

The eukaryotic Pso2/Snm1/Artemis proteins and their function as genomic and cellular caretakers

D. Bonatto¹,
L.F. Revers²,
M. Brendel¹ and
J.A.P. Henriques^{1,3,4}

¹Departamento de Biofísica, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

²EMBRAPA Uva e Vinho, Bento Gonçalves, RS, Brasil

³Curso de Farmácia, Universidade Luterana do Brasil, Canoas, RS, Brasil

⁴Instituto de Biotecnologia, Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil

Abstract

DNA double-strand breaks (DSBs) represent a major threat to the genomic stability of eukaryotic cells. DNA repair mechanisms such as non-homologous end joining (NHEJ) are responsible for the maintenance of eukaryotic genomes. Dysfunction of one or more of the many protein complexes that function in NHEJ can lead to sensitivity to DNA damaging agents, apoptosis, genomic instability, and severe combined immunodeficiency. One protein, Pso2p, was shown to participate in the repair of DSBs induced by DNA inter-strand cross-linking (ICL) agents such as cisplatin, nitrogen mustard or photo-activated bi-functional psoralens. The molecular function of Pso2p in DNA repair is unknown, but yeast and mammalian cell line mutants for *PSO2* show the same cellular responses as strains with defects in NHEJ, e.g., sensitivity to ICLs and apoptosis. The Pso2p human homologue Artemis participates in V(D)J recombination. Mutations in Artemis induce a variety of immunological deficiencies, a predisposition to lymphomas, and an increase in chromosomal aberrations. In order to better understand the role of Pso2p in the repair of DSBs generated as repair intermediates of ICLs, an *in silico* approach was used to characterize the catalytic domain of Pso2p, which led to identification of novel Pso2p homologues in other organisms. Moreover, we found the catalytic core of Pso2p fused to different domains. In plants, a specific ATP-dependent DNA ligase I contains the catalytic core of Pso2p, constituting a new DNA ligase family, which was named *LIG6*. The possible functions of Pso2p/Artemis/Lig6p in NHEJ and V(D)J recombination and in other cellular metabolic reactions are discussed.

Key words

- Non-homologous end joining
- Double-strand breaks
- V(D)J
- *PSO2*
- Artemis
- *Saccharomyces cerevisiae*

Correspondence

J.A.P. Henriques
Departamento de Biofísica
Centro de Biotecnologia, UFRGS
Av. Bento Gonçalves, 9500
91507-970 Porto Alegre, RS
Brasil
Fax: +55-51-3316-7603
E-mail: pegas@cbiot.ufrgs.br

Presented at the XI Congresso Brasileiro de Biologia Celular, Campinas, SP, Brazil, July 15-18, 2004.

Research supported by FAPERGS, CAPES, and GENOTOX (Genotoxicity Laboratory, UFRGS-Brazil).

Received June 30, 2004

Accepted December 7, 2004

Introduction

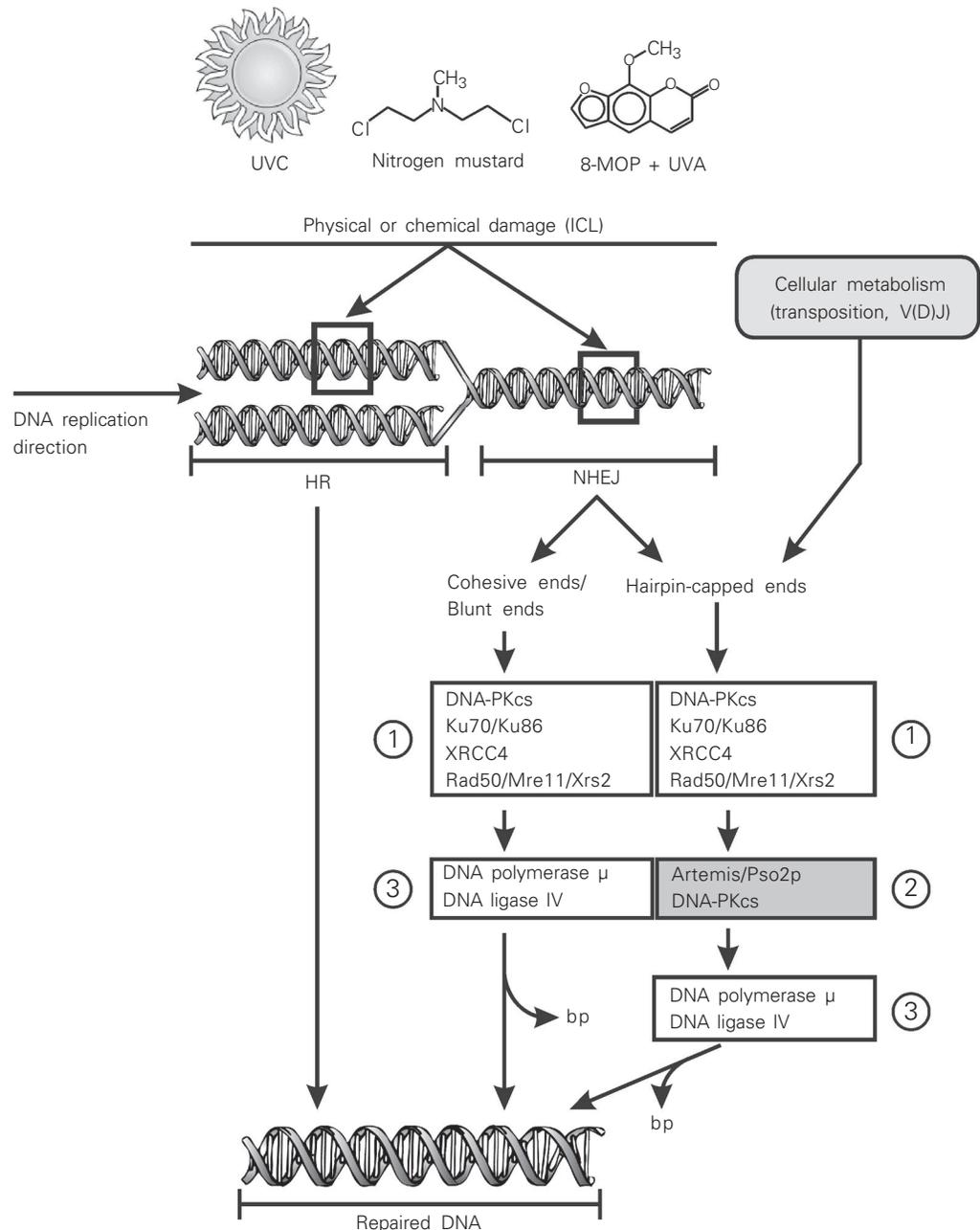
The chromatin of all eukaryotic cells, without exception, is a special target for chemical or physical agents that can induce different kinds of DNA damage, including

base-pairing mismatches, abasic sites, chemically modified bases, single- and double-strand breaks (DSBs), and intra- and/or inter-strand cross-links (ICLs) (1). Depending on the extent of chromatin damage, these alterations may have a profound effect on

cellular well being, leading to cell cycle arrest, tumorigenesis, cell death, or severe combined immunodeficiency disease (SCID) in mammals (1). Among the various forms of DNA lesions that are induced by physical or chemical agents, probably the most dangerous are the DNA DSBs (1,2). DSBs can occur in response to ionizing radiation, to radiomimetic agents or chemical substances

that induce DNA ICLs such as bi-functional nitrogen mustards or 8-methoxypsoralen plus UVA (Figure 1). DSBs also arise as a consequence of natural processes such as V(D)J recombination (a lymphoid-specific process required for gene rearrangement and maturation of T and B cells), and as a by-product of normal cellular metabolism (Figure 1) (3). If not repaired prior to DNA replication or

Figure 1. Schematic drawing of double-strand break (DSB) repair in mammalian cells. DSB induced by inter-strand cross-link (ICL) generated by physical agents (UVC), chemical substances (nitrogen mustard, 8-MOP + UVA), or even cellular metabolism (gray box) on DNA during replication can be repaired by two biochemical pathways: homologous recombination (HR) or non-homologous end joining (NHEJ). HR is the major DNA repair pathway used when two homologous DNA strands are present. NHEJ is used when the homologous DNA strand is not present. The protein complexes that are used for NHEJ repair depend on the type of DNA ends present in the DSB (cohesive ends, blunt ends, or hairpin-capped ends). Protein complexes 1 and 3 repair both cohesive and blunt ends, while hairpin-capped ends are repaired by Artemis/Pso2p/DNA-PKcs (complex 2). The final result is the restitution of high molecular weight DNA, with loss (NHEJ) or without loss (HR) of DNA base pairs (bp). UVC = UV_{254 nm}; 8-MOP + UVA = 8-methoxypsoralen plus UVA; UVA = UV_{365 nm}; DNA-PKcs = DNA-dependent protein kinase catalytic subunit



mitosis, DSBs can induce cell death (4) and, if misrepaired, DSBs have the potential to lead to chromosome translocations, genomic instability and predisposition to cancer (2,5). Interestingly, only one DSB can kill a cell if it leads to the inactivation of an essential gene or triggers apoptosis (2,4,6). Moreover, mutations in many of the factors involved in sensing and repair of DSB damage lead to increased pre-disposition to cancer in man and in animal models (2,7).

In yeast and mammalian cells, DSBs are predominantly repaired by one of two pathways (1), i.e., homologous recombination (HR), or non-homologous end joining (NHEJ) (Figure 1). In addition, NHEJ is also used to repair DSBs that arise during early mammalian lymphocyte development in the context of V(D)J recombination (8). HR and NHEJ have overlapping roles in maintaining chromosomal integrity (9) and can act together to preserve genomic integrity in eukaryotic cells (10). Yeast, unlike multicellular eukaryotes, repairs most of its DSBs using HR, a process that occurs without the loss of genetic information (11). However, NHEJ can be detected in yeast when the mechanisms of HR are inactivated (11). Multicellular eukaryotes use NHEJ as the predominant DNA repair system and this preference could be intrinsic to their genomic organization. The genomes of multicellular eukaryotes contain a substantial fraction of repetitive DNA and, therefore, the homology search process for repair of DSBs by HR is inviable when the breaks occur in the portion of the genome that is repetitive, further leading to chromosomal translocations or cell death (11). Except during late S, G2 and M, when a sister chromatid is physically positioned optimally, homology partners for repetitive regions might be chosen inappropriately from any of the chromosomes (11).

Cells with a defect in NHEJ age in culture more quickly when compared to NHEJ-proficient cells (12). Mouse mutants in either component of the DNA ligase complex (XRCC4 or DNA ligase IV) show defects in

V(D)J recombination (13,14), just as human pre-B cells do (15). These mice die during the final days of gestation, showing an increased apoptotic death of neurons at specific locations in the nervous system at specific times during gestation (11). It is still unclear why some cells die and others do not. Interestingly, Ku70-deficient mice show a depletion of enteric neurons (16). Presumably this apoptotic cell death is triggered by an inability to repair DSBs. Also, the inactivation of NHEJ leads to increased sensitivity to ionizing radiation, genomic instability, and SCID, resulting from the inability to join Rag-cleaved gene segments in progenitor (pro)-B and T lymphocytes (17). Despite their inability to repair DSBs, NHEJ-deficient mice show, at most, a modest predisposition to lymphomas, because cells with unrepaired breaks are eliminated by the checkpoint protein p53 (17). Inactivation of p53 restores pro-B lymphocyte numbers, although it does not rescue NHEJ or lymphocyte development (18). Combined deficiencies for p53 and all NHEJ factors have been analyzed and all were found to lead to consistent development of early-onset pro-B lymphomas (18).

NHEJ basically involves modification of the two broken ends to make them compatible prior to rejoining, resulting in the loss of some information between the two DNA ends. Hence, NHEJ is an imperfect process from the standpoint of preserving genetic information (11). Proteins known to be involved in NHEJ include the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), XRCC4, Ku70 and Ku86, DNA ligase IV, and the Rad50/Xrs2/Mre11 complex (19). These proteins, to be described in more detail below, form complexes with specific functions in the modification of DNA ends for rejoining, or in the stabilization of DNA extremities for further processing.

DNA-PKcs

The DNA-PKcs, which is activated by

double stranded DNA ends, phosphorylates proteins bound to the same DNA molecule. Apart from its large size (469 kDa), the most noticeable feature of DNA-PKcs is a carboxy-terminal catalytic domain which bears amino acid similarity to the catalytic domain of the phosphoinositide-3,4-kinase family of lipid kinases (20). The presence of this conserved region classifies DNA-PKcs as a member of the phosphatidylinositol-3-kinase-related protein kinases (21,22). Ku70 and Ku86 are proteins that form a heterodimer with high affinity for DNA ends and are generally considered to comprise the DNA-binding “subunit” of DNA-PK. However, their association with DNA-PKcs appears not to be obligatory and there is clear evidence for DNA-PKcs-independent functions (Table 1) (23).

Ku70/Ku86

Cells that lack Ku are radiosensitive and defective in DSB repair, and animals lacking

either one of the Ku subunits share many characteristics with DNA-PKcs null animals, e.g., radiosensitivity, immune deficiency, and defective DNA DSB repair (Table 1). In addition, Ku70 and Ku80 null animals have growth defects and premature senescence, indicating that Ku and DNA-PKcs have distinct and overlapping functions (2,11). In plants, specifically *Arabidopsis thaliana*, the expression of both Ku70 and Ku80 genes is up-regulated in response to the induction of DSBs in chromosomal DNA by either bleomycin or methylmethanesulfonate. Mutant lines of *A. thaliana* for Ku80 showed hypersensitivity to the DNA-damaging agents bleomycin and menadione which cause single- and DSBs in DNA, a phenotype consistent with a role in the NHEJ pathway (Table 1) (24,25).

DNA ligase IV

DNA ligase IV, an ATP-dependent DNA ligase that has a special role in NHEJ and

Table 1. Eukaryotic non-homologous end joining proteins.

Proteins	Organisms	Tissues/cells ^a	Cellular process	Substrates	Phosphorylation ^b	Deficiencies ^c
DNA-PKcs	Metazoa	All	NHEJ, V(D)J, telomeric maintenance	DSB	Yes	Sensitivity to cross-links, SCID, senescence
DNA ligase IV (Lig4)	Eukaryotes	All	NHEJ, V(D)J, telomeric maintenance	DSB	Yes	Sensitivity to cross-links, SCID, senescence, neuronal apoptosis
Mre11	Eukaryotes	All	NHEJ, V(D)J, telomeric maintenance	DSB, DNA hairpins	Inconclusive	Sensitivity to cross-links, senescence
Rad50	Eukaryotes	All	NHEJ, V(D)J, telomeric maintenance	DSB, DNA hairpins	Inconclusive	Sensitivity to cross-links, senescence
Xrs2	Metazoa, fungi	All	NHEJ, V(D)J, telomeric maintenance	DSB, DNA hairpins	Inconclusive	Sensitivity to cross-links, senescence
Ku70/Ku80	Eukaryotes	All	NHEJ, V(D)J, telomeric maintenance	DSB	Yes	Sensitivity to cross-links, SCID, senescence
XRCC4	Eukaryotes	All	NHEJ, V(D)J, telomeric maintenance	DSB	Yes	Sensitivity to cross-links, SCID, neuronal apoptosis, senescence
Pso2	Eukaryotes	All	NHEJ, possible other functions in cellular processes	Unknown	Unknown	Sensitivity to cross-links
Artemis	Metazoa	Lymphocytes	NHEJ, V(D)J	DSB, DNA hairpins	Yes	Sensitivity to cross-links, SCID
DNA ligase VI (Lig6)	Plants	Unknown	Unknown	Unknown	Unknown	Unknown

^aProteins present in different types of tissues or cells. ^bIndicates if protein activity is induced or modified by site-specific phosphorylation.

^cPhysiological deficiencies induced by partially functional or non-functional proteins related to NHEJ, V(D)J recombination, and telomeric maintenance. NHEJ = non-homologous end joining; DSB = double-strand break; SCID = severe combined immunodeficiency disease; DNA-PKcs = DNA-dependent protein kinase catalytic subunit.

V(D)J, is present in eukaryotes as diverse as yeast, plants, and metazoa (26). The homologue of the mammalian gene for DNA ligase IV was isolated from *A. thaliana*, and its expression profile indicates that this gene is regulated by ionizing radiation-induced DSBs (26). Deletion of mammalian DNA ligase IV results in death during embryogenesis due to massive neuronal apoptosis (Table 1) (14). A highly radiation-sensitive human cell line isolated from a leukemia patient was found to express a dysfunctional form of DNA ligase IV (Table 1) (14).

XRCC4

XRCC4 exists in a tight complex with DNA ligase IV (27), which is essential for the ligation step in NHEJ and may also be involved in alignment or gap filling prior to ligation (28). In mammalian cells, XRCC4 can interact with DNA, DNA-PKcs, Ku, and DNA polymerase μ , but its precise role in NHEJ is unknown (1). Cells that lack XRCC4 are radio-sensitive, defective in V(D)J recombination and DSB repair, and disruption of XRCC4 in mice is embryonically lethal due to neuronal apoptosis (Table 1) (14). A plant gene with high homology to mammalian XRCC4, that also interacts with DNA ligase IV and has its expression pattern modulated by DSBs, was identified in *A. thaliana* (29).

Rad50/Xrs2/Mre11

The Rad50/Xrs2/Mre11 complex is also very well conserved in all eukaryotes studied so far. These three physically interacting gene products were best characterized in yeast, where they participated in Ku-dependent end joining *in vitro* (30). Mammalian homologues for Rad50p and Mre11p have been identified, but due to the lethality of the mutations no mutants exist (Table 1). In human cells the Mre11p, Rad50p, Nbs1p (MRN complex) is involved in DNA damage signaling, possibly by holding opposing ends

of a DSB in proximity, or participating, via its exonuclease activity, in processing DNA ends prior to ligation (30). It is interesting to note that many proteins participating in NHEJ or V(D)J recombination share a high homology from yeasts to plants and animals, indicating the essentiality of these mechanism to cellular well-being. One protein that participates in NHEJ and V(D)J recombination, and whose function is still largely unknown, is Pso2p/Artemis, which belongs to the metallo- β -lactamase associated CPSF Artemis SNM1/PSO2 (β -CASP) family.

The β -CASP family

The β -CASP family comprises a group of related proteins that use nucleic acids as substrate and function in DNA repair, RNA processing, and V(D)J recombination (31). Hydrophobic cluster analysis (HCA) recently allowed this group to be identified in all three life domains (31). HCA is a sensitive method of sequence comparison that detects 2- and 3-dimensional similarities between protein domains showing very limited amino acid relatedness, typically below the so-called "twilight zone" (25-30%) (31). The method consisted of displaying the primary protein structure on a duplicated α -helical net, where the hydrophobic residues are automatically contoured. The positions of these hydrophobic clusters within the protein correspond well to the secondary protein structures and thus are extremely valuable for phylogenetic inferences. Moreover, conserved protein domains can be mapped with HCA using orthologous sequences from different species. Characteristically, all the proteins of the β -CASP family use as substrate a compound containing an ester linkage and a negative charge in its molecular structure and catalyze the hydrolysis of the former. They are composed of five domains and have an evolutionarily highly conserved HxHxDH signature and a binuclear Zn(II) center, necessary for the ester cleavage (31). In the β -CASP fam-

ily, a conserved carboxy-terminal region, defined as the “ β -CASP” motif, contains the three domains A, B and C, where C plays an important role in nucleic acid metabolism (31). The best-characterized member of this group is Artemis, a protein isolated from cells

of patients suffering from a special type of SCID associated with radiosensitivity (RS-SCID) (32). This disease was found in a group of Athabaskan-speaking American Indians and has been genetically characterized (33). An Artemis/DNA-PKcs complex, with endonucleolytic activity on DSBs or hairpins generated by the Rag1/Rag2 proteins, might act on NHEJ and V(D)J recombination, respectively (34,35). Preliminary protein sequence analyses, including the Artemis/Pso2 sequences, Elac1, Elac2, Cpsf 73-, and Cpsf 100-kDa proteins, indicate similar functions (31). The activity of Elac1/Elac2 proteins is unknown, but sequence analysis suggests a hydrolase function (36,37). Elac1/Elac2 mutant variants have been associated with human prostate cancer (36). Cpsf 100 kDa and Cpsf 73 kDa hydrolyze mRNA, and this protein group has conserved domains in eukaryotes as well as in archaea (38). They are important components of the eukaryotic machinery that processes the 3' end of mRNAs, acting together with two other Cpsf proteins (30/160 kDa), as well as with the cleavage stimulation factor, poly(ADP-ribose) polymerase, two additional cleavage factors (I_m and II_m), and poly(A)-binding protein II (38). Of the three motif domains A, B and C of β -CASP, domain C, according to HCA, has a conserved hydrophobic residue typical of proteins that use DNA as substrate and a histidine residue conserved in proteins that bind RNA (31). Our phylogenetic analysis indicates that Elac1/Elac2, Cpsf 73/Cpsf 100 and Artemis/Pso2 proteins are paraphyletic, not sharing a recent common ancestor. Moreover, the phylogeny of these proteins shows only a functional homology, based on nucleic acid phosphodiesterase activity (Bonatto D, Revers LF, Brendel M and Henriques JAP, unpublished results).

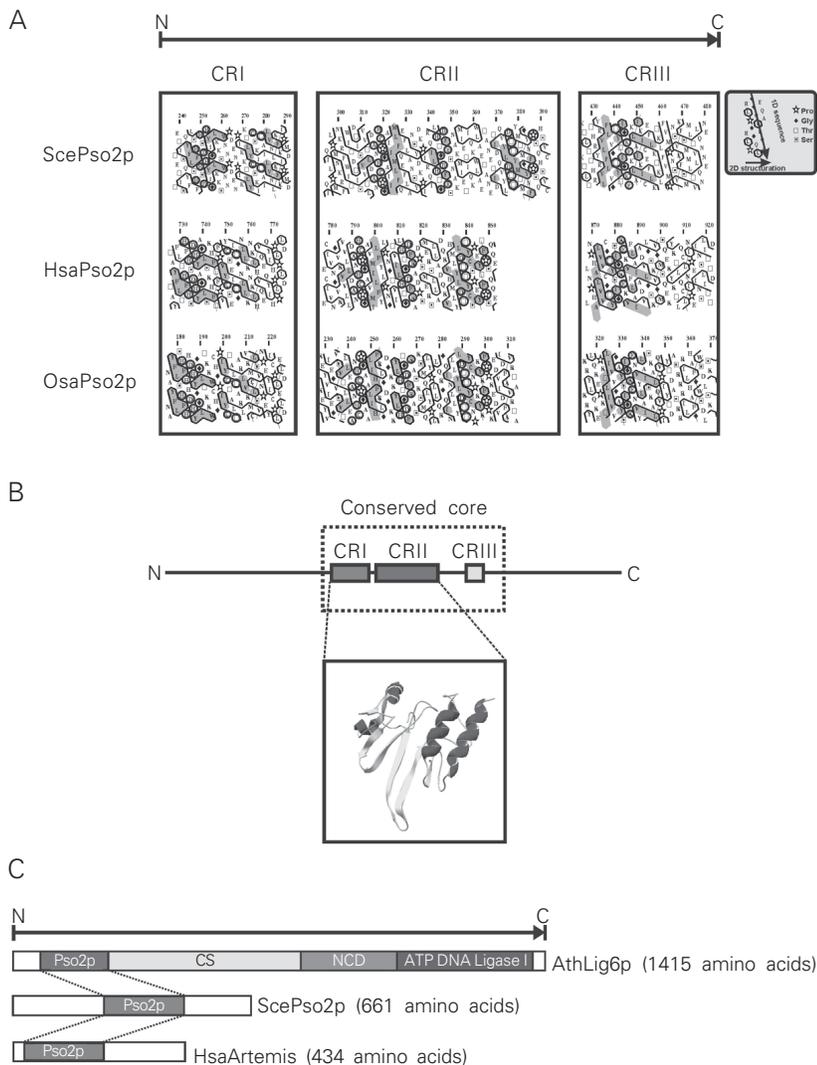


Figure 2. Hydrophobic cluster analysis of yeast (*Saccharomyces cerevisiae*, Sce), human (*Homo sapiens*, Hsa), and rice (*Oriza sativa*, Osa) Pso2p sequences (A). The three conserved regions (CRI-CRIII) of the Pso2 catalytic core are indicated. Conserved hydrophobic amino acid residues appear in gray and conserved hydrophilic amino acid residues are contoured. The way to read the sequence and special symbols is indicated in the gray inset. In B, the Pso2p catalytic core is represented by a dotted box containing the three CRs. A three-dimensional model of CRI and CRII is shown inside the box. A comparison of ScePso2p, *Arabidopsis thaliana* Lig6p (AthLig6p), and HsaArtemis domains is shown in C. The length of sequences is given in parentheses and the direction of proteins, from N-terminus to C-terminus, is indicated by an arrow. CS = conserved sequence; NCD = non-catalytic domain.

20 years since the isolation and characterization of *pso2/snm1* mutants of *S. cerevisiae* (39-43; for reviews, see 44,45) so far give no clue to the function of the Pso2p/Snm1p in ICL repair (Table 1). Clearly, *pso2/snm1* mutants are extremely sensitive to ICL-inducing agents, irrespective of their chemical composition (e.g., ICL induced by 8-MOP + UVA, nitrogen or sulfur mustards, cisplatin, and many others; 39-42); however, they are only mildly sensitive to UVC and not sensitive to ionizing radiation (41,42). Furthermore, *S. cerevisiae pso2/snm1* mutants, though incapable of forming high molecular weight DNA (data from neutral sucrose gradient assays) during repair of ICL, are not defective in repair of DSBs (40,43). Stability of the mitochondrial DNA is also affected in these mutants, as they have a higher-than-wild-type phenotype frequency of spontaneous “petit” mutations (46). This suggests a possible function for Pso2p/Snm1p in mtDNA recombination or repair in yeast. Pso2p/Snm1p mutants also have lower induced mutagenesis when compared to the wild-type strain (41).

In order to better understand the possible functions of Pso2p in DNA repair of *S. cerevisiae*, we have used an *in silico* analysis combining a phylogenetic approach and HCA to characterize the conserved regions (CRs) found between Pso2p and its orthologues. All sequences were obtained directly from GenBank in the National Center for Biotechnological Information web page [http://www.ncbi.nlm.nih.gov/] followed by global pair-wise multiple-alignments. The results of the alignments were then used for HCA (DRAWHCA program, available as a freeware at http://www.lmcp.jussieu.fr). Using the closest species of *S. cerevisiae*, as well as more distant fungal species, we could identify three CRs that are also found in the Artemis/Pso2p/Lig6p sequences of metazoa, protozoa, and plants (Figures 2A-C and 3). These three CRs, which share many conserved amino acid residues (Figures 2A and 3), compose the Pso2p conserved core (CRI, CRII, and CRIII; Figures 2B and 3). It is interesting to note that both CRI and CRII could be three-dimensionally modeled with the Swiss-Pdb Viewer software (http://

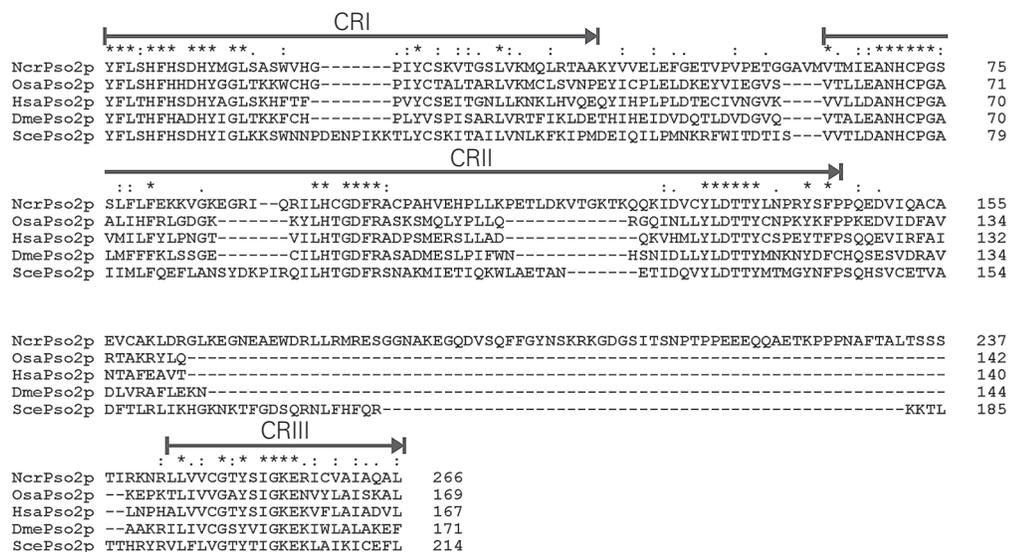


Figure 3. Multiple alignment of Pso2p conserved region sequences (CRI to CRIII) from yeast (*Saccharomyces cerevisiae*, ScePso2p), humans (*Homo sapiens*, HsaPso2p), filamentous fungi (*Neurospora crassa*, NcrPso2p), fruit flies (*Drosophila melanogaster*, DmePso2p), and rice (*Oriza sativa*, OsaPso2p). Identical amino acid residues are indicated by an asterisk and amino acid residues with similar physico-chemical characteristics by one or two dots in CRs. The positions of the CRs are indicated by arrows.

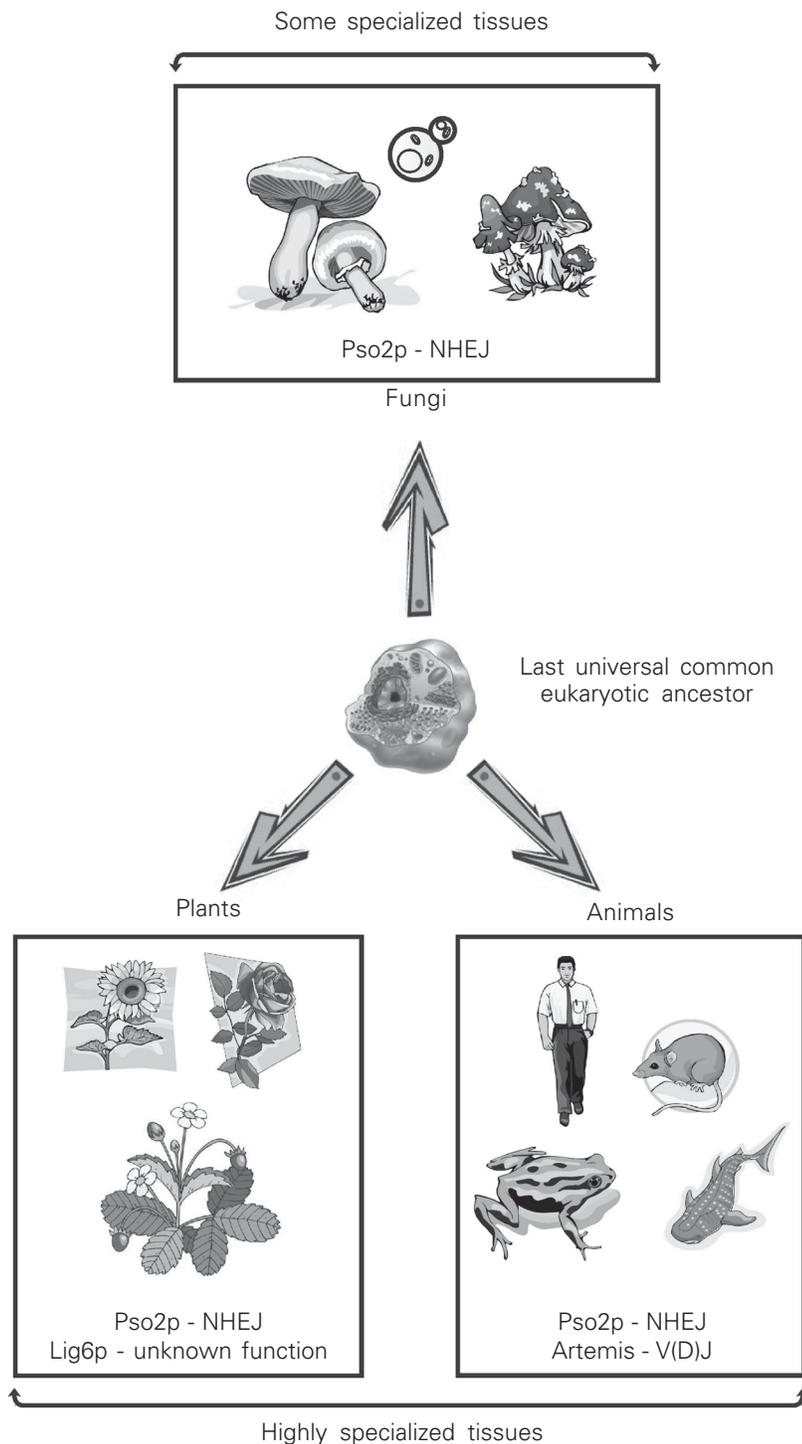


Figure 4. Evolutionary diversification of Pso2p in fungi, animals, and plants from a last universal common eukaryotic ancestor. Artemis and Lig6p are represented within animals and plants, respectively. Animals and plants contain paralogous *PSO2* genes, but they are represented by a single sequence for clarity. Fungi contain only one Pso2p sequence. This diversification might be linked to the tissue diversity found in higher eukaryotes (animals and plants). NHEJ = non-homologous end joining.

www.expasy.org/spdbv) (Figure 2B) using as template the penicillinase sequence of *Pseudomonas aeruginosa*, which belongs to the metallo- β -lactamase superfamily (Protein Data Bank accession number 1dd6) and exhibited some degree of similarity with Pso2p. All Pso2p sequences analyzed so far show highly divergent N- and C-termini, indicative of different types of enzymatic regulations (Bonatto D, Brendel M and Henriques JAP, unpublished results). Moreover, the conserved Pso2p core was found to be associated with other functional domains, e.g., plant-specific DNA Lig6p, which contains a DNA ligase I domain in its C-terminus (Figure 2C), and the Pso2p of *Aspergillus nidulans*, which has a cytochrome P450 domain also in its C-terminus (data not shown). The biochemical significance of these fused domains is still unknown, but we may speculate that these proteins have specific roles in DNA repair or even in chromatin remodeling.

The phylogenetic data indicate the presence of multiple paralogous *PSO2* genes that arise from a last universal common eukaryotic ancestor of metazoa and plants. Again we can speculate that the presence of paralogous *PSO2* genes in multicellular eukaryotes may be associated with the tissue diversity unknown for fungi, suggesting a more specialized function for DNA repair or genome caretaking in plants or metazoa (Figure 4).

The deletion of the *PSO2* gene in *Schizosaccharomyces pombe*, an evolutionarily distant yeast, generates mutant cells that are only modestly sensitive to a variety of cross-linking agents (47). In comparison to yeast, there is much less information available for mammalian Pso2p, making it difficult to predict a physiological function for this protein family. In terms of molecular data, human *PSO2/SNMI* (*hPSO2/hSNMI*) mRNA contains an unusually long 5' UTR which is predicted to form an extensive secondary structure, and which is interspersed with 16 translation initiation codons.

In fact, the function of this long 5' UTR may be to maintain hPso2p at low levels since over-expression should be highly toxic to mammalian cells and appears to result in apoptosis (48). Nevertheless, the regulation of *hPSO2/hSNM1* during mitosis suggests that this gene may play a role in mitotic progression, particularly in response to ICL-inducing agents, and especially during the G2/M transition. In this regard, it is interesting to note that cisplatin-treated cells of the *S. cerevisiae pso2* mutant arrest permanently during the G2/M transition (49). The prolonged arrest in G2/M suggests that the cell is attempting repair or initiating repair in this phase of the cell cycle but cannot complete it without a functional Pso2p (49). Recent data reported by Yu et al. (50) indicate a possible function of Pso2p in DNA repair of hairpins induced by transposition of *Ac/Dc* elements from *Zea mays* in *S. cerevisiae*. In this case, the expression of *Ac/Dc* elements in *S. cerevisiae* allows to assay the repair of excision sites in a variety of yeast mutant backgrounds, specifically of DNA hairpins that appear to form in the host DNA during transposition. This indicates that Pso2p may recognize a DNA hairpin as a structure similar to a covalent ICL lesion and may bind to it, as the Artemis protein of vertebrates does during V(D)J recombination (50).

The Artemis protein

The best-characterized member of the β -CASP family is Artemis (Table 1), which was isolated from cells of patients suffering from a special type of RS-SCID (33). SCID is clinically characterized by opportunistic infections, frequent diarrhea, and failure to thrive. Patients generally die within the first year of life unless treated with, e.g., bone marrow transplantation.

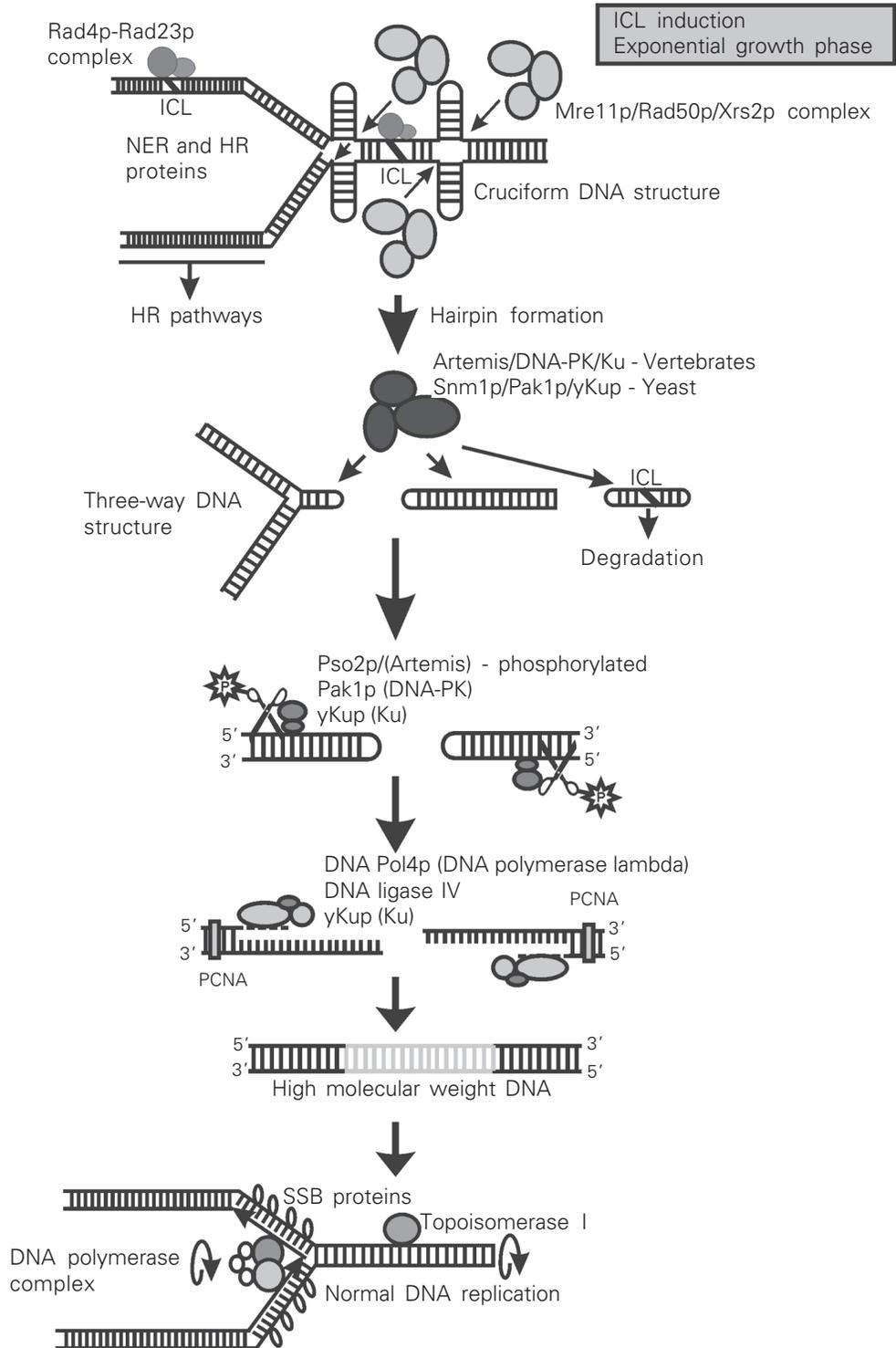
Artemis has 5' to 3' exonucleolytic activity with single-strand DNA specificity and, when associated with DNA-PKcs, forms a phosphorylated complex with endonucle-

olytic activity on both 5' and 3' DNA overhangs; furthermore, it can cleave hairpins generated by the Rag1/Rag2 proteins in V(D)J recombination (34,35). It has been shown that Artemis cooperates with p53 to suppress chromosomal translocations and tumor development in mice. Therefore, it can be considered a tumor suppressor gene. Like other NHEJ/p53 doubly deficient mice, most Artemis-deficient mice succumb to pro-B cell lymphomas by 11-12 weeks of age (10). Despite the striking relationship between NHEJ deficiencies and tumorigenesis in mouse models, potential roles for NHEJ in tumor suppression in humans have remained unclear (10). However, inactivating mutations of Ku70, Ku80, DNA-PKcs, XRCC4, and ligase IV have not been observed in the context of human immunodeficiencies, possibly because of a more severe impact of NHEJ mutations on human cells (10). In contrast, mutations in Artemis have been identified in several cohorts of human SCID patients (10). Therefore, the finding that Artemis functions as a tumor suppressor in mice raises the possibility of a similar function in humans. In this regard, hypomorphic alleles of Artemis have been identified in humans and have been associated with a predisposition to lymphomas (18).

Richardson and Jasin (7) observed that Artemis-deficient mice have increased numbers of chromosomal aberrations, e.g., chromosomal fragmentation, detached centromeres, fusions, and translocations. Artemis thus seems to play an important role as a genomic caretaker (10,18). In addition, Artemis may also function in telomere capping. This hypothesis is based on the increased levels of telomere fusions observed in Artemis-deficient embryonic stem cells (10). Although the precise function of Artemis with respect to telomeres remains unclear, it is highly probable that the Artemis-DNA-PKcs complex may not only function in V(D)J recombination and general DNA DSB repair, but also in telomere maintenance (10).

Interestingly, the use of a transposon system named *Sleeping Beauty* in an Artemis-deficient mammalian cell line does not increase the cell's sensitivity to DSB (51). *Sleeping Beauty* is a Tc1/mariner-like transposable element that, like retroviral integrases

Figure 5. Non-homologous end joining mediated by Snm1p/Pso2p in growing cells after inter-strand cross-linking (ICL) induction during DNA replication. In the presence of a sister strand, DNA repair may proceed via homologous recombination (HR pathways) mediated by Rad4p-Rad23p and HR proteins. Alternatively, the ICL can induce the formation of cruciform DNA structures, especially when palindromic sequences are present. These cruciform structures are recognized by Mre11p/Rad50p/Xrs2p complex that cuts the single-strand DNA regions and induces the formation of DNA hairpins. These DNA hairpins are cleaved by the phosphorylated Artemis (Snm1p-like) DNA-PK/Ku protein complex in metazoa or by phosphorylated Snm1/Pak1p/yKup in fungi, generating a substrate for DNA polymerase λ (Pol4p in yeast) and DNA ligase IV, which perform, together with Ku and PCNA, the rejoining of non-homologous DNA fragments (gray DNA chain) and restitute the DNA replication process. NER = nucleotide excision repair; PCNA = proliferating cell nuclear antigen; SSB = single-strand binding proteins.



and the Rag1 V(D)J recombinase, catalyzes a remarkably similar “overall chemistry” of DNA recombination. However, the structure of *Sleeping Beauty* transposition intermediates is unknown, and they probably do not comprise DNA hairpins, as was seen in *Ac/Ds* elements of maize (51).

Artemis protein was recently used by Poinsignon et al. (52) for site-specific mutagenesis in order to dissect the role of the metallo- β -lactamase and β -CASP domains of Artemis with regard to V(D)J recombination and DNA repair after ionizing radiation. This study demonstrated that Artemis can be divided into two critical regions, with the COOH-terminal region probably playing an important role in protein stabilization and in DNA repair after ionizing radiation (52). However, the authors concentrated their efforts on the study of the CRI and CRII of the Pso2p catalytic core (which encompasses the metallo- β -lactamase and β -CASP domains), necessary for V(D)J recombination but not for DNA repair. In this case, the CRIII should be required for DNA repair functions induced by ionizing radiation or even by ICLs.

The Pso2p/Snm1p of plants: a special case

In contrast to animals, plants are constantly being challenged by sunlight-contained UV radiation because of their obligatory requirement of sunlight for photosynthesis (53). This radiation penetrates plant surface tissues and damages their genome and other cellular targets such as photosystem II and plasma membrane ATPase (53). Characteristically, plants also show endophytic fungi living asymptotically within their tissues (54), where they can produce potentially DNA-damaging mycotoxins (55). Moreover, secondary metabolites (e.g., furocoumarin) can be photo-activated by sunlight and induce DNA ICLs in leaves or aerial parts (55). It is thus likely that different

DNA repair systems are required to repair the errors induced by biotic or abiotic factors in a plant's genome. The NHEJ process in plant tissues is largely unknown, and the DSB repair products have been characterized as excision products of transposable elements, or insertion products of *Agrobacterium* sp T-DNA (56). Interestingly, the analysis of NHEJ proteins in *A. thaliana* (e.g., DNA ligase IV, Ku80, and XRCC4) indicates the conservation of basic DSB repair mechanisms (26).

Using the available genomic information from public databases, we have carried out a phylogenetic study with the aim to find plant-specific Pso2p sequences. Interestingly, we detected paralogous *PSO2* genes in the complete genomes of *A. thaliana* and *O. sativa*, and also a new group of ATP-dependent DNA ligases that contain a Pso2p catalytic core (Table 1, Figure 4) (57). The sequence analyses of these proteins show that the Pso2p catalytic core is localized within the N-terminal part of the protein, while a DNA ligase I domain can be detected in the C-terminal end (Figure 2C), with both domains displaying homology with Pso2p and DNA ligase I of animals and yeasts. Moreover, additional data of microsynteny analysis indicate that these genes of the new DNA ligase family are linked to the *S* and *SLL2* loci of *Brassica* sp and *A. thaliana*, respectively. It should be noted that the *Brassica S* and the *Arabidopsis SLL2* loci consist of a gene complex with distinct stigma-expressed and anther-expressed sequences that determine i) self-incompatibility specificity, ii) some plant defense mechanisms, and iii) floral development (58). Taking into account all of the data obtained, we propose the definition of a new family of DNA ligases, named *LIG6*. Our present knowledge, sustained by theoretical data, suggests that these Lig6-orthologous proteins could be necessary to conserve genomic integrity in plant tissues, especially in reproductive organs with high DNA turnover, where the DNA ligase func-

tion seems to be essential. Biochemical analysis as well as mutational studies are currently in progress in order to determine the roles of these plant-specific DNA ligases in DNA metabolism.

Unfortunately, little is known about the Pso2 proteins in plants. However, the presence of paralogous *PSO2* genes in *A. thaliana* and *O. sativa* is a good indication that, like the tissue diversity found in metazoa, the presence of specialized plant tissues may have specific requirements for repair of DSB or ICL DNA repair.

Concluding remarks

The studies of Pso2p functions in DNA repair or in genome maintenance are just beginning. Since most of the information on putative Pso2p functions comes from its human homologue Artemis, more research is necessary in order to clarify the exact role of Pso2p in DNA metabolism. Since its first genetic studies using mutants of *S. cerevisiae* sensitive to photo-addition of bi-functional psoralens and to nitrogen mustards (39-44), little information has been obtained by conventional genetical approaches. If Pso2p is necessary for reconstitution of high molecular weight DNA, why do yeast mutants, cell lines, or even animal models knocked-out for *PSO2* show a wild-type response phenotype to DNA damaging agents, except ICL-generating chemicals? The answer to this question may be found in the structure of DNA, more specifically in the secondary structures like DNA hairpins that

can arise from palindromic regions during DNA replication slippage or stalled DNA replication forks (59). Recently, we proposed a model where Pso2p would act on DNA hairpin substrates induced by ICLs during DNA replication (46), a feature also shown in the present review (Figure 5). This model proposes that the potential endonucleolytic function of Pso2p is activated via Pak1p-induced phosphorylation. The specific function in DNA repair of this potential protein kinase of *S. cerevisiae* is unknown, but when over-expressed, Pak1p acts as a suppressor of thermo-labile DNA polymerase α mutations (60). Pak1p was identified in a two-hybrid screening of potential protein partners of Pso2p (Revers LF, Strauss M, Bonatto D, Brendel M and Henriques JAP, unpublished results). Our model helps to explain the specific function of Pso2p in repair of DSB that are generated during repair processing of ICLs. Moreover, it also helps to explain the evolution of Artemis in terms of its function on V(D)J recombination. Since Artemis also binds hairpin-capped DNA ends induced by RAG proteins, it may also have the ability to bind hairpin-intermediates generated during some step(s) of DNA ICL repair.

The existence of multiple *PSO2* paralogous genes in metazoa and plants suggests tissue-specific NHEJ functions that are not found in fungi, and this deserves the attention of all researchers interested in NHEJ. Possible new and exciting mechanisms of DNA repair, especially repair of DNA hairpins, could arise from the studies of Pso2p.

References

1. Lees-Miller SP & Meek K (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie*, 85: 1161-1173.
2. Jackson SP (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis*, 23: 687-696.
3. Kuzminov A (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proceedings of the National Academy of Sciences, USA*, 98: 8241-8246.
4. Bernstein C, Bernstein H, Payne CM & Garewal H (2002). DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutation Research*, 511: 145-178.
5. Khanna KK & Jackson SP (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nature Genetics*, 27: 247-254.
6. Rich T, Allen RL & Wyllie RH (2000). Defying death after DNA

- damage. *Nature*, 407: 777-783.
7. Richardson C & Jasin M (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature*, 405: 697-700.
 8. Rooney S, Alt FW, Lombard D, Whitlow S, Eckersdorff M, Fleming J, Fugmann S, Ferguson DO, Schatz DG & Sekiguchi J (2003). Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. *Journal of Experimental Medicine*, 197: 553-565.
 9. Takata M, Sasaki MS, Sonoda E, Morrison C, Hashimoto M, Utsumi H, Yamaguchi-Iwai Y, Shinohara A & Takeda S (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO Journal*, 17: 5497-5508.
 10. Richardson C & Jasin M (2000). Coupled homologous and non-homologous repair of a double-strand break preserves genomic integrity in mammalian cells. *Molecular and Cellular Biology*, 20: 9068-9075.
 11. Lieber MR, Ma Y, Pannicke U & Schwarz K (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nature Reviews. Molecular Cell Biology*, 4: 712-720.
 12. Gu Y, Seidl K, Rathbun GA et al. (1997). Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity*, 7: 653-665.
 13. Frank KM, Sekiguchi JM, Seidl KJ, Swat W, Rathbun GA, Cheng HL, Davidson L, Kangaloo L & Alt FW (1998). Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature*, 396: 173-177.
 14. Gao Y, Sun Y, Frank K et al. (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell*, 95: 891-902.
 15. Grawunder U, Zimmer D, Fugmann S, Schwarz K & Lieber MR (1998). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Molecular Cell*, 2: 477-484.
 16. Li GC, Ouyang H, Li X et al. (1998). Ku70: a candidate tumor suppressor gene for murine T cell lymphoma. *Molecular Cell*, 2: 1-8.
 17. Ma Y, Pannicke U, Schwarz K & Lieber MR (2002). Hairpin end opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in non-homologous end joining and V(D)J recombination. *Cell*, 108: 781-794.
 18. Rooney S, Sekiguchi J, Whitlow S, Eckersdorff M, Manis JP, Lee C, Ferguson DO & Alt FW (2004). Artemis and p53 cooperate to suppress oncogenic *N-myc* amplification in progenitor B cells. *Proceedings of the National Academy of Sciences, USA*, 101: 2410-2415.
 19. Labhart P (1999). Nonhomologous DNA end joining in cell-free systems. *European Journal of Biochemistry*, 265: 849-861.
 20. Smith GC & Jackson SP (1999). The DNA-dependent protein kinase. *Genes and Development*, 13: 916-934.
 21. Hartley KO, Gell D, Smith GC, Zhang H, Divecha N, Connelly MA, Admon A, Lees-Miller SP, Anderson CW & Jackson SP (1995). DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell*, 82: 849-856.
 22. Poltoratsky VP, Shi X, York JD, Lieber MR & Carter TH (1995). Human DNA-activated protein kinase (DNA-PK) is homologous to phosphatidylinositol kinases. *Journal of Immunology*, 155: 4529-4533.
 23. Dynan WS & Yoo S (1998). Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Research*, 26: 1551-1559.
 24. West CE, Waterworth WM, Story GW, Sunderland PA, Jiang Q & Bray CM (2002). Disruption of the *Arabidopsis AtKu80* gene demonstrates an essential role for AtKu80 protein in efficient repair of DNA double-strand breaks *in vivo*. *Plant Journal*, 31: 517-528.
 25. Tamura K, Adachi Y, Chiba K, Oguchi K & Takahashi H (2002). Identification of *Ku70* and *Ku80* homologues in *Arabidopsis thaliana*: evidence for a role in the repair of DNA double strand breaks. *Plant Journal*, 29: 771-781.
 26. van Attikum H, Bundock P, Overmeer RM, Lee L-Y, Gelvin SB & Hooykaas PJJ (2003). The *Arabidopsis AtLIG4* gene is required for the repair of DNA damage, but not for the integration of *Agrobacterium* T-DNA. *Nucleic Acids Research*, 31: 4247-4255.
 27. Lee KJ, Huang J, Takeda Y & Dynan WS (2000). DNA ligase IV and XRCC4 form a stable mixed tetramer that functions synergistically with other repair factors in a cell-free end-joining system. *Journal of Biological Chemistry*, 275: 34787-34796.
 28. Lee JW, Yannone SM, Chen DJ & Povirk LF (2003). Requirement for XRCC4 and DNA ligase IV in alignment-based gap filling for nonhomologous DNA end joining *in vitro*. *Cancer Research*, 63: 22-24.
 29. West CE, Waterworth WM, Jiang Q & Bray CM (2000). *Arabidopsis* DNA ligase IV is induced by γ -irradiation and interacts with an *Arabidopsis* homologue of the double-strand break repair protein XRCC4. *Plant Journal*, 24: 67-78.
 30. D'Amours D & Jackson SP (2002). The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Reviews in Molecular and Cellular Biology*, 3: 317-327.
 31. Callebaut I, Moshous D, Mornon J-P & de Villartay J-P (2002). Metallo- β -lactamase fold within nucleic acids processing enzymes: the β -CASP family. *Nucleic Acids Research*, 30: 3592-3601.
 32. Noordzij JG, Verkaik NS, Van der Burg M et al. (2003). Radiosensitive SCID patients with *Artemis* gene mutations show a complete B-cell differentiation arrest at the pre-B-cell receptor checkpoint in bone marrow. *Blood*, 101: 1446-1452.
 33. Moshous D, Li L, Chasseval R, Philippe N, Jabado N, Cowan MJ, Fischer A & de Villartay J-P (2000). A new gene involved in DNA double-strand break repair and V(D)J recombination is located on human chromosome 10p. *Human Molecular Genetics*, 9: 583-588.
 34. Jeggo P & O'Neill P (2002). The Greek goddess, Artemis, reveals the secrets of her cleavage. *DNA Repair*, 1: 771-777.
 35. Schlissel MS (2002). Does Artemis end the hunt for the hairpin-opening activity in V(D)J recombination? *Cell*, 109: 1-4.
 36. Tavtigian SV, Simard J, Teng DHF et al. (2001). A candidate prostate cancer susceptibility gene at chromosome 17p. *Nature Genetics*, 27: 172-180.
 37. Vogel A, Schilling O, Niecke M, Bettmer J & Meyer-Klaucke W (2002). *ElaC* encodes a novel binuclear zinc phosphodiesterase. *Journal of Biological Chemistry*, 277: 29078-29085.
 38. Chanfreau G, Noble SM & Guthrie C (1996). Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science*, 274: 1511-1514.
 39. Henriques JA & Moustacchi E (1980). Isolation and characterization of *pso* mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae*. *Genetics*, 95: 273-288.
 40. Magaña-Schwencke N, Henriques JAP, Chanet R & Moustacchi E (1982). The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair-deficient strains. *Proceedings of the National Academy of*

- Sciences, USA*, 79: 1722-1726.
41. Cassier C, Chanet R, Henriques JA & Moustacchi E (1980). The effects of three *PSO* genes on induced mutagenesis: a novel class of mutationally defective yeast. *Genetics*, 96: 841-857.
 42. Ruhland A, Haase E, Siede W & Brendel M (1981). Isolation of yeast mutants sensitive to the bifunctional alkylating agent nitrogen mustard. *Molecular and General Genetics*, 181: 346-351.
 43. Wilborn F & Brendel M (1989). Formation and stability of inter-strand cross-links induced by cis- and trans-diamminedichloroplatinum (II) in the DNA of *Saccharomyces cerevisiae* strains differing in repair capacity. *Current Genetics*, 16: 331-338.
 44. Brendel M & Henriques JA (2001). The *pso* mutants of *Saccharomyces cerevisiae* comprise two groups: one deficient in DNA repair and another with altered mutagen metabolism. *Mutation Research*, 489: 79-96.
 45. Brendel M, Bonatto D, Strauss M, Revers LF, Pungartnik C, Saffi J & Henriques JAP (2003). Role of *PSO* genes in repair of DNA damage of *Saccharomyces cerevisiae*. *Mutation Research*, 544: 179-193.
 46. Da Silva KVCL & Henriques JAP (1987). Induction of the cytoplasmic 'petite' mutation in *pso* mutants of *Saccharomyces cerevisiae* by photoaddition of furocoumarins or by ultraviolet radiation. *Mutagenesis*, 2: 187-197.
 47. Lambert S, Mason SJ, Barber LJ, Hartley JA, Pearce JA, Carr AM & McHugh PJ (2003). *Schizosaccharomyces pombe* checkpoint response to DNA inter-strand cross-links. *Molecular and Cellular Biology*, 23: 4728-4737.
 48. Zhang X, Richie C & Legerski RJ (2002). Translation of *hSNM1* is mediated by an internal ribosome entry site that upregulates expression during mitosis. *DNA Repair*, 1: 379-390.
 49. Grossmann KF, Ward AM & Moses RE (2000). *Saccharomyces cerevisiae* lacking *Snm1*, *Rev3* or *Rad51* have a normal S-phase but arrest permanently in G2 after cisplatin treatment. *Mutation Research*, 461: 1-13.
 50. Yu J, Marshall K, Yamaguchi M, Haber JE & Weil CF (2004). Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 24: 1351-1364.
 51. Izsvák Z, Stüwe EE, Fiedler D, Katzer A, Jeggo PA & Ivics Z (2004). Healing the wounds inflicted by *Sleeping Beauty* transposition by double-strand break repair in mammalian somatic cells. *Molecular Cell*, 13: 279-290.
 52. Poinson C, Moshous D, Callebaut I, de Chasseval R, Villey I & de Villartay J-P (2004). The metallo- β -lactamase/ β -CASP domain of Artemis constitutes the catalytic core for V(D)J recombination. *Journal of Experimental Medicine*, 199: 315-321.
 53. Tuteja N, Singh MB, Misra MK, Bhalla PL & Tuteja R (2001). Molecular mechanisms of DNA damage and repair: progress in plants. *Critical Reviews in Biochemistry and Molecular Biology*, 36: 337-397.
 54. Faeth SH (2002). Are endophytic fungi defensive plant mutualists. *OIKOS*, 98: 25-36.
 55. Choi JJ, Klosterman SJ & Hadwiger LA (2001). A comparison of the effects of DNA-damaging agents and biotic elicitors on the induction of plant defense genes, nuclear distortion, and cell death. *Plant Physiology*, 125: 752-762.
 56. Scott L, LaFoe D & Weil CF (1996). Adjacent sequences influence DNA repair accompanying transposon excision in maize. *Genetics*, 142: 237-246.
 57. Bonatto D, Brendel M & Henriques JAP (2005). A new group of plant specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis, and three-dimensional modeling. *Functional Plant Biology* (in press).
 58. Matton DP, Nass N, Clarke AE & Newbigin E (1994). Self-incompatibility: how plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences, USA*, 91: 1992-1997.
 59. Lobachev KS, Gordenin DA & Resnick MA (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell*, 108: 183-193.
 60. Hovland PG, Tecklenberg M & Sclafani RA (1997). Overexpression of the protein kinase Pak1 suppresses yeast DNA polymerase mutations. *Molecular and General Genetics*, 256: 45-53.