moreover, no trace of ku and ligD expression was observed at the transcriptional (Nicolas et al., 2012) and translational (Sinai et al., 2015) levels during the revival of spores. Ku (Ku-GFP fusion) expressed during sporulation is detectable associated to the nucleoid during the first 30 min of the spore revival (Wang et al., 2006), i.e. before initiation of DNA replication and expression of DNA repair genes such as recA (Sinai et al., 2015). Then, in B. subtilis, NHEJ and HR do not seem to compete for the repair of two dsDNA ends.

NHEJ is not a hallmark of bacterial sporulation since ku and ligD genes are also present in non-spore forming bacteria and all spore-forming bacteria do not encode these proteins (Aravind and Koonin, 2001; Bowater and Doherty, 2006; McGovern et al., 2016). In Mycobacteria and in S. meliloti, ku and ligD inactivation leads to a sensitization against IR of mutant bacteria grown in stationary phase and have no or little effect in logarithmic phase, even if recA is inactivated (Stephanou et al., 2007; Pitcher et al., 2007b; Kobayashi et al., 2008; Dupuy et al., 2017). On the other hand, ku and/or ligD inactivation strongly reduces the capacity of logarithmically grown Mycobacteria or S. meliloti to repair enzymatically linearized plasmid DNA upon transformation (Gong et al., 2004; Gong et al., 2005; Aniukwu et al., 2008; Dupuy et al., 2018) as well as enzymatically induced DSB on the chromosome (Stephanou et al., 2007; Gupta et al., 2011; Dupuy et al., 2018). Moreover, ligD is required in exponentially growing M. smegmatis bacteria for infection by the mycobacteriophages Omega and Corndog, where it is thought to promote circularization at the cos site of the phage genomes, thanks to the Ku expressed by the phages (Pitcher et al., 2006). Altogether, these observations show that although not involved in the repair of DSB induced by IR in logarithmic phase, Ku- and LigD-dependent NHEJ is well active in
this growth stage in several bacterial species. This is consistent with the detection of similar amounts of Ku and similar activity of its promoter in logarithmic and early stationary phases in *M. smegmatis* (Bhattarai *et al.*, 2014; Zhou *et al.*, 2015), even if an increased activity of the Ku promoter was observed in late stationary phase cells (Bhattarai *et al.*, 2014). This is also consistent with the expression of a *S. meliloti* ku2-lacZ transcriptional fusion in both log and stationary phase, and the visualization of Ku2-YFP foci after X-ray treatment of cells in both phases (Kobayashi *et al.*, 2008). Together, these observations suggest that the action of NHEJ for repair of IR induced DSBs, which have most probably similar molecular configurations in logarithmic and stationary phases, is negatively regulated in logarithmic phase. The mechanism of this inhibition does not seem dependent on RecA since the deletion of recA in *M. smegmatis* does not increase the NHEJ (Gupta *et al.*, 2011), however it could involve the competition with presynaptic actors of the HR.

In *S. meliloti*, NHEJ activity as measured using a linear plasmid repair assay, displays a significant stimulation when bacteria enter stationary phase, or are exposed to a thermic stress (Dupuy *et al.*, 2018). This stimulation mainly results from the upregulation of *ku* gene expression. The *ku*3 *ku*4 operon, as well as the *ligD4* gene, appeared to be under transcriptional control of the general stress response EcFG sigma factor RpoE2. Examination of *ku* promoter regions in α-Proteobacteria revealed the presence of putative EcFG binding sites upstream at least one *ku* gene in about 90 % of the species, suggesting that stress control of *ku* and *ligD* transcription is widespread in this bacterial class. Accordingly, stress induction of *ku* and/or *ligD* gene expression was described in *Mesorhizobium loti* and *Bradyrhizobium japonicum* (Cytryn *et al.*, 2007; Alexandre *et al.*, 2014). How *ku*2 is up-regulated in stress conditions is currently unknown, but it appears not to be part of the general stress response, the heat shock response, the SOS response,
the stringent response nor to be regulated by quorum sensing (Kobayashi et al., 2008; Dupuy et al., 2018). Whether stress control of NHEJ gene expression is conserved beyond α-Proteobacteria remains to be tested experimentally. In Streptomyces, major NHEJ genes are expressed in late-growth phase (Hoff et al., 2016) and gene products accumulate in spores. However, no induction of the NHEJ-like genes could be detected in S. ambofaciens in response to DNA damaging agents (i.e., Mitomycin C, Hoff et al., 2016).

Besides transcriptional regulation, post-translational modification of NHEJ proteins could also regulate NHEJ activity in bacteria. Indeed the deacetylase Sir2 from M. smegmatis is required for an efficient NHEJ repair in vivo (Li et al., 2011). Overexpression of sir2 leads to a decreased amount of Ku in bacteria, without altering the quantity of the corresponding mRNA. In line with this interaction, Ku<sub>Msme</sub> is acetylated at least at the residue K29, which is conserved among bacterial Ku, and this modification is mainly observed in stationary phase (Zhou et al., 2015). Even if the mechanism by which Sir2 regulates NHEJ proteins remains to be clarified, it seems likely that acetylation could regulate the activity of NHEJ (Zhou et al., 2015).

IV Cross talk with other DNA repair mechanisms

As mentioned in the section III, DSB repair in quiescent cells would essentially rely on NHEJ, but NHEJ and HR repair mechanisms may both co-exist, at least in some NHEJ proficient bacteria during replicating phases. Moreover, in Streptomyces, recent studies (Hoff et al., 2018) on the consequences of DSB repair in the terminal region of the linear chromosome revealed events of HR repair involving homologous sequences at distinct loci suggesting that HR can occur even in the absence of a replicated copy of the chromosome (quiescent phases). This raises the question of how the choice between the two pathways is performed and if there is a competition between
both mechanisms. In eukaryotes, it is known that Ku proteins rapidly sequester DSB free ends and shield them from unscheduled exonucleases, preventing HR to proceed. The removal of Ku involves multiple pathways including ubiquitination, neddylation, phosphorylation and the nuclease activity of the MRE11-dependent complex MRX (yeast) /MRN (mammals) which plays a pivotal role in NHEJ/HR balance (for review see Shibata, 2017). Similarly, in *Mycobacterium*, in *vitro* analyses revealed that Ku<sub>Mtub</sub> protects DSBs from resection by AdnAB helicase-nuclease (Sinha *et al*., 2009) and Ku<sub>Mtub</sub>, but not LigD<sub>Mtub</sub> inactivation is compensated by an elevated HR in *vivo* (Gupta *et al*., 2011). Likewise, in *S. avermitilis*, the deletion of ku<sub>Save1</sub> and ku<sub>Save2</sub> seemed associated with an increased efficiency of HR (Zhang *et al*., 2012). Recent electronic beam irradiation assays in *S. ambofaciens* (Hoff *et al*., 2016, Leblond *et al*., unpublished) also suggest that NHEJ inactivation (*ku*<sub>Samb</sub> triple deletion or *ligC*<sub>Samb</sub> and *ligD*<sub>Samb</sub> double deletion) is compensated for by an increased HR repair. However, unlike in *Mycobacterium* where loss of HR is not compensated by increased NHEJ, a reciprocal effect is observed in *S. ambofaciens* since a recA mutant is 100 times more affected by NHEJ inactivation than the wild-type strain.

Ku and LigD proteins may help coupling NHEJ to other DNA repair pathways. IR-induced cellular damages on DNA include oxidized bases, abasic sites and single-strand breaks (SSB). However, due to the presumed “clustered damage” induced by IR (Ward, 1994), several DNA lesions can be present in close proximity and, if they are not repaired efficiently, can lead to DSB (Ward, 1994; Blaisdell and Wallace, 2001; Cox and Battista, 2005). Then, DSBs induced by IR are potentially surrounded by multiple other DNA lesions. Part of these lesions are taken in charge by the base excision repair (BER) pathway, the nicking of the phosphodiester bond 3’ to an apurinic/apyrimidinic (AP) site by an AP-lyase being one possible step of this pathway (for review see Almeida and Sobol, 2007). Human Ku, besides its function as a scaffold of the NHEJ
machinery, is also an AP-lyase suggested to be involved in the repair of AP site near DSBs (Roberts et al., 2010). This activity is conserved in Ku from B. subtilis and P. aeruginosa (de Ory et al., 2014). As Ku_{Bsub} is able to thread inward DNA molecules (McGovern et al., 2016), via loading on their broken ends, this could facilitate the repair of abasic sites surrounding a DSB. The interplay between NHEJ and BER in bacteria is reinforced by the discovery of an AP-lyase activity driven by the other NHEJ factor: LigD (de Ory et al., 2016). The involvement of LigC1_{Msme} in BER during stationary phase of M. smegmatis (Płociński et al., 2017) shows that the link between NHEJ and BER can even be extended to alternative NHEJ proteins. This multiplicity of NHEJ actors involved in BER is not specific to bacteria, human DNA polymerase λ is one example (see Bebenek et al., 2014).

Remarkably, Ku_{Bsub} is also able to linearize a supercoiled plasmid containing AP-sites (de Ory et al., 2014). This counter-intuitive effect for a protein intended to repair DSB is supposed to be dependent on its AP-lyase activity and its ability to bind end-free dsDNA molecules, DSB arising by the nicking of phosphodiester bonds of proximal AP-sites on opposite strands of a dsDNA molecules containing numerous such lesions (de Ory et al., 2014). Then it is tempting to propose that NHEJ could have evolved from a Ku-dependent BER pathway, primarily independent on the presence of DSB, that has acquired the ability to join Ku-catalyzed broken DNA ends, via the recruitment of LigD, in conditions where the bacterial genome contains numerous AP-sites. This hypothesis and the fact that most of the bacterial Ku proteins contain an extended C-ter tail that is required for their interaction on end-free DNA (Kushwaha and Grove, 2013b; McGovern et al., 2016) but for which the deletion has only a mild effect in vitro on the NHEJ efficiency (Zhu and Shuman, 2010; McGovern et al., 2016) suggest that the extended C-ter of Ku could play an important role in a Ku-dependent BER acting on AP-sites embedded in duplex regions of DNA.
A recent study has functionally linked the NHEJ of *P. aeruginosa* to the MutL protein, a component of the mismatch repair (MMR) pathway. Searching for the molecular mechanism at the basis of a large genomic deletion in the *P. aeruginosa* chromosome allowing to prevent infection by a bacteriophage, Shen et al. (2018) have shown that MutL<sub>Paer</sub> is required for getting the deletion. MutL<sub>Paer</sub> is an endonuclease promoting SSB and, directly or indirectly, DSB (as a consequence of multiple SSB in dsDNA). Surprisingly, the MutL<sub>Paer</sub>-dependent genomic deletion observed after the bacteriophage infection is not dependent on MutS<sub>Paer</sub>, the partner of MutL<sub>Paer</sub> in MMR, suggesting a new function of MutL in this bacterium. Whatever the mechanism, the reduced frequency of the corresponding deletion in a NHEJ deficient mutant led to the proposal that DSB generated by MutL<sub>Paer</sub> are repaired by the NHEJ pathway in this bacterium (Shen et al., 2018). On the other hand, the majority of archaeal species and many actinobacteria lack MutHLS system but display genes putatively involved in a novel mismatch repair system, called NucS/EndoMS, which has been shown in *Corynebacterium glutamicum* to cleave both strands at double-strand DNA mismatched substrate during the repair process (Castañeda-García et al., 2017; Ishino et al., 2018). Whether these induced DSBs are thereafter processed by NHEJ or HR is not yet known, but one can suggest that chromosomes of bacteria possessing a NucS/EndoMS system are more prone to DSBs and consequently to mutations and rearrangements, especially if they also feature a NHEJ machinery, like many actinomycetes.

Another mechanism that induces DSBs is the type II CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and associated proteins) system which is conferring adaptive immunity against phages and exogenous mobile genetic elements (for review see Marraffini, 2015) and which has been engineered to provide an efficient genome editing method (Jinek et al., 2012). Since the CRISPR-Cas system generates DSBs, it could display antagonistic interactions
with DSB repair systems. Bernheim and co-authors (2017) revealed that NHEJ and type II-A CRISPR-Cas systems do not co-occur in all but one of more than 5000 sequenced bacterial genomes. On the other hand, this strong avoidance of co-occurrence was not observed with type II-C system, which is consistent with the fact that, unlike type II-A system, type II-C system lacks Csn2, a multimeric toroidal binding protein that binds double stranded DNA ends and has similar properties to Ku protein (Arslan et al., 2013). The binding of Csn2 at DNA ends could block access to Ku and inhibit efficient repair by NHEJ. In contrast, B. subtilis NHEJ system does not impair Streptococcus pyogenes or Streptococcus thermophilus type II-A CRISPR-Cas9 system (Bernheim et al., 2017).

V- NHEJ and genome evolution

NHEJ is inherently unfaithful, as DNA ends of DSB are often remodeled before ligation, thus generating short insertions or deletions at the repair junction (although repair fidelity seems to vary depending on the tested organism, the type of DNA ends being repaired or the enzymes involved). Interestingly, NHEJ is generally more active in stationary phase or during sporulation, and in α-proteobacteria, is activated under stress conditions, in a process dependent on the general stress response sigma factor EcfG. This is reminiscent of the so-called stress-induced mutagenesis of enterobacteria, where the highly faithful HR-dependent DSB repair becomes mutagenic in a process dependent on the general stress response sigma factor RpoS (Shee et al., 2011). NHEJ was also recently suggested to participate in stationary phase mutagenesis in Pseudomonas putida (Paris et al., 2015). NHEJ could then significantly contribute to DSB-dependent stress-induced mutagenesis in bacteria and could thus participate in their adaptation to adverse environmental conditions.
In addition to the short, non-templated nucleotide insertions usually observed at NHEJ repair junctions, insertion of longer DNA sequences (several tens of nucleotides) of either plasmid or chromosomal origin, was described to have occurred in rare instances at plasmid or chromosome DSB repair junctions in *M. smegmatis* and *S. ambofaciens* (Akey *et al.*, 2006; Aniukwu *et al.*, 2008; Hoff *et al.*, 2018). Although the mechanism of such DNA insertions was not clearly established, it may have resulted from the incidental capture of indigenous DNA fragments at the break by the NHEJ machinery. In eukaryotes, DNA inserted into the genome by NHEJ during evolution include foreign DNA fragments such as mitochondrial DNA, transposable elements, and viral DNA (Moore and Haber, 1996; Ricchetti *et al.*, 1999; Lin and Waldman, 2001a; Lin and Waldman, 2001b; Nakai *et al.*, 2003; Hazkani-Covo and Covo, 2008). In yeast, capture of extranuclear DNA at repaired DSBs was shown to be dependent on DNA end resection and NHEJ complexes (Decottignies, 2005). In the same way, the low level of genome rearrangements in the yeast *Lachancea kluyveri* since its separation from its ancestor with *S. cerevisiae* was assigned to the loss of NHEJ genes in this phylum (Gordon *et al.*, 2011). Interestingly, Dupuy *et al.* demonstrated that linear DNA of heterologous origin, introduced in *S. meliloti* by transformation, could be effectively captured in a NHEJ-dependent way at an artificially generated DSB in the chromosome (Dupuy *et al.*, 2018).

In *Streptomyces*, since the variable NHEJ genes encode actors of the response to DNA damage (Hoff *et al.*, 2016) that presumably participate to the NHEJ and other DNA repair pathways, it is likely that different combinations of NHEJ genes result in variable recombination profiles, and consequently in variable abilities to integrate DNA from horizontal gene transfer. Like for *E. coli* mutator strains (Giraud *et al.*, 2001), such recombination proficient lineages could provide a flexible adaption property to the population in the ecological context by allowing rapid and intense diversification of the accessory gene set. This speculation is substantiated by the strong variability of the NHEJ-gene set revealed within a *Streptomyces* population originating from a
forest soil (Leblond et al., unpublished). Altogether, these data suggest that NHEJ could participate in the acquisition of heterologous DNA during lateral transfer events, overcoming the lack of sequence homology between distantly related species. Interestingly, in support of this hypothesis, Popa et al. (2011) found that gene transfer events from distant donors are more frequent when recipient genomes encode NHEJ proteins. In addition, Kidane et al. (2009) have observed that the absence of Ku reduces natural plasmid transformation and phage DNA transfection in *B. subtilis*.

**Concluding remarks**

As mentioned within this review, many NHEJ proficient bacteria contain several stand-alone genes displaying similarities with genes coding for Nucdom, Poldom, LigDom and/or *ku* genes and their multiplicity renders difficult the identification of the *bona fide* NHEJ genes. The use of two-hybrid screening coupled with affinity purification of the products of these genes in their host and in conditions where NHEJ is active should help to characterize actual NHEJ complexes, and potentially identify new components of the NHEJ machinery, or new links with other DNA repair pathways (e.g. role of AP-lyase activities in response to genotoxic stresses). Structural information (i.e. cryo-electron microscopy and super-resolution microscopy analyses) associated to 3D structures of stand-alone NHEJ actors (LigD, Ku) and of the complete bacterial NHEJ machinery will be of crucial importance for our basic understanding of this mechanism (e.g. NHEJ synapsis step, DNA extremities tethering, processing of variable DNA ends). Furthermore, alternative NHEJ machineries may be involved depending on the growth phase or growth conditions (see above the stress-induced NHEJ machineries of *S. meliloti*). This raises the questions of the competition between distinct NHEJ machineries, and between HR and NHEJ, to
repair DSBs in conditions where both mechanisms coexist. It is clear, based on our knowledge of
the eukaryotic NHEJ, that DNA ends accessibility should be a key point that affects the HR/NHEJ
balance. Genetic, biochemical and cellular localization studies of proteins involved in early steps
of these pathways over time after DSB induction should help to elucidate preferred routes of DSB
repair in different life states. Further, experimental evolution could constitute a powerful
approach for testing the involvement of NHEJ in integration of DNA acquired by horizontal gene
transfer and its impact on the capacity to create and maintain genetic diversity within bacterial
populations in their ecosystems.

Finally, understanding the bacterial NHEJ may have therapeutic interest. Indeed, many human
pathogens possess a NHEJ repair system. This is for instance the case for the most deadly human
pathogen *M. tuberculosis* and the nosocomial pathogen *P. aeruginosa*. Although NHEJ is not
directly involved in survival or virulence in different mammal hosts (Brzostek *et al.*, 2014; Heaton
*et al.*, 2014), it could provide a way to survive to genotoxic conditions in certain non-dividing
states. Identifying these states and developing anti-NHEJ compounds based on the high-
resolution structure of a NHEJ machinery could help to fight against these pathogens.
Acknowledgments

This work was supported by INRA (Institut National de la Recherche Agronomique) -including by the Département Santé des Plantes et Environnement-, by the French Laboratory of Excellence project TULIP (ANR-10-LABX-4), by the French National Research Agency program Streptoflux (ANR Blanc 0096_01), and by ANR through the Laboratory of Excellence ARBRE (ANR-12-LABXARBRE-01). The authors declare that they have no competing interests.

References


Figure 1: Repertoires of NHEJ presumed actors in model bacteria from minimal NHEJ to the more complex situations. Ku, LigDom (ATP-dependent), NucDom, PolDom identified in the genome of the bacterial species (as described in the literature) are highlighted whatever present within a single polypeptide or as standalone domains. Proteins having the same name in different organisms are not necessarily encoded by orthologous genes. \( \checkmark \) indicate actors which have been experimentally involved in DSB repair. \( \diamond \) in M. tuberculosis denote accessory actors involved in NHEJ, based on studies of their homologs in M. smegmatis coding for different LigD with mutated domain (see text). In S. ambofaciens, accessory actors involved in response to DNA damage are highlighted by an asterisk (Hoff et al., 2016, 2018). For the sake of clarity, only one dimer of Ku is shown at DNA ends and the ability of Ku to thread inward the DNA molecule has not been represented.