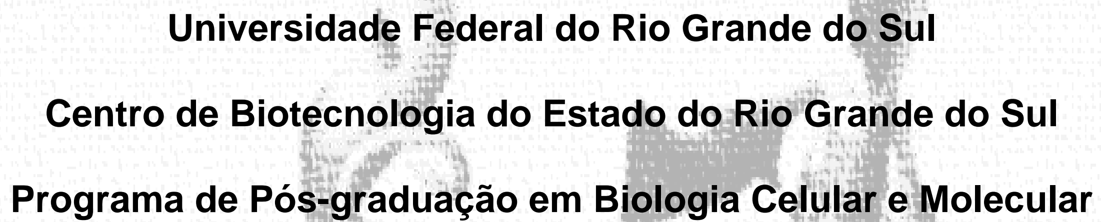


Universidade Federal do Rio Grande do Sul

Centro de Biotecnologia do Estado do Rio Grande do Sul

Programa de Pós-graduação em Biologia Celular e Molecular

A FAMÍLIA PSO2/SNM1 E SUAS POSSÍVEIS FUNÇÕES NA REPARAÇÃO DE DNA E NA MANUTENÇÃO GENÔMICA DE EUCARIOTOS

TESE DE DOUTORADO

DIEGO BONATTO

Porto Alegre, 2005.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO DE BIOTECNOLOGIA DO ESTADO DO RIO GRANDE DO SUL

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

A FAMÍLIA PSO2/SNM1 E SUAS POSSÍVEIS FUNÇÕES NA REPARAÇÃO DE DNA E NA MANUTENÇÃO GENÔMICA DE EUCARIOTOS

Diego Bonatto

Tese submetida ao Programa de Pós-graduação em Biologia Celular e Molecular da Universidade Federal do Rio Grande do Sul, como requisito para a obtenção do grau de Doutor em Ciências

Orientador: Prof. Dr. João Antonio Pêgas Henriques

Porto Alegre, março de 2005.

"NOTHING IN BIOLOGY MAKES SENSE,

EXCEPT IN THE LIGHT OF EVOLUTION"

THEODOSIUS DOBZHANSKY

(1900-1975)

ESTA TESE FOI ESCRITA NO VERÃO DE 2005, ANO EM QUE SE COMEMORA:

203 anos da criação do termo "biologia" por Gottfried Reinhold Treviranus (*Biologie oder Philosophie der lebenden Natur*) e por Jean-Baptiste Lamarck (*Hydrogéologie*) (1802);

146 anos da publicação de *A origem das espécies*, de Charles Darwin (1859);

105 anos da redescoberta do trabalho de Gregor Mendel (1900);

99 anos da criação do termo "genética", por William Bateson (1906);

96 anos da criação do termo "gene", por Wilhelm Johannsen (1909);

61 anos da identificação do DNA como elemento constitutivo do gene por Oswald T. Avery (1944);

61 anos da publicação do ensaio *What is life?*, de Erwin Schrödinger (1944);

52 anos da descrição da estrutura em dupla hélice do DNA por James D. Watson e Francis Crick (1953);

9 anos da decifração completa do genoma de *Saccharomyces cerevisae* (1996);

4 anos da decifração completa do genoma de *Homo sapiens* (2001).

APRESENTAÇÃO

Este trabalho foi desenvolvido nas dependências do Laboratório de Radiobiologia Molecular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul. O projeto foi subsidiado pela Fundação de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo Conselho Nacional de Auxílio a Pesquisa e ao Desenvolvimento (CNPq), pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) e pelo Laboratório de Genotoxicidade do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul (GENOTOX).

AGRADECIMENTOS

Ao Mestre e Prof. Dr. João Antonio Pêgas Henriques pela confiança e amizade despendida ao longo destes dez anos em que estou trabalhando em seu laboratório, pelos ensinamentos científicos, pela orientação livre, por sua dedicação em me tornar um profissional habilitado para fazer pesquisas e por mostrar que a Ciência é, antes de tudo, a interação de muitas pessoas, com formações acadêmicas e visões de mundo diferentes, todos em busca de um ideal comum. Saber liderar esta "sociedade" é uma arte que poucas pessoas tem, e o Prof. Henriques é um líder nato. Não vejo exemplo melhor para seguir profissionalmente.

Ao Prof. Dr. Martin Brendel, por todas as suas sugestões para melhorar a qualidade dos artigos, seja na correção da língua inglesa ou sugerindo novas formas para expressar uma idéia que não estava clara.

Aos Profs. Drs. Giancarlo Pasquali e Arthur G. Fett-Neto pelas valiosas sugestões e colaborações como membros da minha Comissão de Acompanhamento.

Ao Programa de Pós-graduação em Biologia Celular e Molecular (PPGBCM) pelas oportunidades e pelo apoio financeiro.

A Sílvia e ao Luciano, da secretaria do PPGBCM, pela rapidez e pela alta qualidade de seus serviços, estando sempre prontos a ajudar, sejam quais forem os problemas...

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro nestes quatro anos em que realizei meu doutorado.

A Márcia Vaz, secretária do Prof. João A. P. Henriques e colega bióloga. O que seria de nós sem aquela voz no telefone dizendo "Tem café?" ou "Poderia vir aqui na sala do Prof. Henriques rapidinho?". Muito obrigado por todas as conversas, alegres ou não, nestes anos de convívio.

A todas as pessoas, o qual eu não vou citar nomes sem correr o risco de esquecer alguém, que passaram ou que ainda estão no laboratório 219 do Centro de Biotecnologia. Muito obrigado por todas as ajudas e discussões, tanto no âmbito científico quanto no âmbito pessoal. Foi também neste laboratório que eu comecei a minha longa jornada pela Ciência. Às vezes vem aquela saudade dos tempos em que aí trabalhava...(bons tempos aqueles!)

É óbvio que, se eu agradeci ao pessoal do 219, também agradeço ao pessoal do laboratório 210 e 212 do Departamento de Biofísica. É muito divertido trabalhar com todos vocês, e muitas das situações que surgem no dia-a-dia são dignas de uma comédia do tipo "pastelão". Mas nem tudo é comédia, e as pessoas que aí trabalham são exemplos profissionais a serem seguidos. Em especial, agradeço a meu amigo e colega Renato por todas as "indiadas", festas da Farmácia, discussões, conselhos, fofocas e por aí segue a lista...

A todos os professores e funcionários do Departamento de Biofísica os meus sinceros agradecimentos pelo agradável e divertido convívio em todos estes anos.

A minha família, meu porto seguro e meu apoio para todos os momentos, bons ou ruins. Não há palavras para agradecer o suporte que eu continuo recebendo.

A Kátia, minha amada e companheira de todas as horas, de todas as viagens, de todas as conversas, bom, de tudo! A Kátia é o complemento da minha alma e dos meus pensamentos. Somos um só...

Por fim, agradeço a mim mesmo, por ter conseguido chegar até este ponto e, espero, chegar muito mais além! Não é fácil trilhar a via do conhecimento, pois ela exige um nível de abnegação das coisas particulares que poucas pessoas tem em nome da pesquisa científica. Contudo, nada substitui o sabor e o valor de uma descoberta científica, por mais difícil que ela seja (na maioria das vezes, as descobertas são feitas, literalmente, com sangue, suor e lágrimas!). Além disso, são poucas as pessoas que tem a sorte (como eu tive)

de fazer uma Pós-graduação em uma excelente universidade, em um ótimo laboratório e com um orientador que é considerado uma referência nacional e internacional na área de reparação de DNA e mutagênese. É incrível como uma escolha, mesmo que aparentemente pequena ou sem importância para o momento, se torna fundamental com o passar do tempo.

ESTRUTURAÇÃO DA TESE

A presente tese encontra-se estruturada da seguinte forma: uma introdução geral, os objetivos (gerais e específicos), os quatro capítulos principais escritos na forma de artigos científicos (conforme as normas das revistas para as quais foram submetidos), uma discussão separada em subitens, conclusões (gerais e específicas) e perspectivas.

A introdução aborda alguns conceitos importantes para o entendimento dos processos de reparação de DNA em eucariotos, tais como os tipos de danos ou lesões encontrados no DNA (bases modificadas, adutos mono- e bifuncionais e quebras simples e duplas de DNA), os mecanismos de reparação mais conhecidos, com ênfase para a reparação recombinacional e os genes *PSO*. Na introdução também são abordados os principais tipos de técnicas de análises de seqüências e de filogenia, ressaltando a sua importância para o estudo de proteínas de função ainda não conhecida ou pouco caracterizada.

O Capítulo 1 trata de uma revisão geral sobre a família Pso2p/Snm1p, utilizando dados conhecidos até aquele momento. Nesta revisão são incluídos alguns resultados de análises filogenéticas e de seqüências, assim como um possível mecanismo de atuação das proteínas Pso2p/Snm1p por uma via recombinacional do tipo NHEJ. Este capítulo foi aceito para a publicação no periódico *Brazilian Journal of Medical and Biological Research* e encontra-se no prelo.

O Capítulo 2 descreve um estudo filogenético e de seqüências da família Pso2p/Snm1p, detalhando os principais grupos encontrados e suas funções na reparação de DNA, na manutenção da cromatina e na geração de diversidade biológica. Este capítulo foi submetido à publicação para *Proteins: Structure, Function, and Bioinformatics.*

Uma análise filogenética e de seqüências do grupo Ártemis, pertencentes à família Pso2p/Snm1p, é mostrado no Capítulo 3 desta tese. As análises de um importante grupo de

proteínas para a reparação de DNA e para o sistema imunológico adaptativo de metazoário permitiu identificar novas seqüências semelhantes à Ártemis no genoma de algumas espécies de fungos e metazoários. Este trabalho foi submetido para a publicação no periódico *The Protein Journal* e encontra-se na etapa de avaliação pelos revisores.

O Capítulo 4 introduz e analisa uma nova família de DNA ligases exclusivas de plantas. Estas DNA ligases caracterizam-se por possuírem dois domínios: o Pso2p/Snm1p na região N-terminal e o domínio DNA ligase I na região C-terminal. As possíveis funções destas DNA ligases são discutidas no âmbito da fisiologia vegetal, assim como para a reparação de DNA em plantas. Este capítulo foi aceito para a publicação no periódico *Functional Plant Biology*.

Na seqüência apresenta-se uma discussão geral inter-relacionando os resultados descritos nos capítulos acima, as conclusões e as perspectivas geradas por esta tese.

No item 'Anexos' encontram-se: (i) uma tabela descrevendo os principais genes de reparação conhecidos e as vias em que atuam e (ii) uma revisão sobre os genes *PSO* de *Saccharomyces cerevisiae*, publicado no periódico *Mutation Research*.

ÍNDICE

LISTA DE ABREVIATURAS .... *XIV*

RESUMO .... *XVII*

ABSTRACT .... *XIX*

INTRODUÇÃO 1

1. INTRODUÇÃO GERAL *2*

2. TIPOS DE LESÕES MAIS COMUNS AO DNA *4*

2.1. Bases modificadas *6*

2.2. Adutos mono- e bifuncionais *8*

2.3. Quebras simples e duplas de DNA *13*

3. MECANISMOS DE REPARAÇÃO DE DNA *14*

3.1. Vias de reparação de DNA *14*

3.2. Reparação recombinacional *19*

3.2.1. Recombinação homóloga 20

3.2.2. Recombinação não homóloga 24

3.2.3. Recombinação V(D)J 26

4. OS GENES PSO *29*

5. O USO DA ANÁLISE DE SEQÜÊNCIAS E DA FILOGENIA PARA A PREDIÇÃO DE FUNÇÕES PROTÉICAS *32*

5.1. Análise de seqüências *32*

5.1.1. Alinhamento múltiplo de seqüências 32

5.1.2. Análise de agrupamentos hidrofóbicos (HCA) 34

5.2. Análises filogenéticas *37*

OBJETIVOS 40

1. Objetivo geral *41*

2. Objetivos específicos *41*

CAPÍTULO UM 42

The eukaryotic Pso2/Snm1/Artemis Proteins and their function as genomic and cellular caretakers *42*

CAPÍTULO DOIS 57

The Eukaryotic Pso2p/Snm1p Family Revisited: In Silico Analyses of Pso2p A, B and Plasmodium Groups *57*

CAPÍTULO TRÊS 97

In silico Identification and Analysis of New Artemis/Artemis-Like Sequences from Fungal and Metazoan Species *97*

CAPÍTULO QUATRO 129

A new group of plant-specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis, and 3-dimensional modeling *129*

DISCUSSÃO GERAL 144

Discussão geral *145*

O papel da família Pso2p/Snm1p na reparação de ICLs e na manutenção da cromatina *147*

O papel dos grupos Ártemis e Plasmodium na geração de diversidade biológica em metazoários, fungos e protozoários *156*

As seqüências Pso2p/Snm1p de plantas *160*

CONCLUSÕES 164

Conclusão geral *165*

Conclusões específicas *165*

PERSPECTIVAS 168

Perspectivas *169*

REFERÊNCIAS 172

Referências bibliográficas *173*

ANEXOS 195

CURRICULUM VITAE 220

LISTA DE ABREVIATURAS

3-CPS	3-carbetoxipsoraleno
8-MOP	8-metoxipsoraleno
8-OH-Gua	8-oxoguanina
ATM	*Ataxia Telangiectasia Mutated protein* (proteína mutada da ataxia telangiectasia)
ATR	*ATM-/Rad3-related protein* (proteína relacionada a ATM e Rad3p)
BER	*Base excision repair* (reparação por excisão de bases)
BLAST	*Basic local alignment search tools* (ferramentas básicas de alinhamento local)
BLOSUM	*Blocks substitution matrix* (matriz de blocos de substituição)
CAII	*Carbonic anhydrase II* (anidrase carbônica do tipo II)
CPDs	*Ciclobutane pyrimidine dimers* (dímeros de pirimidina)
CPSF 73 kDa/100 kDa	*Cleavage and polyadenylation specificity factor 73 kDa/100 kDa* (fator específico de clivagem e de poliadenilação de 73 kDa/100 kDa)
CRs	*Conserved regions* (regiões conservadas)
DNA-PKcs	*DNA-dependent protein kinase catalytic subunit* (subunidade catalítica da proteína cinase dependente de DNA
DSBR	*Double strand break repair* (reparação de quebras duplas de fita de DNA)

DSBs	*Double strand breaks* (quebras duplas de fita de DNA)
EMS	Etil-metanossulfonato
EROS	Espécies reativas de oxigênio
HATs	*Histone acetiltransferase* (acetiltransferases de histonas)
HCA	*Hydrophobic cluster analysis* (análise de agrupamentos hidrofóbicos)
HMGs	*High mobility group proteins* (proteínas de alta mobilidade)
HN2	Mostarda nitrogenada bifuncional
HR	*Homologous recombination* (recombinação homóloga)
ICLs	*Interstrand cross-links* (pontes intercadeias)
Ig	Imunoglobulinas do tipo g
IR	*Ionizing radiation* (radiação ionizante)
LUCA	*Last universal common ancestor* (ultimo ancestral comum universal)
ME	*Minimum evolution* (evolução mínima)
MMR	*Mismatch repair* (reparação de erros de emparelhamento de bases)
NER	*Nucleotide excision repair* (reparação por excisão de nucleotídeos)
NHEJ	*Non-homologous end joining recombination* (recombinação não homóloga)
NJ	*Neighbor-joining*

PAM	*Point accepted mutation* (mutações pontuais aceitáveis)
PHR	*Photoreactivation repair* (reparação de DNA por fotorreativação)
PSI-BLAST	*Position specific Iterated-BLAST* (BLAST com interação de posições específicas)
PUVA	Psoralenos + UVA
Rags	*Recombination activating genes* (genes ativadores de recombinação)
RS-SCID	*Radiosensitivity-SCID*
RSSs	*Recombination signal sequences* (sequências sinalizadoras de recombinação)
SCID	*Severe combined immunodeficiency syndrome* (sindrome da imunodeficiência severa combinada)
SDSA	*Sinthesis-dependent strand annealing* (síntese dependente de anelamento de fita de DNA)
SSBs	*Single strand breaks* (quebras simples de fita de DNA)
TCR/BCR	*T or B cell receptor* (receptor das células B ou T)
Tg	Timina glicol
TOR	*Target of rapamicyn protein* (proteína alvo da rapamicina)
UV	Luz ultravioleta
UVA	Luz ultravioleta de 365 nm
UVB	Luz ultravioleta de 295-320 nm

RESUMO

O genoma das células eucarióticas é um dos principais alvos para danos induzidos por inúmeros fatores ambientes, sejam estes de origem biótica ou abiótica. Considerando a complexidade da molécula de DNA, não é supreendente que existam diferentes tipos de lesões com os mais variados graus de severidade. Dentre todas as lesões que podem ser induzidas no DNA, as pontes intercadeias (ICLs) estão entre as mais graves. Se não forem reparadas, a presença de apenas um ICL pode ser letal para a célula. Além disso, as lesões do tipo ICLs são quimicamente heterogêneas, podendo modificar a estrutura do DNA de forma permanente ou temporária.

Os mecanismos relacionados à reparação de ICLs ainda são pouco conhecidos em eucariotos. Apesar de várias proteínas terem sido descritas como essenciais ao processo, não há um modelo único que explique esta reparação. Contudo, dentre as diferentes proteínas que participam na reparação de ICLs, destacam-se as nucleases Pso2/Snm1.

A forma de atuação das proteínas Pso2/Snm1 não é conhecida, mas inúmeros dados obtidos com mutantes de *Saccharomyces cerevisiae* e, recentemente, com células de mamífero, mostram que a ausência de Pso2p/Snm1p bloqueia a restituição do DNA de alta massa molecular. Por outro lado, tem sido mostrado que o Pso2p/Snm1p provavelmente atua na manutenção da cromatina, mas de uma forma ainda não completamente esclarecida. Uma das proteínas pertencentes à família Pso2p/Snm1p, Ártemis, possui um papel importante no desenvolvimento do sistema imunológico adaptativo de metazoários e parece ser essencial para outros processos relacionados ao metabolismo de DNA eucariótico.

Desta maneira, este trabalho teve como objetivo principal o estudo da família Pso2p/Snm1p por meio da análise filogenética e de seqüências, comparando-a com proteínas homólogas já descritas em outros organismos. Além disso, esta comparação

permitiu estabelecer uma correlação funcional entre as proteínas em termos de reparação de DNA e manutenção da cromatina eucariótica.

As análises de filogenia e de seqüências claramente demonstraram que as proteínas Pso2/Snm1 podem ser agrupadas em quatro grupos principais ao invés de três, ao contrário do que se conhecia previamente. Três destes grupos, por sua vez, são formados por subgrupos específicos, que possivelmente atuam de forma diferenciada na reparação de DNA, na manutenção da cromatina e na geração de diversidade biológica. Por outro lado, os estudos das seqüências Pso2/Snm1, baseados principalmente na técnica de análises de agrupamentos hidrofóbicos (HCA), revelaram um alto grau de similaridade de estruturas primárias e secundárias entre os diferentes grupos, um indicativo da importância estrutural para a função destas proteínas no metabolismo de DNA. A técnica de HCA permitiu mapear regiões conservadas (CRs) em todas as seqüências estudadas, compondo o chamado domínio Pso2p/Snm1p. Em alguns casos, o domínio Pso2p/Snm1p encontra-se fusionado a outros domínios catalíticos. Neste caso, destaca-se o estudo de uma nova família de DNA ligases dependentes de ATP que são exclusivas de plantas. Esta nova família, denominada de Lig6p, parece ter funções importantes no metabolismo do DNA de plantas, sendo esta a primeira DNA ligase eucariótica com função nucleásica identificada.

Usando os dados obtidos neste trabalho em conjunto com os resultados de outros autores, é sugerido um possível modo de atuação das proteínas Pso2p/Snm1p na reparação de danos do tipo ICL, na manutenção da cromatina e na geração de diversidade biológica.

ABSTRACT

The genome of eukaryotic cells is a major target for damages induced by environmental abiotic and biotic factors. Taking into account the molecular complexity of DNA, it is not surprising that there are different types of lesions, each lesion with a variable degree of severity. Considering all lesions that are found in DNA, the interstrand cross links (ICLs) are the most dangerous. If not correctly repaired, one ICL can induce cell death. Moreover, ICLs are chemically heterogeneous lesions, and ICLs have the ability to induce conformational changes in the DNA structure, in a temporary or permanent way.

The biochemical pathways related to ICL repair are still poorly studied in eukaryotes. Despite the discover of many proteins that seem to be essential for ICL repair, there is not a unique model or pathway that explains the eukaryotic ICL repair processes. However, amongst all proteins described, the nuclease Pso2/Snm1 proteins have an important role in ICL repair.

The mode of action of Pso2/Snm1 proteins are not known, but data obtained with *Saccharomyces cerevisiae* mutant strains and more recently with mammalian cells indicate that the absence of Pso2p/Snm1p blocks the reconstitution and restablishment of high molecular weigth DNA. On the other hand, it was shown that Pso2p/Snm1p probably acts as an eukaryotic genome caretaker in a not fully understood manner. One protein that belongs to the Pso2p/Snm1p family, Artemis, has a central role in the development of the adaptive immunological system of metazoans, and seems to be essential for other DNA metabolic processes.

Therefore, the purpose of this work was to study the Pso2p/Snm1p family by means of phylogenetic and sequences analyses, comparing the Pso2/Snm1 sequences with homologous proteins already described for other organisms. Moreover, this comparison

allows to estabilish a functional correlation between them for DNA repair and eukaryotic genome caretaking.

The phylogenetic and sequence analyses clearly demonstrated that Pso2/Snm1 proteins can be grouped in four major groups instead of three, as previously known. Three of four groups are composed of specific subgroups that probably act in DNA repair, genome caretaking and generation of biological diversity in a non-redundant manner. On the other hand, the hydrophobic cluster analysis (HCA) of Pso2/Snm1 sequences from different groups indicated a high degree of primary and secondary structure similarities, indicating that the structure is important for protein functions in DNA metabolism. The HCA allowed to map four conserved regions (CRs) in all Pso2/Snm1 proteins studied, which together compose the Pso2p/Snm1p domain. It was seen that, in some cases, the Pso2p/Snm1p is found fused to other catalytic domains. A new family of plant-specific ATP-dependent DNA ligases characterizes an example of a fusion between Pso2p/Snm1p with other domains. This new family, named Lig6p, probably has important functions in plant DNA metabolism. Moreover, this is the first DNA ligase with nucleasic function identified in eukaryotes.

Taking the results obtained in this work together with the data of other authors, a mode of action of Pso2p/Snm1p in the repair of ICLs, genome caretaking, and generation of biological diversity is suggested.

I

Introdução

1. Introdução geral

As células eucarióticas, de uma maneira geral, são constantemente desafiadas pelo ambiente que as rodeia. Sendo formadas por inúmeras moléculas orgânicas, por íons metálicos e não-metálicos, as células são estruturas bioquímicas complexas e dinâmicas, que podem interagir e reagir com agentes ambientais de natureza química ou física, resultando na geração de respostas celulares apropriadas para cada agente.

Duas grandes classes de moléculas orgânicas compõem a vasta rede bioquímica da célula eucariótica. A primeira classe de compostos orgânicos é formada por moléculas químicas simples, mas fundamentais para a manutenção da vitalidade celular, tais como açúcares redutores, aminoácidos, ácidos graxos, bases nitrogenadas, nucleosídeos e nucleotídeos, entre outros (Nelson & Cox, 2000). A segunda classe de compostos orgânicos é formada por macromoléculas provenientes da polimerização controlada de substâncias simples (Nelson & Cox, 2000). Esta classe pode ser dividida em três grupos principais de macromoléculas: (i) os polissacarídeos (formados pela polimerização de açúcares), (ii) as proteínas (um polímero de aminoácidos) e (iii) os ácidos nucléicos [polímeros formados por ribonucleotídeos (RNA) ou por desoxirribonucleotídeos (DNA)] (Nelson & Cox, 2000).

As proteínas e os ácidos nucléicos são moléculas que carregam informação biológica que, em última análise, geram as respostas celulares aos estímulos/agentes provenientes do ambiente (Nelson & Cox, 2000). Estes estímulos, de uma maneira geral, podem sinalizar para o início ou a parada da divisão celular, para processos de diferenciação tecidual, para a captação de nutrientes do ambiente, para a excreção de moléculas sinalizadoras, para a reparação de danos ou mesmo para induzir a célula à morte em determinadas condições (Nelson & Cox, 2000). Contudo, muitos destes estímulos também são capazes de gerar danos químicos (temporários ou permanentes) em proteínas e em ácidos nucléicos, resultando na modificação ou na perda total da informação original. A

perda destas informações produz uma modificação do fenótipo celular o qual, em sua maioria, inviabiliza o organismo e resulta na incapacidade do mesmo em sobreviver às mudanças ambientais.

Neste sentido, os ácidos nucléicos são alvos fáceis para agentes indutores de danos, sejam estes de origem biológica, química ou física. Por exemplo, mais de 10^4 lesões diárias ocorrem no genoma de mamíferos, provenientes do tautomerismo e da oxidação de bases, assim como por erros de replicação (Slupphaug *et al.,* 2003). Os agentes químicos e físicos externos, como a fumaça de tabaco, os derivados petroquímicos, a radiação ultravioleta (UV) e a radiação ionizante são importantes para a indução de danos no DNA eucariótico (Slupphaug *et al.,* 2003). Por sua vez, os agentes químicos de origem biológica, como as espécies reativas de oxigênio (EROs), também são uma fonte importante de danos. Neste caso, cabe salientar que as modificações que ocorrem no equilíbrio redox celular aumentam enormemente os danos causados ao DNA (Slupphaug *et al.,* 2003).

Entretanto, os danos podem ser propositalmente introduzidos no genoma de uma determinada célula por processos biológicos distintos. Tais modificações possuem o propósito único de gerar diversidade biológica como uma forma de aumentar a sobrevivência e a resposta do organismo a agentes patogênicos (como é o caso do sistema imunológico adaptativo de vertebrados mandibulados) ou mesmo para gerar variabilidade genética, tal como acontece na meiose eucariótica (Pâques & Harber, 1999; Jeggo, 2002).

Em todos os casos, os danos gerados no DNA, sejam estes de forma acidental ou proposital, necessitam ser reparados. A molécula de DNA não pode ser descartada ou reciclada quando danificada, enquanto que outras macromoléculas, como RNA, proteínas e polissacarídeos são regularmente degradadas e ressintetizadas com base nas informações contidas no DNA (Mitchell *et al.*, 2003). Por este motivo, a ausência de reparação é uma das causas principais para o surgimento de doenças hereditárias, de câncer e de

envelhecimento precoce, além de estar relacionada a disfunções imunológicas em mamíferos ou a deficiências neurológicas (Mitchell *et al.*, 2003).

A reparação do DNA é um processo bioquimicamente complexo, o qual emprega inúmeras proteínas com funções distintas, tanto espacialmente quanto temporalmente. Conforme a natureza do dano ou a sua extensão, determinados complexos protéicos serão preferencialmente utilizados. Estes complexos protéicos definem as chamadas 'vias de reparação de DNA' e são responsáveis pela manutenção do genoma em quaisquer condições fisiológicas. Interessantemente, uma sobreposição funcional entre as diferentes vias de reparação de DNA é a regra na célula em vez da exceção (Lindahl & Wood, 1999; Slupphaug *et al.,* 2003), de forma que vários complexos atuam ao mesmo tempo na reparação de diferentes tipos de danos.

2. Tipos de lesões mais comuns ao DNA

Os danos na molécula de DNA compreendem modificações químicas que ocorrem nas bases nitrogenadas (purinas e pirimidinas), na desoxirribose e na ponte fosfodiéster (Figuras 1 e 2). Estas modificações, na sua totalidade, são provenientes da radiação solar (na forma de luz UV e radiação ionizante), da geração de EROs através de diferentes processos metabólicos ou mesmo de substâncias químicas de origem natural e/ou sintética (Figura 1).

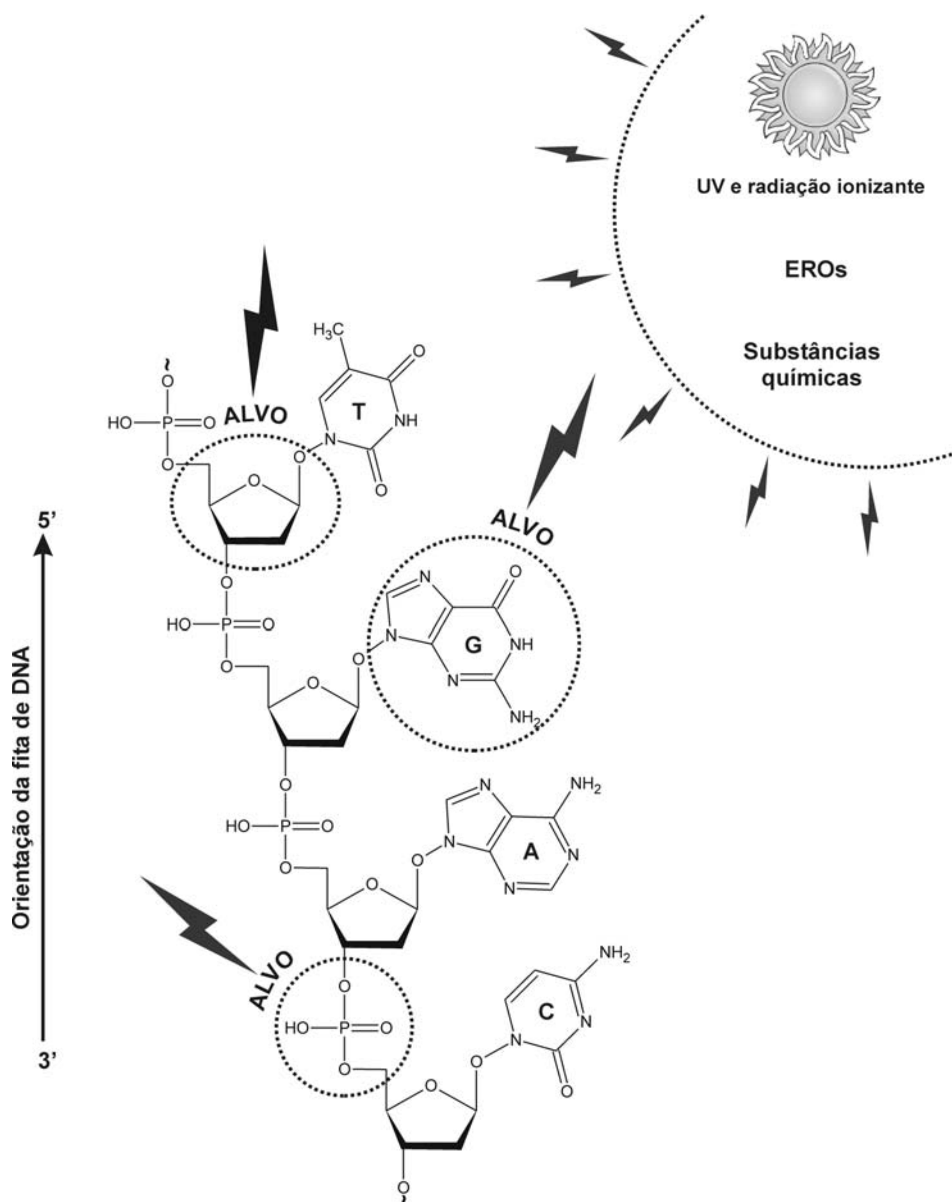


Figura 1. Principais alvos para danos na molécula de DNA. Os danos, que podem acontecer nas bases, na desoxirribose ou na ponte fosfodiéster (indicados pelos círculos pontilhados) são originados a partir da radiação solar (na forma de radiação UV e ionizante), de EROs e de substâncias químicas. A orientação da fita de DNA é mostrada na figura. Siglas: timina, guanina, adenina e citosina (T, G, A e C, respectivamente).

2.1. Bases modificadas

De todas as alterações químicas que podem ocorrer no DNA, as bases modificadas são as mais comuns e também as mais perigosas, desde que não reparadas corretamente (Figura 2; Slupphaug *et al.,* 2003; Evans *et al.,* 2004). Os processos químicos que originam as modificações nas bases nitrogenadas são complexos em sua natureza (Slupphaug *et al*., 2003), e resultam de uma maior ou menor interação dos agentes indutores de danos *per se* com moléculas de importância biológica, tanto de natureza enzimática (superóxido dismutases, tiorredoxinas, citocromo P450, entre outros), como de natureza não-enzimática (por exemplo, glutationa, aminoácidos sulfurados, água ou O_2) (Slupphaug *et al.,* 2003; Evans *et al.,* 2004).

As bases modificadas que mais prevalecem no DNA são a 8-oxoguanina (8-OH-Gua) e a timina glicol (Tg) (Slupphaug *et al.,* 2003). A 8-OH-Gua é uma base fracamente mutagênica (freqüência de mutação de 2,5% a 4,8% em células de mamíferos; Leadon *et al.,* 1997), e resulta, em grande parte, em substituições do tipo G $\rightarrow$ T. Alternativamente, a 8-OH-Gua pode emparelhar erroneamente com uma adenina, produzindo uma substituição do tipo A $\rightarrow$ T (Cheng *et al.,* 1992). Por outro lado, a Tg é considerada como uma base pouco mutagênica (freqüência de mutação de 0,3% em células de mamíferos), causando substituições do tipo T $\rightarrow$ C (Basu *et al.,* 1989). Contudo, a principal alteração causada pela Tg é a modificação da estrutura do DNA, o que leva a um bloqueio do processo de replicação (McNulty *et al.,* 1998). Este mesmo fenômeno é observado em outra alteração de base: os dímeros de pirimidinas ou CPDs (*ciclobutane pyrimidine dimers;* Figura 2). Diversos tipos de CPDs podem ser gerados *in vivo* na presença de luz solar, tais como os dímeros de timina-timina, a pirimidina-pirimidona (6,4)-fotoprodutos e os isômeros de 'Dewar' (Vink & Roza, 2000). Os CPDs podem provocar modificações na estrutura do DNA, causando bloqueios nos processos de replicação e transcrição. Além disso, grande

parte da atividade mutagênica observada em células de mamíferos deve-se à formação de CPDs (Ravanat *et al.,* 2001). Interessantemente, os CPDs foram as primeiras lesões de DNA estudadas (Beukers *et al.,* 1960) e ainda possuem um papel de destaque nas pesquisas relacionadas com as modificações de bases provocadas por agentes químicos e físicos, em especial a luz solar (Vink & Roza, 2001; Douki *et al.,* 2003). Recentemente, a atuação da luz UVA na formação de CPDs ganhou importância ao ser mostrado que esta radiação é a principal fonte de indução de CPDs em células de mamíferos, o que explicaria a alta incidência de tumores de pele após a exposição prolongada a radiação solar em seres humanos (Douki *et al.,* 2003).

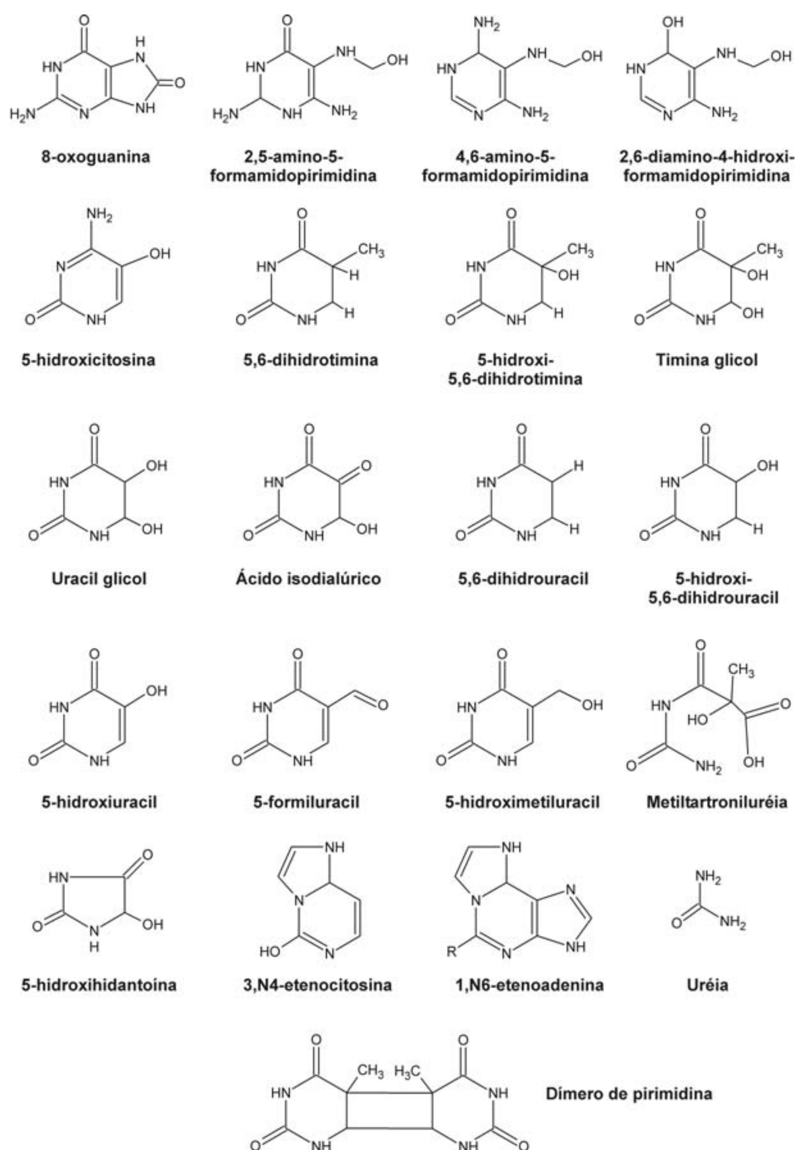


Figura 2. Exemplos de bases modificadas, comumente encontradas na molécula de DNA, provenientes da ação de radiação UV, ionizante, EROs e de agentes químicos (adaptado de Slupphaug *et al.,* 2003).

2.2. Adutos mono- e bifuncionais

A formação de ligações covalentes entre determinadas classes de substâncias químicas e as bases nitrogenadas do DNA é conhecida como adutos (Sharer, 2005). Quimicamente, os compostos ou agentes formadores de adutos podem ser separados em dois grandes grupos: (i) agentes monofuncionais, quando induzem a formação de ligação

covalente entre um composto químico e apenas uma base nitrogenada (adutos monofuncionais) e (ii) agentes bi- ou polifuncionais, quando a substância química possui a habilidade de se ligar covalentemente a duas bases nitrogenadas (adutos bifuncionais), estejam estas situadas na mesma fita de DNA (pontes intracadeia) ou em fitas separadas (pontes intercadeias – *interstrand cross-links* ou ICLs) (Sharer, 2005).

Os agentes monofuncionais compreendem várias classes de substâncias químicas dos quais destacam-se os agentes alquilantes. Estes agentes possuem a propriedade de se ligarem às bases nitrogenadas do DNA por meio de grupos metila ou etila (Sanderson & Shield, 1996). Neste grupo predominam vários compostos com finalidades terapêuticas, especialmente para o tratamento de tumores. Um exemplo importante deste grupo é a ecteinascidina-743 (ET-743; Figura 3a), um alcalóide tetraidroisoquinolínico isolado do urocordata *Ecteinascidia turbinata*, que possui alta eficiência no combate de sarcomas de tecidos moles (D'Incalci *et al.,* 2002; Soares *et al.,* manuscrito submetido para a publicação).

Por outro lado, os agentes bi- ou polifuncionais podem atuar tanto na indução de adutos monofuncionais quanto de pontes intracadeias e ICLs (Dronkert & Kanaar, 2001). Uma ampla variedade de compostos químicos são considerados agentes bifuncionais, sendo que muitos destes possuem aplicações diretas na terapia clínica para tratamento de tumores ou patologias de pele (Dronkert & Kanaar, 2001). Dentre os agentes bifuncionais com importância médica estão os psoralenos, a mitomicina C, a cisplatina e as mostardas nitrogenadas (Figura 3a). Os psoralenos são representados por moléculas pertencentes à classe das furocumarinas, que consistem de metabólitos secundários isolados de plantas das famílias *Umbelliferae, Rutaceae, Moraceeae e Leguminosae* (Guo & Yamazoe, 2004). Vários psoralenos sintéticos estão hoje disponíveis para o uso clínico, dos quais destaca-se o 8-metoxipsoraleno (8-MOP, Figura 3a), especialmente para o tratamento de psoríase, de vitiligo e de mais 30 tipos diferentes de patologias de pele (Morison, 2004) O modo como

os psoralenos, em especial o 8-MOP, formam ICLs no DNA é bastante conhecido (Bethea *et al.,* 1999) e consiste de duas etapas (Figura 3b). Na primeira etapa, o 8-MOP intercala-se entre as bases pirimídicas (com alta afinidade por timinas) sem formar ligações covalentes (Figura 3b). Na segunda etapa, os adutos são formados por uma fotocicloadição entre a dupla ligação 4,5 do grupo furano ou 3,4 do grupo cumarínico com a ligação dupla 5,6 da timina. Um segundo fóton de luz UVA induz a formação de ICLs (Figura 3b; Bethea *et al.,* 1999).

A mitomicina C (Figura 3a), por sua vez, é um antibiótico natural que possui a propriedade de formar monoadutos nas posições N-7 e N-2 da guanina, pontes intracadeias e ICLs (Tomasz, 1995; Kumar *et al.,* 1997). Neste último caso, a formação de ICLs se dá entre as posições N-2 da guanina em sequências do tipo CpG, e constituem cerca de 5% a 13% dos adutos totais em células de mamíferos (Warren *et al.,* 1998). A mitomicina C possui uma ampla aplicabilidade médica, sendo bastante utilizada em combinação com outras drogas para o tratamento de tumores de mama, de pulmão, de próstata e de bexiga (Cummings *et al.,* 1998).

A cisplatina (Figura 3a) é um dos compostos mais utilizados para o tratamento de diferentes tipos de tumores (Gupta *et al.,* 2005). As chances de cura para pacientes com tumores de testículos e de ovários pode chegar a 90% quando a cisplatina é administrada em conjunto com outros quimioterápicos (Bosl & Motzer, 1997). A ação genotóxica da cisplatina deve-se à formação de adutos bifuncionais de DNA do tipo pontes intracadeia (65% GpG e 25% ApG) e ICLs entre guaninas situadas em regiões do DNA ricas em GpC (5% a 8% do total de adutos) (Dronkert & Kanaar, 2001). Em comparação aos psoralenos e a mitomicina C, a cisplatina provoca grandes distorções na estrutura do DNA, as quais podem resultar em quebras da fita e bloqueios nos processos de replicação e transcrição de DNA (Malinge *et al.,* 1999).

As mostardas nitrogenadas, em especial as bifuncionais (HN2; Figura 3a) possuem um importante papel histórico, visto o seu uso durante a Primeira e a Segunda Guerra Mundiais como agentes químicos (Sanderson & Shield, 1996). Contudo, as mostardas constituem a principal classe de drogas antitumorais hoje disponíveis (Souliotis *et al.,* 2003). Seu mecanismo de ação está baseado na alta reatividade que possui com macromoléculas biologicamente importantes, tais como DNA, RNA e proteínas, induzindo múltiplos tipos de lesões (Souliotis *et al.,* 2003). Entretanto, o seu principal alvo é o DNA, alquilando a posição N-7 da guanina ou a posição N-3 da adenina (Osborne *et al.,* 1995). Além disso, as mostardas nitrogenadas bifuncionais podem induzir a formação de ICLs que modificam radicalmente a estrutura do DNA (Rink & Hopkins, 1995), e que constituem cerca de 5% de todos os danos gerados (Dronkert & Kanaar, 2001).

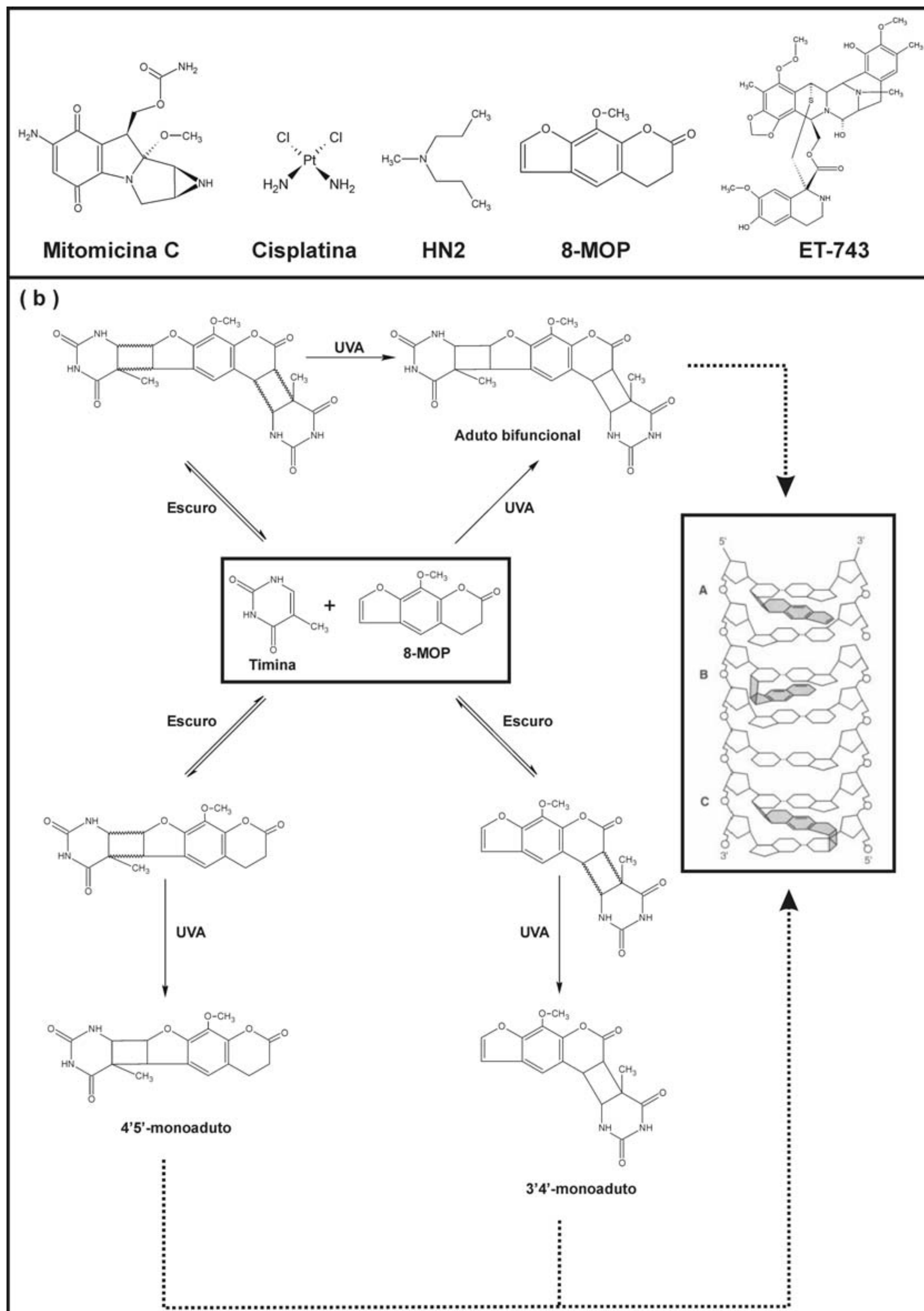


Figura 3. Em (a) exemplos de agentes mono- e/ou bifuncionais. Em (b), esquema da indução de adutos mono- e bifuncionais com 8-MOP e timina, bem como sua disposição espacial na molécula de DNA (quadro interno – A, B, C). Neste caso, quando a reação é feita na ausência de luz UVA, a molécula de 8-MOP tem a propriedade de interagir não covalentemente com a timina. Na presença de UVA,

formam-se as ligações covalentes do tipo 4'5'- e 3'4'-monoadutos (A e B) ou um aduto bifuncional quando duas timinas estão espacialmente próximas (C).

2.3. Quebras simples e duplas de DNA

As quebras de DNA podem acontecer como resultado do ataque de um agente químico ou físico à ponte fosfodiéster (Caldecott, 2001; Dudáš & Chovanec, 2004; Purdy & Su, 2004), por um bloqueio no processo de replicação do DNA (Michel *et al.,* 2004) ou podem ser induzidas por enzimas específicas com funções de reparação e por outras proteínas que atuam no metabolismo geral do DNA (Haber, 1999; Pastink & Lohman, 1999; Flores-Rozas & Kolodner, 2000; Pastink *et al.,* 2001; van den Bosch *et al.,* 2002).

As lesões do tipo quebras de DNA podem ser separadas em dois grupos: simples (*single strand breaks* ou SSBs) e duplas (*double strand breaks* ou DSBs) (Caldecott, 2001; Dudáš & Chovanec, 2004; Purdy & Su, 2004). As quebras simples são consideradas as menos tóxicas para a célula e milhares de SSBs são geradas por dia em células de mamíferos como resultado das atividades de reparação ou metabólicas do DNA (Caldecott, 2001). Contudo, esta falta de toxicidade se deve ao fato de que as SSBs são rapidamente reparadas, e que, de outra forma, poderiam converter-se em DSBs (Caldecott, 2001). A reparação de SSBs é realizada por uma via bioquímica especializada, denominada de 'SSBR' (*SSB repair;* Caldecott, 2001). Interessantemente, a ausência de reparação de SSBs está relacionada a um conjunto de patologias, tais como a ataxia espinocerebelar (Caldecott, 2003) e outros processos neurodegenerativos (Caldecott, 2004).

As lesões do tipo DSBs são consideradas as mais tóxicas para a célula (Dudáš & Chovanec, 2004). Quando não reparadas corretamente, as DSBs podem originar mutações e rearranjos cromossômicos (Jeggo, 1998; Kanaar *et al.,* 1998; Haber, 1999). Por outro lado, a ausência de reparação de DSBs pode levar à morte celular ou à perda de cromossomos

(Kaina, 2003; Dudáš & Chovanec, 2004). Apesar de inúmeros fatores fisiológicos, tais como a geração intracelular de EROs ou a atuação de agentes físico-químicos serem considerados como os fatores primordiais na formação de DSBs, estas lesões também são geradas propositalmente durante a recombinação meiótica (Pâques & Harber, 1999) ou durante a recombinação V(D)J, a qual ocorre em linfócitos B e T em desenvolvimento e provém a base molecular para a formação de imunoglobulinas (Jeggo, 2002).

3. Mecanismos de reparação de DNA

3.1. Vias de reparação de DNA

Como visto no item 2, a enorme complexidade dos diferentes tipos de lesões que se acumulam no DNA podem induzir a instabilidades genômicas que, se não reparadas corretamente, levam à morte celular ou a processos tumorais. Desta maneira, não é surpreendente que a reparação de todas estas lesões requer a atuação de várias proteínas que, juntas, compõem vias ou mecanismos discretos, mas que se mostram espacialmente e temporalmente interconectados (Eisen & Hanawalt, 1999). Esta diversidade de vias de reparação de DNA pode ser observada quando as mesmas são comparadas entre si. Por exemplo, algumas vias são usadas para reparar apenas um tipo de dano, enquanto que outras são mais abrangentes em seus requerimentos enzimáticos. Da mesma forma, algumas vias de reparação são mecanisticamente simples, requerendo apenas algumas enzimas, enquanto que outras são bastante complexas, envolvendo muitos passos enzimáticos e diferentes tipos de complexos protéicos atuando, não só no processo de reparação em si, mas também em outros processos celulares (Eisen & Hanawalt, 1999). Esta diversidade de vias pode ser melhor compreendida quando as mesmas são agrupadas levando-se em conta os mecanismos de ação. Assim, três grandes vias são atualmente conhecidas: a reparação

direta, a reparação por excisão e a reparação recombinacional (Figura 4; Eisen & Hanawalt, 1999).

A reparação direta envolve dois mecanismos principais: (i) a fotorreativação, catalisada por enzimas pertencentes à família das fotoliases/criptocromos (Figura 4; Thompson & Sancar, 2002) e (ii) a reversão de bases alquiladas, catalisadas pelas DNA metiltransferases (Christmann *et al.,* 2003). Por sua vez, a reparação por excisão é formada por três mecanismos principais: a excisão de nucleotídeos (*nucleotide excision repair* ou NER, revisado em Prakash & Prakash, 2000; Christmann *et al.,* 2003 e Costa *et al.,* 2003), a excisão de bases (*base excision repair* ou BER, revisado em Boiteux & Guillet, 2004) e a reparação de erros de emparelhamento de bases (*mismatch repair* ou MMR, revisado em Christmann *et al.,* 2003). Tanto as vias de reparação direta quanto a de excisão são essenciais para a remoção de bases modificadas e adutos, de forma que respondem por grande parte da atividade de reparação de DNA na célula (Huang *et al.,* 1994; Reardon *et al.,* 1997; Memisoglu & Samson, 2000). A reparação recombinacional (tratada em maiores detalhes no item 3.2) é recrutada para as lesões do tipo quebra de DNA, geradas tanto por agentes fisiológicos quanto por agentes físico-químicos.

Interessantemente, os dados genômicos e filogenéticos permitem a comparação interespécies das três vias principais de reparação de DNA e ajudam no entendimento da evolução dos mecanismos de reparação e de tolerância aos danos. Assim, uma perspectiva evolucionária se faz necessária em qualquer estudo comparativo porque ajuda a focar no ‘como’ e no ‘porquê’ do surgimento das similaridades e diferenças na história evolutiva de um determinado grupo de organismos ou de uma família de proteínas (Eisen & Hanawalt, 1999). Em termos de reparação de DNA, a perspectiva evolucionária é a chave para a compreensão das generalidades e particularidades existentes entre as diferentes vias em diferentes espécies.

Além da perspectiva comparativa e evolutiva, é preciso considerar a estruturação da cromatina eucariótica quando se trata de mecanismos de reparação de DNA. A cromatina eucariótica é uma estrutura altamente ordenada, formada por proteínas histônicas e não-histônicas condensadas com o DNA (Allard *et al.,* 2004). Sendo ordenada, a cromatina eucariótica apresenta níveis de organização, com o primeiro nível composto pelo enrolamento do DNA ao redor de um octâmero de histonas (duas cópias das histonas H2A, H2B, H3 e H4), o que define a unidade básica da cromatina: o nucleossomo (Luger *et al.,* 1997). O segundo nível de organização consiste no arranjo de nucleossomos na forma de solenóide, estabilizado pela histona H1 (Wolffe, 1997). Desta maneira, sendo uma estrutura condensada e ordenada, a cromatina necessita sofrer modificações a fim de que a maquinaria de reparação de DNA tenha acesso às lesões (Allard *et al.,* 2004). Estas modificações ou 'remodelagens' da cromatina são realizadas por dois mecanismos: (i) modificação química das histonas nucleossomais por meio da sua fosforilação, acetilação, metilação, ubiquitinação e ADP-ribosilação (Peterson, 2002; Koundrioukoff *et al.,* 2004) e (ii) uso da energia liberada pela hidrólise do ATP para alterar a estrutura da cromatina (Turner, 2002). Em presença de danos, a remodelagem da cromatina se processa pela fosforilação de uma variante da histona H2A denominada de H2AX (Fernandez-Capetillo, 2004; Anexo 1). Esta fosforilação é induzida por um grupo de proteínas, evolutivamente conservadas, denominadas de ATM (*Ataxia Telangiectasia Mutated*) e ATR (*ATM-/Rad3-related proteins*) em mamíferos (Mec1p/Tel1p em leveduras), as quais pertencem à família das fosfatidilinositol-3-cinases (PI3K) (Fernandez-Capetillo, 2004; Anexo 1). A H2AX fosforilada ou γ-H2AX ativa uma cascata de cinases, também pertencentes à família das PI3K, que rapidamente coordenarão os eventos de reparação de DNA, bem como induzirão as paradas do ciclo celular (Zhou & Elledge, 2000; Rouse & Jackson, 2002). Além da γ-H2AX, diversas outras proteínas remodeladoras da cromatina são recrutadas conforme a natureza do dano, seja uma quebra de DNA ou uma modificação de base. Por exemplo, as

acetiltransferases de histonas (HATs), que compõem 6 famílias distintas de proteínas envolvidas na acetilação de histonas H3 e H4, são recrutadas para sinalizar danos provocados por luz UV na cromatina (Moore & Krebs, 2004). Por fim, o recrutamento de determinados tipos de remodeladores de cromatina também determinará quais das três vias serão usadas durante a reparação (Allard *et al.,* 2004).

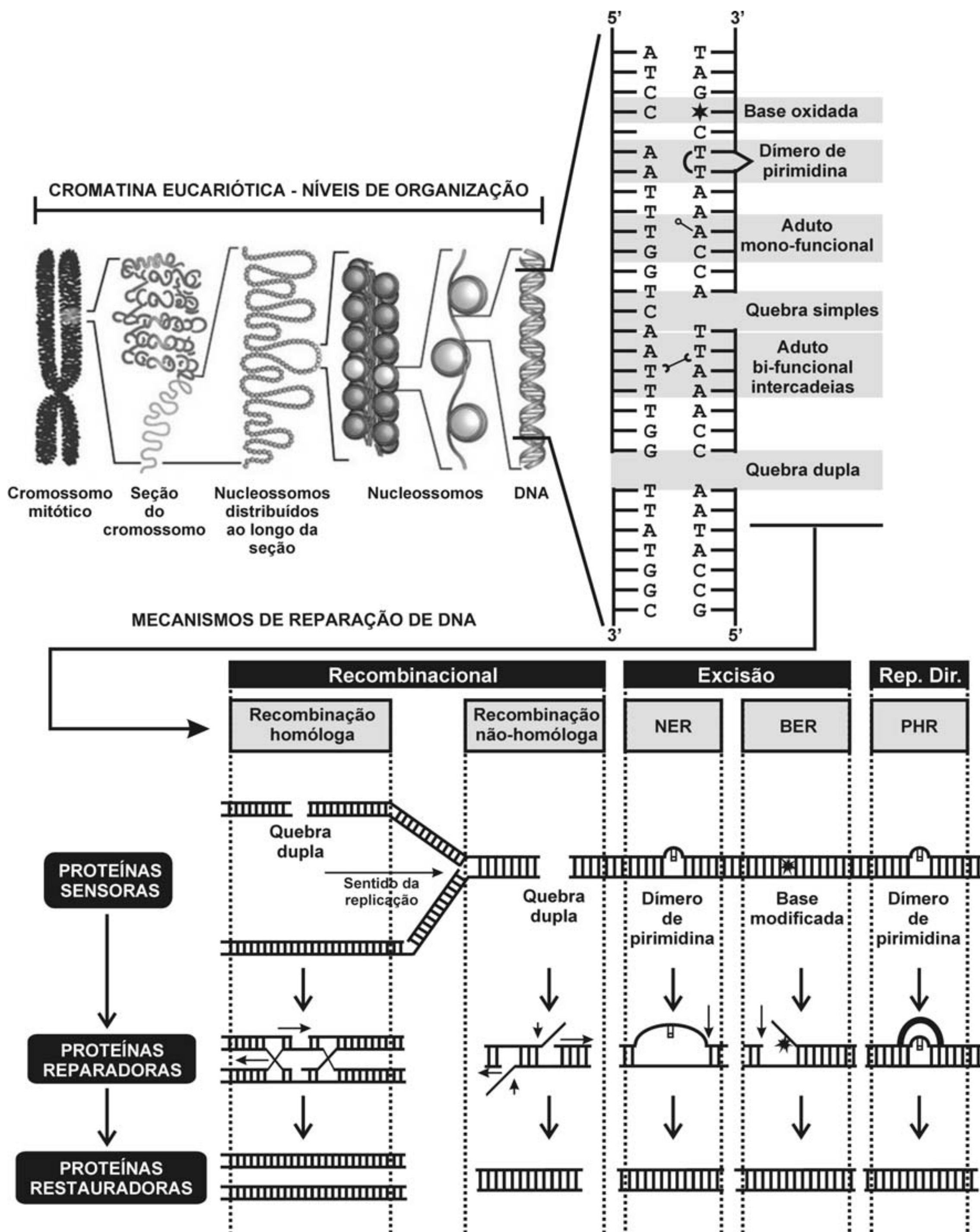


Figura 4. Diagrama esquemático dos três mecanismos principais de reparação de DNA (recombinacional, excisão e reparação direta), bem como de algumas subvias de reparação e sua atuação na cromatina eucariótica. Os diferentes níveis estruturais da cromatina (de cromossomo a fita dupla de DNA) são mostrados na figura. A presença de várias lesões em uma fita de DNA que está replicando (tais como bases modificadas, quebra simples e duplas, e adutos mono- e bifuncionais) são reconhecidas

por subvias diferentes por meio de proteínas sensoras. Estas, por sua vez, recrutam as proteínas que irão executar os reparos necessários, conforme cada situação (flechas menores). A dupla fita de DNA, bem como a replicação, são restituídas graças à ação de proteínas restauradoras, tais como as DNA polimerases e as DNA ligases.

3.2. Reparação recombinacional

Como visto previamente, a geração de bases modificadas no DNA bem como adutos mono- e bifuncionais podem resultar em quebras altamente genotóxicas que, se não forem reparadas, podem ser letais (Kaina, 2003; Dudáš & Chovanec, 2004). Estas quebras são reparadas por um mecanismo especializado denominado de reparação recombinacional.

A reparação recombinacional abrange três mecanismos: a recombinação homóloga (*homologous recombination* ou HR, não mutagênica), a recombinação não-homóloga (*non-homologous end joining recombination* ou NHEJ, mutagênica) e a recombinação V(D)J (mutagênica). Cada uma destas vias possui requerimentos enzimáticos únicos, sendo que o recrutamento de uma ou de outra via é dependente de uma série de fatores fisiológicos celulares (Lieber, 1999). Interessantemente, as vias HR e NHEJ podem ter funções sobreponíveis para a manutenção da integridade cromossomal em eucariotos de uma maneira geral (Takata *et al.,* 1998). Contudo, diferentes espécies podem usar uma via em preferência à outra. Por exemplo, as leveduras utilizam a via HR para reparar DSBs, enquanto que os eucariotos multicelulares usam, preferivelmente, a via NHEJ para lidar com as DSBs (Jackson, 2002). Um motivo que possivelmente explicaria o uso preferencial de uma via em relação à outra em leveduras e eucariotos multicelulares é a organização genômica distinta que cada um destes apresenta. O genoma de eucariotos multicelulares, na grande maioria das vezes, contém uma fração substancial de DNA repetitivo, o que dificulta a reparação por HR (Lieber *et al.,* 2003). Exceto durante as fases S, G_2 e M, quando existe

um posicionamento ótimo de cromátides irmãs, a reparação por HR de regiões repetitivas cromossomais pode resultar em translocações e/ou morte celular (Lieber *et al.,* 2003).

3.2.1. Recombinação homóloga

A recombinação homóloga em eucariotos é uma via evolutivamente conservada, sendo que os homólogos protéicos, necessários para realizar esta função, estão presentes tanto em eucariotos unicelulares quanto multicelulares (Jackson, 2002; Anexo 1). Os eventos relacionados à via HR são complexos em sua natureza, mas um modelo simplificado permite a compreensão das principais proteínas envolvidas no processo (Figura 5). Este modelo representa apenas um de dois conhecidos, que são a reparação de quebras duplas (*double strand breaks repair* ou DSBR) e a síntese dependente de anelamento de fita (*synthesis-dependent strand annealing* ou SDSA) (Krogh & Symington, 2004). Neste caso, o modelo descrito para esta seção representa a DSBR, pois esta é a principal via HR na célula (Krogh & Symington, 2004).

A função primária da HR em células mitóticas, independente do modelo proposto, é o reparo de DSBs ou SSBs que se formam em conseqüência do colapso da forquilha de replicação de DNA e que é resultante da ação de diferentes mecanismos de reparação de bases modificadas ou adutos (Krogh & Symington, 2004). A HR também é requerida para a manutenção telomérica e para a meiose, sendo essencial para estabelecer uma conexão física entre cromossomos homólogos a fim de assegurar a correta disjunção dos mesmos na primeira divisão meiótica (Krogh & Symington, 2004). Adicionalmente, a alta freqüência de recombinação meiótica promovida pela HR contribui para a geração de diversidade genética observada nos gametas (Krogh & Symington, 2004).

Basicamente, a via HR envolve um conjunto de proteínas pertencentes ao chamado ‘grupo Rad52p’ (Wood *et al.,* 2001; Krogh & Symington, 2004). Estas proteínas, tais como Rad50, Rad51, Rad52, Rad54, Rad55, Rad57, Rad59, Rdh54, Mre11, Xrs2, entre outras

(Anexo 1), possuem funções diversas, mas altamente sincronizadas, o que permite a correção das quebras de DNA quando uma fita homóloga está presente.

O primeiro evento relacionado à HR é a ressecção nucleolítica da fita 5'-3', promovida por um complexo protéico que contém Rad50, Mre11p e Xrs2p (Complexo MRX; Jackson, 2002; Krogh & Symington, 2004). Esta ressecção possui, como objetivo, a geração de extremidades invasivas do tipo 3' fita simples que resultarão no processo recombinacional (Figura 5). Em *E. coli,* os homólogos da Rad50p/Mre11p denominados de SbcC/SbcD formam um grande complexo com atividades 3'→5' exonucleásica ATP-dependente e endonucleásica fita simples-dependente (Sharples & Leach, 1995; Connelly *et al.,* 1997). O repertório nucleolítico da Mre11p inclui: (i) atividade exonucleásica 3'→5' em extremidades cegas e rescindidas, (ii) atividade endonucleásica em fitas simples de DNA, tanto circulares quanto lineares e (iii) atividade endonucleásica em estruturas secundárias de DNA, como *hairpins*. A observação de que homopolímeros fita simples de DNA são resistentes à clivagem por Mre11p sugere que esta enzima reconhece estruturas secundárias transientes de DNA, clivando especialmente em regiões de transição fita dupla/fita simples (Krogh & Symington, 2004). Uma vez que as extremidades 3' fita simples estão disponíveis, a proteína Rad51 (um filamento nucleoprotéico) liga-se a estas e inicia o evento de invasão a uma fita homóloga (Figura 5). Esta invasão é grandemente facilitada pelas proteínas mediadoras Rad54 e RPA, as quais são responsáveis pela eliminação de estruturas secundárias da região 3' terminal. Desta maneira, a região 3´ terminal da fita danificada é restaurada por uma DNA polimerase, que copia as informação da fita intacta. Terminada a extensão, as extremidades livres são novamente reunidas pela DNA ligase I (Figura 5). Entretanto, este processo gera junções do tipo Holliday, que necessitam ser resolvidas a fim de gerarem duas moléculas intactas de DNA (Figura 5; Krogh & Symington, 2004). A resolução das junções de Holliday é realizada pelas enzimas Mus81 e Mms4 (Eme1), as quais pertencem à família XPF de endonucleases estrutura-

específicas (Heyer, 2004). O resultado final da resolução das junções de Holliday pode ser ou não a permuta de fitas (*crossover*), especialmente durante o processo de recombinação meiótica (Krogh & Symington, 2004).

Em mamíferos, as deficiências na via HR resultantes de alguma mutação são compensadas com um aumento de atividade das vias de recombinação não-homóloga (Rassool, 2003). Como resultado da atividade diminuída da HR, duas patologias são conhecidas em seres humanos: a anemia de Fanconi, cujos pacientes apresentam uma alta instabilidade cromossômica e propensão a leucemias de origem mielóide (Faivre *et al.,* 2000), e a síndrome de Bloom, cujos pacientes apresentam uma alta propensão a desenvolver diferentes tipos de tumores, além da instabilidade cromossômica associada (Ellis & German, 1996).

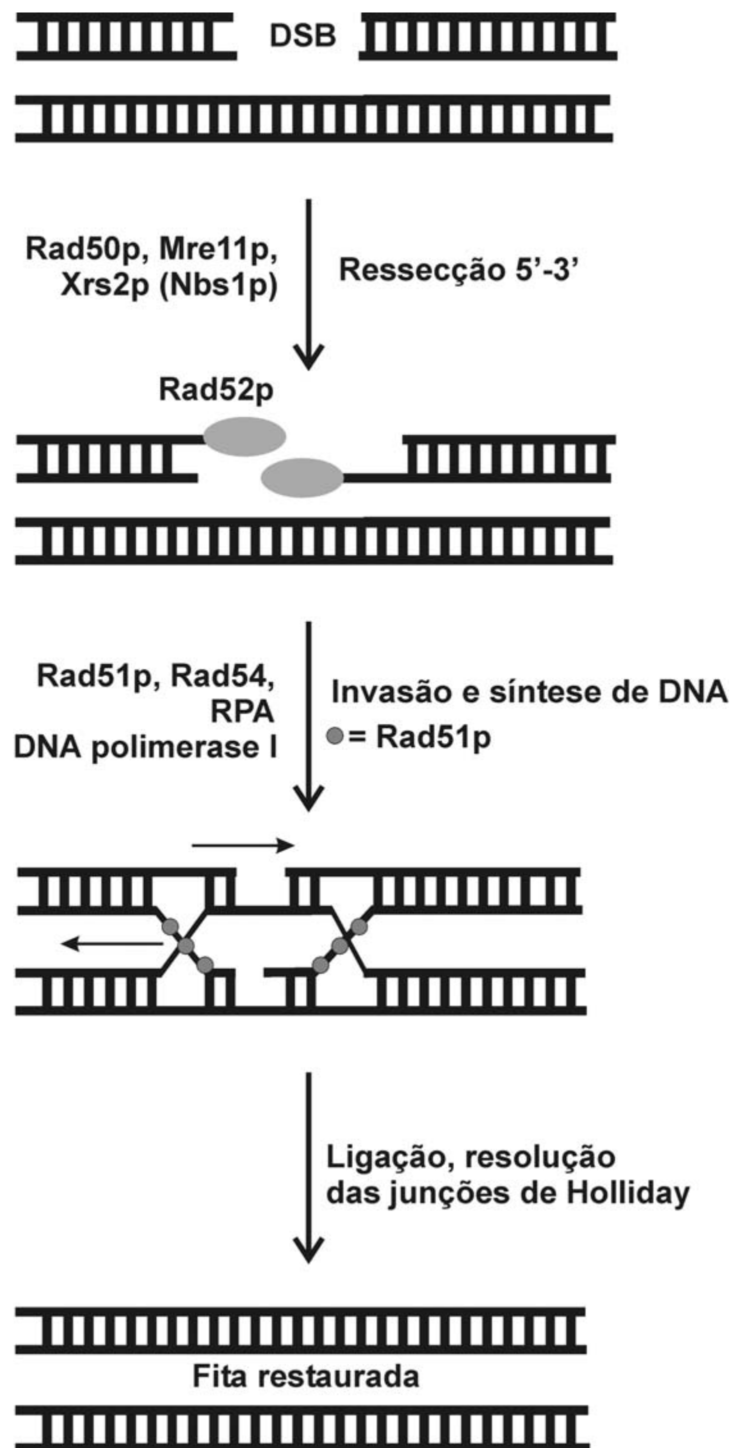


Figura 5. Modelo simplificado da via HR. A quebra dupla é reconhecida pelo complexo Rad50p/Mre11p/Xrs2p, o qual induz a ressecção da fita 5'→3'. Com a ajuda das proteínas Rad51, Rad54 e RPA ocorre a invasão das extremidades 3' danificadas na fita homóloga não danificada, a qual é utilizada como molde pela DNA polimerase. Após a ligação dos fragmentos, as junções de Holliday resultantes são resolvidas e ambas as fitas restauradas (adaptado de Jackson, 2002).

3.2.2. Recombinação não homóloga

A recombinação não homóloga eucariótica ou NHEJ representa um dos ramos principais da via recombinacional. Presente em todos os eucariotos, a via NHEJ é recrutada quando há a necessidade de religar duas extremidades de DNA, resultantes de uma quebra, e que não possuem homologia entre si. Além disso, pequenas regiões de homologia (ou microhomologias, quando abrangem apenas alguns nucleotídeos) existentes entre duas fitas não homólogas de DNA podem ser utilizadas pela via NHEJ para a reparação (Labhart, 1999). Assim, diversas subvias de reparação pertencentes à via NHEJ são conhecidas, cada qual com características próprias, mas cujo resultado final é a reparação sujeita a erros de DSBs.

Inúmeras patologias associadas à deficiências na via NHEJ já foram descritas em seres humanos e outros mamíferos. Estas deficiências, em grande parte, resultam no aumento da incidência de tumores malignos, no envelhecimento precoce, na inviabilidade embrionária e nas instabilidades cromossômicas de diferentes naturezas (Iliakis *et al.,* 2004). Embora os mecanismos moleculares da via NHEJ ainda não sejam conhecidos em detalhes, a junção de duas extremidades de DNA requer, pelo menos, quatro passos: (1) a detecção da DSB; (2) a formação de uma ponte molecular que mantenha as duas extremidades próximas uma da outra; (3) o processamento das extremidades, de forma a torná-las compatíveis para a ligação e (4) a ligação das extremidades (Weterings & van Gent, 2004). Os requerimentos protéicos que são necessários para estes quatro passos estão detalhados no Capítulo 1 assim como no Anexo 1 desta tese. Contudo, a maquinaria básica da via NHEJ consiste nas proteínas Xrcc4, na DNA ligase IV, no heterodímero Ku70/Ku80 bem como na subunidade catalítica da cinase dependente de DNA (DNA-PKcs) (Weterings & van Gent, 2004). Um modelo simplificado desta via é apresentado a fim de facilitar a compreensão do processo (Figura 6).

Em um primeiro momento, a detecção da DSB é realizada por um holocomplexo formado pela DNA-PKcs e pelo heterodímero Ku70/Ku80. Este holocomplexo, em conjunto com as extremidades de DNA, constitui a chamada sinapse (DeFazio *et al.,* 2002). Uma vez formada a sinapse, diferentes endonucleases são recrutadas para modificar as extremidades de DNA, das quais destacam-se o complexo MRX e a proteína homóloga de Pso2p/Snm1p conhecida como Ártemis (Lees-Miller & Meek, 2003). Após a modificação, DNA polimerases e ligases restituem a dupla fita de DNA (Figura 6).

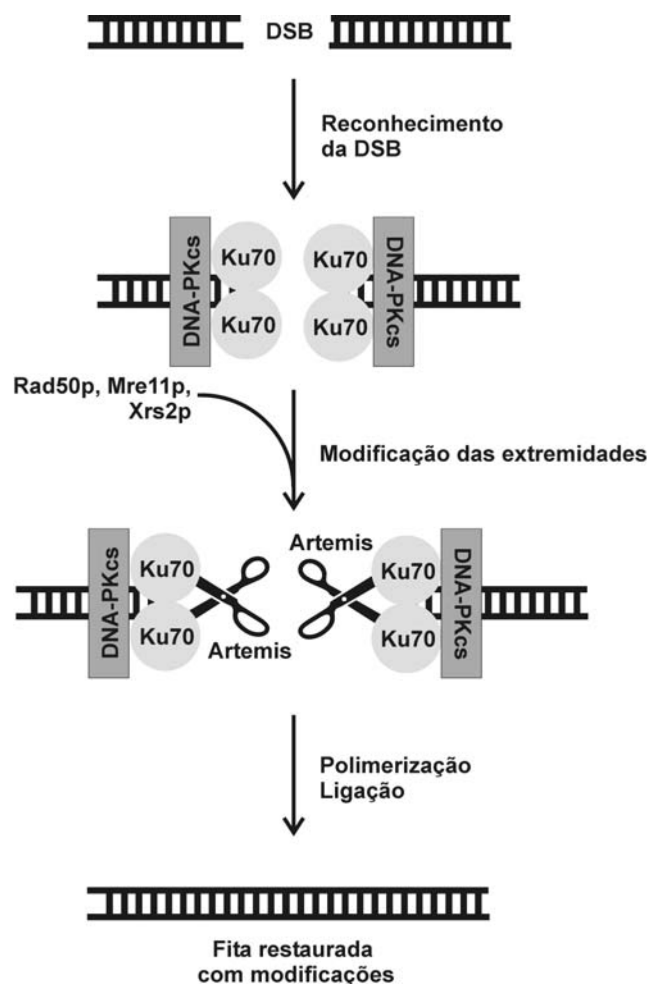


Figura 6. Modelo simplificado da via NHEJ. Após a detecção da DSB, forma-se o holocomplexo ao redor das extremidades livres, constituído pela DNA-PKcs e pelos heterodímeros Ku70/80. Uma vez formado o holocomplexo, diferentes endonucleases (Ártemis e/ou complexo MRX) são recrutadas para modificar as extremidades livres a

fim de torná-las compatíveis para ligação. Por fim, as extremidades modificadas servem de substrato para o complexo formado pela enzimas Xrcc4 e DNA ligase IV, as quais restauram a fita de DNA (adaptado de Rassool, 2003).

3.2.3. Recombinação V(D)J

Em mamíferos, a geração de DSBs pode ser originada a partir de mecanismos fisiológicos, tal como acontece durante a meiose (item 3.2.1). Contudo, um segundo mecanismo de indução de DSBs está relacionado ao aumento da diversidade imunológica adaptativa em vertebrados mandibulados, sendo que a reparação destas quebras é realizada por uma via recombinacional especializada denominada de recombinação V(D)J. Este tipo de recombinação é característico de células linfóides, e baseia-se na introdução de DSBs em seqüências sinalizadoras de recombinação (*recombination signal sequences* ou RSSs) que flanqueiam os locos que codificam para os receptores das células T e B (TCR e BCR, respectivamente), e para as imunoglobulinas (Ig) (Gellert, 2002). Estes locos não contêm genes maduros que codificam para o TCR/BCR ou para a Ig, mas sim um arranjo de genes, formados por três segmentos separados denominados de V (para *variable*), de D (para *diversity*) e J (*joining*) (Fugmann *et al.,* 2000). Três passos, altamente regulados e coordenados entre si, são essenciais para a recombinação V(D)J (Figura 7): (1) a iniciação, caracterizada pela indução de DSBs nas RSSs e formação de estruturas secundárias de DNA do tipo *hairpins* nos segmentos V e J; (2) a indução de mecanismos de reparação, especialmente a via NHEJ e (3) a reparação dos danos seguida da formação das seqüências maduras (Rooney *et al.,* 2003).

A iniciação é realizada por um conjunto de endonucleases linfócito-específicas com características de transposases denominadas de Rag1 e Rag2 (*recombination activating genes* ou Rags, Fugmann *et al.,* 2000). Tanto a Rag1 quanto a Rag2 possuem a função de reconhecer e introduzir DSBs nas regiões próximas às RSSs, sendo esta ação facilitada por

proteínas de alta mobilidade (*high mobility group proteins* ou HMGs), uma família de proteínas não-histônicas que induzem modificações espaciais na cromatina (Figura 7, Fugmann *et al.,* 2000). Uma vez introduzidas as DSBs, as Rags geram dois tipos de extremidades (Figura 7): cegas (que flanqueiam as RSSs; também denominadas nesta fase de *signal ends*) e *hairpins*, sendo que esta última flanqueia os segmentos V e J (*coding ends*) (Rooney *et al.,* 2003). A geração de extremidades cegas e *hairpins* serve de base para o recrutamento de proteínas de reparação da via NHEJ. Neste caso, o holocomplexo DNA-PK liga-se em ambas as extremidades e recruta a proteína Ártemis, responsável pela abertura dos *hairpins* (Rooney *et al.,* 2003). Uma vez abertos os *hairpins*, as extremidades são processadas pela enzima TdT (uma DNA polimerase independente de DNA), a qual adiciona nucleotídeos de forma aleatória, e pelas enzimas DNA ligase IV e Xrcc4, que promovem a ligação dos segmentos V e J (*coding joint*) e das RSSs (*signal joint*) (Figura 7).

As mutações que alteram o funcionamento de qualquer uma destas enzimas, sejam linfóide-específicas ou relacionadas à via NHEJ, originam um conjunto de patologias imunológicas conhecidas como SCID (*severe combined immunodeficiency syndrome*). A SCID possui como principal característica a ausência de Ig e de linfócitos B/T circulantes, sendo que a severidade da patologia varia conforme a proteína afetada. Em geral, a SCID tende a ser fatal logo nos primeiros meses de vida (Buckley, 2004).

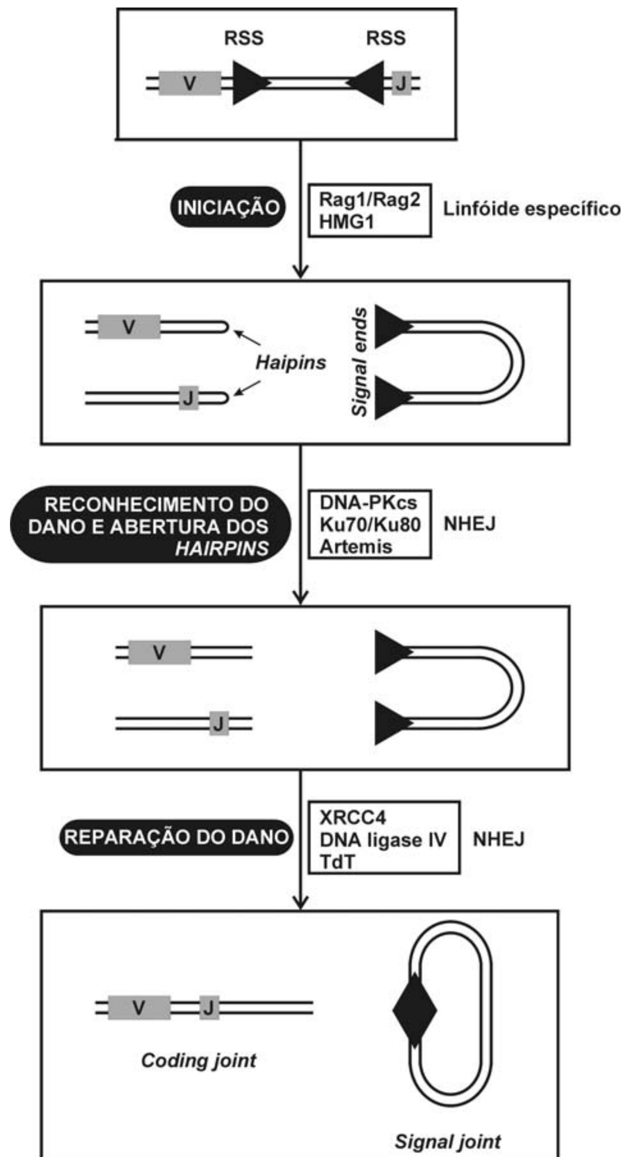


Figura 7. Diagrama esquemático da recombinação V(D)J. Três passos são descritos para este processo. O primeiro passo (iniciação) é caracterizado pelo reconhecimento das seqüências RSSs que flanqueiam os segmentos V e J pelas endonucleases Rag1/Rag2 e pelas proteínas de alta mobilidade 1 e 2 (HMG1/HMG2). As Rags induzem DSBs nas RSSs e originam extremidades do tipo cegas e *hairpins*. No segundo passo, caracterizado pelo reconhecimento dos danos, as proteínas DNA-PKcs, Ku70/Ku80 e Ártemis são recrutadas para a abertura dos *hairpins* sendo que, no terceiro passo, as extremidades geradas pela abertura dos *hairpins* são processadas

pelas enzimas TdT, Xrcc4 e DNA ligase IV, originando as cadeias maduras de Ig e TCR/BCR.

4. Os genes *PSO*

A fotoquimioterapia empregando psoralenos e luz UVA (PUVA terapia) é largamente utilizada para tratar desordens dermatológicas (ver item 2.2). Os produtos formados pela fotoativação de psoralenos nos fibroblastos da pele e em células tumorais são responsáveis pela eficácia deste tratamento. Contudo, embora a PUVA terapia seja efetiva na terapia clínica, os fotoprodutos resultantes são altamente citotóxicos, gerando espécies reativas de oxigênio e inibindo a replicação e a transcrição do DNA genômico e mitocondrial. Neste sentido, as alterações promovidas pela PUVA terapia no metabolismo de DNA podem resultar em diferentes tipos de mutações, em quebras simples e duplas, e danos ao nível cromossomal. Como conseqüências gerais, a PUVA terapia pode induzir a apoptose celular e elevar os riscos de surgimento de tumores, de um modo dose-dependente (Dronkert & Kanaar, 2001; Greenberg *et al.,* 2001).

Desta maneira, e com o objetivo de estudar as conseqüências fotoquímicas, biológicas e genéticas da PUVA terapia em células vivas, Henriques & Moustacchi (1980) isolaram, a partir de uma população de células mutagenizadas com etil-metanossulfonato (EMS), uma nova classe de mutantes (*pso*) de *S. cerevisiae,* sensíveis à fotoadição de psoralenos mono- (3-CPs) e bi-funcionais (8-MOP). A análise molecular desses genes e a caracterização fenotípica de seus alelos mutantes têm progredido consideravelmente (revisado em Brendel *et al.,* 2003; Anexo 1). Atualmente, dez genes *PSO* estão caracterizados fenotipicamente, sendo que nove foram identificados molecularmente. Os dados mostram que sete genes *PSO* codificam para proteínas que estão diretamente envolvidas na reparação de lesões no DNA geradas pela fotoativação de psoralenos e por

outros mutágenos. Interessantemente, dois genes *PSO* não estão envolvidos diretamente na reparação de lesões de DNA, mas participam na manutenção do equilíbrio redox intracelular (Brendel *et al.,* 2003). As características dos genes *PSO1* a *PSO10* estão resumidas na Tabela 1 (para maiores detalhes sobre os genes *PSO*, ver Anexo 2).

Dentre todos os genes *PSO* conhecidos, o *PSO2/SNM1* destaca-se por suas funções específicas na reparação de danos do tipo ICLs. As características da proteína Pso2/Snm1, bem como a sua presença e formas de atuação em diferentes organismos, são revistas no Capítulo 1 desta tese.

Tabela 1. Propriedades das proteínas codificadas pelos genes *PSO* e os fenótipos dos seus alelos mutantes

Gene/Alelo	Proteína (kDa)/Função	Fenótipo do mutante
PSO1/REV3	173; DNA polimerase translesão	Sensível à radiação e a agentes mutagênicos químicos; baixa mutabilidade
PSO2/SNM1	72; endo/exonuclease	Sensível ao tratamento com todos os agentes indutores de ICLs
PSO3/RNR4	40; ribonucleotídeo redutase	Redução na mutabilidade induzida e recombinação; as células não atingem a fase estacionária de crescimento
PSO4/PRP19	57; proteína associada ao *spliceosome*	Sensível a agentes mutagênicos, não recombina em mitose, baixa mutabilidade e defectivo em esporulação
PSO5/RAD16	91; DNA helicase	Sensível à UVC e lesões oxidativas
PSO6/ERG3	43; ergosterol dessaturase	Sensível a danos oxidativos
PSO7/COX11	28; associa-se ao complexo da citocromo oxidase	Células sensíveis a 4-NQO
PSO8/RAD6	19; enzima conjugada à ubiquitina	Sensível à UVC e a vários agentes mutagênicos; mutagênese reduzida
PSO9/MEC3	53; controle específico de parada de ciclo celular em resposta a danos no DNA na fase G_2	Sensível a vários agentes mutagênicos; mutagênese reduzida
PSO10	Desconhecida	Sensível a vários agentes mutagênicos; mutagênese reduzida; bloqueado em esporulação

5. O uso da análise de seqüências e da filogenia para a predição de funções protéicas

5.1. Análise de seqüências

5.1.1. Alinhamento múltiplo de seqüências

Uma das pedras fundamentais que estabeleceram o campo da Bioinformática é a comparação ou o alinhamento múltiplo de seqüências, tanto de proteínas quanto de ácidos nucléicos. Com a ajuda de diferentes metodologias de alinhamento de seqüências é possível estudar, por exemplo, os padrões estruturais de proteínas que são evolutivamente conservadas, detectar ou demonstrar homologias entre seqüências e/ou famílias de seqüências, estabelecer correlações funcionais entre grupos ou famílias protéicas e determinar a relação ancestral existente entre organismos de diferentes espécies (Thompson *et al.,* 1994; Chenna *et al.,* 2003).

Atualmente, são conhecidas duas formas principais de alinhamento múltiplos: o alinhamento local e o alinhamento global (Figura 8; Chenna *et al.,* 2003). Cabe salientar que os princípios matemáticos que operam ambos os métodos de alinhamentos são bastante complexos, de forma que apenas uma visão geral será apresentada neste item.

Os métodos baseados em alinhamentos locais tentam determinar se subsegmentos de uma seqüência estão presentes em outras seqüências. Em outras palavras, estes métodos utilizam apenas determinadas regiões de uma seqüência polipeptídica para o processo comparativo (Phillips *et al.,* 2000). Os alinhamentos locais possuem aplicações na busca de seqüências em bancos de dados genômicos e/ou proteômicos, sendo que diversos programas são usados rotineiramente para esta função, tais como o BLAST (*Basic Local Alignment Search Tools*) e/ou o PSI-BLAST (*Position-Specific Iterated*-BLAST) (Altschul *et al.,* 1997; Jones & Swindells, 2002).

Embora os métodos de alinhamento local possam ter utilidade para a detecção de seqüências que possuam um certo grau de similaridade, os métodos de alinhamento global são utilizados, principalmente, para a análise filogenética, sejam estas de proteínas ou de ácidos nucléicos (Phillips *et al.,* 2000). Os métodos globais comparam seqüências em sua totalidade, ou seja, cada elemento de uma seqüência (que pode ser um resíduo de aminoácido ou de nucleotídeo) é comparado com os elementos de uma outra seqüência. Diferentes programas de alinhamento global têm sido descritos, sendo que o programa ClustalW é o mais utilizado (Thompson *et al.,* 1994).

Figura 8. Exemplos de um alinhamento global e um alinhamento local de duas seqüências polipeptídicas. Os *gaps* (os quais representam inserções e/ou deleções de resíduos de aminoácidos) estão indicados pelos círculos pontilhados.

Independente do método usado para o alinhamento de seqüências, há a necessidade de se avaliar estatisticamente a significância do resultado obtido (Baxevanis & Ouellete, 2001). Diferentes metodologias estatísticas têm sido propostas para este fim, de forma que a maioria destas levam em conta a presença e o número de *gaps* originados a partir do alinhamento de duas seqüências (Figura 9). O *gap* é o resultado de um processo de inserção e/ou de deleção (*indels*) de resíduos de aminoácidos de uma proteína ao longo de sua história evolutiva, sendo que o uso destes *gaps* na análise estatística de um alinhamento permite identificar o quão semelhantes são duas seqüências. Ou seja, quanto maior o número de *gaps,* maior é a chance de que o alinhamento seja fortuito (Baxevanis &

Ouellete, 2001). Além dos *gaps*, também são aplicadas as chamadas matrizes de substituição a fim de aumentar a sensibilidade e a qualidade dos alinhamentos obtidos (Mount, 2001). É fato conhecido de que certos aminoácidos podem ser facilmente substituídos por outros em proteínas relacionadas, sem a perda da função (por exemplo, a substituição de um aminoácido apolar por outro apolar ou a troca de um aminoácido básico por outro básico). Contudo, estas substituições também podem levar a modificações estruturais em proteínas com a conseqüente perda de função, especialmente quando a substituição afeta resíduos de aminoácidos com propriedades físico-químicas distintas. Assim, uma matriz de substituição leva em conta esta propriedade e atribui um valor, negativo ou positivo, para cada par comparativo de aminoácidos (Mount, 2001). Os exemplos mais comuns de matrizes de substituição são a BLOSUM (*Blocks Substitution Matrix*) e a PAM (*Point Accepted Mutation*) (Mount, 2001).

Entretanto, deve ser salientado que estes métodos de alinhamento, apesar de seu poder analítico, não são capazes de inferir as relações estruturais e/ou evolutivas existentes entre duas seqüências quando as mesmas possuem um índice de similaridade abaixo de 25% (também conhecido como *twiligth zone*) (Gaboriaud *et al.,* 1987). Assim, um outro método para o estudo de seqüências, a análise de agrupamentos hidrofóbicos, permite combinar a detecção de homologias entre estruturas primárias com a análise de estruturas secundárias de proteínas mesmo quando estas possuem um índice baixo de similaridade (Gaboriaud *et al.,* 1987).

5.1.2. Análise de agrupamentos hidrofóbicos (HCA)

A análise de agrupamentos hidrofóbicos ou HCA é uma poderosa ferramenta de alinhamento de seqüências e detecção de domínios globulares evolutivamente conservados (Callebaut *et al.,* 1997). A metodologia de HCA baseia-se no fato de que a maioria dos aminoácidos hidrofóbicos possui a tendência de formar domínios compactos e estáveis,

característicos de estruturas secundárias regulares, tais como as α-hélices e as β-estruturas (Callebaut *et al.,* 1997). A identificação destas estruturas em um diagrama de HCA, seja de famílias de proteínas correlatas ou não, permite mapear regiões conservadas que, em última análise, podem estar associadas a domínios catalíticos ou estruturais necessários para a atividade da proteína (Gaboriaud *et al.,* 1987; Callebaut *et al.,* 1997).

Basicamente, um diagrama de HCA é formado a partir da seqüência primária de uma determinada proteína, onde esta é disposta ao redor de um cilindro (Figura 9a). Este cilindro é cortado ao longo de seu eixo e o diagrama é desenrolado, formando um diagrama bi-dimensional (Figura 9b). O diagrama bi-dimensional é então compactado e duplicado (Figura 9c), de forma que os resíduos hidrofóbicos são agrupados e contornados (Figura 9d). As formas dos agrupamentos correspondem a estruturas secundárias regulares que, por sua vez, comporão a estrutura terciária da proteína (Figuras 9 e 10).

Uma vez que os diagramas de HCA de um determinado grupo de proteínas estão disponíveis, a comparação de regiões e de resíduos de aminoácidos conservados, bem como a disposição das regiões globulares, é realizada visualmente. Os dados gerados pelo HCA são utilizados, então, para a modelagem de proteínas ou para gerar informação funcional sobre as mesmas.

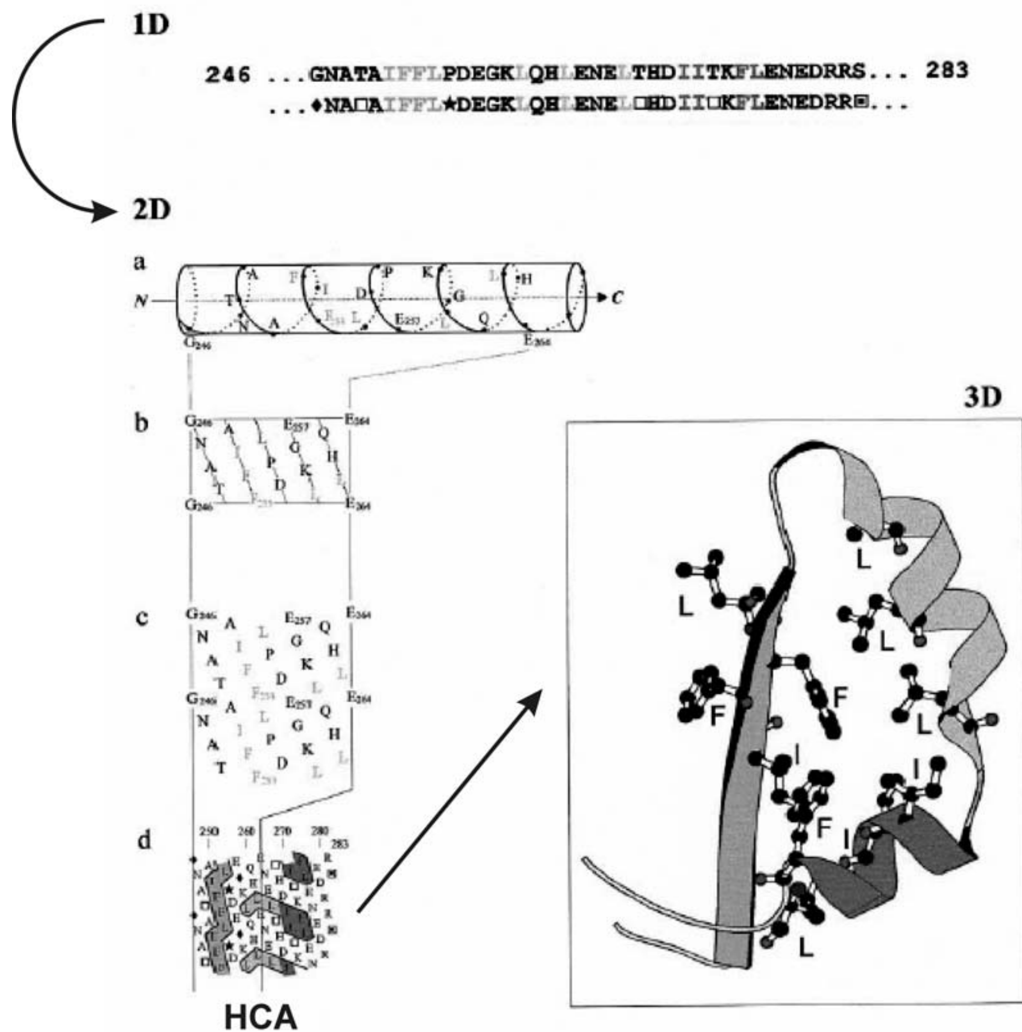
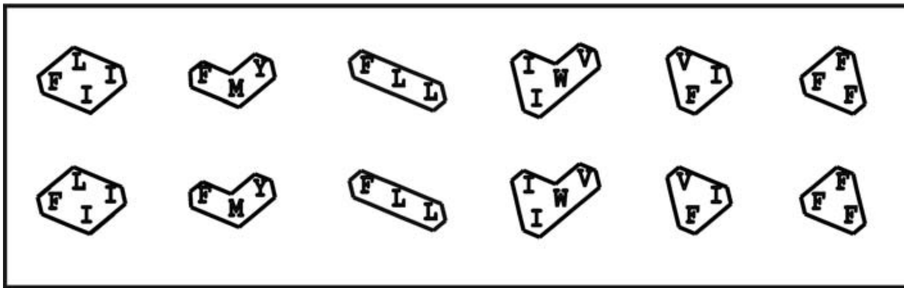


Figura 9. Conversão de uma seqüência primária de aminoácidos (1D) em um diagrama de HCA (2D). A estrutura primária é enrolada ao redor de um cilindro (a) que é cortado ao longo de seu eixo originando um diagrama bi-dimensional (b). Este diagrama é duplicado e compactado (c), de forma que os agrupamentos hidrofóbicos são contornados (d). A forma que estes agrupamentos possuem correspondem a estruturas secundárias regulares que, por sua vez, definirão a estrutura tri-dimensional da seqüência (3D). (Adaptado de Callebaut *et al.,* 1997).



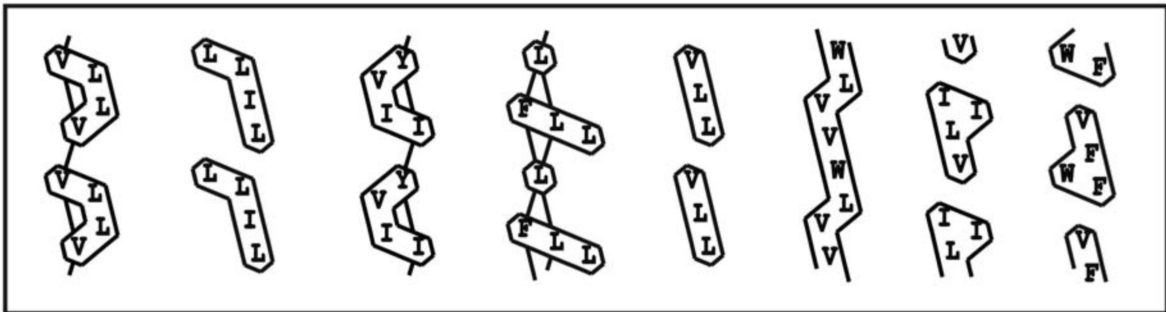


Figura 10. Exemplos de diferentes formas de agrupamentos hidrofóbicos que correspondem a α-hélices e β-estruturas.

5.2. Análises filogenéticas

A filogenia busca determinar as relações evolutivas existentes entre as diferentes famílias de proteínas, de ácidos nucléicos ou de organismos. Estas relações são comumente representadas em gráficos matemáticos conhecidos como ‘árvores filogenéticas’ (Baxevanis & Ouellete, 2001). Uma árvore filogenética possui ‘ramos’ (Figura 11), onde estão situadas as seqüências ou os organismos em estudo, sendo que a proximidade de duas ou mais seqüências/organismos determina a sua relação evolutiva. Em outras palavras, quanto mais próximo estiverem duas seqüências, maior é a probabilidade de que ambas compartilhem um ancestral comum (também conhecido como ‘nó’ em uma árvore filogenética, Figura 11). Um grupo de proteínas que compartilha um ancestral comum constitui um clado (Figura 11). Por este motivo, a filogenia de proteínas continua sendo a mais importante ferramenta de análise de seqüências (Baxevanis & Ouellete, 2001). Uma vez que uma

família de proteínas é encontrada em um organismo ou em um grupo de organismos, a relação filogenética entre elas pode ajudar a prever quais proteínas terão funções equivalentes, previsões estas que podem ser testadas e validadas por experimentos bioquímicos e genéticos (Baxevanis & Ouellete, 2001).

Assim, o estudo da evolução de proteínas envolve a comparação de seqüências homólogas, ou seja, seqüências que possuem origens comuns, mas que podem ter ou não a mesma função. Estas seqüências homólogas podem ser classificadas como: (i) ortólogas (proteínas de diferentes organismos que derivam de um ancestral comum, sendo que, normalmente, apresentam a mesma função), (ii) parálogas (proteínas homólogas originadas a partir de um evento de duplicação em um determinado organismo; na maioria das vezes, as proteínas parálogas possuem funções diferenciadas) e (iii) xenólogas (proteínas homólogas originadas a partir de um evento de transferência horizontal entre dois organismos) (Figura 11; Baxevanis & Ouellete, 2001; Mount, 2001).

Uma análise filogenética requer três passos: (i) o alinhamento das seqüências, (2) a determinação do modelo evolutivo e (3) o desenho da árvore a partir dos resultados obtidos com o modelo evolutivo. Os métodos utilizados para o alinhamento de seqüências foram detalhados no item 5.1.1., sendo que um bom alinhamento é fundamental para a geração de dados de qualidade para a análise filogenética.

A determinação do modelo evolutivo também é essencial para a obtenção de dados que correspondam à história evolutiva de uma família de seqüências e que permitam o desenho ideal da árvore filogenética. Atualmente são descritos quatro modelos evolutivos (Li, 1997): (1) métodos de distância [dos quais destacam-se *Neighbor-Joining* (NJ) e a Evolução Mínima (ME)], (2) métodos de máxima parcimônia, (3) método da máxima verossimilhança e (4) inferência Bayesiana. Cada modelo possui características matemáticas únicas, sendo que a escolha do modelo depende do tipo de dado a ser analisado (Li, 1997; Baxevanis & Ouellete, 2001). Contudo, para análises filogenéticas de

proteínas, os métodos de distância e de máxima parcimônia são os mais utilizados, considerando-se a velocidade com que os dados são analisados e a qualidade dos resultados gerados (Li, 1997).

Por fim, a robustez de uma árvore filogenética, independente do modelo evolutivo usado, deve ser testada estatisticamente a fim de garantir que o padrão evolutivo de uma árvore filogenética representa a realidade biológica (Li, 1997). Diferentes metodologias estatísticas têm sido propostas para este fim, destacando-se o *bootstrap* e a análise de ramos internos (Baxevanis & Ouellete, 2001). Ambos os métodos baseiam-se na reamostragem dos resultados obtidos a partir de um modelo filogenético, de forma que o número de vezes que um determinado ramo se forma, a partir de novas árvores filogenéticas geradas aleatoriamente pelo método, é considerado como um teste de confiabilidade. Em outras palavras, quanto maior o número de vezes que o ramo se forma, maiores são as chances de que este represente uma verdadeira relação filogenética (Mount, 2001)

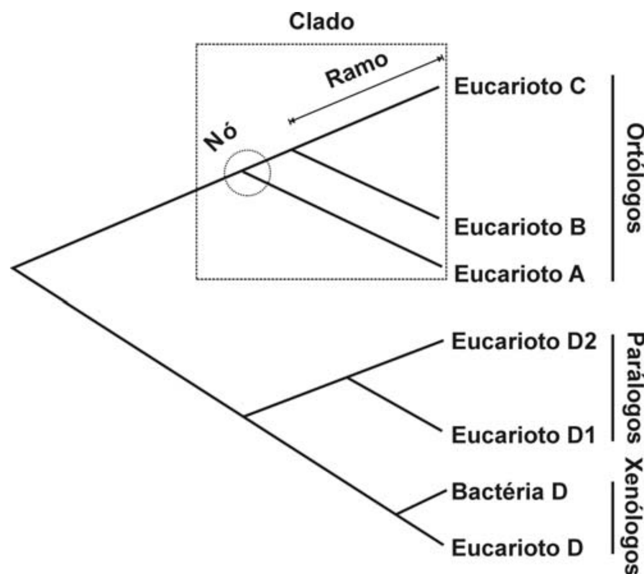
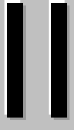


Figura 11. Exemplo de uma árvore filogenética contendo seqüências ortólogas (eucarioto A, B e C), parálogas (eucarioto D1 e D2) e xenólogas (bactéria e eucarioto D). Nesta árvore também estão representados os principais elementos de um filograma como o nó (círculo pontilhado), o ramo e o clado (quadrado pontilhado).



Objetivos

1. Objetivo geral

Estudar a proteína Pso2/Snm1 de *Saccharomyces cerevisiae* por meio da análise filogenética e de seqüências, comparando-a com proteínas homólogas já descritas para outros organismos e cujas funções, tanto na reparação de DNA quanto na manutenção da cromatina eucariótica, já estejam estabelecidas.

2. Objetivos específicos

- Analisar a presença de seqüências ortólogas e parálogas a Pso2p em organismos pertencentes aos filos *Viridiplantae, Metazoa, Fungi* e *Protozoa* utilizando, para este fim, HCA, modelagem de proteínas e análises filogenéticas;
- Determinar a presença de domínios conservados nas proteínas Pso2 por meio da técnica de HCA;
- Determinar a natureza e o número de grupos que compõem a família Pso2p/Snm1p por meio de análises filogenéticas;
- Caracterizar novas seqüências do grupo Ártemis, identificadas em metazoários e fungos, por meio da análise de seqüências, filogenia e modelagem de proteínas;
- Caracterizar, por filogenia e análise de seqüências, uma nova família de DNA ligases de plantas a qual possui um domínio Pso2p/Snm1p em sua extremidade N-terminal.

1

Capítulo Um

THE EUKARYOTIC PSO2/SNM1/ARTEMIS PROTEINS AND THEIR FUNCTION AS GENOMIC AND CELLULAR CARETAKERS

Brazilian Journal of Medical and Biological Research **(2005) 38:**

Brazilian Journal of Medical and Biological Research (2005) 38: 321-334
ISSN 0100-879X Review

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

D. Bonatto[1], L.F. Revers[2], M. Brendel[1] and J.A.P. Henriques[1,3,4]

[1]Departamento de Biofísica, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
[2]EMBRAPA Uva e Vinho, Bento Gonçalves, RS, Brasil
[3]Curso de Farmácia, Universidade Luterana do Brasil, Canoas, RS, Brasil
[4]Instituto de Biotecnologia, Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil

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

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*

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

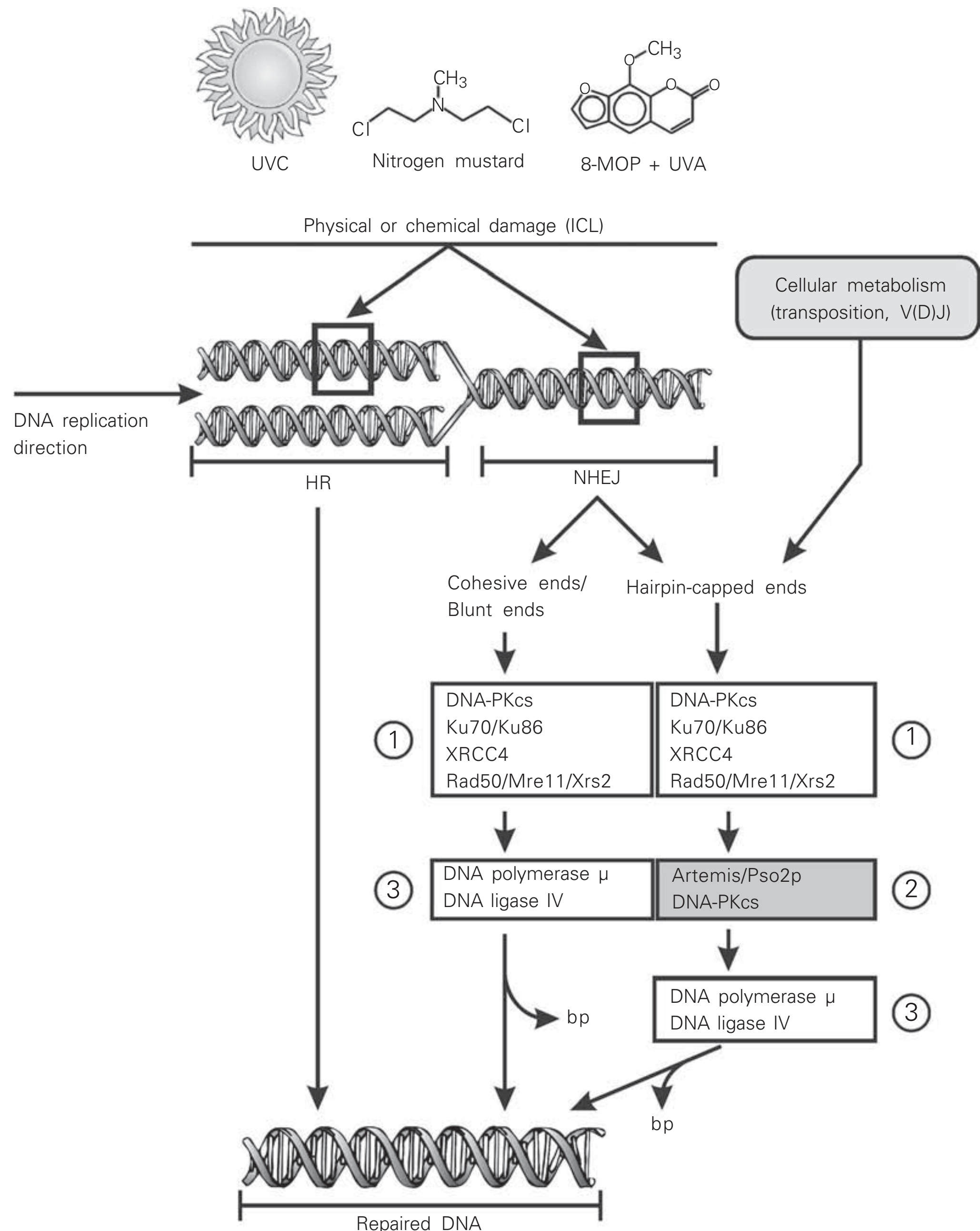


Figure 1. Schematic drawing of double-strand break (DSB) repair in mammalian cells. DSB induced by inter-strand cross-link (ICL) generated by physical agents (UVC), chemical substances (nitrogen mustard, 8-MOP + UVA), or even cellular metabolism (gray box) on DNA during replication can be repaired by 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

[a]Proteins present in different types of tissues or cells. [b]Indicates if protein activity is induced or modified by site-specific phosphorylation. [c]Physiological 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-ß-lactamase associated CPSF Artemis SNM1/PSO2 (ß-CASP) family.

The ß-CASP family

The ß-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 ß-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 ß-CASP fam-

ily, a conserved carboxy-terminal region, defined as the "ß-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 Athabascan-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 ß-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).

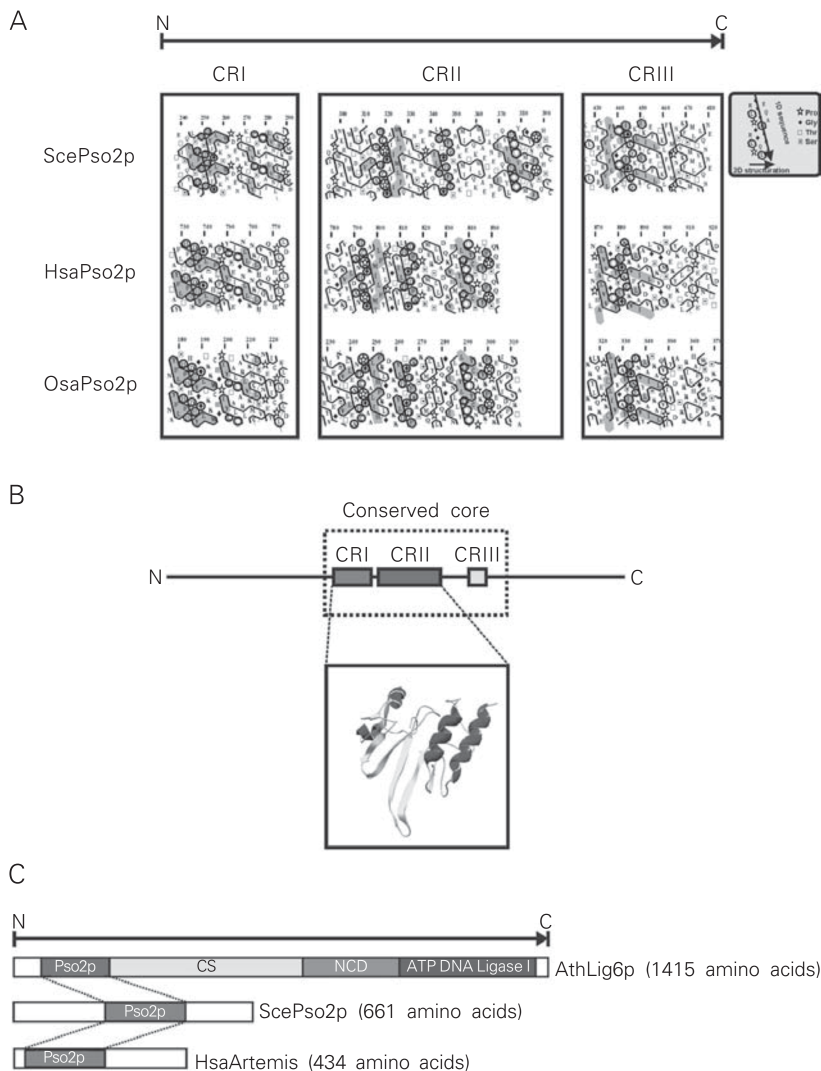


Figure 2. Hydrophobic cluster analysis of yeast (*Saccharomyces cerevisiae*, Sce), human (*Homo sapiens*, Hsa), and rice (*Oriza sativa*, Osa) Pso2p sequences (A). The three conserved regions (CRI-CRIII) of the Pso2 catalytic core are indicated. Conserved hydrophobic amino acid residues appear in gray 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.

The Pso2/Snm1 protein

Experimental data accumulated over the

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://

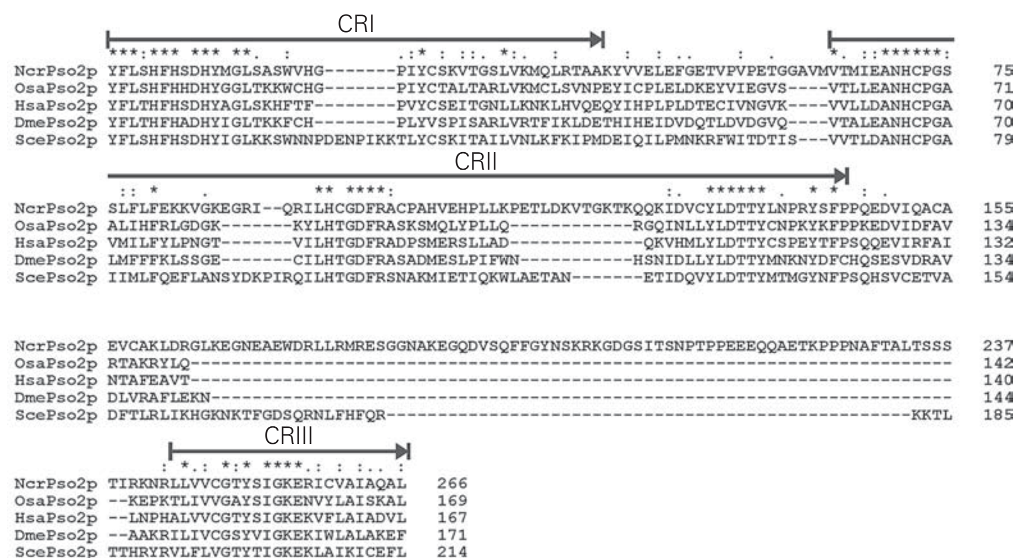


Figure 3. Multiple alignment of Pso2p conserved region sequences (CRI to CRIII) from yeast (*Saccharomyces cerevisiae*, ScePso2p), humans (*Homo sapiens*, HsaPso2p), filamentous fungi (*Neurospora crassa*, NcrPso2p), fruit flies (*Drosophila melanogaster*, DmePso2p), and rice (*Oriza sativa*, OsaPso2p). 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.

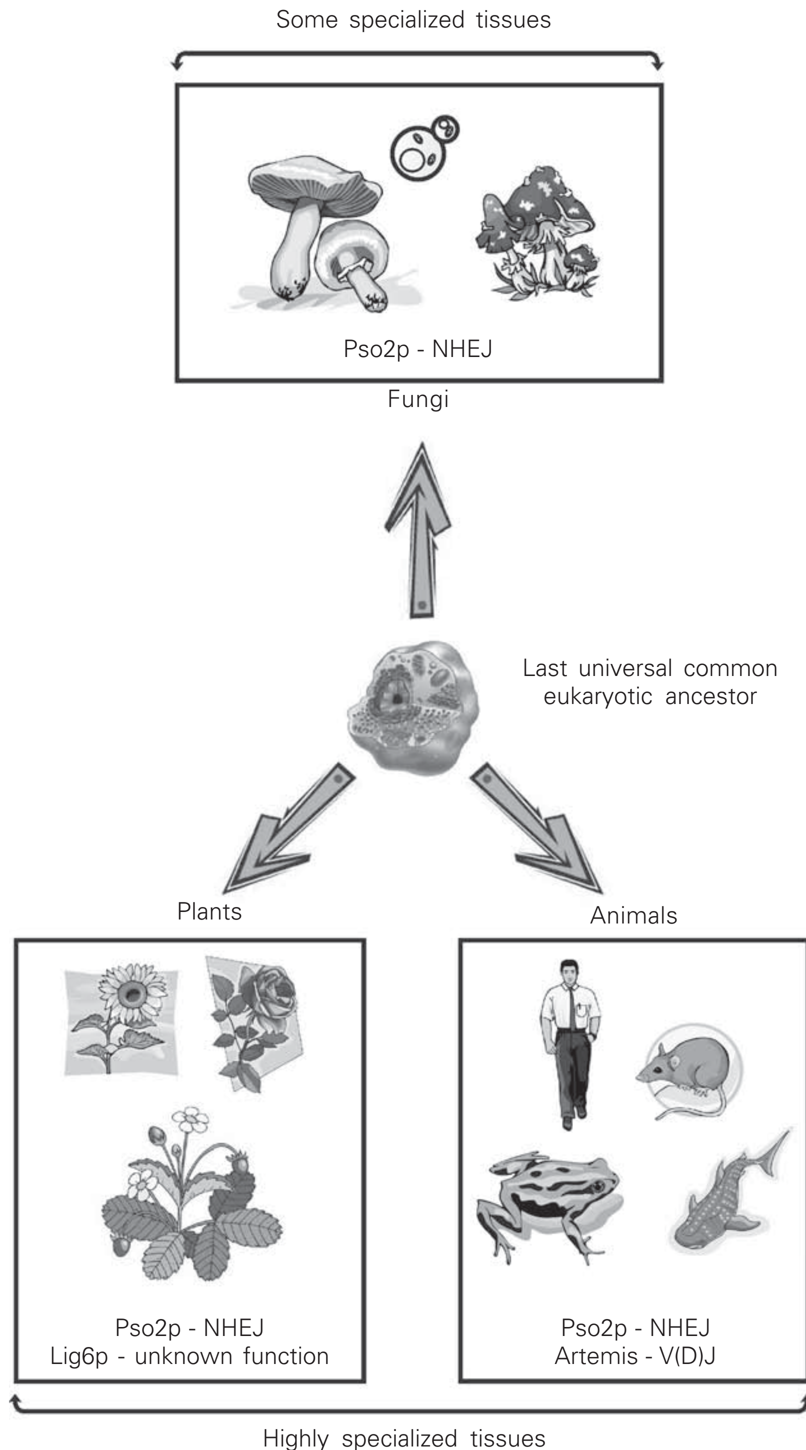


Figure 4. Evolutionary diversification of Pso2p in fungi, animals, and plants from a last universal common eukaryotic ancestor. Artemis and Lig6p are represented within animals and plants, respectively. Animals and plants contain paralogous *PSO2* genes, but they are represented by a single sequence 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-ß-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*/*SNM1* (*hPSO2*/*hSNM1*) 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 ß-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 endonucleolytic 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 transposition 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

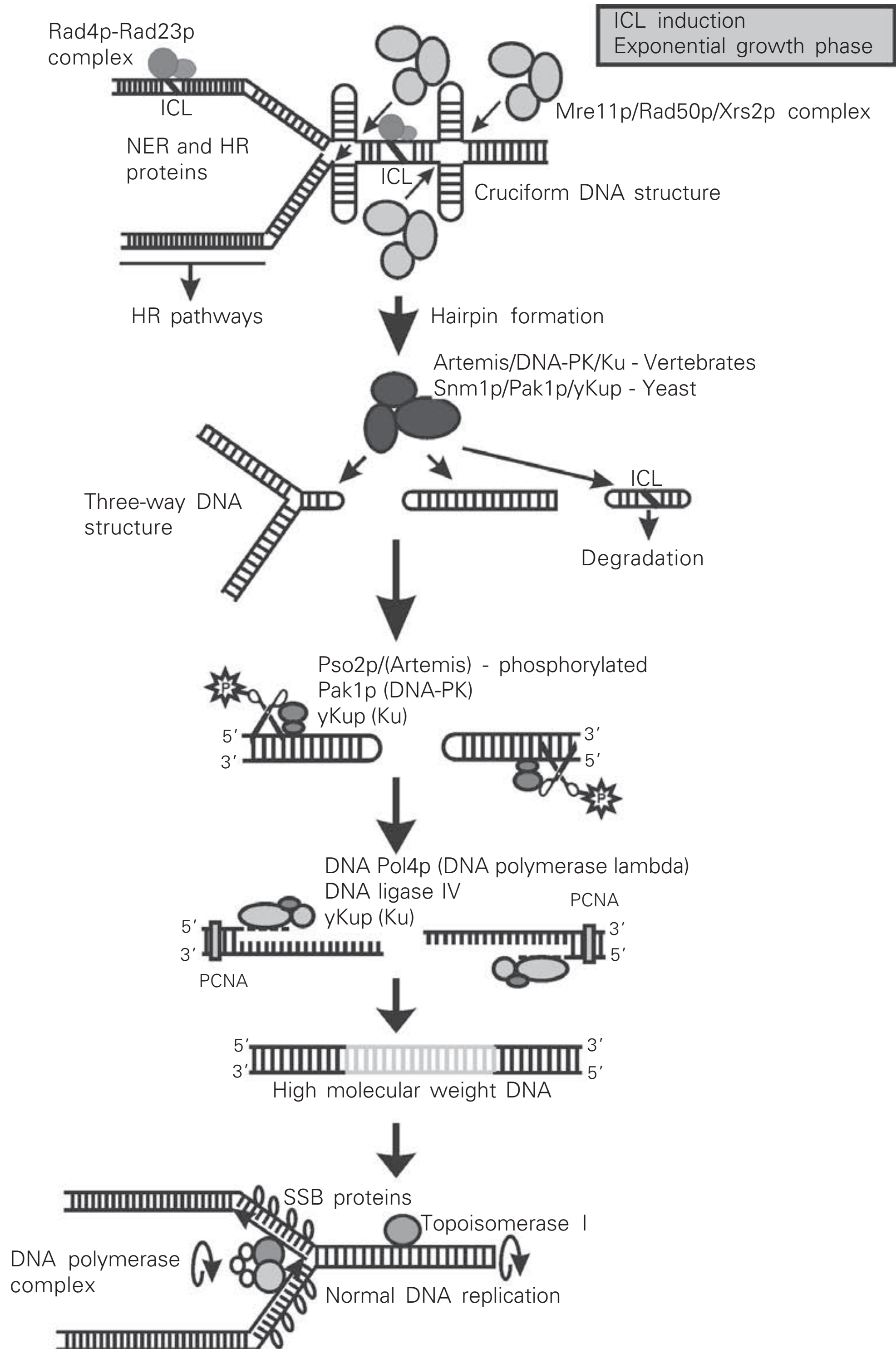


Figure 5. Non-homologous end joining recombination mediated by Snm1p/Pso2p in growing cells after inter-strand cross-linking (ICL) induction during DNA replication. In the presence of a sister strain, DNA repair may proceed via homologous recombination (HR pathways) mediated by Rad4p-Rad23p and HR proteins. 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-ß-lactamase and ß-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-ß-lactamase and ß-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 asymptomatically 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. Kuziminov 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-ß-lactamase fold within nucleic acids processing enzymes: the ß-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 interstrand 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. Poinsignon C, Moshous D, Callebaut I, de Chasseval R, Villey I & de Villartay J-P (2004). The metallo-ß-lactamase/ß-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 illegimate 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.

2

Capítulo Dois

THE EUKARYOTIC PSO2P/SNM1P FAMILY REVISITED: *IN SILICO* ANALYSES OF PSO2P A, B AND *PLASMODIUM* GROUPS

Proteins: Structure, Function, and Bioinformatics **(submetido para publicação)**

The Eukaryotic Pso2p/Snm1p Family Revisited: *In Silico* Analyses of Pso2p A, B and *Plasmodium* Groups

Diego Bonatto[1], Martin Brendel[1,2] and João Antonio Pêgas Henriques[1,3,4] *

[1]Departamento de Biofísica/Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, 91507-970 Porto Alegre, RS, Brasil.
[2]Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz (UESC), Ilhéus, BA, Brasil.
[3]Curso de Farmácia, Universidade Luterana do Brasil (ULBRA), Canoas, RS, Brasil.
[4]Instituto de Biotecnologia, Universidade de Caxias do Sul (UCS), Caxias do Sul, RS, Brasil.

Short title: The Pso2p/Snm1p A, B and *Plasmodium* Groups

[*]Address to which proofs should be sent:

João Antonio Pêgas Henriques

Departamento de Biofísica/Centro de Biotecnologia, UFRGS,

Av. Bento Gonçalves 9500, Porto Alegre, RS, Brasil, 91507-970.

Telephone: +55-51-3316-7602. Fax: +55-51-3316-6084.

E-mail:pegas@cbiot.ufrgs.br

Contract/grant sponsor: FAPERGS, CAPES, GENOTOX laboratory.

ABSTRACT. The eukaryotic family of Pso2/Snm1 exo/endonuclease proteins has important functions in repair of DNA damages induced by chemical interstrand cross-linking agents and ionizing radiation. These exo/endonucleases are also necessary for V(D)J recombination and genomic caretaking. However, despite the growing biochemical data about this family, little is known about the number of orthologous/paralogous Pso2p/Snm1p sequences in eukaryotes and how they are phylogenetically organized. In this work we have characterized new Pso2p/Snm1p sequences from the finished and unfinished eukaryotic genomes and performed an in-depth phylogenetic analysis. The results indicate that four phylogenetically related groups compose the Pso2p/Snm1p family: (i) the Artemis/Artemis-like group, (ii) the Pso2p A group, (iii) the Pso2p B group and (iv), the Pso2p *Plasmodium* group. Using the available biochemical and genomic information about Pso2p/Snm1p family, we concentrate our research in the study of Pso2p A, B and *Plasmodium* groups. The phylogenetic results showed that A and B groups can be organized in specific subgroups, with different functions in DNA metabolism. Moreover, we subjected selected Pso2p A, B and *Plasmodium* proteins to hydrophobic cluster analysis (HCA) in order to map and to compare conserved regions within these sequences. Four conserved regions could be detected by HCA, which are distributed along the metallo-β-lactamase and β-CASP motifs. Interestingly, both Pso2p A and B proteins are structurally similar, while Pso2p *Plasmodium* proteins have an unique domain organization. The possible functions of A, B and *Plasmodium* groups are discussed.

Key words: Pso2p/Snm1p family; Pso2p A group; Pso2p B group; Pso2p *Plasmodium* group; hydrophobic cluster analysis; protein phylogeny.

INTRODUCTION

The metallo-β-lactamase superfamily is composed by proteins with important functions for general metabolism in both prokaryotic and eukaryotic organisms.[1,2] Characteristically, all proteins of the metallo-β-lactamase superfamily use as substrate a compound containing an ester linkage and a negative charge in its molecular structure; they catalyze the hydrolysis of the former.[1,2] Moreover, the proteins of metallo-β-lactamase family have a highly conserved HxHxDH signature and a binuclear Zn^{2+} center (the metallo-β-lactamase motif), which is necessary for the ester cleavage step.[1,2] The characteristic metallo-β-lactamase fold consists of external α-helices and two internal layers of β-sheets.[3] In general, the metallo-β-lactamase fold allows for remarkably different catalytic activities and metal selectivities.[4]

Using the methodology of hydrophobic cluster analysis (HCA), Callebaut *et al.*[2] recently described a new family of metallo-β-lactamases proteins, which is present in all three Domains of Life.[2] This family, the metallo-β-lactamase associated CPSF Artemis SNM1/PSO2 (β-CASP) proteins, has important functions in DNA and RNA metabolism.[2] A detailed analysis of conserved domains showed that all proteins of the β-CASP family contain two conserved regions: (i) a metallo-β-lactamase motif and (ii) a "β-CASP" motif, composed by 3 domains named A, B and C, which are necessary for the protein functions.[2] Preliminary protein sequence analyses, including the Artemis/Pso2 sequences, Elac1/Elac2 and cleavage and polyadenylation specificity factor (CPSF) 73 kDa/100 kDa indicate similar functions.[2] The Elac1/Elac2 proteins, also known as binuclear metallohydrolase zinc phosphodiesterase (ZiPD) enzymes, are essential for tRNA biosynthesis in both prokaryotes and eukaryotes.[5,6] Interestingly, Elac2 mutant variants have been associated

with human prostate cancer[7-9] and studies of mRNA expression revealed that both human proteins were expressed in all tissues analyzed.[10,11] The CPSF 73 kDa hydrolyzes mRNA, and this group is conserved in eukaryotes as well as in archaea.[12] These proteins are important component of the eukaryotic machinery that processes the 3' end of mRNAs, acting together with three other CPSF proteins (30/100/160 kDa), as well as with the cleavage stimulation factor (CstF), poly(ADP-ribose) polymerase (PARP), two additional cleavage factors (CF Im and CF IIm), and poly(A)-binding protein II (PAB II).[12]

However, the best-characterized member of the β-CASP family is Artemis, a protein isolated from cells of patients suffering from a special type of severe combined immunodeficiency syndrome associated with radiosensitivity (RS-SCID).[13] Artemis has 5' to 3' exonucleolytic activity with single-strand DNA specificity and when associated with DNA-dependent protein kinase (DNA-PKcs) forms a phosphorylated complex with endonucleolytic activity on both 5' and 3' DNA overhangs. Furthermore it can cleave hairpins generated by the Rag1/Rag2 proteins.[14,15] The Artemis protein is part of Pso2p/Snm1p family, which has been characterized in eukaryotes. Actually, the Pso2p/Snm1p family can be subdivided into three phylogenetic related groups: Pso2p/Snm1p A, Pso2p/Snm1p B, and Artemis.[16] With exception of Artemis, the function(s) of Pso2p/Snm1p A and B groups in DNA metabolism are not fully understood. However, many authors agree that these proteins act in a post-incision step during the repair of interstrand cross-links (ICLs) induced by chemical compounds in eukaryotic cells.[16-18] Recent data indicate that Artemis and Pso2 A group proteins could have an important role in the maintenance of genomic integrity in vertebrate cells.[19] On the other hand, the B group appears to have a broader function when compared to A group, with

proteins that are recruited for the repair of DNA damages induced by ICLs or ionizing radiation (IR).[20] As the importance of Pso2p/Snm1p family is exponentially growing, it has become clear that it is necessary to re-analyze the phylogenetic position and the structural features of the Pso2p/Snm1p A and B groups. In this work we have identified and characterized new members of Pso2p/Snm1p A and B groups from eukaryotic genomic databases using sensitive methods of phylogenetic analysis. The phylogeny indicates that A and B groups are composed of distinct subgroups. Moreover, our results point to a not previously described *Plasmodium* Pso2p/Snm1p group. Additional HCA allowed us to refine the results obtained from phylogeny and to map conserved domains of these proteins.

MATERIALS AND METHODS

Protein data mining

Eighty-five protein sequences (representing the β-CASP family) were obtained directly from the following databases: (i) GenBank hosted in the National Center for Biotechnological Information (NCBI) web page [http://www.ncbi.nlm.nih.gov/]; (ii) Genolevures Database [http://cbi.labri.fr/Genolevures/index.php]; (iii) Fugu Genome Project [http://www.fugu-sg.org]; (iv) *Chlamydomonas* Genetics Center [http://www.biology.duke.edu/chlamy]; (v) Solanaceae Genomics Network [http://www.sgn.cornell.edu]; and (vi) *Danio rerio* Sequencing Project [http://www.sanger.ac.uk/Projects/D_rerio]. BLAST, PSI-BLAST and TBLASTN programs were used for initial domain screening and comparison[20] with *Sce*Pso2p (Table 1) as query sequence. All searches were made to saturation. Wise2 program at European Bioinformatics Institute (http://www.ebi.ac.uk/wise2/) was used to identify Pso2/Snm1 proteins from unfinished genomic sequences of eukaryotic species. The parameters for

prediction were: local mode; no intron bias; splice site modeled; synchronous model; and GeneWise623 algorithm.

Sequence comparison and phylogenetic inference

Global pair-wise multiple-alignment of members of eukaryotic β-CASP family was performed in the CLUSTALX 1.8 program.[22] The alignment parameters used were: gap open penalty 10.00; gap extension 0.20; sequences >10% diverged delayed; BLOSUM series matrix; residue-specific penalties on; and hydrophilic penalties on. When necessary the alignments were manually adjusted using the BioEdit program.[23] Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 2.1.[24] Neighbor-Joining (NJ) method was used for phylogenetic tree searching and inference. The statistical reliability of the phylogenetic trees was tested by interior branch analysis with 1,000 replications. Moreover, the Poisson correction was applied in NJ for distance estimation. The pair-wise deletion option was used in handling gaps or missing data obtained from the alignment.

Hydrophobic cluster analysis (HCA)

HCA of selected Pso2p/Snm1p sequences was performed as previously published[25] and consisted in displaying the primary protein structure on a duplicated α-helical net, where the hydrophobic residues are automatically contourned. The hydrophobic clusters observed in an HCA plot are not distributed in a random manner, but correspond highly to the secondary protein structures associated with conserved globular domains. Thus, HCA plots are extremely valuable for phylogenetic inferences when protein sequences have a weak homology (<25% of identity/similarity) or to define conserved domains and sequence signatures in a set of homologous proteins.[25] It should be noted that HCA, when compared

to 'linear' or one-dimensional methods of database screening, e.g. BLAST and PSI-BLAST (which need, at least, >30% of aa identity/similarity over a sufficient length), is a more effective tool to detect sequence similarity that reflects a true three-dimensional relationship between two or more proteins.[26] Moreover, HCA has the advantage that sequence conservation can be detected in a group of proteins without any previous one-dimensional alignment.[26,27] The program DRAWHCA, used in such analysis, is available as a freeware at http://www.lmcp.jussieu.fr.

RESULTS AND DISCUSSION

The phylogeny of Pso2p/Snm1p family

In order to identify new potential Pso2 A and B proteins from the available eukaryotic genomic databases, a data mining approach was taken using the *Sce*Pso2p (Table 2) as query in BLAST, PSI-BLAST and TBLASTN programs. We could identify many orthologues/paralogues Pso2 proteins (present as complete and incomplete sequences) in *Protozoa, Fungal, Metazoa,* and *Viridiplantae* databases (Tables 1-3). With the objective of determining the global phylogenetic relationship of the Pso2 proteins (Tables 1 and 2) within the β-CASP family, the sequences obtained from genomic database mining were subjected to a global sequence comparison followed by a phylogenetic analysis with selected eukaryotic Elac2 and CPSF 73 kDa proteins (Table 3). As expected, both Elac2 and CPSF 73 kDa sequences compose two distinct phylogenetic groups, with function on RNA metabolism (Fig. 1). The Elac proteins were previously identified in all three domains, but their function is not very well understood.[4,10] Sequence analyses of Elac proteins grouped these proteins in two major clades: (i) the ubiquitous Elac1 clade, present in eukaryotes, eubacteria, and archaebacteria and (ii) the Elac2 clade, formed by eukaryotic

sequences.[10] Biochemical studies of *Escherichia coli* elaC protein, which belongs to Elac1 clade, showed that this enzyme is a 3' tRNase responsible for the removal of a 3' trailer from precursor tRNA (pre-tRNA).[4] The enzyme cleaves pre-tRNA immediately downstream of a discriminator nucleotide,[28,29] onto which the CCA residues are added to produce mature tRNA. The same enzymatic reaction was observed in metazoans and plants Elac2 proteins, acting in various nuclear and organellar pre-tRNAs.[30,31] Interestingly, human *ELAC2* gene was the first prostate cancer susceptibility gene identified by linkage analysis and positional cloning.[10] Recent data suggest that human Elac2 interact with γ-tubulin, an indicative that Elac proteins could have an active role on cell cycle control.[32] On the other hand, CPSF 73 kDa plays a central role in pre-mRNA processing, specifically in the addition of a poly(A) tail.[33] This mRNA modification is an essential process that promotes transcription termination,[34] transport of the mRNA from the nucleus,[35,36] optimal translation and regulate mRNA stability.[37] The addition of the poly(A) tail *in vivo* is accomplished in two-step, tightly coupled reactions comprising endonucleolytic cleavage of the 'U- or GU-rich' downstream elements and subsequent addition of a poly(A) tail.[38] A set of protein factors is required to reconstitute the complete polyadenylation reaction *in vitro,* and most of these factors consist of several subunits.[39] Until now, four CPSF proteins were described in eukaryotic cells: CPSF 160 kDa, CPSF 100 kDa, CPSF 73 kDa and CPSF 30 kDa.[38] The functions of CPSF 160 kDa and CPSF 30 kDa in pre-mRNA processing are well established, but less is known about the roles of CPSF 100 kDa and CPSF 73 kDa. Recently, Ryan *et al.*[40] showed that the induction of point mutations in the metallo-β-lactamase motif of CPSF 73 kDa inactivated the enzyme. Moreover, they also showed that the 3' processing endonuclease activity is dependent on Zn^{2+} as co-factor, a characteristic of

almost all proteins of β-CASP family.[40] In *Arabidopsis thaliana,* two genes encoding products with significant homology to CPSF 73 kDa subunit were identified: *AtCPSF73-I* and *AtCPSF73-II.*[41] Genetic analysis indicated that the disruption of the *AtCPSF73-II* gene in *A. thaliana* heterozygous mutant lines caused severe reduction in genetic transmission of female gametes due to a loss of fertility, while the transmission of male gametes was normal.[41] As observed in yeast,[40] the homozygous mutant lines for *AtCPSF73-II* were not viable,[41] again corroborating the essentiality of CPSF 73 kDa for cellular function.

The global sequence alignment and phylogenetic analysis of Pso2p/Snm1p family showed that these proteins can be grouped in four distinct clades, all statistically supported by internal branch analysis, with functions on DNA metabolism: (i) the Pso2p *Plasmodium* group (PPG), which is apparently the oldest group of Pso2p/Snm1p family, (ii) the Artemis/Artemis-like group, (iii) the Pso2p B group and (iv) Pso2p A group (Fig. 1). Artemis/Artemis-like group forms a clade that diverged early in the evolution of Pso2p/Snm1p family, while Pso2p A and Pso2p B groups compose two clades that share a recent common ancestor (Fig. 1). Despite the interesting results obtained in this work with Artemis/Artemis-like sequences, this group was analyzed in depth elsewhere (Bonatto *et al.*, manuscript submitted). Thus, we focus our efforts on the study of Pso2p A, B and *Plasmodium* groups using the phylogenetic data associated with HCA of selected sequences.

The Pso2p A group

The Pso2p A group is the largest clade within the Pso2p/Snm1p family (Figs. 1 and 2). Probably all eukaryotic organisms contain at least one gene that codifies for a Pso2p A sequence. The phylogeny allowed us to divide this large group in 4 related subgroups, all

supported by internal branch analysis: (i) the plant-specific A (PSA) subgroup, (ii) the multicellular eukaryotic A (MEA) subgroup, (iii) the protozoan A (PRA) subgroup and (iv) the Pso2p fungal A (FA) subgroup (Table 2, Fig. 2).

The plant Pso2p sequences are the most diversified proteins within the Pso2p A group. The phylogeny shows that these sequences are present in both PSA and MEA subgroups (Fig. 2). The PSA subgroup contains sequences that belong to the plant-specific Pso2p A and to ATP-dependent DNA ligase VI (Lig6) proteins (Fig. 2). The functions of the Pso2p PSA sequences on plant DNA metabolism are unknown. However, we recently described that the Lig6p composes a distinct clade which shares a last universal common ancestor with eukaryotic DNA ligases I.[42] A HCA mapping of conserved regions in Lig6p from different plant species shows two highly conserved domains: (i) a Pso2p/Snm1p domain in the N-terminus of the protein; and (ii) a DNA ligase I domain situated in the C-terminus of the sequence.[42]

The other plant sequences belong to MEA subgroup. We identified two plant members in this group, one *A. thaliana* sequence (*Ath*Pso2p MEA, Table 2 and Fig. 2) and one *Oryza sativa* sequence (*Osa*Pso2p MEA, Table 2 and Fig. 2). A possible function of *Ath*Pso2p MEA on plant DNA metabolism was recently studied, and the results indicated that this protein is required for recombinational repair of DNA lesions induced by reactive oxygen species,[43] a function that was not previously described for the Pso2p/Snm1p family.

The metazoan and protozoan Pso2p form two distinct subgroups (MEA and PA, respectively) within A group. Differently from plant sequences, we could not detect paralogous proteins in these organisms (Fig. 2). The functions of MEA subgroup in DNA repair have been studied in mammalian cells and, recently, in chicken B-cell line.[16,19] The disruption of Pso2p MEA sequences in mouse embryonic stem (ES) cells induces an

increase sensitivity to mitomycin C (MMC) but not to other cross-linking agents or to IR.[44] It has been described that *Hsa*Pso2p MEA is located within multiple punctate foci or forming one or two larger bodies in the nucleus.[45] Moreover, *Hsa*Pso2p MEA colocalizes in a DNA damage-independent manner with the p53 binding-protein (53BP1), a protein that plays a role in the cellular response to IR.[45] The function of this complex is not fully understood, but Akhter *et al.*[19] recently demonstrated that *Hsa*Pso2p MEA and 53BP1 interact with components of the anaphase-promoting complex (APC)/cyclosome, suggesting that *Hsa*Pso2p MEA is a component of a mitotic stress checkpoint that negatively targets the APC prior to chromosome condensation. The disruption of chicken *Gga*Pso2p MEA also results in an increased sensibility to MMC and cisplatin but not to other DNA damage agents.[16] These facts support a role of Pso2p MEA in the repair of ICLs in multicellular eukaryotes. Additionally, it has been shown that chicken *Gga*Pso2p MEA physically interacts with PIAS1, a small ubiquitin-like modifier (SUMO) E3 ligase, an indicative that *Hsa*Pso2p could be sumoylated during the ICL repair.[16] As both plant and metazoan Pso2p MEA sequences form a strongly cohesive subgroup (Fig. 2), and considering that plant MEA sequences have a role in the repair of oxidative damage, it should be interesting to test if Pso2p MEA-defective metazoan cells are also sensitive to oxidative damages in DNA.

In the protozoan species *Dictiostelyum discoideum* and *Entamoeba histolytica* we could identify two Pso2p A sequences, which together compose the PA subgroup (Table 2, Fig. 2). The available data about a possible function of Pso2p A proteins in protozoan cells practically do not exist, but taking into account the closest phylogenetic relationship of PA subgroup with the MEA and PSA subgroups, these proteins probable have a role in the repair of ICL or oxidative DNA damages.

In comparison to PSA, MEA and PA subgroups, the Pso2p sequences of fungi (FA subgroup) appear to have diverged early during the evolution of Pso2p A group, forming a separated clade (Fig. 2). Some proteins of the FA subgroup, e.g. the Pso2p FA of *Saccharomyces cerevisiae*, have been genetically and biochemically characterized. Despite the large experimental data accumulated since the isolation of *pso2/snm1* mutants of *S. cerevisiae*, the function of *Sce*Pso2p FA in ICL repair is unknown.[46-50] The yeast *pso2/snm1* mutants are extremely sensitive to ICL-inducing agents,[46-49] being only mildly sensitive to UVC and not sensitive to IR.[48,49] Furthermore, *S. cerevisiae pso2/snm1* mutants, though incapable of forming high molecular weight DNA during repair of ICL, are not defective in repair of DSBs.[47,50] Stability of the mitochondrial DNA is also affected in these mutants, as they have a higher-than-WT frequency of spontaneous "petit" mutations,[51] suggesting a possible function for Pso2p/Snm1p in mtDNA repair in yeast. Recently, we have proposed a model in which the Pso2p would act on DNA hairpin substrates induced by ICLs during DNA replication.[17] This model was partially corroborated when Yu *et al*. showed that *Sce*Pso2p FA could function in DNA repair of hairpins induced by transposition of Ac/Dc elements from *Zea mays* in *S. cerevisiae*.[53] In this case, the expression of Ac/Dc elements in *S. cerevisiae* allows assaying the repair of excision sites in a variety of yeast mutant backgrounds.[53] This indicates that *Sce*Pso2p FA 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.[53] Recently, it was demonstrated that purified, non-phosphorylated *Sce*Pso2p FA has an *in vitro* single-stranded 5' exonuclease activity.[54]

While biochemical data of Pso2p A proteins is becoming available, much less is known about the structural requirements and essential aa residues for Pso2 A function.

With this fact in mind, we refined the results obtained with the phylogeny and subjected some sequences representing the different Pso2p A subgroups to HCA. Four conserved regions could be detected in selected Pso2p proteins (CRI-CRIV, Fig. 3), which encompass the metallo-β-lactamase and β-CASP motifs.[2] Also, a striking conservation of secondary structures and hydrophilic aa residues was evident in all subgroups analyzed (Fig. 3). CRI, which is part of metallo-β-lactamase motif, displays the HFHxDHxGxxK signature between a β-strand and an α-helix structure (Fig. 3). In all proteins of Pso2p A group observed, an invariant residue of phenylalanine was identified between the two first histidines of HFHxDHxGxxK sequence. As previously described,[1,2] the HFHxDHxGxxK sequence is essential for Zn^{2+} binding, and point mutations in this signature normally abolish the protein function.[54] In CRI a conserved (T,S)xxT sequence could be detected in all Pso2p A sequences. CRII contains three consensus sequences (Fig. 3). The first consensus sequence (E,D)xNHCPG is situated between an α-helix and a β-strand (Fig. 3). The second and third HTGDFR and D(N,T)T consensus sequences, respectively, are both located in the end of a β-strand structure (Fig. 3). Mutational studies in *Hsa*Artemis protein, which contains similar signatures, showed that these aa residues are necessary for protein function.[55] The CRIII and CRIV domains, both belonging to β-CASP motif, contain a GxϕxϕGKE (where ϕ is any hydrophobic aa residue) and a SEHSS sequences, respectively (Fig. 3). The CRIII domain has not been previously described, and the presence of a conserved lysine and glutamic acid residues in this sequence probably indicate a function in metal or structure coordination. On the other hand, point mutations that change the conserved histidine residue in Artemis CRIV domain for alanine disrupt the nucleasic activity of this protein.[55]

One finding that is very interesting is related to the Pso2 FA protein of the filamentous fungi *Aspergillus nidulans* (Table 2, Fig. 2). The *Ani*Pso2p is a very large protein (2408 aa) and using sequence analyses, we could map four independent functional domains (Fig. 4A). The first domain correspond to an UbiE sequence situated around aa 700 and 880 of *Ani*Pso2p FA, which has similarity with UbiE protein of *E. coli* (Fig. 4B). The UbiE is a C-methyltransferase enzyme, being necessary for both ubiquinone (CoQ) and menaquinone biosynthesis.[56] The corresponding C-methyltransferase gene in yeast was identified as *COQ5*.[57] CoQ/UbiE functions in the respiratory electron transport chain of the inner mitochondrial membranes of eukaryotes and in the plasma membrane of prokaryotes.[58] In addition to respiratory electron and proton transport, the redox properties of CoQ/UbiE allow the reduced form (CoQH2) to scavenge lipid peroxyl radicals either directly or indirectly as mediated through α-tocopherol.[59] This antioxidant function of CoQH2 serves to protect cells from the oxidative, damaging effect of polyunsaturated fatty acids.[60] The second conserved sequence, which spans a region between aa 1000 and 1090, shows a high similarity with the active domain of *Sce*Snp3p (Fig. 4B). This protein has an important function in binding and stabilization of the 3' end of the spliceosomal U6 snRNA.[61] *Ani*Pso2p FA shows a highly conserved Pso2p A domain situated around aa 1440 and 1820 (third domain, Fig. 4B). The fourth sequence shows an elevated identity with the active domain of cytochrome P450 enzymes, a large superfamily of haemoprotein monooxygenases present in prokaryotes and eukaryotes that play an important role in the oxidative metabolism of a wide variety of both exogenous and endogenous substrates.[62] It should be noted that *A. nidulans* is a mycotoxigenic fungus, which produces and accumulates the molecular precursor of aflatoxines denominated sterigmatocystin.[63]

Despite being less potent than aflatoxines, sterigmatocystin is able to bind covalently to DNA and induce DNA adducts.[64] Moreover, sterigmatocystin is very carcinogenic in mammalian models.[64] Interestingly, *A. nidulans* mutant strains that produce no sterigmatocystin or accumulate different intermediates of this mycotoxin have been shown to be less fit than wild-type strains, as defined by reduced sporulation.[65] Taking into account all these data, we speculate that *Ani*Pso2p FA could have an important function in protecting the genome of *A. nidulans* during the vegetative growth or sporulation. Both UbiE and P450 domains could be important to keep the redox equilibrium in the vicinity of *A. nidulans* genome, while both Snp3p and Pso2p A domains would be necessary to induce the appropriate mRNA maintenance or DNA repair pathway in sterigmatocystin-rich cells.

The Pso2p/Snm1p B group

The data analyses of Pso2p B group showed very interesting results (Fig. 5A). In comparison to its sister Pso2p A group (Fig. 1), the Pso2p B group is less diversified and is basically restrict to protozoan, metazoan and plant species (Table 3 and Fig. 5A). Additionally, our data prospection was not able to reveal any fungal Pso2p B sequences. Three subgroups within Pso2p B group could be identified on the basis of sequence analysis and phylogeny (Fig. 5A): (i) the *Leishmania* B (LB) sequences, (ii) the Pso2p plant specific B (PSB) sequences, and (iii) the multicellular eukaryotic B (MEB) sequences. Interestingly, the protozoan *Leishmania major* has two paralogous LB sequences (LB1 and LB2, Fig. 5A), but the biological significance of this fact is unknown. In plants, as observed with Pso2p A sequences, we found paralogous proteins that fall in both PSB and MEB subgroups (Fig. 5A). However, the importance of these proteins for plant genome's maintenance is completely ignored. The lack of experimental data is also observed for metazoan MEB sequences. Noteworthy, two Pso2p B sequences were found in a cDNA

database of *M. musculus* (*Mmu*Pso2p MEB1 and *Mmu*Pso2p MEB2, Table 3). Both sequences are codified by one gene situated in the mouse chromosome 3 and it is likely that differential splicing generate these proteins (data not shown). It is possible that other mammalian Pso2p MEB sequences are also subjected to the same cellular process.

While the functions of Pso2p A group in eukaryotic DNA repair or genome caretaking are becoming established, much less is known about the functions of Pso2p B. Some recent works showed that the cellular depletion of *Hsa*Pso2p and *Gga*Pso2p MEB (Table 3) resulted in hypersensitivity to different DNA ICLs agents and to IR.[16,17] Whether these cellular phenotypes arise because the Pso2p B proteins are directly involved in the repair of this DNA damage or play a more indirect role, it is a question that remains unsolved.[17] The HCA of selected Pso2p sequences of LB, PSB and MEB subgroup showed similar secondary structure conservation with Pso2p A sequences and also the same number of CRs (Fig. 5B). However, the signature sequences associated with CRs show fewer conserved aa residues. CRI of Pso2 B group is characterized by the consensus sequence HxHxDHxG (Fig. 5B). CRII, as observed in Pso2p A sequences, contains three signatures: xAxHCPG, $G(x)_4$HTGDFR and D(C,T)T (Fig. 5B), all probably related to the Zn^{2+} atom coordination together with CRI. Otherwise, the signatures of CRIII [G(K,Q)E; Fig. 5B] and CRIV [SxHS(C,S); Fig. 5B] are similar to those Pso2p A group.

The Pso2p/Snm1p *Plasmodium* group family

The Pso2p sequences corresponding to the *Plasmodium* species comprise a very ancient and distinct clade within the Pso2p/Snm1p family, as revealed by phylogeny [the Pso2p *Plasmodium* group (PPG), Figs. 1 and 6A]. Two complete and one incomplete Pso2p PPG sequences, from *Plasmodium falciparum, Plasmodium yoelli* and *Plasmodium chabaudi,* were identified in genomic databases (Table 3). These sequences, when

submitted to HCA, showed a distinct domain organization in comparison to proteins of Pso2p A and B groups (Fig. 6B). While in Pso2p A and B groups the metallo-β-lactamase motif (which comprises CRI and part of CRII domains) is fused to β-CASP motif (comprising part of CRII domain and the totality of CRIII and CRIV domains), in Pso2p PPG these motifs are separated from each other by a stretch of 250-300 aa residues (Fig. 6B). Within this stretch, many secondary structures and hydrophilic aa residues are conserved, indicating that this intermediary sequence probably has an important role on *Plasmodium* Pso2 protein function. The *Pfa*Pso2p PPG and *Pyo*Pso2p PPG also show a hinge sequence situated in the vicinity of aa 485 and 537, respectively (Fig. 6B). This hinge sequence is mainly composed of the neutral but fully polar asparagine and serine residues (data not shown). However, the biological significance of this hinge sequence is unknown.

All *Plasmodium* species belong to the large and diverse Apicomplexa phylum, which has a great economical and medical importance.[66] The members of this phylum are all parasites, with a complex life cycle, and some are important causative agents of human and animal diseases.[66,67] The most important of apicomplexans is *Plasmodium,* the agent of malaria, recognized by the World Health Organization as being one of the top three killers in the world.[66,67] The phylogeny of *Plasmodium* species is still controversial, but many studies point that they form a very ancient group, probably diverging from all other Apicomplexan species before the Cambrian.[66] Moreover, little is known about the mechanisms of DNA repair in *Plasmodium* species. Being intracellular parasites, the *Plasmodium* species employ antigenic variation in their cell surface in order to survive against the host immune pressure.[68] The antigenic variations in the *Plasmodium* species are linked to genetic rearrangements that arise during the repair of DNA breaks in the parasite

genome.[68] Many evidences suggest that a NHEJ pathway could be operating in these species, which partially explain the high level of clonal variation that *Plasmodium* species show, even in the absence of immunological pressure.[68] Like the mammalian Artemis proteins, which compose an ancient group of Pso2p/Snm1p family (Fig. 1), and are necessary for generation of immunological diversity during V(D)J recombination,[15] we speculate that the Pso2p *Plasmodium* group sequences could also have a role in the generation of parasite antigenic variation.

CONCLUSIONS

The Pso2p/Snm1p is an expanding family of DNA repair and genomic caretaking sequences with important biochemical mechanisms that are poorly understood. Until now, we are far from a final conclusion of Pso2p/Snm1p functions in DNA metabolism.

The fast growing genomic data allow us to study the presence and/or absence of Pso2p/Snm1p family in virtually all eukaryotic kingdoms. As we show in this work, the Pso2p/Snm1p family can be divided in four distinct groups, each containing specific subgroups. Our efforts were conducted with the objective to better understand the phylogenetic organization and the structural conservation of the Pso2p A and B groups, together with the previously undescribed Pso2p *Plasmodium* group. The importance of Pso2p A group for eukaryotic DNA repair is apparent, taking into account the experimental data already accumulated and the presence of very conserved sequences in all kingdoms, as shown in this work. On the other hand, the importance of Pso2p B group for DNA repair or genomic maintenance in metazoans, plants and *Leishmania* species is largely unknown, despite its structural homology with Pso2p A sequences. The same conclusion is valid for the Pso2p sequences of *Plasmodium* species, which form a separated ancient group within the Pso2p/Snm1p family. Characteristically, these parasites species show extensive clonal

variation and we cannot rule out a possible function of Pso2p *Plasmodium* sequences in this process. Interestingly, the Artemis sequences, which belong to Pso2p/Snm1p family, generate immunological diversity in jawed vertebrates, a function that was retained by selected Pso2/Snm1 proteins during the irradiation of the metazoan adaptative immune system. Actually, many biochemical data about the functions of Pso2p/Snm1p family came from the studies of mammalian Artemis sequences, a specific protein required for V(D)J recombination and genome maintenance. However, the presence of Pso2p paralogous sequences in almost all eukaryotes studied here indicates that, despite its structural similarities, each Pso2p group could have a more specific function in DNA metabolism, which can be recruited for determined situations in the organism's life cycle, like oxidative DNA damages, ICLs, or recombinational processes. The accumulating biochemical and genomic evidences about the functions of Pso2p/Snm1p family will clarify the importance of each Pso2p group for DNA metabolism in different eukaryotic species.

ACKNOWLEDGEMENTS

This work was supported by research grants from Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and GENOTOX (Laboratório de Genotoxicidade, UFRGS).

REFERENCES

1. Aravind L. An evolutionary classification of the metallo-β-lactamase fold proteins. *In Silico* Biol 1999;1:69-91.

2. Callebaut I, Moshous D, Mornon J–P, de Villartay J–P. Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. Nucleic Acids Res 2002;30:3592-3601.

3. Carfi A, Pares S, Duee E, Galleni M, Duez C, Frère JM, Dideberg O. The 3D structure of a zinc metallo-β-lactamase from *Bacillus cereus* reveals a new type of protein fold. EMBO J 1995;14:4914-4921.

4. Schilling O, Rüggeberg S, Vogel A, Rittner N, Weichert S, Schmidt S, Doig S, Franz T, Benes V, Andrews SC, Baum M, Meyer-Klaucke W. Characterization of an *Escherichia coli elaC* deletion mutant. Biochem Biophys Res Commun 2004;320:1365-1373.

5. Schiffer S, Rösch S, Marchfelder A. Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. EMBO J 2002;21:2769-2777.

6. Vogel A, Schilling O, Niecke M, Bettmer J, Meyer-Klaucke W. *ElaC* encodes a novel binuclear zinc phosphodiesterase. J Biol Chem 2002;277:29078-29085.

7. Meitz JC, Edwards SM, Easton DF, Murkin A, Ardern-Jones A, Jackson RA, Williams S, Dearnaley DP, Stratton MR, Houlston RS, Eeles RA. *HPC2/ELAC2* polymorphisms and prostate cancer risk: analysis by age of onset of disease. Br J Cancer 2002;87:905-908.

8. Severi G, Giles GG, Southey MC, Tesoriero A, Tilley W, Neufing P, Morris H, English DR, McCredie MR, Boyle P, Hopper JL. *ELAC2/HPC2* polymorphisms, prostate-specific antigen levels, and prostate cancer. J Natl Cancer Inst 2003;95:818-824.

9. Takaku H, Minagawa A, Takagi M, Nashimoto M. A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. Nucleic Acids Res 2003;31:2272-2278.

10. Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, Carillo AR, Chen Y, Dayananth P, Desrochers M, Dumont M, Farnham JM, Frank D, Frye C, Ghaffari S, Gupte JS, Hu R, Iliev D, Janecki T, Kort EN, Laity KE, Leavitt A, Leblanc G, McArthur-Morrison J, Pederson A, Penn B, Peterson KT, Reid JE, Richards S, Schroeder M, Smith R, Snyder SC, Swedlund B, Swensen J, Thomas A, Tranchant M, Woodland AM, Labrie F, Skolnick MH, Neuhausen S, Rommens J, Cannon-Albright LA. A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 2001;27:172-180.

11. Yanaihara N, Kohno T, Takakura S, Takei K, Otsuka A, Sunaga N, Takahashi M, Yamazaki M, Tashiro H, Fukuzumi Y, Fujimori Y, Hagiwara K, Tanaka T, Yokota J. Physical and transcriptional map of a 311-kb segment of chromosome 18q21, a candidate lung tumor suppressor locus. Genomics 2001;72:169-179.

12. Chanfreau G, Noble SM, Guthrie C. Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science 1996;274:1511-1514.

13. Noordzij JG, Verkaik NS, Van der Burg M, van Veelen LR, de Bruin-Versteeg S, Wiegant W, Vossen JM, Weemaes CM, de Groot R, Zdzienicka MZ, van Gent DC, van Dongen JJ. 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 2003;101:1446-1452.

14. Jeggo P, O'Neill P. The greek goddess, Artemis, reveals the secrets of her cleavage. DNA Repair 2002;1:771-777.

15. Schlissel MS. Does Artemis end the hunt for the hairpin-opening activity in V(D)J recombination? Cell 2002;109:1-4.

16. Ishiai M, Kimura M, Namikoshi K, Yamazoe M, Yamamoto K, Arakawa H, Agematsu K, Matsushita N, Takeda S, Buerstedde J-M, Takata M. DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. Mol Cell Biol 2004;24:10733-10741.

17. Brendel M, Bonatto D, Strauss M, Revers LF, Pungartnik C, Saffi J, Henriques JAP. Role of *PSO* genes in repair of DNA damage of *Saccharomyces cerevisiae*. Mutat Res 2003;544:179-193.

18. Dronkert ML, de Wit J, Boeve M, Vasconcelos ML, van Steeg HT, Tan L, Hoeijmakers JH, Kanaar R. Disruption of mouse *SNM1* causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. Mol Cell Biol 2000;20:4553-4561.

19. Akhter S, Richie CT, Deng JM, Brey E, Zhang X, Patrick C, Behringer RR, Legerski RJ. Deficiency in *SNM1* abolishes an early mitotic checkpoint induced by spindle stress. Mol Cell Biol 2004;24:10448-10455.

20. Demuth I, Digweed M, Concannon P. Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation. Oncogene 2004;23:8611-8618.

21. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-3402.

22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673-4680.

23. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999;41:95-98.

24. Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 2001;17:1244-1245.

25. Gaboriaud C, Bissery V, Benchetrit T, Mornon J-P. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett 1987;224:149-155.

26. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon J-P. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci 1997;53:621-645.

27. Lemesle-Varloot L, Henrissat B, Gaboriaud C, Bissery V, Morgat A, Mornon J-P. Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D representation of protein sequences. Biochimie 1990;72:555-574.

28. Nashimoto M. Conversion of mammalian tRNA 3’ processing endoribonuclease to four-base-recognizing RNA cutters. Nucleic Acids Res 1995;23:3642-3647.

29. Nashimoto M. Distribution of both lengths and 5’ terminal nucleotides of mammalian pre-tRNA 3’ trailers reflects properties of 3’ processing endoribonuclease. Nucleic Acids Res 1997;25:1148-1155.

30. Dubrovsky EB, Dubrovskaya VA, Levinger L, Schiffer S, Marchfelder A. *Drosophila* RNase Z processes mitochondrial and nuclear pre-tRNA 3' ends *in vivo.* Nucleic Acids Res 2004;32:255-262.
31. Schiffer S, Helm M, Théobald-Dietrich A, Giegé R, Marchfelder A. The plant tRNA 3' processing enzyme has a broad substrate spectrum. Biochemistry 2001;40:8264-8272.
32. Korver W, Guevara C, Chen Y, Neuteboom S, Bookstein R, Tavtigian S, Lees E. The product of the candidate prostate cancer susceptibility gene *ELAC2* interacts with the γ-tubulin complex. Int J Cancer 2003;104:283-288.
33. Jenny A, Minvielle-Sebastia L, Preker PJ, Kellerdagger W. Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. Science 1996;274:1514-1517.
34. Beelman CA, Parker R. Degradation of mRNA in eukaryotes. Cell 1995;81:179-183.
35. Eckner R, Ellmeier W, Birnstiel ML. Mature mRNA 3' end formation stimulates RNA export from the nucleus. EMBO J 1991;10:3513-3522.
36. Huang Y, Carmichael GG. Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol Cell Biol 1996;16:1534-1542.
37. Wickens M, Anderson P, Jackson RJ. Life and death in the cytoplasm: messages from the 3' end. Curr Opin Genet Dev 1997;7:220-232.
38. Wahle E, Rüegsegger U. 3'-end processing of pre-mRNA in eukaryotes. FEMS Microbiol Rev 1999;23:277-295.

39. Zhao J, Hyman L, Moore C. Formation of mRNA 3’ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev 1999;63:405-445.

40. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA 2004;10:565-573.

41. Xu R, Ye X, Li QQ. *AtCPSF73-II* gene encoding an *Arabidopsis* homolog of CPSF 73 kDa subunit is critical for early embryo development. Gene 2004;324:35-45.

42. Bonatto D, Brendel M, Henriques, JAP. A new group of plant-specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis, and 3-dimensional modeling. Funct Plant Biol 2005;32:1-4.

43. Molinier J, Stamm M-E, Hohn B. SNM-dependent recombinational repair of oxidatively induced DNA damage in *Arabidopsis thaliana.* EMBO Rep 2004;5:1-6.

44. Dronkert MLG, de Wit J, Boeve M, Vasconcelos ML, van Steeg H, Tan TL, Hoeijmakers JHJ, Kanaar R. Disruption of mouse *SNM1* causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. Mol Cell Biol 2000;20:4553-4561.

45. Richie CT, Peterson C, Lu T, Hittelman WN, Carpenter PB, Legerski RJ. hSnm1 colocalizes and physically associates with 53BP1 before and after DNA damage. Mol Cell Biol 2002;22:8635-8647.

46. Henriques JA, Moustacchi E. Isolation and characterization of *pso* mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae*. Genetics 1980;95:273-288.

47. Magaña-Schwencke N, Henriques JAP, Chanet R, Moustacchi E. The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA:

comparison of wild-type and repair-deficient strains. Proc Natl Acad Sci USA 1982;79:1722-1726.

48. Cassier C, Chanet R, Henriques JA, Moustacchi E. The effects of three *PSO* genes on induced mutagenesis: a novel class of mutationally defective yeast. Genetics 1980;96:841-857.

49. Ruhland A, Haase E, Siede W, Brendel M. Isolation of yeast mutants sensitive to the bifunctional alkylating agent nitrogen mustard. Mol Gen Genet 1981;181:346-351.

50. Wilborn F, Brendel M 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. Curr Genet 1989;16:331-338.

51. Da Silva KVCL, Henriques JAP. Induction of the cytoplasmic 'petite' mutation in *pso* mutants of *Saccharomyces cerevisiae* by photoaddition of furocoumarins or by ultraviolet radiation. Mutagenesis 1987;2:187-197.

52. Brendel M, Henriques JA. The *pso* mutants of *Saccharomyces cerevisiae* comprise two groups: one deficient in DNA repair and another with altered mutagen metabolism. Mutat Res 2001;489:79-96.

53. Yu J, Marshall K, Yamaguchi M, Haber JE, Weil CF. Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae.* Mol Cell Biol 2004;24:1351-1364.

54. Li X, Hejna J, Moses RE. The yeast Snm1 protein is a DNA 5'-exonuclease. DNA Repair 2004;4:163-170.

55. Pannicke U, Ma Y, Hopfner K-P, Niewolik D, Lieber MR, Schwarz K. Functional and biochemical dissection of the structure-specific nuclease ARTEMIS. EMBO J 2004;23:1987-1997.

56. Lee PT, Hsu AY, Ha HT, Clarke CF. A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: isolation and identification of the *Escherichia coli ubiE* gene. J Bacteriol 1997;179:1748-1754.

57. Barkovich RJ, Shtanko A, Shepherd JA, Lee PT, Myles DC, Tzagaloff A, Clarke C. F. Characterization of the *COQ5* gene from *Saccharomyces cerevisiae*. Evidence for a C-methyltransferase in ubiquinone biosynthesis. J Biol Chem 1997;272:9182-9188.

58. Brandt U, Trumpower B. The protonmotive Q cycle in mitochondria and bacteria. Crit Rev Biochem Mol Biol 1994;29:165-197.

59. Bowry VW, Mohr D, Cleary J, Stocker R. Prevention of tocopherol-mediated peroxidation in ubiquinol-10-free human low density lipoprotein. J Biol Chem 1995;270:5756-5763.

60. Do TQ, Schultz JR, Clarke CF. Enhanced sensitivity of ubiquinone-deficient mutants of *Saccharomyces cerevisiae* to products of autoxidized polyunsaturated fatty acids. Proc Natl Acad Sci U S A 1996;93:7534-7539.

61. Kufel J Allmang C, Petfalski E, Beggs J, Tollervey D. Lsm proteins are required for normal processing and stability of ribosomal RNAs. J Biol Chem 2003;278:2147-2156.

62. Degtyarenko KN, Archakov AI. Molecular evolution of P450 superfamily and monooxygenase systems. FEBS Lett 1993;332:1-8.

63. Sweeney M.J, Dobson ADW. Mycotoxin production by *Aspergillus, Fusarium* and *Penicillium* species. Int J Food Microbiol 1998;43:141-158.

64. Wang J-S, Groopman JD. DNA damage by mycotoxins. Mutat Res 1999;424:167-181.

65. Calvo AM, Wilson RA, Bok JW, Keller NP. Relationship between secondary metabolism and fungal development. Microbiol Mol Biol Rev 2002;66:447-459.

66. Escalante AA, Ayala FJ. Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. Proc Natl Acad Sci U S A 1995;92:5793-5797.

67. Maréchal E, Cesbron-Delauw M-F. The apicoplast: a new member of the plastid family. Trends Plant Sci 2001;6:200-205.

68. Bhattacharyya MK, Norris DE, Kumar N. Molecular players of homologous recombination in protozoan parasites: implications for generating antigenic variation. Infect Genet Evol 2004;4:91-98.

TABLES

Table 1. Pso2p/Snm1p A Group proteins

Subgroup	Species	Protein name	Accession number
Pso2p PSA			
	Arabidopsis thaliana	*Ath*Lig6p	gi|12597768|
	Brassica napus	*Bna*Lig6p	gi|7657880|
	Brassica rapa	*Bra*Lig6p	gi|30314605|
	Chlamydomonas reinhardtii	*Cre*Lig6p	scaffold 704|1|38579
	Medicago truncatula	*Mtr*Lig6p	gi|38257195|
	Oryza sativa	*Osa*Lig6p	gi|20805031|
	Arabidopsis thaliana	*Ath*Pso2p PSA	gi|15225548|
	Oryza sativa	*Osa*Pso2p PSA	gi|51091393|
	Petunia hybrida	*Phy*Pso2p PSA	SGN-U210424
Pso2p MEA			
	Anopheles gambiae	*Aga*Pso2p MEA	gi|31205277|
	Arabidopsis thaliana	*Ath*Pso2p MEA	gi|15231597|
	Drosophila melanogaster	*Dme*Pso2p MEA	gi|21357063|
	Homo sapiens	*Hsa*Pso2p MEA	gi|23618907|
	Mus musculus	*Mmu*Pso2p MEA	gi|9055350|
	Oryza sativa	*Osa*Pso2p MEA	gi|21740626|
	Pan troglodytes	*Ptr*Pso2p MEA	gi|55634570|
	Takifugu rubripes	*Tru*Pso2p MEA	scaffold_384
Pso2p PA			
	Dictyostelium discoideum	*Ddi*Pso2p PA	gi|28828251|
	Entamoeba histolytica	*Ehi*Pso2p PA	gi|56469541|
Pso2p FA			
	Ashbya gossypii	*Ago*Pso2p FA	gi|54299553|
	Aspergillus nidulans	*Ani*Pso2p FA	gi|50257057|
	Candida albicans	*Cal*Pso2p FA	gi|46443342|
	Coprinopsis cinerea	*Cci*Pso2p FA	gi|33347022|
	Cryptococcus neoformans	*Cne*Pso2p FA	gi|50257057|
	Debaryomyces hansenii	*Dha*Pso2p FA	gi|50418215|
	Fusarium virguliforme	*Fvi*Pso2p FA	gi|38262419|
	Gibberella zeae	*Gze*Pso2p FA	gi|49243354|
	Kluyveromyces lactis	*Kla*Pso2p FA	gi|50311643|
	Kluyveromyces waltii	*Kwa*Pso2p FA	gi|45444537|
	Magnaporthe grisea	*Mgr*Pso2p FA	gi|38101788|
	Neurospora crassa	*Ncr*Pso2p FA	gi|32423615|
	Saccharomyces bayanus	*Sba*Pso2p FA	gi|29365317|
	Saccharomyces castellii	*Sca*Pso2p FA	gi|30988063|
	Saccharomyces cerevisiae	*Sce*Pso2p FA	gi|6323786|
	Saccharomyces mikatae	*Smi*Pso2p FA	gi|29363588|
	Schizosaccharomyces pombe	*Spo*Pso2p FA	gi|19113674|
	Ustilago maydis	*Uma*Pso2p FA	gi|49083214|
	Yarrowia lipolytica	*Yli*Pso2p FA	gi|50550517|

Table 2. Pso2p/Snm1p B and Plasmodium Group

Group	**Subgroup**	***Species***	**Protein name**	**Accession number**
Pso2p/Snm1p B group				
	Pso2p LB			
		Leishmania major	*Lma*Pso2p LB1	gi|11061630|
		Leishmania major	*Lma*Pso2p LB2	gnl|TIGR 5693|
	Pso2p PSB			
		Arabidopsis thaliana	*Ath*Pso2p PSB	gi|25402874|
		Lycopersicon esculentum	*Les*Pso2p PSB	SGN-U236949
		Oryza sativa	*Osa*Pso2p PSB	gi|20161430|
	Pso2p MEB			
		Arabidopsis thaliana	*Ath*Pso2p MEB	gi|20148601|
		Homo sapiens	*Hsa*Pso2p MEB	gi|12383082|
		Lotus corniculatus	*Lco*Pso2p MEB	gi|29122758|
		Mus musculus	*Mmu*Pso2p MEB1	gi|26338211|
		Mus musculus	*Mmu*Pso2p MEB2	gi|26339774|
		Oryza sativa	*Osa*Pso2p MEB	gi|34910110|
		Pan troglodytes	*Ptr*Pso2p MEB	gi|55587855|
		Rattus norvegicus	*Rno*Pso2p MEB	gi|27660542|
		Solanum tuberosum	*Stu*Pso2p MEB	SGN-U262534
		Tetraodon nigroviridis	*Tni*Pso2p MEB	gi|47228753|
Pso2p *Plasmodium* Group				
		Plasmodium chabaudi	*Pch*Pso2p PPG	gi|56520598|
		Plasmodium falciparum	*Pfa*Pso2p PPG	gi|23509933|
		Plasmodium yoelii	*Pyo*Pso2p PPG	gi|23490291|

Table 3. Artemis/Artemis-like, Elac2 and CPSF 73 proteins

Group	Species	Protein name	Accession number
Artemis/Artemis-like			
	Anopheles gambiae	*Aga*Artemis-like	gi|31203062|
	Apis mellifera	*Ame*Artemis-like	gi|48096919|
	Candida albicans	*Cal*Artemis-like	gi|46438865|
	Canis familiaris	*Cfa*Artemis	gi|50205195|
	Ciona intestinalis	*Cin*Artemis-like	gi|23587642|
	Cryptococcus neoformans	*Cne*Artemis-like	gi|50255259|
	Debaryomyces hansenii	*Dha*Artemis-like	gi|50424115|
	Gallus gallus	*Gga*Artemis	gi|50764059|
	Gibberella zeae	*Gze*Artemis-like	gi|42551205|
	Homo sapiens	*Hsa*Artemis	gi|13872809|
	Mus musculus	*Mmu*Artemis	gi|22023549|
	Pongo pygmaeus	*Ppy*Artemis	gi|55731498|
	Rattus norvegicus	*Rno*Artemis	gi|22023557|
	Tetraodon nigroviridis	*Tni*Artemis-like	gi|47218348|
	Yarrowia lipolytica	*Yli*Artemis-like	gi|49650441|
CPSF 73			
	Bos taurus	*Bta*CPSF 73 kDa	gi|1707412|
	Homo sapiens	*Hsa*CPSF 73 kDa	gi|7706427|
	Mus musculus	*Mmu*CPSF 73 kDa	gi|9055194|
	Saccharomyces kudriavzevii	*Sku*Ysh1p	gi|6323307|
	Saccharomyces cerevisiae	*Sce*Ysh1p	gi|6323307|
	Schizosaccharomyces pombe	*Spo*Ysh1p	gi|19112240|
Elac2p			
	Caenorhabditis elegans	*Cel*Elac2p	gi|42559555|
	Gorilla gorilla	*Ggo*Elac2p	gi|10946489|
	Homo sapiens	*Hsa*Elac2p	gi|10880933|
	Macaca fascicularis	*Mfa*Elac2p	gi|26000220|
	Mus musculus	*Mmu*Elac2p	gi|13540343|
	Saccharomyces cerevisiae	*Sce*Trz1p	gi|6322932|
	Saccharomyces kluyveri	*Skl*Trz1p	gi|30987496|
	Schizosaccharomyces pombe	*Spo*Trz1p	gi|19112306|

FIGURE LEGENDS

Fig.1. Topological view of an unrooted phylogenetic tree obtained from the global alignment of all proteins listed in Tables 1 to 3. Both Elac2 and CPSF 73 kDa proteins compose two groups phylogenetically related to RNA metabolism, as indicated by the external dashed bar. The Pso2p/Snm1p family, related to eukaryotic DNA metabolism (external solid line) is formed by four groups (each group indicated by different gray shades): the PPG sequences (showed by a triangle), the Artemis/Artemis-like sequences (losangle), the Pso2p B group (square) and the Pso2p A group (circle). The unrooted tree is supported by an internal branch analysis value of > 90%.

Fig. 2. Topological view of an unrooted phylogenetic tree of Pso2p A group. Numbers at nodes indicate internal branch analysis values (1,000 replicates). The Pso2p PSA, MEA, PA and FA subgroups are indicated in the figure. Symbol: LUCA (last universal common ancestor).

Fig. 3. Hydrophobic cluster analysis (HCA) of selected Pso2p A group proteins. Conserved hydrophobic aa residues are shaded in gray, while conserved hydrophilic aa residues are circled. The boundaries of the metallo-β-lactamase and β-CASP motifs as well as the conserved regions I to IV (CRI to CRIV) of Pso2p A sequences are shown in the figure. The way to read the sequences and special symbols are indicated in the inset.

Fig. 4. Schematic representation (A) of *Ani*Pso2p FA sequence and its four conserved domains (UbiE, Snp3p, Pso2p A and P450). The direction of protein, from N-terminus to C-terminus, is indicated by an arrow. In (B), HCA of the four domains identified in *Ani*Pso2p FA sequence is indicated. The GenBank accession number of the sequences used

for HCA is: *Sce*Pso2p (gi|6323786|), *Sce*Snp3p (gi|536027|), *Eco*UbiE (gi|17380538|) and *Ani*bphA (gi|117178|).

Fig. 5. Topological view (A) of an unrooted phylogenetic tree of Pso2p B group. Numbers at nodes indicate internal branch analysis values (1,000 replicates). The Pso2p LB, PSB and MEB subgroups are indicated. In (B), hydrophobic cluster analysis (HCA) of selected Pso2p B group proteins are shown. Conserved hydrophobic aa residues are shaded in gray, while conserved hydrophilic aa residues are circled. The boundaries of the metallo-β-lactamase and β-CASP motifs are shown by a dashed line. Conserved regions I to IV (CRI to CRIV) of Pso2p B sequences as well as the consensus sequence derived from HCA alignment are indicated in the figure. The way to read the sequences and special symbols of HCA are indicated in the inset.

Fig. 6. Topological view (A) of an unrooted phylogenetic tree of Pso2p PPG group. Numbers at nodes indicate internal branch analysis values (1,000 replicates). In (B), hydrophobic cluster analysis (HCA) of *Pfa*Pso2p PPG and *Pyo*Pso2p PPG proteins. The four conserved regions (CRI to CRIV) were highlighted by an empty box in both sequences. Moreover, a separated box shows the HCA of the 300 (*Pfa*Pso2p PPG) and 250 aa stretch (*Pyo*Pso2p PPG). An empty box also highlighted the conserved secondary structures within this stretch. A hinge region is indicated in both sequences. The boundaries of the metallo-β-lactamase and β-CASP motifs are shown by a dashed line. The way to read the sequences and special symbols are indicated in the inset.

Figure 1.

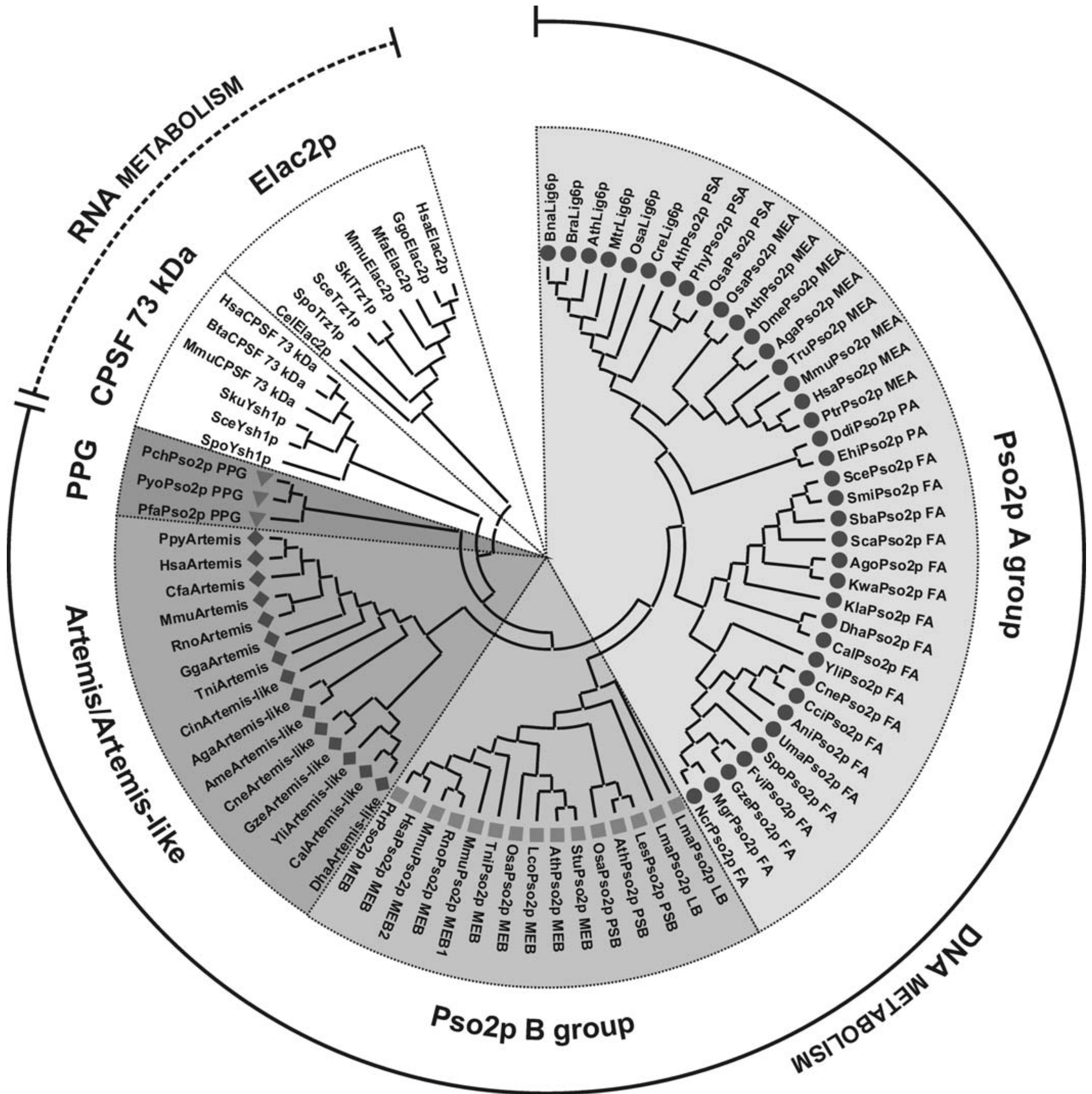


Figure 2.

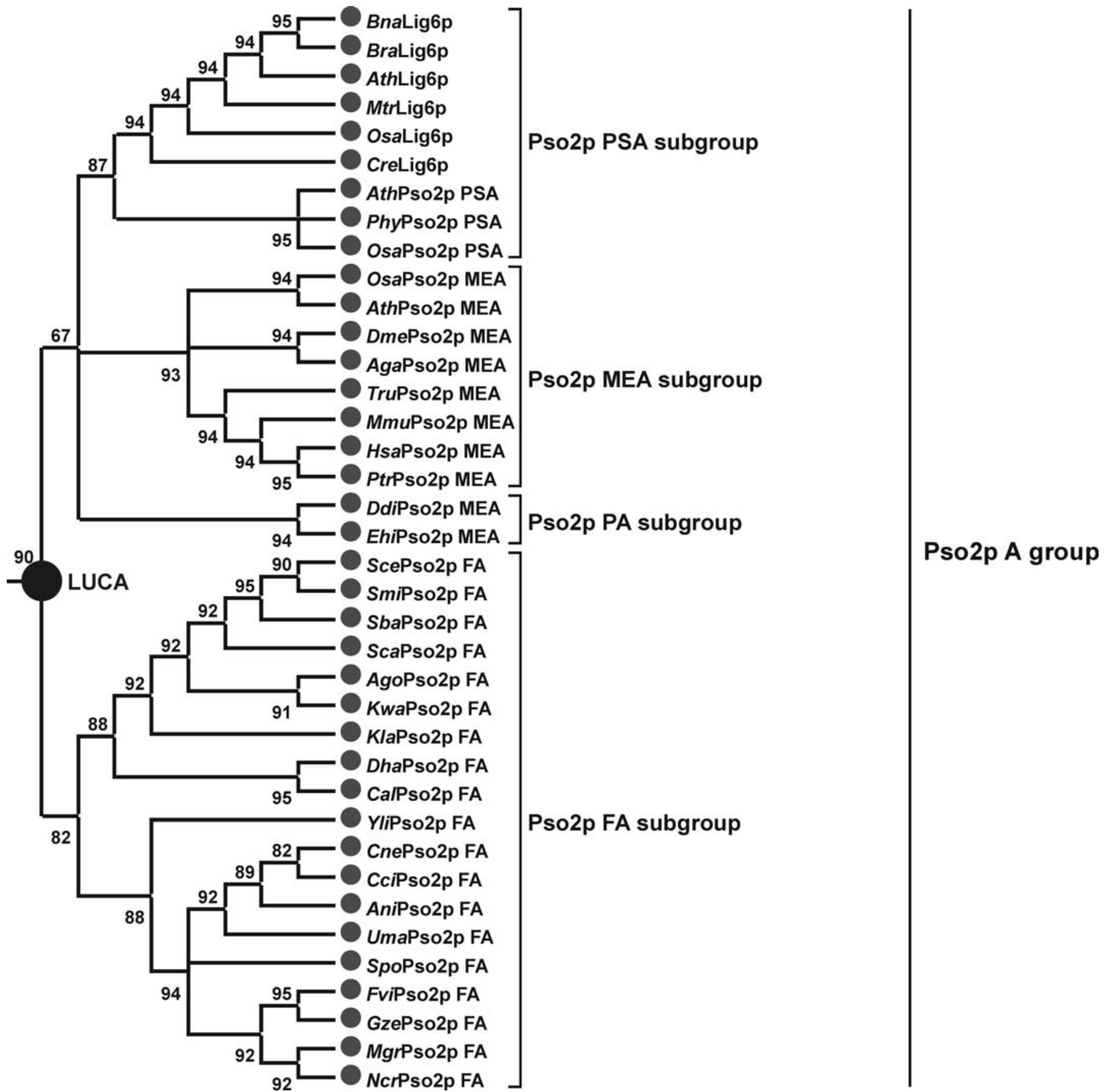


Figure 3.

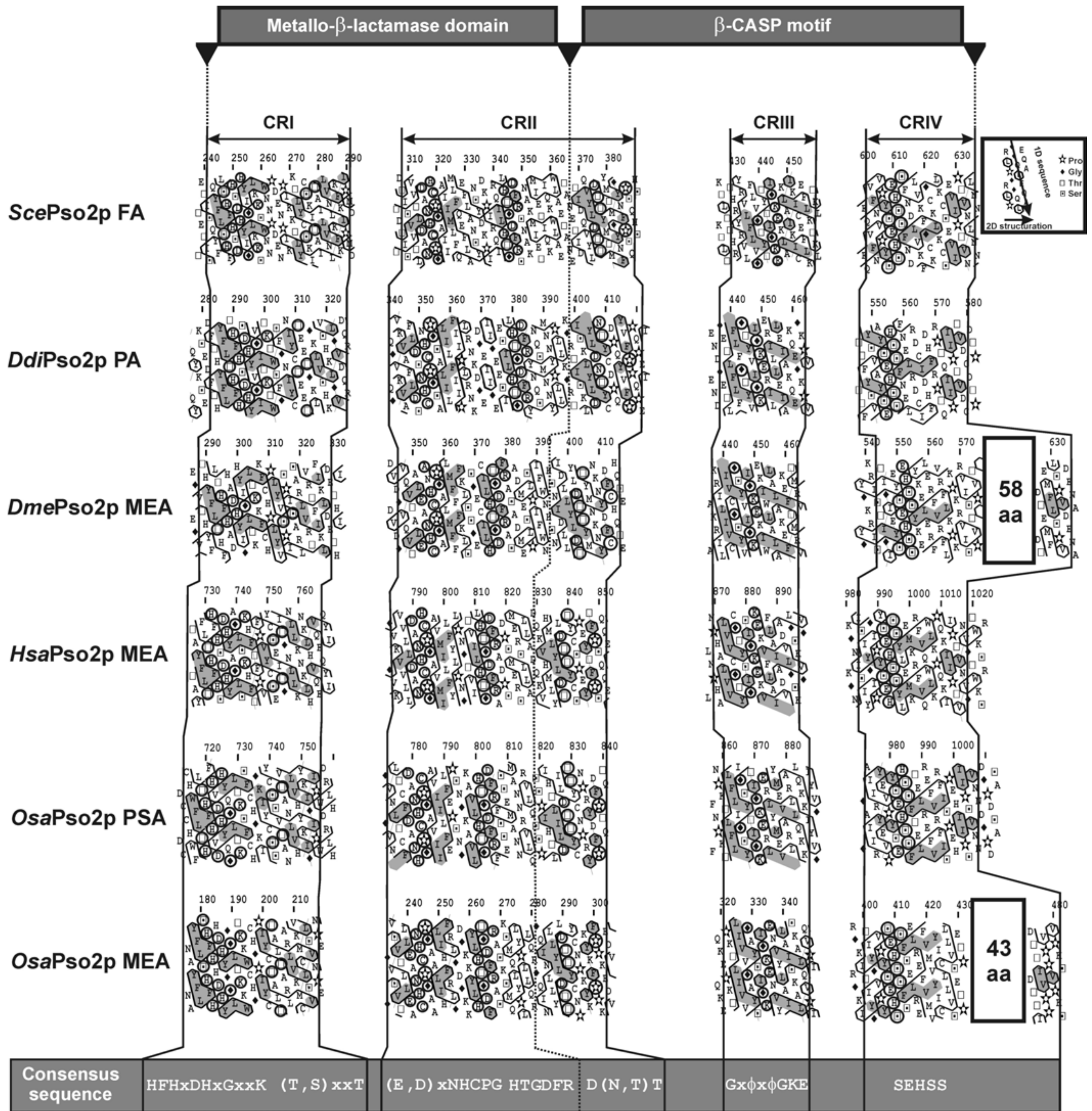
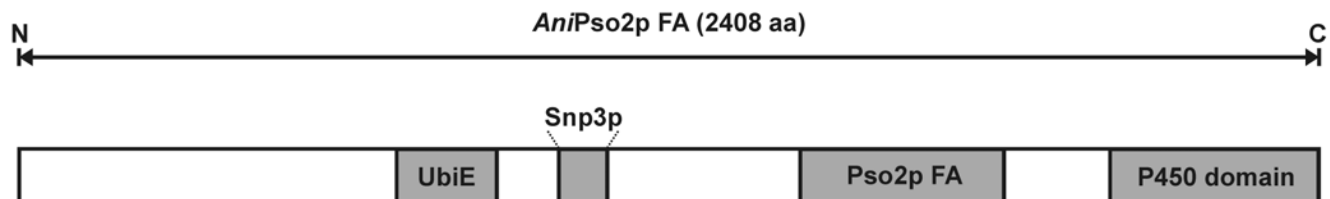


Figure 4.



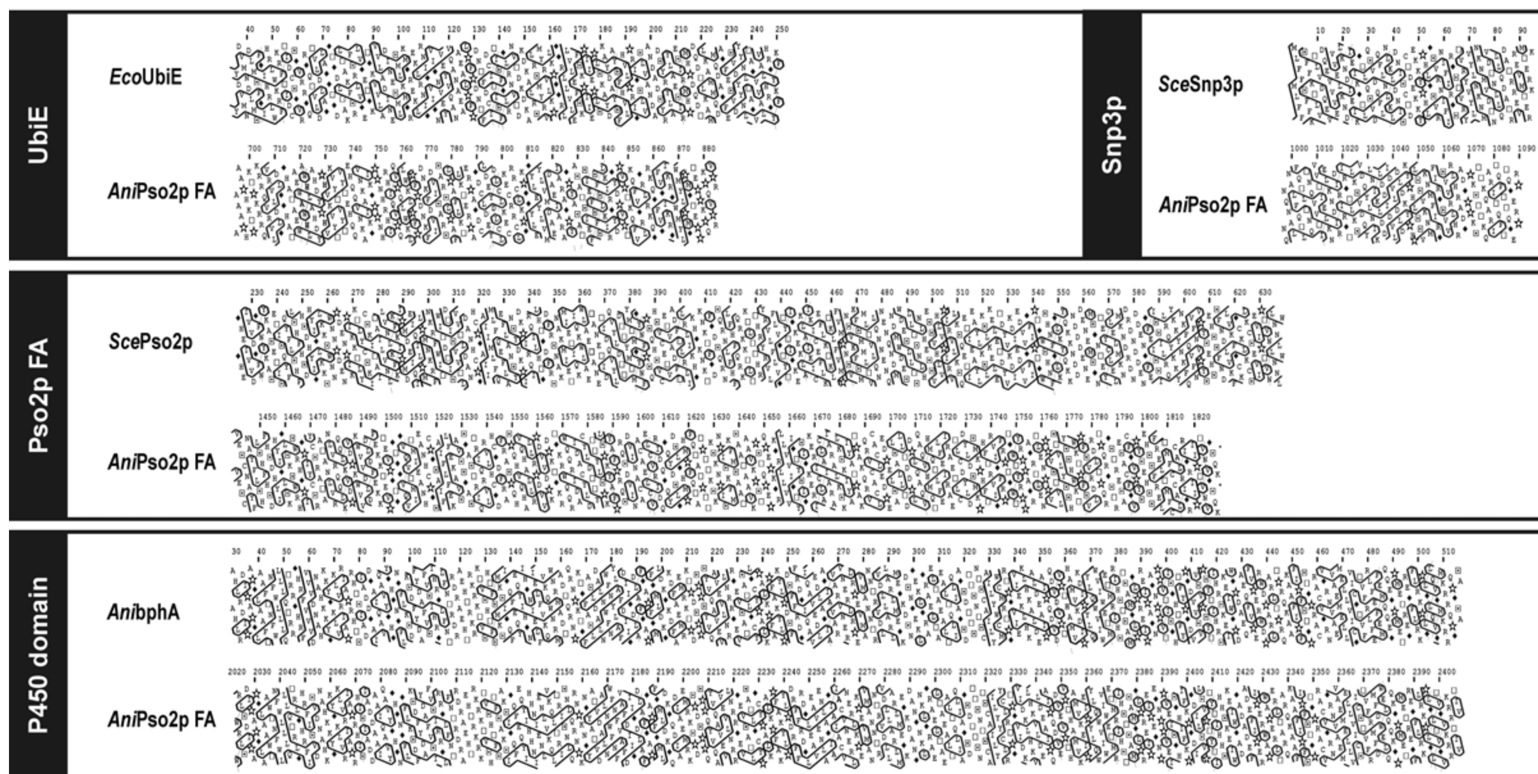
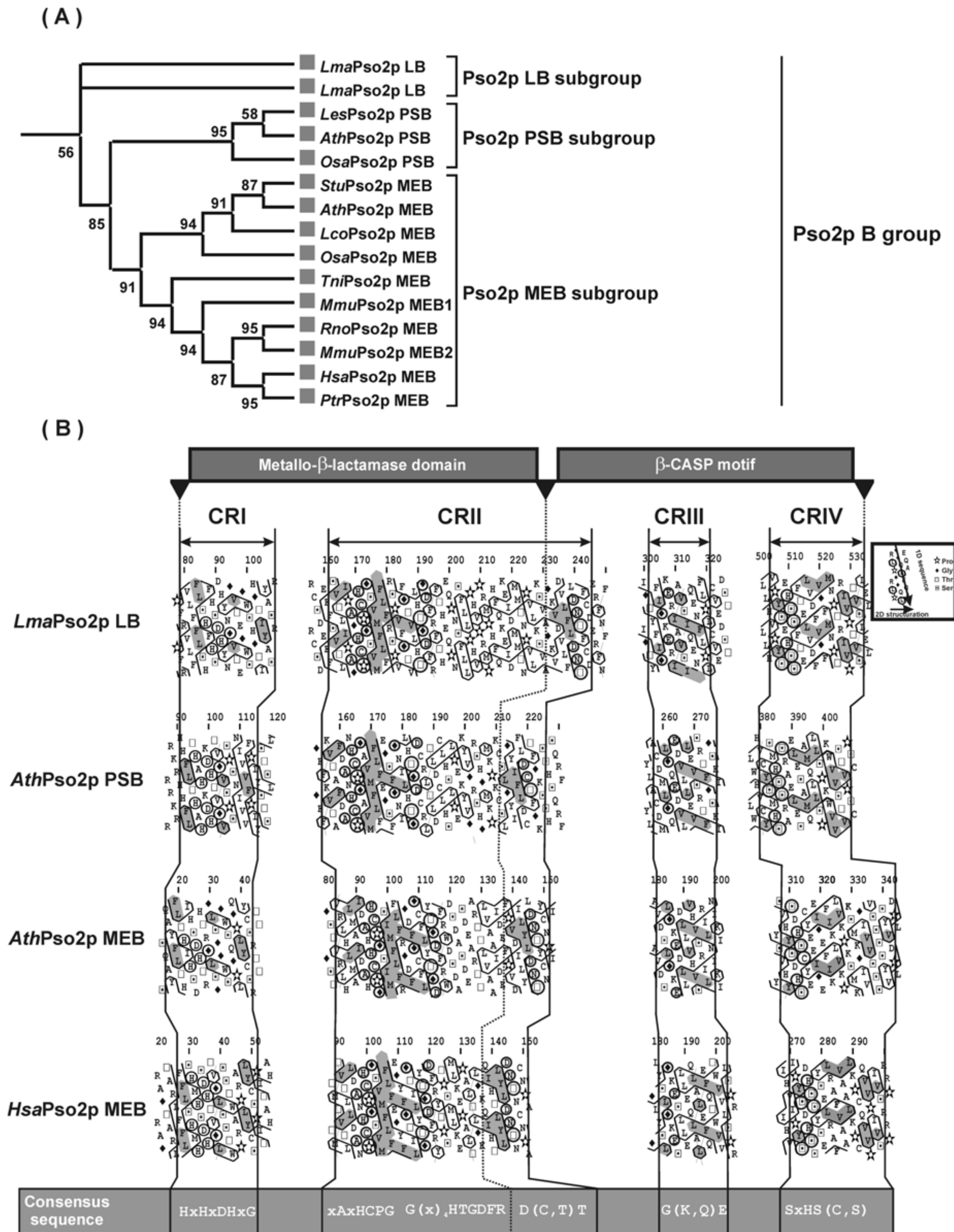
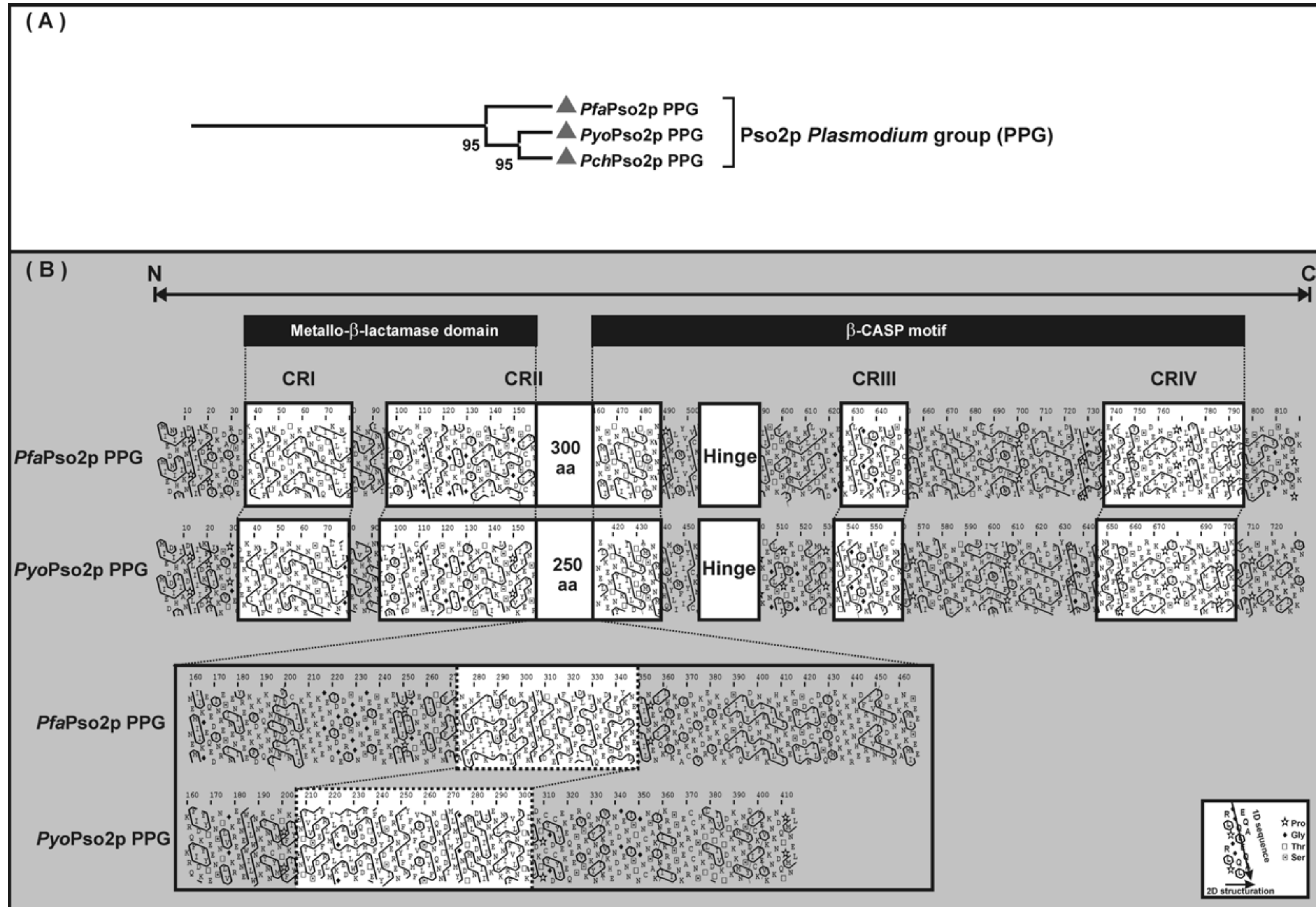


Figure 5.



95

Figure 6.



3

Capítulo Três

IN SILICO IDENTIFICATION AND ANALYSIS OF NEW ARTEMIS/ARTEMIS-LIKE SEQUENCES FROM FUNGAL AND METAZOAN SPECIES

The Protein Journal **(submetido para publicação)**

In silico Identification and Analysis of New Artemis/Artemis-Like Sequences from Fungal and Metazoan Species

Diego Bonatto[1], Martin Brendel[1,2] and João Antonio Pêgas Henriques[1,3,4 *]

[1]Departamento de Biofísica/Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, 91507-970 Porto Alegre, RS, Brasil.

[2]Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz (UESC), Ilhéus, BA, Brasil.

[3]Curso de Farmácia, Universidade Luterana do Brasil (ULBRA), Canoas, RS, Brasil.

[4]Instituto de Biotecnologia, Universidade de Caxias do Sul (UCS), Caxias do Sul, RS, Brasil.

Short title: New Artemis/Artemis-Like Sequences

[*]Address to which proofs should be sent:

João Antonio Pêgas Henriques

Departamento de Biofísica/Centro de Biotecnologia, UFRGS,

Av. Bento Gonçalves 9500, Porto Alegre, RS, Brasil, 91507-970.

Telephone: +55-51-3316-7602. Fax: +55-51-3316-6084.

E-mail: pegas@cbiot.ufrgs.br

Contract/grant sponsor: FAPERGS, CAPES, GENOTOX laboratory.

Abbreviations: aa, aminoacids; ATM, *ataxia telangiectasia* mutated protein; ATR, ATM-/Rad3-related proteins; CAII, carbonic anhydrase II, CR, conserved region; DNA-PKcs, DNA-dependent protein kinase catalytic subunit; DSBs, double-strand breaks; HCA, hydrophobic cluster analysis; HR, homologous recombination; ICL, interstrand cross links; IR, ionizing radiation; LUCA, last universal common ancestor; ME, minimum evolution; NHEJ, non-homologous end joining recombination; SSBs, single-strand breaks.

ABSTRACT

The Artemis Group comprises mammalian proteins with important functions in the repair of ionizing radiation-induced DNA double-strand breaks and in the cleavage of DNA hairpin extremities generated during V(D)J recombination. Little is known about the presence of Artemis/Artemis-like proteins in non-mammalian species. We have characterized new Artemis/Artemis-like sequences from the genomes of some unconventional fungi and from non-mammalian metazoan species. An in-depth phylogenetic analysis of these new Artemis/Artemis-like sequences showed that they form a distinct clade within the Pso2p/Snm1p A and B Groups. Hydrophobic cluster analysis and three-dimensional modeling allowed to map and to compare conserved regions in these Artemis/Artemis-like proteins. The results indicates that Artemis probably belongs to an ancient DNA recombination mechanism that diversified with the evolution of multi-cellular eukaryotic lineage.

KEY WORDS: Artemis/Artemis-like proteins, Pso2p/Snm1p, DNA recombination, hydrophobic cluster analysis, protein phylogeny

1. INTRODUCTION

Eukaryotic chromatin is a relatively easy target for reactive chemical and physical agents, including cross-linking substances and ionizing radiation, respectively. Both DNA and the nucleoproteins that compose chromatin can be irreversibly modified by these agents, resulting in chromosomal rearrangements, deletions and other genetic alterations (Lees-Miller and Meek, 2003). As chromosomal DNA contains most of an organism's genetic information, modifications introduced in this molecule are potentially lethal if not repaired. Amongst all of these DNA lesions the double strand breaks (DSB) are the most dangerous lesions (Jackson, 2002; Lees-Miller and Meek, 2003). Interestingly, the generation of DSB in genomic DNA is a common process in eukaryotic cells, occurring during certain stages of the life cycle, e.g. in meiosis or in DNA re-arrangements for antibody production in B cells (Kuzminov, 2001; Lee *et al.,* 2004; Pâques and Haber, 1999). During evolution, eukaryotic cells have developed a complex network of proteins that, by sensing all types of DNA-damage and inducing the appropriate response, maintain the genome's integrity. This network can be sub-divided into different DNA repair pathways, each controlled by cell cycle, damage types and substrate requirements (Bernstein *et al.,* 2002; Cline and Hanawalt, 2003). DSBs are primarily repaired by homologous recombination (HR) and/or by non-homologous end joining recombination (NHEJ). In the case of HR, the presence of a DSB elicits a genomic search for similar (homologous) sequences and the repair involves base pairing of long stretches of matched base pairs (Aylon and Kupiec, 2004). In contrast, NHEJ is a mechanism able to join DNA ends with no, or minimal, homology (Aylon and Kupiec, 2004). In addition, NHEJ is also used to repair DSBs that arise during early lymphocyte development in the context of V(D)J recombination (Lieber *et al.,* 2003; Rooney *et al.,* 2003). The NHEJ pathway contains six protein members namely Ku70, Ku80, XRCC4, DNA ligase 4 (Lig4), DNA-dependent protein kinase catalytic subunit (DNA-PKcs),

and Artemis (Rooney *et al.,* 2003). Many proteins that participate in NHEJ or V(D)J recombination share a high homology, from yeasts to plants and animals, indicating the essentiality of this mechanism for cellular well-being (Lees-Miller and Meek, 2003).

Artemis is a group of proteins that belongs to the β-CASP family, a member of the metallo-β-lactamase superfamily (Callebaut *et al.,* 2002). Artemis has 5' to 3' exonucleolytic activity with single-strand DNA specificity and, when associated with DNA-PKcs, forms a phosphorylated complex with endonucleolytic activity on both 5' and 3' DNA overhangs. Furthermore, it can cleave hairpins generated by the Rag-1/Rag-2 proteins in V(D)J recombination (Jeggo and O'Neill, 2002; Schlissel, 2002). Artemis cooperates with p53 to suppress chromosomal translocation and tumor development in mice and, therefore, can be considered a tumor suppressor. Like other NHEJ/p53 doubly-deficient mice, most Artemis-deficient mice succumb to pro-B cell lymphomas at the age of 11–12 weeks (Richardson and Jasin, 2000). Moreover, Artemis interacts with the checkpoint kinase *ataxia telangiectasia* mutated protein (ATM) and ATM-/Rad3-related proteins (ATR) after exposure of cells to ionizing radiation (IR) or UV irradiation, respectively (Zhang *et al.,* 2004). These findings indicate that Artemis is required for the maintenance of a normal DNA damage-induced G2/M cell cycle arrest (Zhang *et al.,* 2004). However, despite the data obtained with mammalian cells on Artemis, little is known about how and when Artemis protein is recruited for DNA repair. Due to intrinsic difficulties in constructing mammalian cell lines with more than one knockout or knockdown gene, an alternative biological model allowing the study of Artemis in DNA repair would be welcome. Yeasts, especially the conventional species *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, have many advantages as model organisms when compared to plants or metazoans (Aylon and Kupiec, 2004). A large number of yeast mutant strains for many metabolic pathways and cellular components can be easily isolated, using a combination of sophisticated genetic and

biochemical analyses. Also, yeast cells can grow rapidly in defined or complete culture media, their cell cycle can be synchronized, and many mutant strains can be tested for different phenotypes at the same time (Aylon and Kupiec, 2004).

An Artemis-like protein has not been discovered in conventional yeast species until now. But fungi, plants and metazoans contain an Artemis orthologue protein known as Pso2p/Snm1p. The family of Pso2p/Snm1p is divided in two groups: A and B, both associated with the recombinational repair of DSBs induced by chemical agents (Brendel *et al.,* 2003; Demuth *et al.,* 2004; Molinier *et al.,* 2004). Artemis and Pso2p/Snm1p have low aa sequence homology (Moshous *et al.,* 2001), indicating that both proteins possibly have different functions in DNA repair in metazoan cells (Bonatto *et al.,* 2005).

In this work, we have identified and characterized new members of the Artemis protein family, by searching in eukaryotic genomic databases using sensitive methods of phylogenetic analysis. Additional hydrophobic cluster analysis (HCA) allowed us to refine the results obtained from phylogeny and to map conserved domains in these new Artemis/Artemis-like proteins. HCA data was further confirmed by three-dimensional sequence modeling.

2. MATERIALS AND METHODS

2.1. Protein data mining and sequence analyses

Sixty-four protein sequences (representing the eukaryotic Pso2p family, which includes the Pso2p A, B, and Artemis proteins) were obtained directly from GenBank hosted in the National Center for Biotechnological Information (NCBI) web page [http://www.ncbi.nlm.nih.gov/]. BLAST and PSI-BLAST programs were used for initial domain screening and comparison (Altschul *et al.*, 1997). Moreover, the *Hsa*Artemis sequence (Table 1) was used as template in the Genolevures Database [http://cbi.labri.fr/Genolevures/index.php] in order to identify possible new Artemis

homologous sequences in unconventional yeast species. All searches were made to saturation. Wise2 program at European Bioinformatics Institute (http://www.ebi.ac.uk/wise2/) was used to identify Artemis-like/Artemis proteins from the unfinished genomic sequences of fungi and metazoan species. The parameters for prediction were: local mode; no intron bias; splice site modeled; synchronous model; and GeneWise623 algorithm. The theoretical pI and molecular weight of selected Artemis proteins (complete sequences) were calculated with COMPUTE pI/Mw program hosted in ExPASy Molecular Biology Server at Swiss Institute of Bioinformatics [http://us.expasy.org/tools/pitool.html].

2.2. Algorithms for sequence comparison and phylogenetic inference

Sequence to sequence BLAST (BL2SEQ; Tatusova *et al.,* 1999) was used for local alignment of Artemis proteins with default options (program blastp; matrix BLOSUM62; open gap penalty 11; and gap extension 1). Global pair-wise multiple-alignment of members of eukaryotic Pso2p family was performed in the CLUSTALX 1.8 program (Thompson *et al.,* 1994). The following alignment parameters were used: gap open penalty 10.00; gap extension 0.20; sequences >10% diverged delayed; PAM series matrix; residue-specific penalties on; and hydrophilic penalties on. When necessary, the alignments were manually adjusted using the BioEdit program (Hall, 1999).

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 2.1 (Kumar *et al.,* 2001). Minimum Evolution (ME) method was used for phylogenetic tree searching and inference. The statistical reliability of the phylogenetic trees was tested by interior branch analysis with 1,000 replications. Moreover, the Poisson correction was applied in ME for distance estimation. The pair-wise deletion option was used in handling gaps or missing data obtained from the alignment. The branches representing the Pso2p A and B Groups sequences were condensed. From

these 64 proteins we have chosen 13 Artemis-like/Artemis orthologous proteins for subsequent sequence analyses (Table 1).

2.3. Hydrophobic cluster analysis (HCA)

HCA of selected Artemis sequences was performed as previously published (Gaboriaud *et al.*, 1987). Briefly, it consisted in displaying the primary protein structure on a duplicated α-helical net, where the hydrophobic residues are automatically contourned. The hydrophobic clusters observed in an HCA plot are not distributed in a random manner, but correspond highly with the secondary protein structures associated with conserved globular domains. Thus, HCA plots are extremely valuable for phylogenetic inferences when protein sequences have a weak homology (<25% of identity/similarity) or to define conserved domains and sequence signatures in a set of homologous proteins (Gaboriaud *et al.,* 1987). It should be noted that HCA, when compared with 'linear' or one-dimensional methods of database screening, e.g. BLAST and PSI-BLAST (which needs, at least, >30% of aa identity/similarity over a sufficient length), is a more effective tool to detect sequence similarity that reflects a true three-dimensional relationship between two or more proteins (Callebaut *et al.,* 1997). Moreover, HCA has the advantage that sequence conservation can be detected in a group of proteins without any previous one-dimensional alignment (Callebaut *et al.,* 1997; Lemesle-Varloot *et al.,* 1990). The program DRAWHCA, used in such analysis, is available as a freeware at http://www.lmcp.jussieu.fr.

2.4. Three-dimensional modeling

In silico three-dimensional modeling of *Dha*Artemis-like and *Hsa*Artemis sequences was performed with 3D-JIGSAW (Bates *et al.,* 2001) algorithm using as template the penicillinase sequence of *Stenotrophomonas maltophilia* (*Sma*Penicillinase, EC number 3.5.2.6), which belongs to the metallo-β-lactamase superfamily (Protein Data Bank accession number 1sml). Three-dimensional

comparative models were generated by UCSF Chimera software (Huang *et al.,* 1996). The zinc-binding sites of *Dha*Artemis-like, *Hsa*Artemis and *Sma*Penicillinase sequences were determined with MetSite software (Sodhi *et al.,* 2004).

3. RESULTS

3.1. Identification of new metazoan and fungal Artemis/Artemis-like sequences

With the aim of identifying new Artemis sequences in metazoan and in other unrelated groups, the human Artemis sequence (*Hsa*Artemis, Table 1) was used as query in BLAST and PSI-BLAST programs, available in GenBank and in Genolevures databases. The data mining performed in metazoan sequences allowed us to identify new Artemis/Artemis-like proteins in the unfinished genome of two arthropods [*Apis mellifera* (*Ame*Artemis-like) and *Anopheles gambiae* (*Aga*Artemis-like)], one urochordate [*Ciona intestinalis* (*Cin*Artemis-like)], one puffer fish species [*Tetraodon nigroviridis* (*Tni*Artemis)], and the domestic dog [*Canis familiaris* (*Cfa*Artemis)] (Table 1). A BL2SEQ analysis of the new metazoan Artemis/artemis-like sequences indicated that these proteins have a high similarity when compared with the best-studied *Hsa*Artemis, with the following ranking: *Ame*Artemis-like, 57%; *Aga*Artemis-like, 45%; *Cin*Artemis-like, 56%; *Tni*Artemis, 76%; and *Cfa*Artemis, 86%. When the same data mining was performed in different fungal genomes, we could identify four complete Artemis-like sequences that show some degree of homology with *Hsa*Artemis. These sequences were identified in a phytopathogenic fungus [*Gibberela zeae* (*Gze*Artemis-like)], in an alkane-using yeast [*Yarrowia lipolytica* (*Yli*Artemis-like)], in a human yeast pathogen [*Candida albicans* (*Cal*Artemis-like)], and in an halotolerant/cryotolerant yeast [*Debaryomyces hansenii* (*Dha*Artemis-like)] (Table 1). The comparative BL2SEQ with *Hsa*Artemis gave the following ranking: *Gze*Artemis-like, 47%; *Yli*Artemis-like, 46%; *Cal*Artemis-like, 40%; and *Dha*Artemis-like, 40%. A sequence screening of more conventional fungi species (*S. cerevisiae, S. pombe,* and

Neurospora crassa) using *Hsa*Artemis as template did not reveal any Artemis-like protein. Moreover, we could not identify Artemis/Artemis-like sequences in complete or genomic databases of plants or protozoans.

These new metazoan and fungal Artemis/Artemis-like sequences were subjected to a global sequence comparison followed by a phylogenetic analysis in order to determine their relationship with the family of eukaryotic Pso2/Snm1 proteins. As expected, the Pso2p A and B Groups, together with Artemis Group, compose three phylogenetically distinct clades highly supported by interior branch analysis (Figure 1). Both Pso2p A and B Groups share a common ancestor, with Artemis Group composing a more ancient clade (Figure 1). An in-depth phylogenetic analysis of Artemis Group indicated that the clade can be subdivided in two subgroups (supported by internal branch analysis) with both sharing a last universal common ancestor (LUCA) (Figure 1). The first subgroup comprises the previously un-characterized fungal Artemis-like sequences while the second contains the best-characterized members of metazoan Artemis/Artemis-like sequences (Figure 1).

This phylogenetic result led us to analyze the theoretical physico-chemical properties of the new fungal Artemis-like proteins by comparing them with the best-known *Hsa*Artemis and the Artemis sequence of *Mus musculus* (*Mmu*Artemis, Table 2). The physico-chemical data indicated that fungal Artemis-like proteins have similar acidic pI (between 5.72 to 6.70), as *Mmu*Artemis and *Hsa*Artemis (5.77 and 5.69, respectively) (Table 2). The exception is *Yli*Artemis, with a basic pI of 8.14 (Table 2). In terms of size and m.w., the fungal Artemis-like proteins have polypeptide chains varying from 537 aa to 839 aa (m.w. from 67.5 kDa to 94 kDa, Table 2). For all subsequent studies, we chose the *Dha*Artemis-like protein because of its good agreement of sequence homology with metazoan Artemis/Artemis-like proteins.

3.2. HCA of Artemis/Artemis-like sequences

HCA is a very sensitive method of sequence comparison that detects 2- and 3-dimensional similarities between protein domains showing very limited aa relatedness, typically below the so-called “twilight zone” (25-30%) (Gaboriaud *et al.*, 1987). The method uses aa hydrophobic cluster plots to identify globular domains associated with structurally or functionally conserved features. In this case, moderate or highly hydrophobic aa form clusters of various shapes and sizes. Thus, a horizontal diamond-shaped cluster corresponds to the hydrophobic face of an α-helix; internal β-strands are represented by a cluster that alters hydrophobic residues on the left and on the right of an HCA plot; and surface strands are shown by ‘zig-zag’ patterns. Moreover, conserved protein domains can be mapped with HCA using orthologous sequences from different species. An HCA comparison between metazoan Artemis/Artemis-like and *Dha*Artemis-like sequences indicated that the N-terminus of the Artemis/Artemis-like proteins can be subdivided in four conserved regions (CR, Figures 2 and 3). The first CR (CRI) is the smallest globular domain of the four CR, and is basically composed by β-structures (Figure 2). It contains the characteristic signature of the metallo-β-lactamase family, the highly conserved sequence H(V,C)HxDH(M,L)xG situated between a β-strand (β1) and an α-helix (α1) structures (Figure 2). Three signatures compose the CRII domain: (i) a consensus sequence (S,P)AxHCPG found between two β-strands (β4 and β5), (ii) the signature TGD(IF)R, situated between a β-strand (β6) and an α-helix (α2) and, (iii) the D(S,T)T signature, located closest to an α-β structure (β-α1) (Figure 2). CRIII is composed mainly by α-helices and has only a small signature (xGxEx) located between two-conserved α-helices (α4 and α5). CRIV is a large globular domain characterized by the SxH(A,S)S signature situated closest to a highly conserved α-helix (α9, Figure 3). The average HCA and global alignment homology

scores calculated for each CR gave similar results, with CRII containing the highest number of conserved hydrophilic and hydrophobic aa residues (Figures 2 and 3).

3.3. Three-dimensional modeling of Artemis/Artemis-like sequences

Three-dimensional modeling of *Dha*Artemis-like and *Hsa*Artemis was done in order to confirm the results obtained by HCA. When compared with other proteins of the metallo-β-lactamase superfamily, the *Dha*Artemis-like and *Hsa*Artemis CRI domain exhibited structural similarity with the N-terminus domain of *S. maltophilia* penicillinase enzyme (E-values of 3×10^{-51} and 2×10^{-46}, respectively). Using this information, we modeled the N-terminus of *Hsa*Artemis and *Dha*Artemis-like. As expected, this resulted in a similar three-dimensional conformation of *Dha*Artemis-like, *Hsa*Artemis and *Sma*Penicillinase N-termini, with the CRI and CRII domains present in both proteins (Figures 4A to C). CRIII and CRIV domains are only present in the Artemis Group and could not be modeled using the *Sma*Penicillinase template. When the results of three-dimensional modeling were refined, taking into account the potential zinc-binding properties of CRI domain of *Sma*Penicillinase, *Hsa*Artemis, and *Dha*Artemis-like protein (Figure 5A to C), it became clear that all analyzed CRI domains have similar tertiary conformation. In the case of *Sma*Penicillinase, the aa residues H84, H86 and H160 (not shown) coordinate a zinc atom (Figure 5A). Similar aa residues could coordinate a zinc atom in CRI domain of *Hsa*Artemis [aa residues H33, H35 and H115 (not shown); Figure 5B] and *Dha*Artemis-like protein [aa residues H39, H41 and H107 (not shown); Figure 5C]. Interestingly, both CRI domains of *Hsa*Artemis and *Dha*Artemis-like protein have a cysteine residue between the two first histidines (C34 and C40, respectively; Figure 5B and 5C). This cysteine residue is characteristic of metazoan Artemis/Artemis-like protein, and it is not found in other proteins of Pso2p A and B Groups (data not shown).

4. DISCUSSION

The results obtained by primary sequence analysis and phylogeny, combined with HCA and protein three-dimensional modeling of selected Artemis sequences, clearly indicated that Artemis can be found in non-mammalian organisms. A comparison of these sequences showed that, despite their difference in length, all CR studied have similar disposition, being located within the N-terminus of the Artemis/Artemis-like proteins (Figure 6). This similar location indicates structural conservation during the evolution of the Artemis Group. Moreover, both HCA and three-dimensional modeling highlight conserved hydrophilic and hydrophobic residues that are considered to be necessary for Artemis function (Figures 2 and 3). For example, the presence of a cysteine residue between two important histidine of the metallo-β-lactamase domain (CRI, Figures 5B and 5C) is characteristic for metazoan Artemis/Artemis-like protein, and it is also present in fungal Artemis-like sequences. The histidine residues of CRI seem necessary for the coordination of a zinc atom (Callebaut *et al.,* 2002), but little importance has been given for this cysteine residue. In an engineered human carbonic anhydrase II (CAII, EC number 4.2.1.1), a zinc-containing metalloenzyme that catalyzes the reversible hydration of CO_2 to HCO_3^-, a non-natural cysteine together with three histidine residues could bind to a zinc atom (His3Cys pocket) (Ippolito and Christianson, 1993; Kiefer *et al.,* 1993). In this case, the cysteine induces the formation of a zinc-thiolate bond, which optimally coordinates the zinc cofactor (Ippolito and Christianson, 1993). The His3Cys pocket is also found in the zinc binding sites of β-lactamase II from *Bacillus cereus* (Sutton *et al.,* 1987) and in matrix metalloproteinase pro-enzymes (Holz *et al.,* 1992; Salowe *et al.,* 1992). Probably, the His3Cys pocket has an important role in the zinc cofactor coordination in metazoan Artemis and fungal Artemis-like proteins. The role of the other CR is still subject of biochemical studies. Recently, Pannicke *et al.* (2004) mutated nine evolutionary conserved histidine and aspartic acid residues within the *Hsa*Artemis

protein. With the exception of one of them (H38), all other residues seem essential for overhang endonucleolytic and hairpin-opening activities. Identical aa residues can be found in non-mammalian metazoan and in fungal Artemis/Artemis-like sequences, again suggesting that the Artemis Group is not exclusive to jawed vertebrates.

The mammalian Artemis proteins are the best-characterized members of the Pso2p/Snm1p family, which itself is a member of the β-CASP family (Callebaut *et al.,* 2002). The Pso2p/Snm1p family is composed of three distinct groups (Callebaut *et al.,* 2002): (i) the A Group, whose function is associated exclusively with the repair of chemically-induced ICLs in DNA (Brendel *et al.,* 2003; Ishiai *et al.,* 2004); (ii) the B Group, composed by proteins that are recruited for recombinational repair of lesions induced by oxidative agents (Molinier *et al.,* 2004), by ICLs or IR (Demuth *et al.,* 2004); and (iii) the proper Artemis Group. Multi-cellular organisms like plants and metazoans, contain one or more proteins that belong to Pso2/Snm1 A and B Groups, while fungi and other lower eukaryotes apparently have only one protein of the Pso2/Snm1 A Group (Bonatto *et al.,* 2005). This fact could reflect the tissue complexity of higher eukaryotes, where proteins with redundant or specific functions are required for maintenance of genomic integrity (Bonatto *et al.,* 2005). However, the presence of an Artemis-like sequence in different fungi, in two arthropods and in an urochordate species indicates that the Artemis Group proteins could have more functions than are presently known. Many mammalian organisms and cells, e.g. homozygous knockout murine models or human Artemis-deficient cell cultures, have been largely used to study the function(s) of Artemis (Noordzij *et al.,* 2003; Rooney *et al.,* 2002; Rooney *et al.,* 2003). Until now, all data clearly indicate that the mammalian Artemis proteins play a role in consolidating the adaptative immune system, participating in V(D)J recombination (Le Deist *et al.,* 2004). In vertebrates, adaptive immune responses also play a crucial role in defense against pathogens and depend on the generation of a vast

repertoire of soluble and membrane-bound antigen receptors presented by B-lymphocytes and T-lymphocytes, respectively. This assembly process is highly conserved throughout evolution in all jawed vertebrates, starting with sharks (Le Deist *et al.,* 2004). A consistent phenotype in all types of Artemis-deficient cells is the low efficiency in opening of hairpin coding ends, as reflected by the accumulation of hairpin intermediates generated by Rag-1/Rag-2 proteins in Artemis-deficient thymocytes (Rooney *et al.,* 2002). Experimental data accumulated on mammalian Artemis indicate that these proteins are genomic caretakers, as Artemis-deficient murine cells have increased chromosomal instability, including telomeric fusions (Rooney *et al.,* 2003). Thus, Artemis might function in a subset of NHEJ reactions that require end processing (Rooney *et al.,* 2003). Artemis clearly functions in a common repair pathway of IR-induced DSB in which the proteins γ-H2AX, 53BP1, Nbs1, Mre11, and DNA-PK also play a role (Riballo *et al.,* 2004).

As Artemis is an important protein for vertebrates, it was not surprising to find an Artemis protein in the puffer fish *T. nigroviridis*. Both B (Ig) and TCR genes of *Danio rerio* (zebrafish) genes undergo V(D)J recombination (Haire *et al.,* 2000). Mechanistically, this process of somatic diversification of the rearranging antigen receptors found in fish appears to resemble that seen in mammals (Yoder *et al.,* 2002), and both Rag-1 and Rag-2 proteins have been identified in zebrafish and thus may function in this process (Willett *et al.,* 1997).

However, the presence of an Artemis protein in an urochordate, *C. intestinalis* (*Cin*Artemis-like), is not easily explained. *C. intestinalis* is an organism that occupies a key phylogenetic position in vertebrate evolution, as it belongs to the subphylum Urochordata, that together with the subphyla Vertebrata and Cephalochordata composes the phylum Chordata (Cameron *et al.,* 2000). Despite the complexity of the immune repertoire of the Vertebrata subphylum, genomic analysis of the immunity-related genes

in *C. intestinalis* suggests the presence of a well and uniquely developed innate immune system (Azumi *et al.,* 2003). Although some possible precursors of the jawed vertebrate adaptive immune system were identified, they still appear distant from functional adaptive immunity that is equipped with somatic mechanisms for generation of diversity (Azumi *et al.,* 2003). Moreover, the specific genes that are required for V(D)J recombination as Ig, TCR, MHC class I and II, *RAG-1* and *2*, and AID, were not found in the *C. intestinalis* genome (Azumi *et al.,* 2003). Thus, we can speculate that the *Cin*Artemis-like protein probably has a more specific function in DNA repair, possibly acting in NHEJ recombination. In this sense, the presence of Artemis-like protein in two arthropods, *A. mellifera* (*Ame*Artemis-like) and *A. gambiae* (*Aga*Artemis-like), could give some clues about the functions of Artemis in lower metazoa. A DNA-PKcs gene was recently identified in the genomes of *Drosophila melanogaster,* in *A. gambiae* and in *A. mellifera* (Doré *et al.,* 2004). DNA-PKcs plays a key role in NHEJ, in V(D)J recombination and in telomeric maintenance (Blunt *et al.,* 1995; Smith and Jackson, 1999). It seems likely that the DNA-PKcs and Artemis proteins are not a late evolutionary add-on as previously thought (Doré *et al.,* 2004). The arthropod *Ame*Artemis-like and *Aga*Artemis-like proteins give a strong support for the hypothesis of an ancient NHEJ machinery that later developed into a V(D)J recombination in jawed vertebrates, necessary for an adaptative immune system.

The same hypothesis could be raised for fungal Artemis-like proteins. As these proteins were found in four fungal species, we think that Artemis could be a reminiscent protein of an ancestral DNA recombination process that, by some evolutive reason, was kept in a few lower eukaryotes.

The species *G. zeae* is a mycotoxigenic filamentous fungus that causes destructive epidemics of maize ear rot and wheat and barley head blight in temperate regions (Desjardins, 2003); *Y. lipolytica* is a dimorphic yeast species that is able to use

several unusual carbon sources like parafins, various alcohols and acetate (Casaregola *et al.,* 2000); *C. albicans* is also a dimorphic fungus and is the most common human fungal pathogen, causing both mucosal and systemic infections, particularly in immuno-compromised people (Johnson, 2003); and *D. hansenii* is a cryotolerant, marine yeast, which can tolerate salinity levels up to 24% (Lépinglea *et al.,* 2000). Despite their biological differences, all fungal species listed here are capable of surviving in highly stress-inducing environments. It is possible that the metazoan and fungal Artemis/Artemis-like proteins could be recruited by some other genetic mechanism that also generates biological diversity, a necessary condition for survival in stressing environments.

In sexually reproducing organisms, genetic diversity and viability of the gametes are ensured by recombination occurring during meiosis (Borde *et al.,* 2004). Meiotic recombination can be divided into three successive stages: (i) initiation, which consists in the formation of programmed DNA double strand breaks (DSBs) and their subsequent processing by 5´ to 3´ resection to form single-strand tails; (ii) the repair of these initiating lesions through homologous recombination; and (iii) the formation of the crossover products that create a physical interhomolog connection ensuring their proper segregation during meiosis I (Borde *et al.,* 2004). Interestingly, many proteins that are related to DSB repair of somatic cells (e.g. Mre11p, Rad50p, and Xrs2p) also have an important role in the repair of programmed DSB during meiosis (Borde *et al.,* 2004). Proteins involved in generation of meiotic DSB in higher eukaryotes have not been identified and, in lower eukaryotes, the data indicates that DSB are created by endonuclease Spo11p (Pâques and Harber, 1999). However, the exact role of Spo11p in DSB formation is not fully understood (Bhuiyan and Schmekel, 2004). A Rag-2 like protein, named Peas, which is evolutionarily conserved among metazoans, was analyzed by expression and immuno-histochemical assays and shown to be specifically expressed

in testis, particularly in pachytene spermatocytes, pointing to its putative involvement in meiotic recombination (Ohinata *et al.,* 2003). The same was found for *ARTEMIS* where of all human tissues expression is highest in ovarian and testis cells (Moshous *et al*., 2001). Taking into account our data and the results of others, we may speculate that the Artemis Group could also act in the repair of DSB introduced during meiosis, thus helping in generating the observed genetic diversity. During evolution of multi-cellular organisms, with the increase of tissue diversity and the dawn of refined DNA recombination systems [V(D)J and NHEJ], the role of Artemis also changed and became a more specialized one, as is now seen in mammalian cells.

The identification of fungal Artemis-like proteins is important in many aspects. Firstly, in yeast advanced molecular techniques can be used to study the function of these proteins and to compare them with their mammalian counterparts. Secondly, as the yeast genome is easily manipulated, single, double or multiple mutants of genes encoding Artemis-like proteins can be combined with defect DNA repair genes and the resulting phenotypes studied under stressing or DNA-damaging conditions.

ACKNOWLEDGMENTS

This work was supported by research grants from Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and GENOTOX (Laboratório de Genotoxicidade, UFRGS).

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). *Nucleic Acids Res.* **25:** 3389-3402.

Aylon, Y., and Kupiec, M. (2004). *DNA Repair* **3:** 797-815.

Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M.,·Ikeda, M., Arai, M., Inoue, Y.,·Shimizu, T., Satoh,

N., Rokhsar, D. S.,·Du Pasquier, L., Kasahara, M., Satake, M., and Nonaka, M. (2003). *Immunogenetics* **55:** 570-581.

Bates, P. A., Kelley, L. A., MacCallum, R. M., and Stemberg, M. J. E. (2001). *Proteins* **5:** 39-46.

Bernstein, C., Bernstein, H., Payne, C. M., and Garewal, H. (2002). *Mutat. Res.* **511:** 145-178.

Bhuyan, H., and Schmekel, K. (2004). *Genetics* **168:** 775-783.

Blunt, T., Finnie, N. J., Taccioli, G. E., Smith, G. C., Demengeot, J., Gottlieb, T. M., Mizuta, R., Varghese, A. J., Alt, F. W., and Jeggo, P. A. (1995). *Cell* **80:** 813-823.

Bonatto, D., Revers, L. F., Brendel, M., and Henriques, J. A. P. *Braz. J. Med. Biol. Res.* (2005), *in press.*

Borde, V., Lin, W., Novikov, E., Petrini, J. H., Lichten, M., and Nicolas, A. (2004). *Mol. Cell* **13:** 389-401.

Brendel, M., Bonatto, D., Strauss, M., Revers, L. F., Pungartnik, C., Saffi, J., and Henriques, J. A. P. (2003). *Mutat. Res.* **544:** 179-193.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., and Mornon, J-P. (1997). *Cell. Mol. Life Sci.* **53:** 621-645.

Callebaut, I., Moshous, D., Mornon, J–P., and de Villartay, J-P. (2002). *Nucleic Acids Res.* **30:** 3592-3601.

Cameron, C. B., Garey, J. R., and Swalla, B. J. (2000). *Proc. Natl. Acad. Sci. USA* **97:** 4469-4474.

Casaregola, S., Neuvéglisea, C., Lépinglea, A., Bona, E., Feynerola, C., Artiguenaveb, F., Winckerb, P., and Gaillardina, C. (2000). *FEBS Lett.* **487:** 95-100.

Cline, S. D., and Hanawalt, P. C. (2003). *Nat. Rev. Mol. Cell Biol.* **4:** 361-372.

Demuth, I., Digweed, M., and Concannon, P. (2004). *Oncogene* **23:** 8611-8618.

Desjardins, A. E. (2003). *Annu. Rev. Phytopathol.* **41:** 177-198.

Doré, A. S., Drake, A. C. B., Brewerton, S. C., and Blundell, T. L. (2004). *DNA Repair* **3:** 33-41.

Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J-P. (1987). *FEBS Lett.* **224:** 149-155.

Haire, R. N., Rast, J. P., Litman, R. T., and Litman, G.W. (2000). *Immunogenetics* **51:** 915-923.

Hall, T. A. (1999). *Nucleic Acids Symp. Ser.* **41:** 95-98.

Holz, R. C., Salowe, S. P., Smith, C. K., Cuca, G. C., and Que, L. (1992). *J. Am. Chem. Soc.* **114:** 9611-9614.

Huang, C.C., Couch, G.S., Pettersen, E.F., and Ferrin, T.E. (1996). *Pac. Symp. Biocomput.* **1:** 724.

Ippolito, J. A., and Christianson, D. W. (1993). *Biochemistry* **32:** 9901-9905.

Ishiai, M., Kimura, M., Namikoshi, K., Yamazoe, M., Yamamoto, K., Arakawa, H., Agematsu, K., Matsushita, N., Takeda, S., Buerstedde, J-M., and Takata, M. (2004). *Mol. Cell. Biol.* **24:** 10733-10741.

Jackson, S. P. (2002). *Carcinogenesis* **23:** 687-696.

Jeggo, P., and O'Neill, P. (2002). *DNA Repair* **1:** 771-777.

Johnson, A. (2003). *Nat. Rev. Microbiol.* **1:** 106-116.

Kiefer, L. L., Krebs, J. F., Paterno, S. A., and Fierke, C. A. (1993). *Biochemistry* **32:** 9896-9900.

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software, Arizona State University, Tempe, Arizona, USA.

Kuzminov, A. (2001). *Proc. Nat. Acad. Sci. USA* **98:** 8241-8246.

Le Deist, F., Poinsignon, C., Moshous, D., Fischer, A., and de Villartay, J-P. (2004). *Immunol. Rev.* **200:** 142-155.

Lee, G. S., Neiditch, M. B., Salus, S. S., and Roth, D. B. (2004). *Cell* **117:** 171-184.

Lees-Miller, S. P., and Meek, K. (2003). *Biochimie* **85:** 1161-1173.

Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A., and Mornon, J-P. (1990). *Biochimie* **72:** 555-574.

Lépinglea, A., Casaregola, S., Neuvéglisea, C., Bona, E., Nguyena, H-V., Artiguenaveb, F., Winckerb, P., and Gaillardina, C. (2000). *FEBS Lett.* **487:** 82-86.

Lieber, M. R., Yunmei, M., Pannicke, U., and Schwartz, K. (2003). *Nat. Rev. Mol. Cell Biol.* **4:** 712-720.

Molinier, J., Stamm, M-E., and Hohn, B. (2004). *EMBO Rep.* **5:** 1-6.

Moshous, D., Callebaut, I., Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., and de Villartay, J-P. (2001). *Cell* **105:** 177-186.

Noordzij, J. G., Verkaik, N. S., Van der Burg, M., van Veelen, L. R., de Bruin-Versteeg, S., Wiegant, W., Vossen, J. M., Weemaes, C. M., de Groot, R., Zdzienicka, M. Z., van Gent, D. C., and van Dongen, J. J. (2003). *Blood* **101:** 1446-1452.

Ohinata, Y., Sutou, S., and Mitsui, Y. (2003). *FEBS Lett.* **537:** 1-5.

Pannicke, U., Ma, Y., Hopfner, K-P., Niewolik, D., Lieber, M. R., and Schwarz, K. (2004). *EMBO J.* **23:** 1987-1997.

Pâques, F., and Haber, J. E. (1999). *Microbiol. Mol. Biol. Rev.* **63:** 349-404.

Riballo, E., Kühne, M., Rief, N., Doherty, A., Smith, G. C. M., Recio, M-J., Reis, C., Dahm, K., Fricke, A., Krempler, A., Parker, A. R., Jackson, S. P., Gennery, A., Jeggo, P. A., and Löbrich, M. (2004). *Mol. Cell* **16:** 1-20.

Richardson, C., and Jasin, M. (2000). *Mol. Cell. Biol.* **20:** 9068-9075.

Rooney, S., Alt, F. W., Lombard, D., Whitlow, S., Eckersdorff, M., Fleming, J., Fugmann, S., Ferguson, D. O., Schatz, D. G., and Sekiguchi, J. (2003). *J. Exp. Med.* **197:** 553-565.

Rooney, S., Sekiguchi, J., Zhu, C., Cheng, H-L., Manis, J., Whitlow, S., DeVido, J., Foy, D., Chaudhuri, J., Lombard, D., and Alt, F. W. (2002). *Mol. Cell* **10:** 1379-1390.

Salowe, S. P., Marcy, A. I., Cuca, G. C., Smith, C. K., Kopka, I. E., Hagmann, W. K., and Hermes, J. D. (1992). *Biochemistry* **31:** 4535-4540.

Schlissel, M. S. (2002). *Cell* **109:** 1-4.

Smith, G. C., and Jackson, S. P. (1999). *Genes Dev.* **13:** 916-934.

Sodhi, J. S., Bryson, K., McGuffin, L. J., Ward, J. J., Wernisch, L., and Jones, D. T. (2004). *J. Mol. Biol.* **342:** 307-320.

Sutton, B. J., Artymiuk, P. J., Cordero-Borboa, A. E., Little, C., Phillips, D. C., and Waley, S. G. (1987). *Biochem. J.* **248:** 181-188.

Tatusova, T. A., and Madden, T. L. (1999). *FEMS Microbiol. Lett.* **174:** 247-250.

Thompson, J. D., Higgins, D.G., and Gibson, T. J. (1994). *Nucleic Acids Res.* **22:** 4673-4680.

Willett, C. E., Cherry, J. J., and Steiner, L. A. (1997). *Immunogenetics* **45:** 394-404.

Yoder, J. A., Nielsen, M. E., Amemiya, C. T., and Litman, G. W. (2002). *Microbes Infect.* **4:** 1469-1478.

Zhang, X., Succi, J., Feng, Z., Prithivirajsingh, S., Story, M. D., and Legerski, R. J. (2004). *Mol. Cell. Biol.* **24:** 9207-9220.

TABLES

Table 1. Protein sequences used in this work.

Protein Group	Sequence name	Species	Acession No. (GenBank)
Fungal Artemis-like			
	*Dha*Artemis-like	*Debaryomyces hansenii*	gi|50424115|
	*Cal*Artemis-like	*Candida albicans*	gi|46438865|
	*Yli*Artemis-like	*Yarrowia lipolytica*	gi|49650441|
	*Gze*Artemis-like	*Gibberella zeae*	gi|42551205|
Metazoan Artemis/Artemis-like			
	*Ame*Artemis-like	*Apis mellifera*	gi|48096919|
	*Aga*Artemis-like	*Anopheles gambiae*	gi|31203062|
	*Cin*Artemis-like	*Ciona intestinalis*	gi|23587642|
	*Tni*Artemis-like	*Tetraodon nigroviridis*	gi|47218348|
	*Gga*Artemis	*Gallus gallus*	gi|50764059|
	*Cfa*Artemis	*Canis familiaris*	gb|AAEX01006067.1|
	*Hsa*Artemis	*Homo sapiens*	gi|13872809|
	*Rno*Artemis	*Rattus norvegicus*	gi|22023557|
	*Mmu*Artemis	*Mus musculus*	gi|22023549|

Table 2. Physico-chemical analyses of fungal Artemis-like proteins and two mammalian Artemis proteins (metazoan Artemis Group).

Protein Group	Protein name	Length	pI	M.W. (Da)
Fungal Artemis-like				
	*Dha*Artemis-like	783	6.45	90218.9
	*Cal*Artemis-like	537	6.70	60890.1
	*Yli*Artemis-like	839	8.14	94004.6
	*Gze*Artemis-like	600	5.72	67467.3
Metazoan Artemis				
	*Hsa*Artemis	692	5.69	78422.3
	*Mmu*Artemis	705	5.77	78920.5

FIGURE LEGENDS

Figure 1. Unrooted phylogenetic tree of Pso2p A, Pso2p B, and Artemis Groups. Numbers at nodes indicate internal branch analysis values (1,000 replicates) and the horizontal bar represents a distance of 0.5 substitutions per site. The Pso2p A and B Group clades are shown in a condensed form. The representative subclades of metazoan and fungal Artemis/Artemis-like sequences are highlighted in gray. Symbol: LUCA (last universal common ancestor).

Figure 2. Hydrophobic cluster analysis (HCA) of the conserved region I (CRI) and conserved region II (CRII) sequences from mammalian Artemis (*Hsa*Artemis), non-mammalian metazoan Artemis/Artemis-like (*Tni*Artemis, *Cin*Artemis-like and *Aga*Artemis-like), and fungal Artemis-like (*Dha*Artemis-like) proteins. Conserved hydrophobic and hydrophilic aa residues are shaded in gray and black, respectively. Secondary structures (α, β or α-β structures) derived from HCA are shown in the figure. A consensus sequence derived from HCA of Artemis CRI and CRII followed by an average homology score (%), calculated from HCA and global alignment for each CR, is also indicated in the figure. The direction of polypeptide chain, from N-terminus to C-terminus, is indicated by an arrow. The way to read the sequences and special symbols are indicated in the inset.

Figure 3. Hydrophobic cluster analysis (HCA) of the conserved region III (CRIII) and conserved region IV (CRIV) sequences from mammalian Artemis (*Hsa*Artemis), non-mammalian metazoan Artemis/Artemis-like (*Tni*Artemis, *Cin*Artemis-like and *Aga*Artemis-like), and fungal Artemis-like (*Dha*Artemis-like) proteins. Conserved hydrophobic and hydrophilic aa residues are shaded in gray and black, respectively. Secondary structures (α, β or α-β structures) derived from HCA are shown in the figure. A consensus sequence derived from HCA of Artemis CRIII and CRIV followed by an average homology score (%), calculated from HCA and global alignment for each

CR, is also indicated in the figure. The direction of polypeptide chain, from N-terminus to C-terminus, is indicated by an arrow. The way to read the sequences and special symbols are indicated in the inset.

Figure 4. Three-dimensional model of *Sma*Penicillinase (A), *Hsa*Artemis (B) and *Dha*Artemis-like (C) proteins. The models were turned in an angle of 180° to show the Artemis and Penicillinase conserved region I (CRI) and conserved region II (CRII). The secondary structures that compose the CRI and CRII are highlighted in black and gray, respectively. CRIII and CRIV could not be modeled due to its low similarity with *Sma*Penicillinase sequence. The N-terminus (NH_2) and the C-terminus (HOOC) are indicated.

Figure 5. Three-dimensional analysis of conserved region I (CRI) of *Sma*Penicillinase (A), *Hsa*Artemis (B) and *Dha*Artemis-like (C). Conserved aa residues that belong to the zinc binding pocket are shown in light gray. Other conserved CRI aa residues are indicated in dark gray.

Figure 6. Schematic representation of *Hsa*Artemis, *Tni*Artemis, *Cin*Artemis-like, *Aga*Artemis-like and *Dha*Artemis-like sequences. The direction of proteins, from N-terminus to C-terminus, is indicated by a line. The length of sequences (in aa residues) is shown on the right side of the figure. Incomplete sequence lengths are shown between parentheses. Abbreviations: conserved region I, II, III, IV (CRI), (CRII), (CRIII), and (CRIV), respectively. The dotted lines between all CRs of Artemis/Artemis-like proteins indicate high level of primary and secondary sequence homologies.

Figure 1.

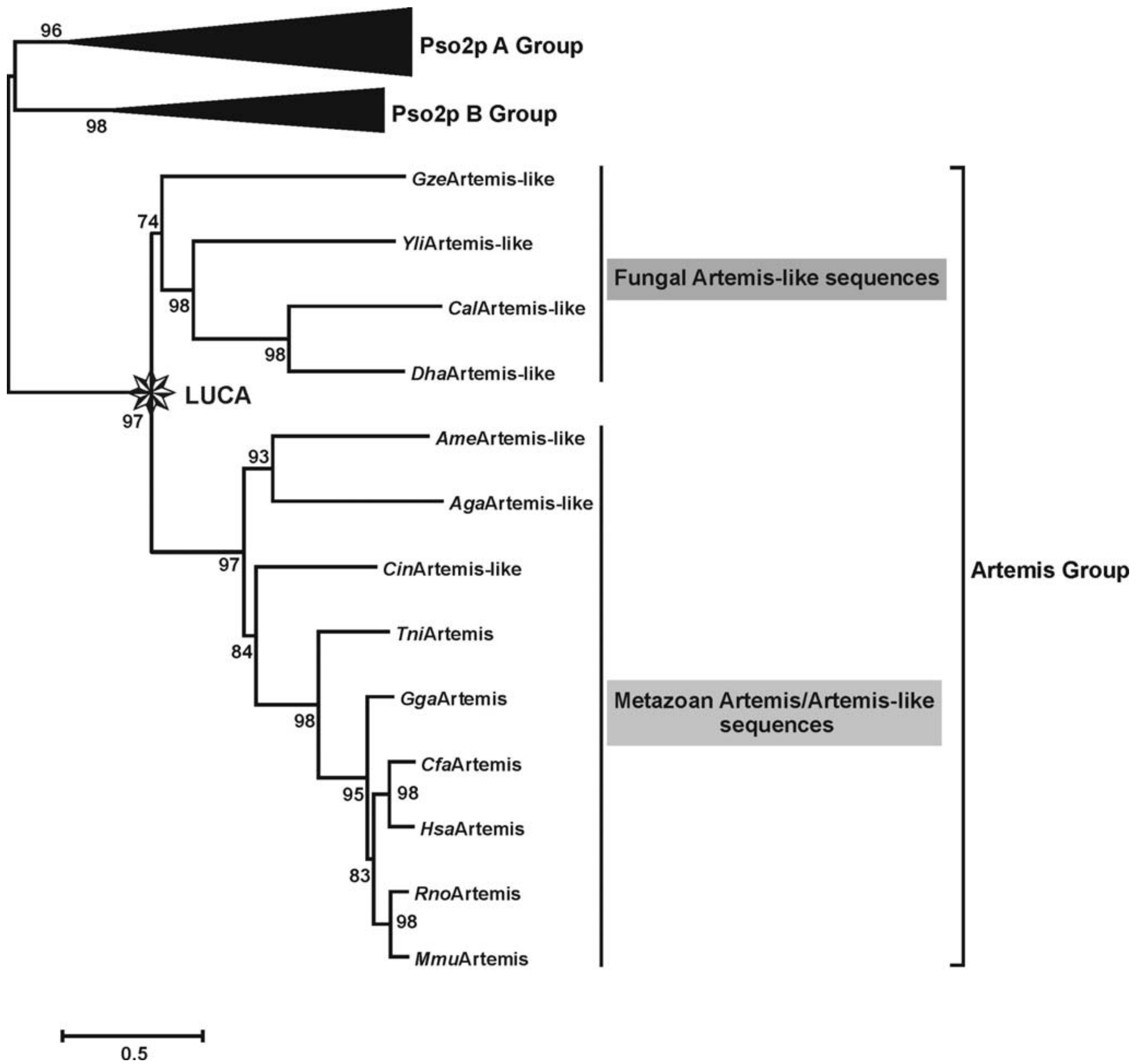


Figure 2.

N ⟵⟶ COOH

CRI: β1 α1 β2 β3

CRII: β4 β5 β6 α2 β–α1 α3

*Hsa*Artemis

*Tni*Artemis

*Cin*Artemis-like

*Aga*Artemis-like

*Dha*Artemis-like

☆ Pro ♦ Gly □ Thr ▣ Ser

1D sequence

2D structuration

Consensus sequence: HCHxDH | (SP)AxHCPG TGD(IF)R D(ST)T

	CRI		CRII
HCA	60		70
Global alignment	57		50

Average homology score (%)

Figure 3.

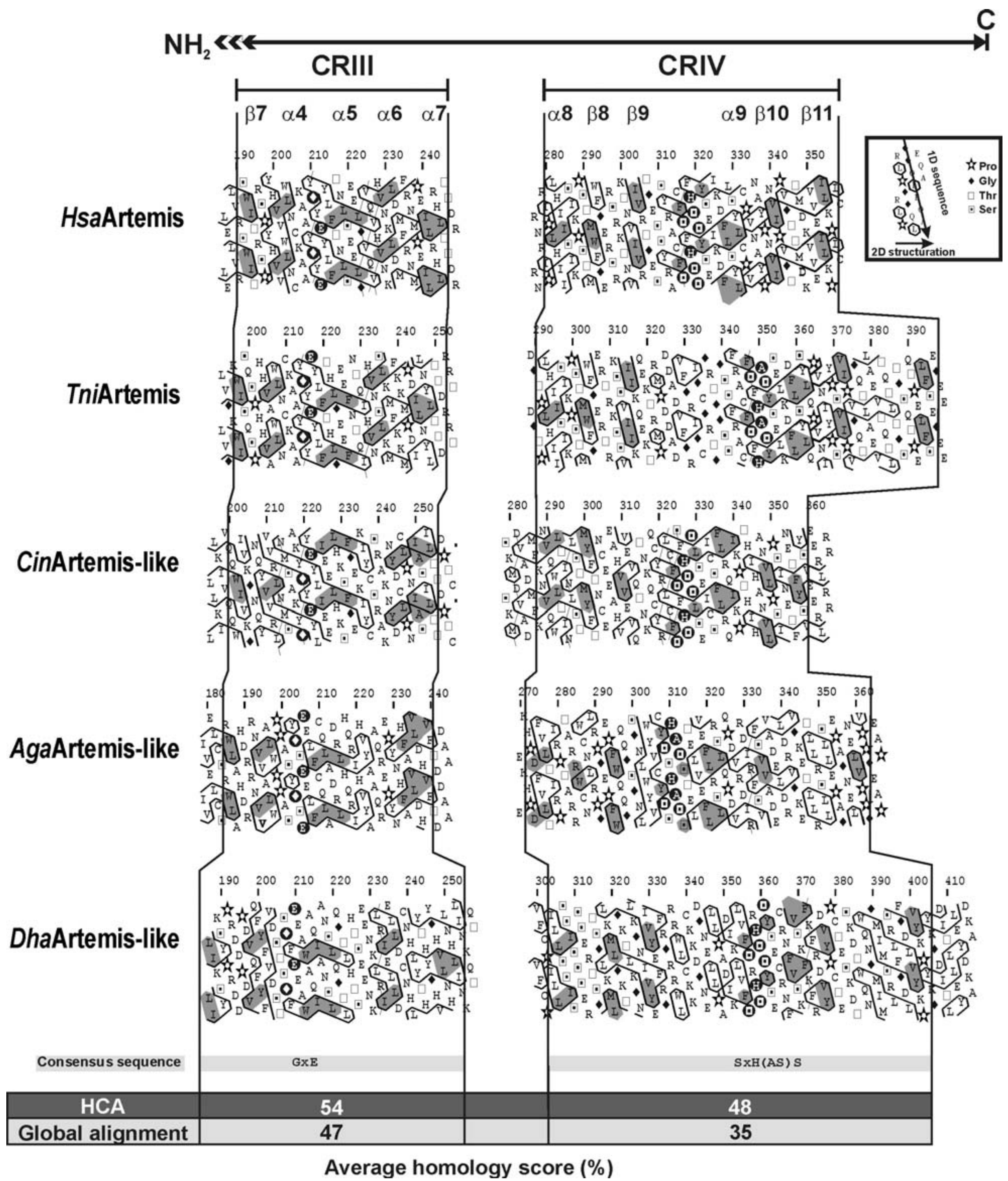


Figure 4.

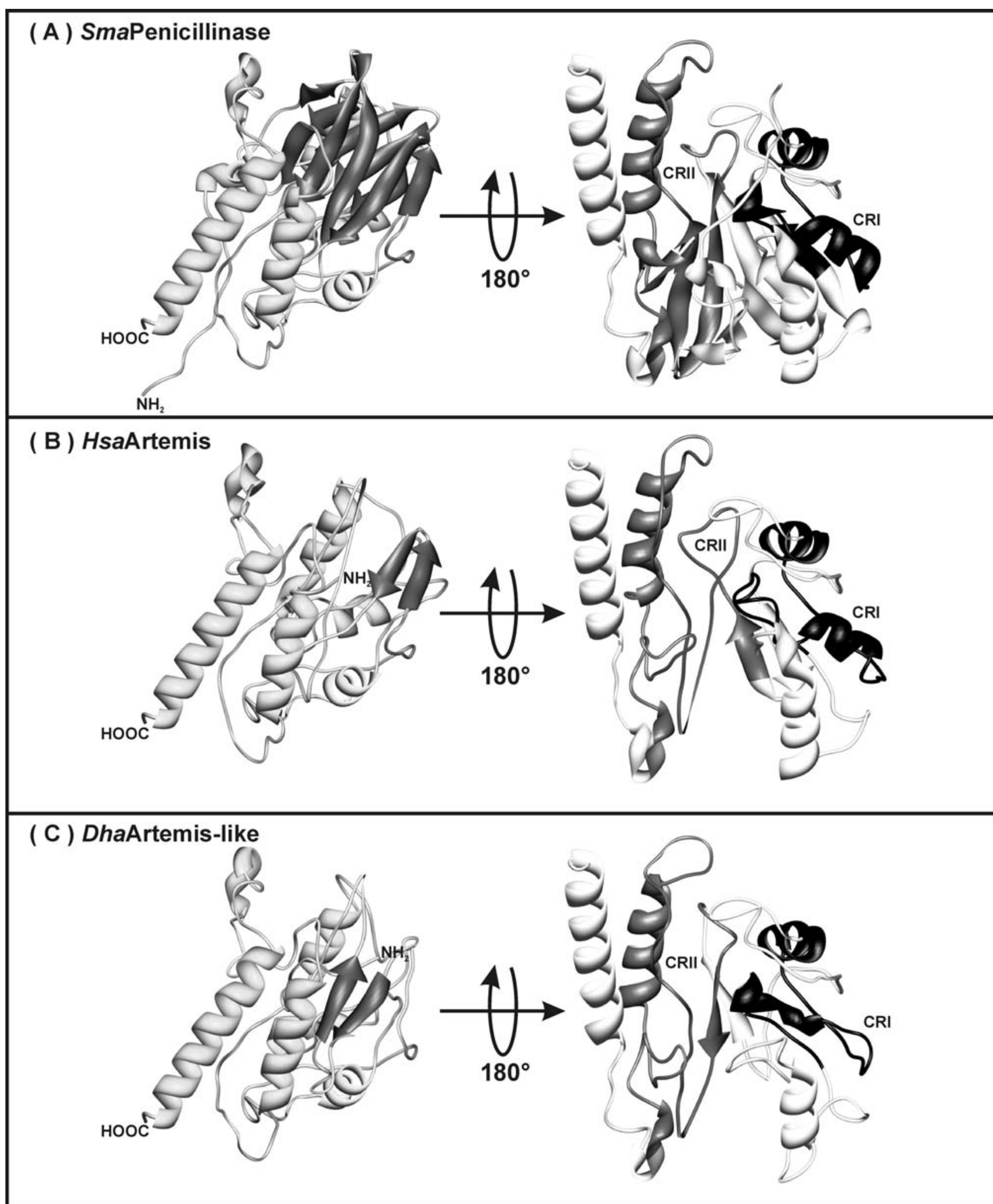


Figure 5.

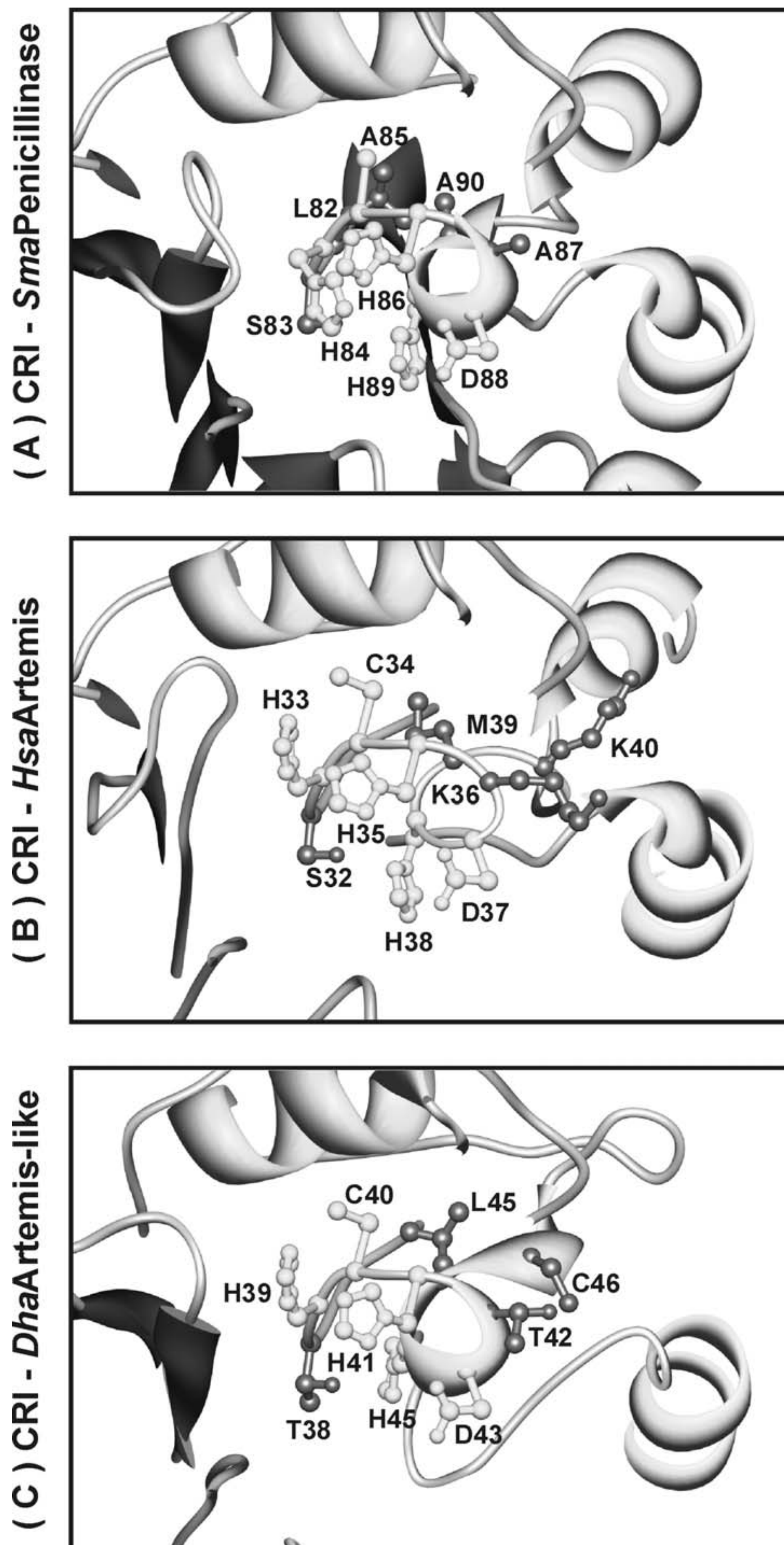
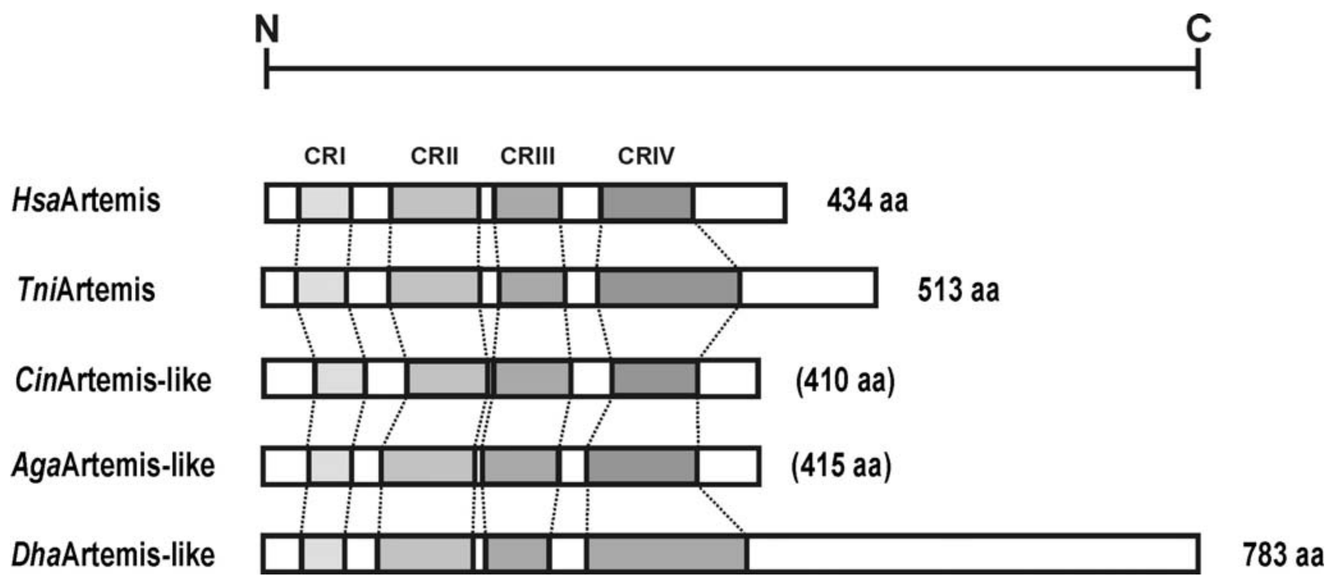


Figure 6.



4

Capítulo Quatro

A NEW GROUP OF PLANT-SPECIFIC ATP-DEPENDENT DNA LIGASES IDENTIFIED BY PROTEIN PHYLOGENY, HYDROPHOBIC CLUSTER ANALYSIS, AND 3-DIMENSIONAL MODELING

Functional Plant Biology **(2005) 32: 1–14**

CSIRO PUBLISHING

www.publish.csiro.au/journals/fpb Functional Plant Biology, 2005, **32**, 161–174

A new group of plant-specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis and 3-dimensional modelling

Diego Bonatto[A], *Martin Brendel*[A,B] *and João Antonio Pêgas Henriques*[A,C,D]

[A]Departamento de Biofísica / Centro de Biotecnologia, UFRGS, Av. Bento Gonçalves 9500, 91507-970 Porto Alegre, RS, Brazil.
[B]Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz (UESC), Ilhéus, BA, Brazil.
[C]Curso de Farmácia, Universidade Luterana do Brasil (ULBRA), Canoas, RS, Brazil.
[D]Instituto de Biotecnologia, Universidade de Caxias do Sul (UCS), Caxias do Sul, RS, Brazil.
Corresponding author. Email: pegas@cbiot.ufrgs.br

Abstract. The eukaryotic ATP-dependent DNA ligases comprise a group of orthologous proteins that have distinct roles in DNA metabolism. In contrast with the well-known DNA ligases of animal cells, the DNA ligases of plant cells are poorly described. Until now, only two DNA ligases (I and IV) genes of *Arabidopsis thaliana* (L.) Heynh were isolated and characterised. Use of the complete genomic sequences of *Oryza sativa* L. and *A. thaliana*, as well as the partially assembled genomic data of *Medicago truncatula* L. and *Brassica* spp., allowed us to identify a new family of ATP-dependent DNA ligases that are found only in the *Viridiplantae* kingdom. An in-depth phylogenetic analysis of protein sequences showed that this family composes a distinct clade, which shares a last universal common ancestor with DNA ligases I. *In silico* sequence studies indicate that these proteins have distinct physico-chemical properties when compared with those of animal and fungal DNA ligases. Moreover, hydrophobic cluster analysis and 3-dimensional modelling allowed us to map two conserved domains within these DNA ligases I-like proteins. Additional data of microsynteny analysis indicate that these DNA ligases I-like genes are linked to the *S* and *SLL2* loci of *Brassica* spp. and *A. thaliana*, respectively. Combining the results of all analyses, we propose the creation of the DNA ligases VI (*LIG6*) family, which is composed by plant-specific DNA ligases.

Keywords: DNA ligase I, eukaryotic DNA ligases, hydrophobic cluster analysis, plant DNA ligases, protein phylogeny, Pso2p / Snm1p / Artemis.

Introduction

Maintenance of genomic integrity is necessary for the viability of cells and the health of prokaryotic and eukaryotic organisms. Some metabolic processes, such as DNA replication, DNA recombination or DNA repair, can yield genotoxic DNA lesions by introducing gaps in one or both DNA strands (Tuteja *et al*. 2001). Normally, these gaps are sealed by a group of specialised and ubiquitous enzymes named DNA ligases. Belonging to the superfamily of nucleotidyltransferases, DNA ligases can be grouped into two classes by virtue of co-factor requirement for either NAD^+ (EC 6.5.1.2, prokaryotic DNA ligases) or ATP (EC 6.5.1.1, mainly eukaryotic DNA ligases).

The family of ATP-dependent DNA ligases is structurally and functionally conserved in all eukaryotes (Martin and MacNeill 2002), sharing a common ancestor with mRNA-capping enzymes (Schuman and Schwer 1995). Their functional structure is well characterised and features at least two conserved domains, the catalytic core domain (CD) and the non-catalytic domain (NCD). The CD is composed of six sequence motifs related to prokaryotic NAD^+-dependent ligases and it is highly conserved between DNA ligases and

Abbreviations used: aa, amino acid; BER, base excision repair; CD, catalytic core domain; CRs, conserved regions; CS, conserved sequence; DSB, double-strand DNA breaks; ESTs, expressed sequence tags; HCA, hydrophobic cluster analysis; MCE, mRNA-capping enzymes; NCD, non-catalytic domain; NHEJ, non-homologous end-joining; NJ, neighbour-joining; NP, nucleoplasmin; PCNA, proliferating cell nuclear antigen; SI, self-incompatibility; SSB, single-strand DNA breaks.

© CSIRO 2005 10.1071/FP04143 1445-4408/05/020161

mRNA-capping enzymes (Schuman and Schwer 1995). The CD is responsible for joining the phosphodiester backbone single-strand breaks in a three-step manner: a complex of AMP and DNA-ligase catalyses the transfer of activated AMP to the 5′ phosphate of the nick, allowing the formation of the phosphodiester bond, and then seals the nick with the concomitant release of AMP (Wilkinson and Bowater 2001). The function of NCD on DNA ligase activity is largely unknown (Martin and MacNeill 2002).

Until now, five subclasses of eukaryotic DNA ligases were described — DNA ligase I to V — with DNA ligase I, III and IV being the best characterised in terms of genetic and biochemical properties. DNA ligases I, III and IV have important roles in DNA metabolism, acting in replication, repair and recombination (Timson *et al*. 2000). The function of DNA ligase V has yet to be identified, and DNA ligase II seems to be a degradation product of DNA ligase III (Martin and MacNeill 2002).

In contrast with mammalian cells, little is known about the number and function of DNA ligases in plant cells. Some early work on plant DNA ligases has partially analysed the activity of these enzymes in extracts of *Pisum sativum* L. (Kessler 1971), in *Daucus carota* L. (Tsukada and Nishi 1971), in *Lilium* microspores (Howell and Hecht 1971), and in *Secale cereale* L. (Elder *et al*. 1987). Recently, two genes of *A. thaliana* corresponding to the DNA ligases I and IV have been isolated and characterised (Taylor *et al*. 1998; West *et al*. 2000). Both proteins show high functional and structural similarity with DNA ligases I and IV of fungi and animals, indicating their evolutionary conservation related to DNA metabolism in eukaryotes (Taylor *et al*. 1998; West *et al*. 2000).

Despite the valuable biological information made available by the completed genomic projects of *A. thaliana* (The *Arabidopsis* Genome Initiative 2000) and *Oryza sativa* (Yu *et al*. 2002), as well as partial genomic analysis of different plant species, e.g. *Medicago truncatula* (Waugh *et al*. 2002) and *Brassica* spp. (*Brassica* Genome Initiative; http://brassica.bbsrc.ac.uk/ validated 19 January 2005), little attention has been paid to screening for potential new DNA ligases that could have different functions in plants, e.g. repair of DNA damage induced by sunlight or phytopathogens, DNA replication or DNA recombination.

In this work we have identified and characterised a new group of plant-specific DNA ligases (which we named DNA ligase VI or *LIG6*) that can be found in plant genomic databases using sensitive methods of phylogenetic analysis. Additional hydrophobic cluster analysis (HCA) and 3-dimensional protein modelling have allowed us to refine the results obtained from phylogeny and to map conserved domains of this new group of eukaryotic ATP-dependent DNA ligases. Microsynteny analysis of *LIG6* points out to specific functions of DNA ligase VI in plant physiology.

Materials and methods

Protein data mining and sequence analyses

Forty protein sequences (representing the superfamily of eukaryotic nucleotidyltransferases) were obtained directly from GenBank hosted in the National Center for Biotechnological Information (NCBI) web page (http://www.ncbi.nlm.nih.gov/ validated 19 January 2005). BLAST and PSI-BLAST programs were used for initial domain screening and comparison (Altschul *et al*. 1997). DNA ligase I of *Arabidopsis thaliana* (L.) Heynh (Table 1) was used as a query to search possible orthologous / paralogous proteins in Plant Genomes Central (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html validated 19 January) and in the *Solanaceae* Genomics Network (http://soldb.cit.cornell.edu/index.html validated 19 January 2005). All searches were made to saturation. The theoretical pI and molecular weight of proteins were calculated with COMPUTE pI / Mw program hosted in ExPASy Molecular Biology Server at Swiss Institute of Bioinformatics (http://us.expasy.org/ validated 19 January 2005).

Algorithms for sequence comparison

Global pair-wise multiple-alignments were performed with the amino acid (aa) sequences in the ClustalX 1.8 program (Thompson *et al*. 1994). The alignment parameters used were: gap open penalty 10.00; gap extension 0.20; sequences >10% diverged delayed; BLOSUM series matrix; residue-specific penalties on; and hydrophylic penalties on. When necessary the alignments were manually adjusted with the BioEdit program (Hall 1999).

Gene prediction and microsynteny analysis

Wise2 program at European Bioinformatics Institute (http://www.ebi.ac.uk/Wise2/ validated 19 January 2005) was used to identify plant-specific DNA ligase genes from the unfinished genomic sequences of *Brassica rapa* L., *Brassica napus* L., and *Medicago truncatula* L. The parameters for prediction were: local mode; no intron bias; splice site modelled; synchronous model; and GeneWise623 algorithm. The DNA ligase VI of *A. thaliana* (AthLig6p) was used as query in all searches. Microsynteny analysis was initially made with TBLASTX programme (NCBI) using as a query a genomic fragment of approximately 30 495 nt from chromosome 1 of *A. thaliana* (GenBank accession number NC_003070.4, nt sequence from 24 938 643 to 24 969 137) which contains the locus of *AthLIG6*. The microsynteny analysis was restricted to *Viridiplantae* database keeping the standard parameters.

Algorithms for phylogenetic inference

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 2.1 (Kumar *et al*. 2001). Neighbour-joining (NJ) method was used for phylogenetic tree searching and inference. The statistical reliability of the phylogenetic trees was tested by bootstrap analysis with 1000 replications. Moreover, the Poisson correction was applied in NJ for distance estimation. The complete deletion option was used in handling gaps or missing data obtained from the alignment.

Hydrophobic cluster analysis

Hydrophobic cluster analysis (HCA) of selected sequences was performed as previously published (Gaboriaud *et al*. 1987), and consisted in displaying the primary protein structure on a duplicated α-helical net, where the hydrophobic residues are automatically contourned. These hydrophobic clusters observed in an HCA plot are not distributed in a random manner, but correspond highly with the secondary protein structures associated with conserved globular domains. Thus, HCA plots are extremely valuable for phylogenetic inferences when protein sequences have a weak homology (<25% of identity/similarity) or to define conserved domains and sequence

Table 1. Protein sequences used in the present study

Protein group	Sequence name	Species	GenBank acession no.
ATP-dependent DNA ligases I			
	DNA ligase I	*Arabidopsis thaliana*	gi|1359495|
	DNA ligase I	*Caenorhabditis elegans*	gi|17562602|
	DNA ligase I	*Schizossacharomyces pombe*	gi|118774|
	DNA ligase I	*Crithidia fasciculata*	gi|312384|
	DNA ligase I	*Homo sapiens*	gi|4557719|
	DNA ligase I	*Mus musculus*	gi|6754544|
	DNA ligase I	*Oryza sativa*	gi|16905197|
	DNA ligase I	*Rattus norvegicus*	gi|13540673|
	DNA ligase I	*Saccharomyces cerevisiae*	gi|3515|
	DNA ligase I	*Xenopus laevi*	gi|2118374|
ATP-dependent DNA ligases VI			
	DNA ligase VI	*Arabidopsis thaliana*	gi|12597768|
	DNA ligase VI	*Oryza sativa*	gi|20805031|
	DNA ligase VI	*Brassica napus*	gi|7657880|
	DNA ligase VI	*Brassica rapa*	gi|30314605|
	DNA ligase VI	*Medicago truncatula*	gi|38257195|
ATP-dependent DNA ligases III			
	DNA ligase III-α	*Homo sapiens*	gi|7710126|
	DNA ligase III-α	*Mus musculus*	gi|1794223|
	DNA ligase III-α	*Xenopus laevi*	gi|18029882|
	DNA ligase III-⇓	*Homo sapiens*	gi|4504995|
	DNA ligase III-⇓	*Mus musculus*	gi|1794221|
	DNA ligase III-⇓	*Xenopus laevi*	gi|18029884|
ATP-dependent DNA ligases IV			
	DNA ligase IV	*Arabidopsis thaliana*	gi|9651815|
	DNA ligase IV	*Candida albicans*	gi|14916982|
	DNA ligase IV	*Drosophila melanogaster*	gi|7292907|
	DNA ligase IV	*Gallus gallus*	gi|15778121|
	DNA ligase IV	*Homo sapiens*	gi|860937|
	DNA ligase IV	*Saccharomyces cerevisiae*	gi|2494163|
	DNA ligase IV	*Schizossacharomyces pombe*	gi|19075388|
	DNA ligase IV	*Xenopus laevi*	gi|18029886|
mRNA-capping enzymes			
	MCE1	*Arabidopsis thaliana*	gi|23306380|
	MCE1	*Candida albicans*	gi|1783271|
	MCE1	*Drosophila melanogaster*	gi|7292972|
	MCE1	*Encephalitozoon cuniculi*	gi|19171287|
	MCE1	*Homo sapiens*	gi|3097308|
	MCE1	*Mus musculus*	gi|6685627|
	MCE1	*Saccharomyces cerevisiae*	gi|1246907|
	MCE1	*Schizossacharomyces pombe*	gi|2239238|
	MCE1	*Xenopus laevi*	gi|7239232|
	MCE1-like	*Arabidopsis thaliana*	gi|15232015|
	MCE1-like	*Arabidopsis thaliana*	gi|15241729|

signatures in a set of homologous proteins (Gaboriaud *et al*. 1987). It should be noted that HCA, when compared with 'linear' or one-dimensional methods of database screening, e.g. BLAST and PSI-BLAST (which needs, at least, >30% of aa identity/similarity over a sufficient length), is a more effective tool to detect sequence similarity that reflects a true 3-dimensional relationship between two or more proteins (Callebaut *et al*. 1997). Moreover, HCA has the advantage that sequence conservation can be detected in a group of proteins without any previous one-dimensional alignment (Lemesle-Varloot *et al*. 1990; Callebaut *et al*. 1997). The program DRAWHCA, used in such analysis, is available as a freeware at http://www.lmcp.jussieu.fr/ (validated 19 January 2005).

3-Dimensional modelling

In silico 3-dimensional modelling of *N*- and *C*-terminal sequences of AthLig6p was performed with 3D-JIGSAW (Bates *et al*. 2001) and Swiss-Model (Guex and Peitsch 1997) algorithms using as template the penicillinase sequence of *Pseudomonas aeruginosa*, which

belongs to the metallo-⇓-lactamase superfamily (Protein Data Bank accession number 1dd6); and the ATP-dependent DNA ligase from bacteriophage T7 (Protein Data Bank accession number 1a0i). In both algorithms the default parameters were applied. 3-Dimensional comparative models were generated by Swiss-PdbViewer software (Guex and Peitsch 1997), available at http://www.expasy.org/spdbv (validated 19 January 2005).

Results

Sequence and phylogenetic analysis of plant-specific DNA ligase I-like proteins

In order to identify new potential DNA ligase sequences in plants, a genomic analysis approach was taken from the complete genomes of *A. thaliana* and *O. sativa*, which are available at Plant Genomes Central. Using the DNA ligase I of *A. thaliana* (AthLig1p; Table 1) as query in BLAST and PSI-BLAST programs, we identified three DNA ligase I-like sequences belonging to *O. sativa*, *A. thaliana*, and *M. truncatula*. Moreover, when the option TBLASTX (protein query against all six frames translated database) was used to search for possible orthologous in the genomes of other plant species, we found two complete DNA ligase I-like sequences in the partially assembled genomic data of *Brassica rapa* and *B. napus*. Also, a short sequence fragment corresponding to the *C*-terminus of a DNA ligase I-like protein was found in five assembled expressed sequence tags (ESTs) of *Solanun tuberosum* and in the green algae *Chlamydomonas reinhardtii* (data not shown).

With the objective of determining the phylogenetic relationship of the DNA ligase I-like proteins with the superfamily of nucleotidyltransferases, these sequences were compared with eukaryotic DNA ligases and mRNA-capping enzymes (MCE). Interestingly, the phylogenetic data indicated the divergence of two clades that are strongly supported by bootstrap analysis from a last common universal ancestor (LUCA; Fig. 1). One group, the clade of plant-specific DNA ligase I-like proteins has diverged early from a LUCA with the classical DNA ligases I of fungi, metazoans, and plants, having a high bootstrap support (Fig. 1). We propose to call these plant-specific DNA ligase I-like proteins as DNA ligase VI. The vertebrate DNA ligase III family composes a distinct group that diverged early from the same common ancestor that gave rise to the DNA ligase I families. As expected, both MCE and DNA ligase IV compose two distinct and ancient clades, which probably arose from an ancestral enzyme with nucleotidyltransferase function to act in the processing of the eukaryotic mRNA and in DNA repair, respectively.

Phylogenetic data lead us to analyse the physico-chemical properties of the plant-specific DNA ligase I-like proteins and to compare them with the best-studied *A. thaliana* DNA ligases (AthLig1p and AthLig4p). The primary sequence analysis of the DNA ligase I-like proteins of *A. thaliana* indicates a protein with high molecular weight (>150 kDa) in comparison with AthLig1p (87.8 kDa) and AthLig4p (137.8 kDa). Moreover, a theoretical pI was calculated for DNA ligase I-like protein of *A. thaliana* and compared with the other two *A. thaliana* DNA ligases. The result indicates that AthLig6p is an acidic protein (pI 5.88), while AthLig1p (pI 8.20) and AthLig4p (pI 8.24) are basic proteins. Similar results were obtained with other plant-specific DNA ligase I-like sequences, including the DNA ligase I-like sequence of *C. reinhardtii* (data not shown).

Hydrophobic cluster analysis of plant DNA ligase I-like proteins

Hydrophobic cluster analysis is a sensitive method of sequence comparison that detects 2- and 3-dimensional similarities between protein domains showing very limited aa relatedness, typically below the so-called 'twilight zone' (25–30%) (Gaboriaud *et al*. 1987). The method uses aa hydrophobic cluster plots to identify globular domains associated with structural or functional conserved features. In this case, moderate or highly hydrophobic aa form clusters of various shape and size. Thus, a horizontal diamond-shaped cluster correspond to the hydrophobic face of an α-helix; internal ⇓-strands are represented by a cluster that alters hydrophobic residues on the left and on the right of an HCA plot; and surface strands are shown by 'zig-zag' patterns. Moreover, conserved protein domains can be mapped with HCA using orthologous sequences from different species. In the case of *O. sativa* and *A. thaliana* DNA ligase I-like proteins we mapped conserved domains using two orthologous sequences of *S. cerevisiae* (Fig. 2*A*, *B*), which were previously obtained using the AthLig6p as a query in the PSI-BLAST program. One of these sequences is Pso2p / Snm1p, a protein belonging to the ⇓-CASP family that is involved in DNA repair of interstrand cross links induced by chemical agents like furocoumarins or bi-functional mustards. The HCA allowed us to identify of three conserved regions (CRs) in the *N*-terminus of Lig6p, which are characteristic of Pso2p / Snm1p group. The first CR shows the typical signature of the metallo-⇓-lactamase family, i.e. the highly conserved HxHxDH sequence situated between α-helix structures (Fig. 2*A*). CRII and CRIII show the same conserved secondary structures typical for the Pso2p / Snm1p family. In the case of CRII, the signature (D, E)ANHCPG could be found between two ⇓-sheet structures (Fig. 2*A*). CRIII shows a conserved (V, I)GKEK signature between ⇓-sheet structures (Fig. 2*A*). In all plant Lig6p analysed so far, CRs are situated at the beginning of the protein *N*-terminus (Fig. 3). The HCA homology score calculated for CRs gives a high level of similarity (>80%) of primary and secondary structures (Table 2). An average score could be obtained when the CRs were globally aligned (Table 2) and a BLAST comparison between Pso2p / Snm1p, AthLig6p, and OsaLig6p indicates an average level of similarity of 49% between the primary sequences.

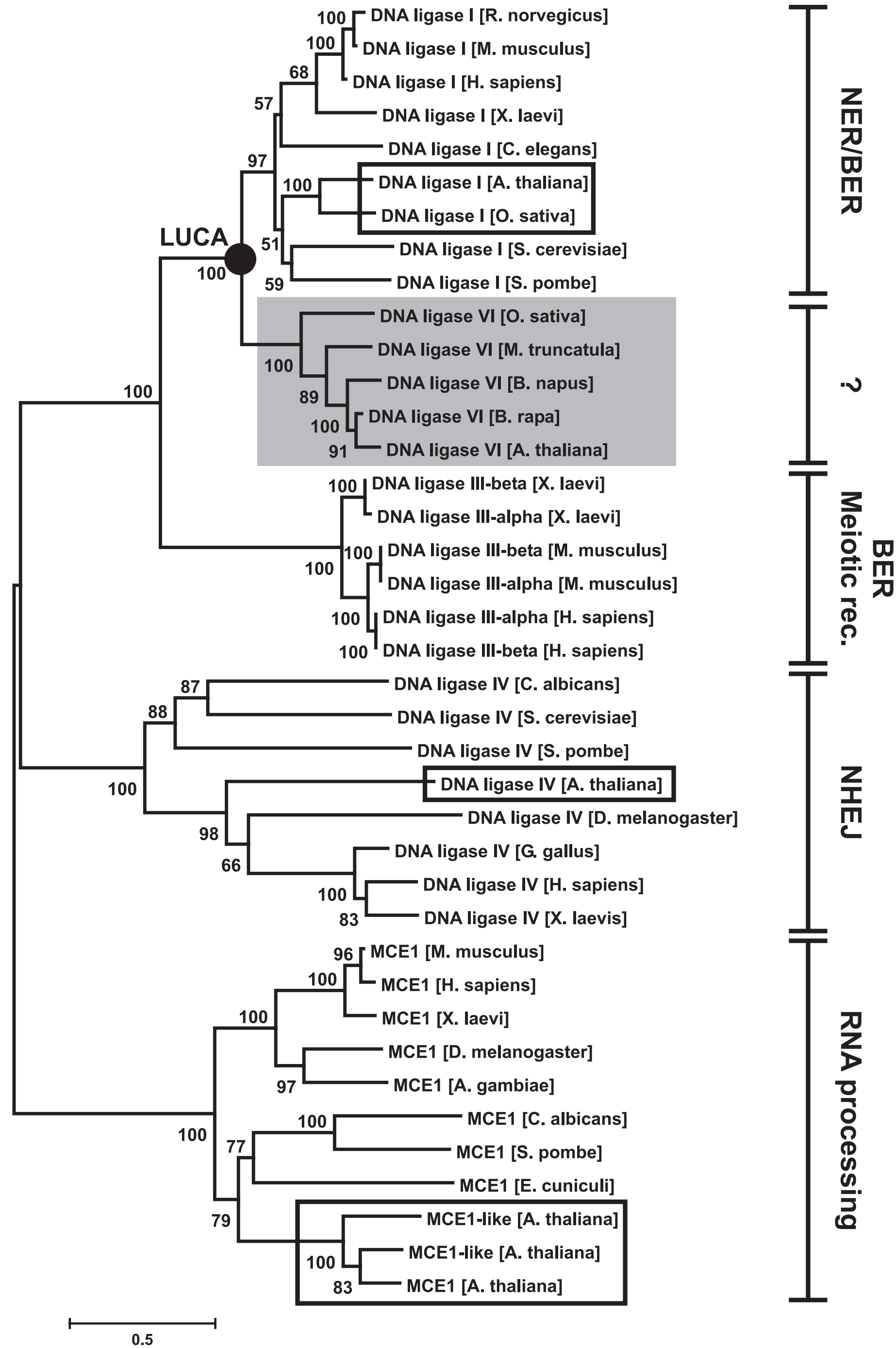


Fig. 1. Unrooted phylogenetic tree for DNA ligases and mRNA-capping enzymes. Numbers at nodes indicate bootstrap values (1000 replicates) and the horizontal bar represents a distance of 0.5 substitutions per site. Clusters with bootstrap values below 50% were collapsed. The clade of plant-specific DNA ligase VI (Lig6p) is indicated by a grey box. AthLig1p, AthLig4p, and OsaLig1p as well as the paralogous mRNA capping enzymes of *A. thaliana* are indicated by an empty box on the phylogram. The known functions of each clade on DNA metabolism are show at the right of the phylogram. LUCA, last universal common ancestor; NER, nucleotide excision repair; BER, base excision repair; NHEJ, non-homologous end joining repair; Meiotic rec, meiotic recombination; ?, unknown function.

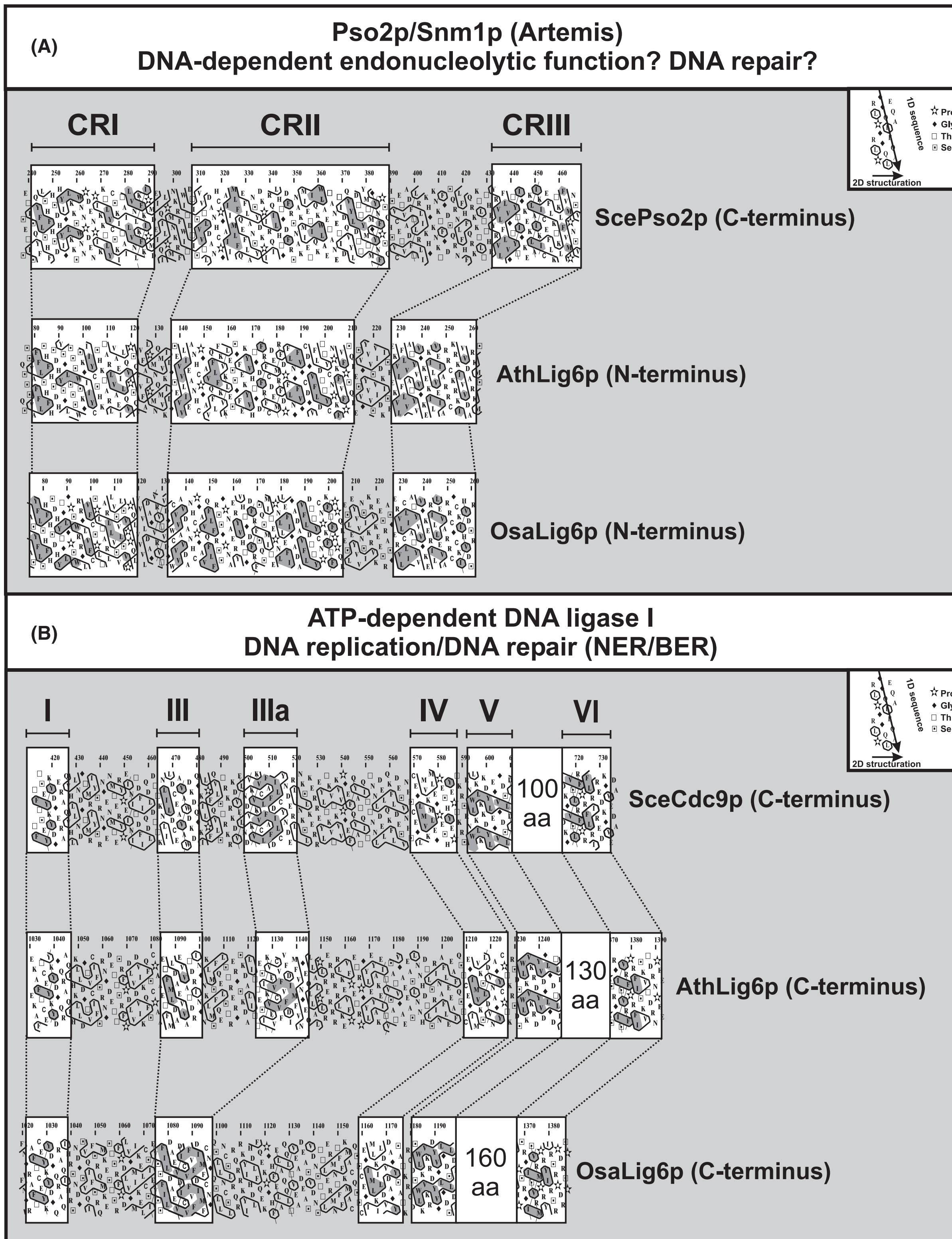


Fig. 2. Hydrophobic cluster analysis of *N*- and *C*-termini of AthLig6p and OsaLig6p. In (*A*) the *N*-termini of AthLig6p and OsaLig6p were compared with the Pso2p / Snm1p of *S. cerevisiae*, and the three conserved regions (CRI-III) were highlighted with an empty box. In (*B*) the *C*-termini of AthLig6p and OsaLig6p were compared with the Cdc9p of *S. cerevisiae*, and the DNA ligase I core (domains I to VI) was highlighted in the same way as described above. Conserved hydrophobic aa residues are shaded in grey. The way to read the sequences and special symbols is indicated in the inset.

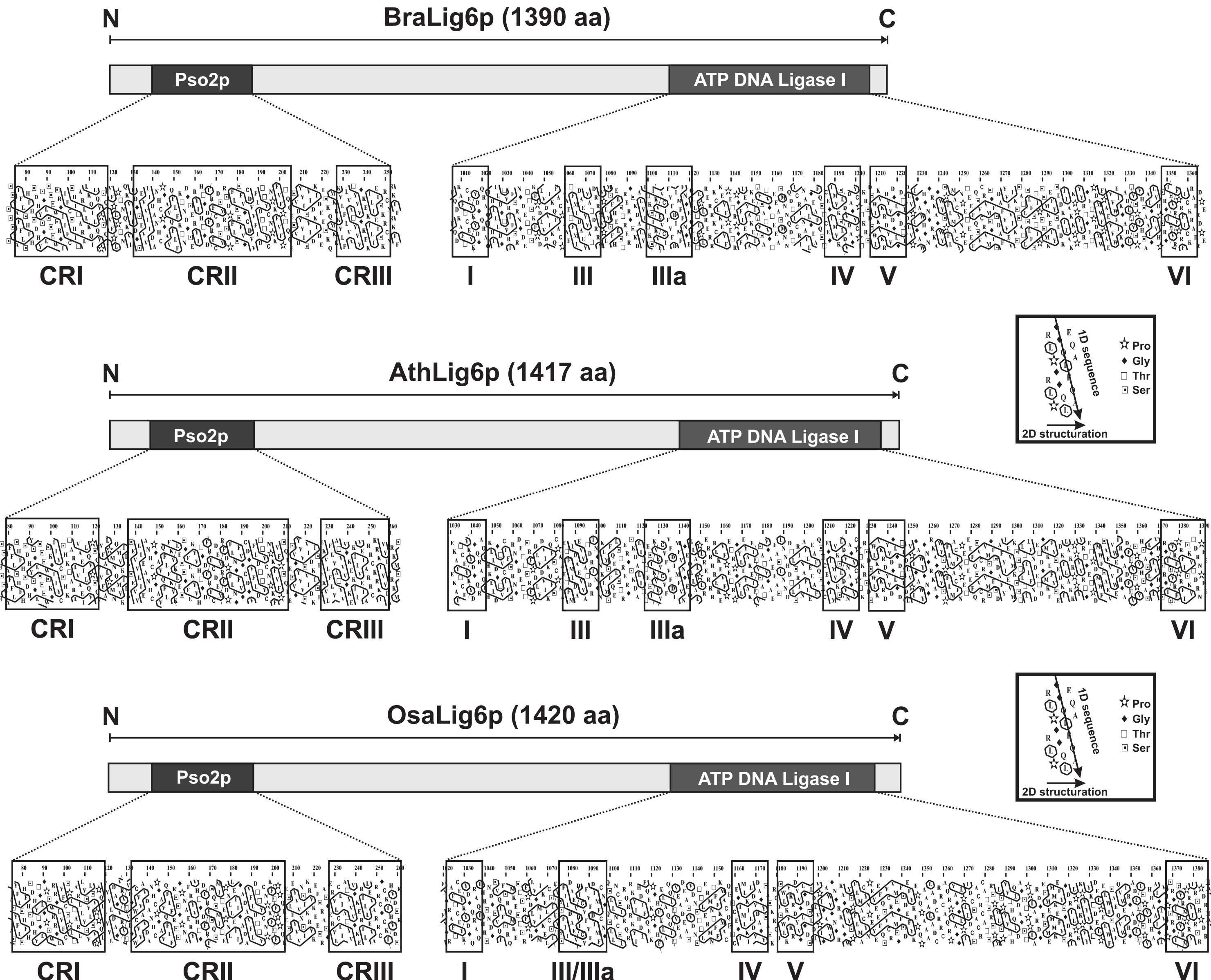


Fig. 3. Comparison of conserved regions of BraLig6p, AthLig6p, and OsaLig6p. The Pso2p / Snm1p in the *N*-terminus of sequence as well as the DNA ligase core domain in the *C*-terminus are indicated by a HCA draw on scheme. The conserved globular domains of Pso2p (CRI-III) and DNA ligase I core (I-VI) are indicated by an empty box. The way to read the sequences and special symbols are indicated in the inset.

Table 2. Homology scores calculated from global alignment and HCA using the Pso2p / Snm1p and DNA ligase I domains from Pso2p, Cdc9p, AthLig6p and OsaLig6p sequences

Domains		Homology score (%) HCA	Global alignment
Pso2p / Snm1p			
	CRI	82	64
	CRII	81	57
	CRII	81	69
DNA ligase I			
	I	100	86
	III / IIIa	84	52
	IV	84	73
	V	100	88
	VI	88	72

The *C*-terminus of plant Lig6p contains the classical domains (or 'core') of a DNA ligase I protein (Figs 2*B*, 3), as found in the Cdc9p of *S. cerevisiae*, in T7 DNA ligase, or in mammalian DNA ligase I (Timson *et al*. 2000). These six domains were easily identified by HCA in Cdc9p, AthLig6p and OsaLig6p and also in all other plant Lig6p analysed, with a high level of identity among aa residues (Fig. 2*B*; Table 2). A BLAST analysis of Cdc9p, AthLig6p and OsaLig6p shows an average level of similarity of 54% within the aa sequences. Interestingly, domains III and IIIa, which are separated by a loop of ± 20 aa residues in Cdc9p and AthLig6p, are fused in OsaLig6p, composing a unique globular domain (Fig. 2*B*). Moreover, this fusion was not found in any other analysed Lig6p.

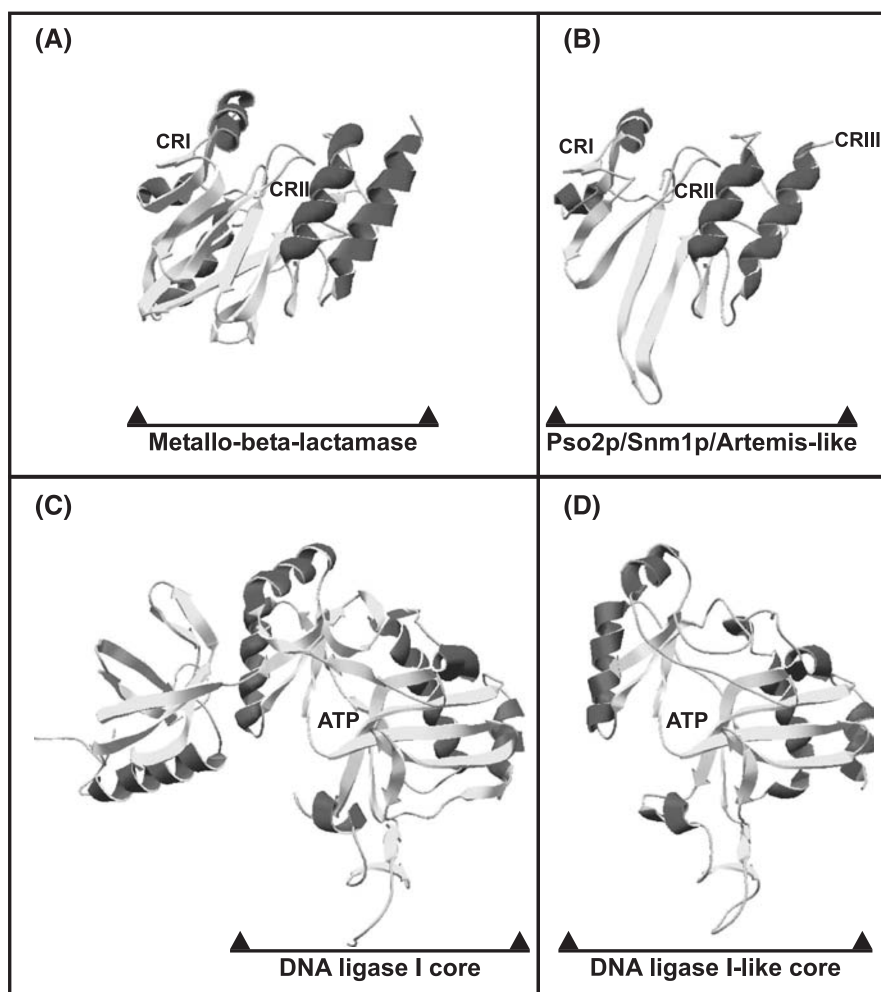


Fig. 4. 3-Dimensional modelling of Pso2p (*A*, *B*) and DNA ligase I core (*C*, *D*) domains of plant-specific Lig6p. The metallo-⇓-lactamase domain of the penicillinase of *P. aeruginosa* (*A*) as well as the Pso2p / Snm1p / Artemis-like domain of AthLig6p (*B*) are shown with their respective conserved regions (I and II). CRIII could not be modelled due to its low similarity with penicillinase sequence. The ATP-dependent DNA ligase of bacteriophage T7 (*C*) and DNA ligase I-like core of AthLig6p show high similarity of tertiary structure. The ATP binding site is indicated in both DNA ligase cores.

3-Dimensional modelling of the N- and C-termini of DNA ligase I-like coding sequences

3-Dimensional modelling was performed using the sequences of the *N*- and *C*-termini of Lig6p to corroborate the data obtained by HCA (Fig. 4). The Pso2p / Snm1p metallo-⇓-lactamase domain exhibited some degree of similarity to the enzyme penicillinase of *Pseudomonas aeruginosa*, also a metallo-⇓-lactamase protein. Using this information, we modelled the *N*-terminus of Lig6p. As expected, this resulted in a similar 3-dimensional conformation of Lig6p *N*-terminus and penicillinase (Fig. 4*A*, *B*), with the CRI and CRII domains present in both proteins. Unfortunately, the CRIII domain is only present in the Pso2p / Smn1p and Lig6p families and it could not be modelled using the penicillinase sequence. The Lig6p *C*-terminus has a good level of tertiary structure homology with the ATP-dependent DNA ligase of bacteriophage T7, a DNA ligase I prototype enzyme (Fig. 4*C*, *D*). The DNA ligase I core could be easily identified in Lig6p using the structural information obtained from the crystallographic studies of T7 DNA ligase, again corroborating the data obtained by HCA.

Genomic location and microsynteny of DNA ligase I-like in different plant species

In an attempt to locate the DNA sequences of *LIG6* in plant chromosomes, we used the information available in the literature and in Plant Genomics Database. Both *Brassica LIG6* genes were found linked to the *S* locus, and *AthLIG6* was also found linked to the *SLL2* locus of *A. thaliana* (Fig. 5), which is known to control the reproductive pattern of members of the *Brassicaceae* family (Matton *et al*. 1994). These regions are highly syntenic with each other. In *O. sativa* and *M. truncatula*, the *LIG6* gene also was found in a region

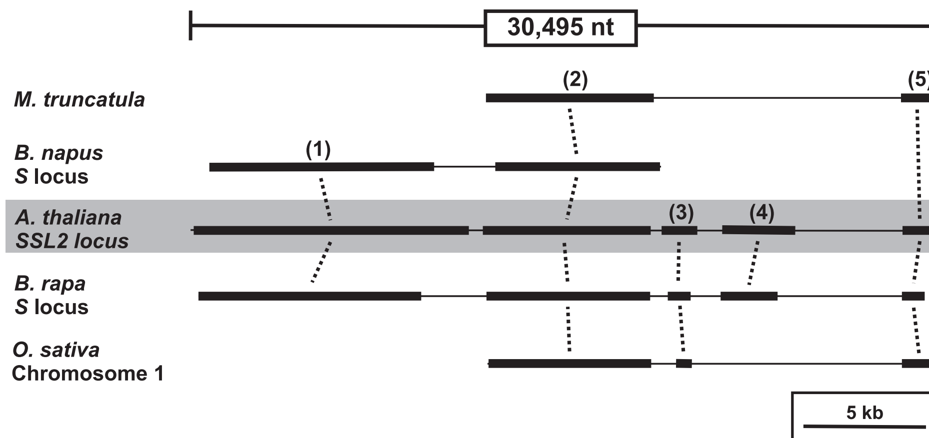


Fig. 5. Schematic microsynteny relationships between *LIG6* gene-containing regions of *M. truncatula*, *B. napus*, *A. thaliana*, *B. rapa*, and *O. sativa*. Regions are drawn approximately to scale (scale bar indicates 5 kb). Similarity between the chromosomal sequences is indicated: - - - -. Predicted genes using *A. thaliana* sequences (grey box) are: (1) At1 g66680, putative *S* locus-linked protein; (2) At1 g66730, *LIG6*; (3) At1 g66750, putative cell division protein kinase; (4) At1 g66760, MATE efflux family protein; and (5) At1 g66780, MATE efflux family protein.

Table 3. Analysis of Lig6p from different plant species

Chromosome no. indicates chromosomal location of DNA ligase VI, *S* locus indicates whether DNA ligase VI are associated with *S* locus or *S* locus-like, compatibility indicates the reproductive compatibility of plant (SC, self-compatible; SI, self-incompatible), microsynteny indicates whether the region where DNA ligase VI is located shows synteny between different plant species. *AthLIG6* is situated in *SSL* locus, which is syntenic with *S* locus of *Brassica*. *BraLIG6* and *BnaLIG6* are situated in *S* locus, which are syntenic between different *Brassica* species and with *SSL* locus of *A. thaliana. OsaLIG6* is situated in a region of chromosome 1 which is syntenic with *SSL* locus of *A. thaliana.* ND, not determined

Gene	Species	Family	Chromosome no.	*S* locus	Compatibility	Microsynteny
AthLIG6	*A. thaliana*	Brassicaceae	1	Yes	SC	Yes
BraLIG6	*B. rapa*	Brassicaceae	ND	Yes	SI	Yes
BnaLIG6	*B. napus*	Brassicaceae	ND	Yes	SI	Yes
OsaLIG6	*O. sativa*	Poaceae	1	ND	SC	Yes
MtrLIG6	*M. truncatula*	Fabaceae	ND	ND	SC	Yes

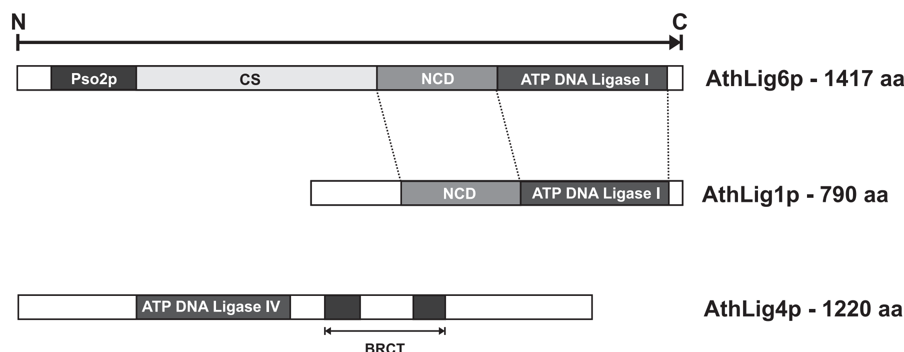


Fig. 6. Schematic representation of Lig6p, Lig1p, and Lig4p of *A. thaliana*. The direction of proteins, from *N*-terminus to *C*-terminus, is indicated by an arrow. CS, conserved sequence; NCD, non-catalytic region; BRCT, BRCA tandem binding domains. The dotted lines between AthLig1p and AthLig6p indicate high level of primary and secondary sequence homologies.

that shows microsynteny with *SLL2* locus of *A. thaliana* (Table 3; Fig. 5).

Discussion

The phylogenetic and primary sequence data analyses, together with HCA and 3-dimensional modelling, indicated that plant DNA ligase I-like proteins comprise a new family of nucleotidyltransferases that diverged early in the evolution of DNA ligase I. This new protein group, which is supported by bootstrap analysis, was called DNA ligase VI (*LIG6*). A comparison of DNA ligases I, IV and VI of *A. thaliana* showed the structural similarities and differences among these enzyme families (Fig. 6). Besides the conserved *C*-termini found between Lig1p and Lig6p, Lig6p also contains an *N*-terminus that shows a high level of conservation of primary and secondary structures with the eukaryotic Pso2p / Snm1p and a conserved sequence (CS), which is unique for these proteins, and probably has a role in the protein's function. It is interesting to note that the CS of all Lig6p contains (around aa 520 or 600, depending on the Lig6p analysed) a typical proliferating cell nuclear antigen (PCNA) interacting protein motif (PIP box, data not shown). The PIP motif, present in the *N*-termini of fungal and metazoan Lig1p, is composed by the consensus signature Qxx(M / I / L)xx(F / Y)(F / Y), which is necessary for association of Lig1p with PCNA, helping to stabilise Lig1p during Okasaki fragment joining and ligation steps of the long patch base excision repair (BER) (Warbrick 1998; Tom *et al*. 2001). Recent studies have indicated that PCNA has important roles not only in DNA replication but also in DNA repair, including nucleotide excision repair, post-replicational mismatch repair, BER, apoptosis and cytosine methylation (Chuang *et al*. 1997). PCNA is also known to interact with DNA polymerase ε and p21, and is probably one of the key proteins involved in controlling cell cycle progression (Warbrick *et al*. 1995). The functions of PCNA in plants are not fully understood, but gene expression assays indicate that PCNA mRNA is present in high copies in rapidly dividing tissues such as flower buds, apical meristems, and young leaves, while mature stems and fully expanded leaves have significantly lower levels of PCNA transcripts (Markley *et al*. 1993; Lopez *et al*. 1997; Kimura *et al*. 2004). Thus, the presence of a PIP box in the Lig6 protein strongly indicates that this new DNA ligase family has an important function in maintaining DNA integrity in rapidly dividing tissues.

Until recently, only two types of DNA ligases were characterised in *A. thaliana*. DNA ligase I (AthLig1p) has the typical CD and NCD domains, with a high level of identity / similarity with DNA ligases I from fungi and animals. In the thermo-conditional *cdc9* mutant of *S. cerevisiae AthLIG1* was able to complement the phenotype at the non-permissive temperature, and studies of tissue-specific expression of *AthLIG1* detected this mRNA in all tissues, being highest in the flowering structures, which contain an elevated proportion of cells undergoing mitosis and meiosis (Taylor *et al*. 1998). On the other hand, DNA ligase IV (AthLig4p) contains the characteristic DNA ligase core of all members of the DNA ligase family, sharing a low similarity with the DNA ligase IV of *S. cerevisiae* and mammals (West *et al*. 2000). The presence of two BRCT domains in the *C*-terminus of AthLig4p is also a characteristic of this group, indicating an interaction with small acidic proteins (Tomkinson and Mackey 1998). Moreover, transcripts of *AthLIG4* were identified in all tissues studied, with the highest levels of expression observed in young flowers at the time of bud opening and in roots. Expression of the *AthLIG4* gene is also regulated by . -irradiation, which causes both single- and double-strand DNA breaks (SSB and DSB, respectively), and this observation is consistent with the involvement of Lig4p in DSB repair (West *et al*. 2000).

The HCA mapping of conserved regions in Lig6p from different plant species shows two highly conserved

domains: (i) a Pso2p / Snm1p domain in the *N*-terminus of the protein (Figs 2*A*, 3); and (ii) a DNA ligase I domain situated in the *C*-terminus of the sequence (Figs 2*B*, 3). Data obtained from three-dimensional modelling of Lig6p *N*- and *C*-termini with penicillinase of *P. aeruginosa* (which belongs to the metallo-⇓-lactamase superfamily), and T7 DNA ligase (the prototype enzyme of DNA ligase I family), also supported the HCA-derived information (Fig. 4*A*–*D*).

The Pso2p / Snm1p belongs to the ⇓-CASP family, a group of related proteins using nucleic acids as substrate, acting in DNA repair, RNA processing, and V(D)J recombination (Callebaut *et al*. 2002). This family is included in the metallo-⇓-lactamase superfamily, whose members show a catalytic action on ester type ligations. HCA allowed this group to be identified in all three Life Domains (Callebaut *et al*. 2002). Using HCA, we have recently mapped three CRs in the *C*-terminus of Pso2p / Snm1p of *S. cerevisiae* that may be associated with protein function (data not shown). The best-characterised member of the ⇓-CASP family is Artemis, which was isolated from cells of patients suffering from a special type of severe combined immunodeficiency syndrome associated with radio-sensitivity (RS-SCID; Noordzij *et al*. 2003). Artemis has 5′ to 3′ exonucleolytic activity with single-strand DNA specificity and, when associated with DNA-PKcs, forms a phosphorylated complex with endonucleolytic activity on both 5′ and 3′ DNA overhangs; moreover it can cleave hairpins generated by the Rag1 / Rag2 proteins (Jeggo and O'Neill 2002; Schlissel 2002). Clearly, all experimental data show that *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; Brendel and Henriques 2001). However, they are not sensitive to UVC and ionising radiation (Ruhland *et al*. 1981; Magaña-Schwencke *et al*. 1982). 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 DSBR (Magaña-Schwencke *et al*. 1982; Wilborn and Brendel 1989). All results presented here give a good support on the role of the *N*-terminus of Lig6p in endonucleolytic processing of DNA, a function that was not described previously for an eukaryotic DNA ligase. Moreover, a new family of ATP-dependent DNA ligases in prokaryotes that contain a nuclease and a primase domain flanking the DNA ligase core of the proteins was discovered recently (Weller and Doherty 2001). Della *et al*. (2004), working with DNA ligase D from *Mycobacterium tuberculosis* (Mt-Lig), show that this ATP-dependent DNA ligase has a gap-filling polymerase, a terminal transferase, a primase, and a 3′ to 5′ exonuclease activities. Moreover, Zhu and Shuman (2004), also working with a DNA ligase D from *Pseudomonas aeruginosa*, analysed its intrinsic DNA polymerase activity function capable of templated and non-templated DNA primer-extension reactions. The experimental data indicate that the multi-domain DNA ligase D is recruited for bacterial DNA non-homologous end-joining (NHEJ) repair, probable processing the DNA extremities, and then proceeding to their rejoining. We suggest that the plant-specific Lig6p could act in an analogous mode as bacterial DNA ligase D, first processing the DNA extremities with its Pso2p / Snm1p domain to generate a substrate for its DNA ligase I-like domain.

The DNA ligase I core (domains I–V), which is situated at the *C*-terminus of Lig6p, comprises two anti-parallel ⇓-sheets surrounded by α-helices and short loops. The sheets are separated by a deep pocket that forms the nucleotide-binding site. The first region contains the sequence signature of the DNA ligase I family, consisting of the residues KYDG(E,Q)RA, essential to form an AMP adduct with DNA, the other motifs contribute with residues that line the walls of the pocket and contact the nucleotide triphosphate co-factor (Doherty and Suh 2000). The cores of Lig6p and Lig1p share a high level of identity in both primary and secondary structures (Fig. 2*B*), suggesting a possible function of Lig6p in DNA metabolism, specifically in BER, meiotic recombination or even DSBR. Participation of DNA ligase I at the replication forks of dividing chromosomes is well documented, there it functions in joining Okazaki fragments (Nasmyth 1979; Waga *et al*. 1994). Ligase I is also required for DNA polymerase ⇓-associated BER (Prasad *et al*. 1996).

In terms of physico-chemical properties, the Lig6p group contains the largest proteins of the nucleotidyltransferase superfamily, with sequence lengths >1390 aa. Moreover, Lig6p are acidic proteins, with a pI between 5.0 and 6.0, while all eukaryotic DNA ligases previously analysed are basic proteins (pI > 8.0), which explains their affinity for DNA molecules (Wu *et al*. 2001). Acidic proteins able to bind to chromatin and to induce changes in its structure have been described in metazoa. For example, nucleoplasmin (NP) is a nuclear protein that mediates the correct association of DNA with histones, enabling the formation of nucleosomes during early development (Bañuelos *et al*. 2003). Nucleomorphin, an acidic nuclear calmodulin-binding protein of *Dictyostelium discoideum*, also shows the property of heterochromatin binding in this organism, inducing significant changes in the structure of chromatin during *D. discoideum* development (Myre and O'Day 2002). Thus, the acidic charge of the Lig6p could also be associated with changes in the chromatin state required for meiotic recombination, DSB / BER repair, plant gametogenesis, or even for the developmental plasticity observed in plants (Wagner 2003).

Microsynteny analysis gave some clues about a possible physiological function for Lig6p by showing that in the *Brassicacea* family, the *LIG6* is linked to the *SLL2* locus of

(*A. thaliana*) and to the *S* locus of *Brassica* spp. Moreover, in *M. truncatula* and *O. sativa*, *LIG6* was located in a homeologous region of *A. thaliana SLL2* locus. The function of the *S* and *SLL2* loci in the reproductive pattern of angiosperms in controlling the self-incompatibility (SI) mechanisms is well known (Matton *et al*. 1994). Molecular analysis of the *Brassica S* locus region has shown that this Mendelian locus is a gene complex consisting of distinct stigma- and anther-expressed genes that determine SI specificity in stigma and pollen, respectively (Nasrallah 2000). Despite the fact that *A. thaliana*, *O. sativa* and *M. truncatula* are described as self-compatible species, the location of *LIG6* in the *S* locus could also be linked to other physiological processes. Evidence of a role for members of the *S* locus in a wide range of processes within the plant, including plant defence (Pastuglia *et al*. 1997) and development (Dwyer *et al*. 1994) is accumulating. Previous studies of DNA ligases of *A. thaliana* describe the highest levels of DNA ligase activity in young flowers (Taylor *et al*. 1998; West *et al*. 2000), especially AthLig1p. Taking into account that AthLig1p and AthLig6p share a common domain (DNA ligase I), and that the molecular assays employed to identify DNA ligase I expression do not have a good discrimination power (Taylor *et al*. 1998), it is probable that this high activity of DNA ligase in flower tissues could result from a combined action of both proteins or from a specific enzymatic activity of AthLig6p alone.

In contrast with animals, plants are constantly being challenged by UV radiation through the sunlight because of their obligatory requirement of sunlight for photosynthesis (Tuteja *et al*. 2001). By penetrating top layers of plant-tissues UV light can damage the cells' genomes, it may also negatively affect other cellular targets such as photosystem II and plasma membrane ATPase (Stapleton 1992). Characteristically, plants also show endophytic fungi living asymptomatically within their tissues (Faeth 2002). Many of these fungi produce mycotoxins that could be DNA damaging agents (Choi *et al*. 2001). Moreover, secondary plant metabolites (e.g. furocoumarins) can be photo-activated by sunlight to induce DNA cross-links in leaves or aerial parts of plants (Choi *et al*. 2001). It is thus likely that different DNA repair systems are required to fix the errors induced by biotic or abiotic factors in the plant's genome, giving DNA ligases an important role in these processes. Three or more DNA ligases are necessary in metazoans to keep their genomic health, and these organisms are less challenged by sunlight or pathogens (Timson *et al*. 2000). The data presented here suggest that Lig6p could be necessary to conserve the genome integrity in plant tissues, especially in reproductive organs with high DNA turnover. Biochemical analysis and mutational studies are currently in progress in order to determine the role of *LIG6* in plant physiology.

Acknowledgments

We thank Drs Arthur G. Fett-Neto and Giancarlo Pasquali for critical reading of this manuscript. This work was supported by research grants from Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS), CAPES and GENOTOX (Laboratório de Genotoxicidade, UFRGS).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402. doi: 10.1093/nar/25.17.3389

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815. doi: 10.1038/35048692

Bañuelos S, Hierro A, Arizmendi JM, Montoya G, Prado A, Muga A (2003) Activation mechanism of the nuclear chaperone nucleoplasmin: role of the core domain. *Journal of Molecular Biology* **334**, 585–593. doi: 10.1016/j.jmb.2003.09.067

Bates PA, Kelley LA, MacCallum RM, Stemberg MJE (2001) Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins: Structure, Function, and Genetics* **5**, 39–46. doi: 10.1002/prot.1168

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. doi: 10.1016/S1383-5742(01)00066-7

Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and Molecular Life Sciences* **53**, 621–645. doi: 10.1007/s000180050082

Callebaut I, Moshous D, Mornon J-P, de Villartay J-P (2002) Metallo-⇓-lactamase fold within nucleic acids processing enzymes: the ⇓-CASP family. *Nucleic Acids Research* **30**, 3592–3601. doi: 10.1093/nar/gkf470

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. doi: 10.1104/pp.125.2.752

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**, 1996–2000. doi: 10.1126/science.277.5334.1996

Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE, Wilson TE, Doherty AJ (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306**, 683–685. doi: 10.1126/science.1099824

Doherty AJ, Suh SW (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research* **28**, 4051–4058. doi: 10.1093/nar/28.21.4051

Dwyer KG, Kandasamy MK, Mahosky DI, Acciai J, Kudish BI, Miller JE, Nasrallah ME, Nasrallah JB (1994) A superfamily of *S* locus-related sequences in *Arabidopsis*: diverse structures and expression patterns. *The Plant Cell* **6**, 1829–1843. doi: 10.1105/tpc.6.12.1829

Elder RH, Dell´Aquila A, Mezzina M, Sarasin A, Osborne DJ (1987) DNA ligase in repair and replication in the embryos of rye *Secale cereale*. *Mutation Research* **181**, 61–71.

Faeth SH (2002) Are endophytic fungi defensive plant mutualist? *Oikos* **98**, 25–36. doi: 10.1034/j.1600-0706.2002.980103.x

Gaboriaud C, Bissery V, Benchetrit T, Mornon JP (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Letters* **224**, 149–155. doi: 10.1016/0014-5793(87)80439-8

Guex N, Peitsch MC (1997) Swiss-Model and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Howell SH, Hecht NB (1971) The appearance of polynucleotide ligase and DNA polymerase during the synchronous mitotic cycle in *Lilium* microspores. *Biochimica et Biophysica Acta* **240**, 343–352.

Jeggo P, O'Neill P (2002) The greek goddess, Artemis, reveals the secrets of her cleavage. *DNA Repair* **1**, 771–777. doi: 10.1016/S1568-7864(02)00096-4

Kessler B (1971) Isolation, characterization and distribution of a DNA ligase from higher plants. *Biochimica et Biophysica Acta* **240**, 496–505.

Kimura S, Tahira Y, Ishibashi T, Mori Y, Mori T, Hashimoto J, Sakaguchi K (2004) DNA repair in higher plants; photoreactivation is the major DNA repair pathway in non-proliferating cells while excision repair (nucleotide excision repair and base excision repair) is active in proliferating cells. *Nucleic Acids Research* **32**, 2760–2767. doi: 10.1093/nar/gkh591

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) 'MEGA2: molecular evolutionary genetics analysis software.' (Arizona State University: Tempe)

Lemesle-Varloot L, Henrissat B, Gaboriaud C, Bissery V, Morgat A, Mornon JP (1990) Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences. *Biochimie* **72**, 555–574. doi: 10.1016/0300-9084(90)90120-6

Lopez I, Khan S, Vazquez J, Hussey PJ (1997) The proliferating cell nuclear antigen (PCNA) gene family in *Zea mays* is composed of two members that have similar expression programs. *Biochimica et Biophysica Acta* **1353**, 1–6.

Magaña-Schwencke N, Henriques JAP, Chanet R, Moustacchi E (1982) The fate of 8-methoxypsoralen photoinduced cross-links 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.

Markley NA, Bonham-Smith PC, Moloney MM (1993) Molecular cloning and expression of a cDNA encoding the proliferating cell nuclear antigen from *Brassica napus* (oilseed rape). *Genome* **36**, 459–466.

Martin IV, MacNeill SA (2002) ATP-dependent DNA ligases. *Genome Biology* **3**, 3005.1–3005.7. doi: 10.1186/gb-2002-3-4-reviews3005

Matton DP, Nass N, Clarke AE, Newbigin E (1994) Self-incompatibility: how plants avoid illegimate offspring. *Proceedings of the National Academy of Sciences USA* **91**, 1992–1997.

Myre MA, O'Day DH (2002) Nucleomorphin: a novel, acidic, nuclear calmodulin-binding protein from *Dictyostelium* that regulates nuclear number. *Journal of Biological Chemistry* **277**, 19735–19744. doi: 10.1074/jbc.M109717200

Nasmyth KA (1979) Genetic and enzymatic characterization of conditional lethal mutants of the yeast *Schizosaccharomyces pombe* with a temperature-sensitive DNA ligase. *Journal of Molecular Biology* **130**, 273–284. doi: 10.1016/0022-2836(79)90541-2

Nasrallah JB (2000) Cell-cell signaling in the self-incompatibility response. *Current Opinion in Plant Biology* **3**, 368–373. doi: 10.1016/S1369-5266(00)00098-4

Noordzij JG, Verkaik NS, van der Burg M, van Veelen LR, de Bruin-Versteeg S, *et al.* (2003) Radiosensitive SCID patients with Artemis gene mutations show a complete B-cell diferentiation arrest at the pre-B-cell receptor checkpoint in bone marrow. *Blood* **101**, 1446–1452. doi: 10.1182/blood-2002-01-0187

Pastuglia M, Roby D, Dumas C, Cock JM (1997) Rapid induction by wounding and bacterial infection of an *S* gene family receptor-like kinase gene in *Brassica oleracea. The Plant Cell* **9**, 49–60. doi: 10.1105/tpc.9.1.49

Prasad R, Singhal RK, Srivastava DK, Molina JT, Tomkinson AE, Wilson SH (1996) Specific interaction of DNA polymerase ⇓ and DNA ligase I in a multiprotein base excision repair complex from bovine testis. *The Journal of Biological Chemistry* **271**, 16000–16007. doi: 10.1074/jbc.271.27.16000

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. doi: 10.1007/BF00425609

Schlissel MS (2002) Does Artemis end the hunt for the hairpin-opening activity in V(D)J recombination? *Cell* **109**, 1–4. doi: 10.1016/S0092-8674(02)00694-3

Schuman S, Schwer B (1995) RNA capping enzymes and DNA ligases: a superfamily of covalent nucleotidyltransferases. *Molecular Microbiology* **17**, 405–410.

Stapleton AE (1992) Ultraviolet radiation and plants: burning questions. *The Plant Cell* **4**, 1353–1358. doi: 10.1105/tpc.4.11.1353

Taylor RM, Hamer MJ, Rosamond J, Bray CM (1998) Molecular cloning and functional analysis of the *Arabidopsis thaliana* DNA ligase I homologue. *The Plant Journal* **14**, 75–81. doi: 10.1046/j.1365-313X.1998.00094.x

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Timson DJ, Singleton MR, Wigley DB (2000) DNA ligases in the repair and replication of DNA. *Mutation Research* **460**, 301–318. doi: 10.1016/S0921-8777(00)00033-1

Tom S, Henricksen LA, Park MS, Bambara RA (2001) DNA ligase I and proliferating cell nuclear antigen form a functional complex. *The Journal of Biological Chemistry* **276**, 24817–24825. doi: 10.1074/jbc.M101673200

Tomkinson AE, Mackey ZB (1998) Structure and function of mammalian DNA ligases. *Mutation Research* **407**, 1–9. doi: 10.1016/S0921-8777(97)00050-5

Tsukada K, Nishi A (1971) Polynucleotide ligase from cultured plant cells. *Journal of Biochemistry* **70**, 541–542.

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.

Waga S, Bauer G, Stillman B (1994) Reconstitution of complete SV40 DNA-replication with purified replication factors. *Journal of Biological Chemistry* **269**, 10923–10934.

Wagner D (2003) Chromatin regulation of plant development. *Current Opinion in Plant Biology* **6**, 20–28. doi: 10.1016/S1369526602000079

Warbrick E (1998) PCNA binding through a conserved motif. *BioEssays* **20**, 195–199. doi: 10.1002/(SICI)1521-1878(199803)20:3<195::AID-BIES2>3.0.CO;2-R

Warbrick E, Lane DP, Glover DM, Cox LS (1995) A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen. *Current Biology* **5**, 275–282. doi: 10.1016/S0960-9822(95)00058-3

Waugh M, Anderson W, Bell C, Inman J, Schilkey F, Sullivan J, May GD (2002) *Medicago* genome initiative (MGI). *Nucleic Acids Research* **30**, 1–12.

Weller GR, Doherty AJ (2001) A family of DNA repair ligases in bacteria? *FEBS Letters* **505**, 340–342. doi: 10.1016/S0014-5793(01)02831-9

West CE, Waterworth WM, Jiang Q, Bray CM (2000) *Arabidopsis* DNA ligase IV is induced by . -irradiation and interacts with an *Arabidopsis* homologue of the double strand break repair protein XRCC4. *The Plant Journal* **24**, 67–78. doi: 10.1046/j.1365-313x.2000.00856.x

Wilborn F, Brendel M (1989) Formation and stability of interstrand 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. doi: 10.1007/BF00340711

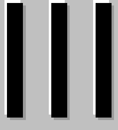
Wilkinson A, Bowater DJ (2001) Bacterial DNA ligases. *Molecular Microbiology* **40**, 1241–1248. doi: 10.1046/j.1365-2958.2001.02479.x

Wu Y-Q, Hohn B, Ziemienowicz A (2001) Characterization of an ATP-dependent type I DNA ligase from *Arabidopsis thaliana. Plant Molecular Biology* **46**, 161–170. doi: 10.1023/A:1010679901911

Yu J, Hu S, Wang J, Wong GK-S, Li S, *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* **296**, 79–92. doi: 10.1126/science.1068037

Zhu H, Shuman S (2004) A primer-dependent polymerase function of *Pseudomonas aeruginosa* ATP-dependent DNA ligase (LigD). *Journal of Biological Chemistry* (In press).

Manuscript received 10 August 2004, accepted 6 January 2005



Discussão geral

Discussão geral

A reparação de ICLs em células eucarióticas compreende a associação de vários mecanismos pouco caracterizados bioquimicamente (Schärer, 2005). Neste sentido, a natureza química de um ICL é bastante variada (ver Introdução, item 2.2.), o que pode resultar em diferentes respostas da célula a fim de garantir a sua sobrevivência (Schärer, 2005). Em leveduras, a presença de apenas um ICL não reparado pode ser letal (Magaña-Schwencke *et al.,* 1982; Brendel & Henriques, 2001), enquanto que, para células de mamíferos, tem sido estimado que 40 ICLs não reparados podem induzir a morte celular (Lawley & Phillips, 1996; Schärer, 2005). Entretanto, vários autores têm demonstrado que o mecanismo recombinacional, e neste caso as vias HR e NHEJ, seria a primeira escolha da célula para a reparação de danos do tipo ICLs (Panasci *et al.,* 2004). Assim, a atuação conjunta de proteínas pertencentes à via HR, como é o caso do complexo MRX, e de proteínas pertencentes à via NHEJ, como o Ku70/Ku80 e a DNA-PKcs, mostrou-se essencial para a reparação de danos do tipo ICL, seja em células de mamíferos ou em leveduras (Dronkert & Kanaar, 2001). Além disso, algumas proteínas pertencentes ao mecanismo de excisão, especialmente enzimas da via NER acoplada à transcrição, poderiam ser recrutadas para a reparação de danos desta natureza (Dronkert & Kanaar, 2001). Por exemplo, em células de mamífero a resolução do ICL (mas não a formação de DSBs) é dependente do complexo XPF-ERCC1, o qual geraria substrato para um processo recombinacional do tipo HR ou NHEJ (de Silva *et al.,* 2000; Niedernhofer *et al.,* 2004). Contudo, vários experimentos feitos com plasmídeos contendo ICLs gerados por psoralenos fotoativados mostram que o padrão de geração de DSB e a resolução do dano não é consistente com a atuação de proteínas da via NER e que, provavelmente, outros mecanismos poderiam ser recrutados para esta finalidade (Li *et al.,* 1999). Em leveduras,

cujos mecanismos de reparação de ICLs são um pouco mais conhecidos do que em células de mamíferos, foi postulado que o uso preferencial da via HR em relação a via NHEJ ou vice-versa é dependente de inúmeros fatores, dos quais predominariam a fase do ciclo celular e a natureza química do ICL (McHugh *et al.,* 2000). Desta maneira, as leveduras que estão na fase exponencial de crescimento e são defectivas para a via HR são tão sensíveis aos ICLs gerados por HN2 quanto leveduras que estão em fase estacionária e são defectivas para a via NHEJ (McHugh *et al.,* 2000). Além disso, os estudos de epistasia de leveduras mostraram que três mecanismos independentes, representados pelas proteínas Rad51, Pso2/Snm1 e Rev3, poderiam estar atuando na reparação de ICLs gerados por 8-MOP fotoativado e cisplatina (Henriques & Moustacchi, 1981; Henriques *et al.,* 1997; Grossman *et al.,* 2001). Também conforme Grossman *et al.* (2001), a sensibilidade dos mutantes da via NER de leveduras para ICLs é muito menor do que a combinação de qualquer um dos mutantes para as três proteínas citadas, indicando que o mecanismo de excisão atuaria somente em algumas condições particulares de reparação, as quais não são ainda conhecidas.

Tendo em vista a importância do Pso2p/Snm1p na reparação de ICLs, buscou-se compreender como esta família de proteínas poderia interagir nos mecanismos de reparação de ICLs e também na manutenção da estabilidade genômica em eucariotos. Para este fim, foram utilizadas diferentes metodologias de análises de seqüências combinadas a informações filogenéticas e funcionais conhecidas sobre a família Pso2p/Snm1p, que permitiram inferir não só um possível mecanismo de reparação de ICLs (Capítulo 1) mas também outras funções não descritas previamente, tais como a reparação de bases oxidadas associadas à recombinação em eucariotos multicelulares (Capítulo 2) e à geração de diversidade biológica em protozoários e fungos (Capítulo 2 e 3). Graças a esta combinação de técnicas, também foi possível identificar uma nova família de DNA ligases ATP-dependentes exclusiva de plantas que possuem, em sua extremidade N-terminal, um

domínio Pso2p/Snm1p (Capítulo 4), sendo esta a primeira DNA ligase eucariótica com função endonucleásica já identificada.

O papel da família Pso2p/Snm1p na reparação de ICLs e na manutenção da cromatina

Os dados bioquímicos conhecidos até o momento sobre a família Pso2/Snm1 claramente demonstram sua atuação em um passo pós-incisão do ICL (Capítulo 1 e Anexo II), sendo que a restituição do DNA de alta massa molecular não é observada para as linhagens mutantes de *S. cerevisiae* defectivas nesta proteína (Magaña-Schwencke *et al.,* 1982; Wilborn & Brendel, 1989). Historicamente, os mutantes de leveduras *pso2/snm1* foram isolados pela sua alta sensibilidade a ICLs induzidos por diferentes agentes químicos, sendo que esta sensibilidade não depende da natureza da base nitrogenada envolvida no ICL (Brendel e Henriques, 2001). Este mecanismo de atuação também foi identificado em células de mamíferos defectivas para Pso2p/Snm1p, especialmente quando tratadas por mitomicina C (Capítulo 1). Interessantemente, em células tronco embrionárias de camundongos onde o gene *PSO2/SNM1* foi disruptado, a sensibilidade a ICLs foi observada somente para a mitomicina C, não sendo possível identificar um aumento de sensibilidade para outros agentes indutores de ICLs ou para a radiação ionizante (Dronkert *et al.,* 2000). A explicação postulada para este fenômeno foi que, em mamíferos, a presença de genes parálogos que codificam para diferentes proteínas Pso2/Snm1 poderiam exercer funções redundantes na reparação de ICLs (Dronkert *et al.,* 2000). De fato, os dados genômicos e bioquímicos conhecidos até o momento apontavam para a presença de três genes *PSO2/SNM1* parálogos em mamíferos: o *PSO2A/SNM1A*, o *PSO2B/SNM1B* e *ARTEMIS*. Contudo, a falta de informações filogenéticas e estruturais sobre a família Pso2p/Snm1p

impôs uma séria restrição na análise dos dados obtidos por experimentos bioquímicos e genéticos, tanto em leveduras quanto em células de mamíferos.

Uma classificação estrutural da proteína Pso2/Snm1 só foi possível recentemente graças ao trabalho de Callebaut *et al.*(2002) que, utilizando a técnica de HCA, conseguiram enquadrar o Pso2p/Snm1p em uma nova família de proteínas denominada de β-CASP, que pertencem, por sua vez, à superfamília das metalo-β-lactamases. As metalo-β-lactamases são proteínas presentes tanto em eucariotos quanto em procariotos, e possuem as mais diferentes funções (Aravind, 1999). De uma forma geral, estas proteínas têm a capacidade de hidrolisar substratos que contêm, em sua estrutura, uma ligação do tipo éster e uma carga geral negativa, como a existente no DNA e no RNA (Aravind, 1999; Callebaut *et al.,* 2002). A hidrólise só é possível por meio do chamado 'domínio metalo-β-lactamásico', que consiste de uma seqüência-assinatura altamente conservada do tipo HxHxDH e um centro binuclear ligante de Zn^{+2}. Adicionalmente, as proteínas pertencentes à família β-CASP possuem, em sua estrutura, um domínio identificado pela técnica de HCA conhecido como 'β-CASP', o qual poderia ser responsável pela ligação da proteína a ácidos nucléicos (Capítulos 1 e 2).

A família β-CASP é formada por proteínas que atuam tanto no metabolismo de RNA, como é o caso das Elac1/Elac2 (RNAse Z) e do fator específico de clivagem e de poliadenilação de 73 kDa e de 100 kDa [*cleavage and polyadenylation specificity factor* (CPSF) 73 kDa/100 kDa], assim como de proteínas que atuam no metabolismo de DNA, como é o caso da família Pso2p/Snm1p (Capítulos 1 e 2).

Contudo, à medida que mais dados genômicos provenientes da análise de diferentes eucariotos eram disponibilizados para a comunidade científica, mais urgente tornava-se a realização de uma classificação filogenética e estrutural da família β-CASP que também levasse em conta as mais recentes informações bioquímicas sobre a proteína Pso2/Snm1.

Ao contrário do que se conhecia previamente(Callebaut *et al.,* 2002; Ishiai *et al.,* 2004), a filogenia e a análise de seqüências revelaram que quatro grupos compõem a família Pso2p/Snm1p (Figura 1, Capítulo 2; Tabela 2). Destes quatro grupos, três são formados por subgrupos distintos (Tabelas 1 a 3, Capítulo 2; Tabela 2) que, provavelmente, atuam de forma não redundante na reparação de DNA e na manutenção da cromatina.

Tabela 2. Resumo dos principais grupos, subgrupos e funções das proteínas pertencentes à família Pso2p/Snm1p.

Grupo	Subgrupo	Função
Pso2p Plasmodium (PPG)		Desconhecida
Pso2p/Snm1p A		
	Pso2p A de fungos [fungal A Pso2p (FA)]	Reparação de ICLs, manutenção da cromatina (?)[a]
	Pso2p A de eucariotos multicelulares [multicellular eukaryotic Pso2p A (MEA)]	Reparação de ICLs gerados por mitomicina C e cisplatina; reparação de bases oxidadas por meio de recombinação (*A. thaliana* e *O. sativa*); manutenção da cromatina
	Pso2p A específico de plantas [plant-specific Pso2p A (PSA)]	Desconhecida
	Pso2p A de protozoários [protozoan Pso2p A (PA)]	Desconhecida
Pso2p/Snm1p B		
	Pso2p B de Leishmania [Leishmania Pso2p B (LB)]	Desconhecida
	Pso2p B de eucariotos multicelulares [multicellular eukaryotic Pso2p B (MEB)]	Reparação de ICLs e de DSBs gerados por radiação ionizante
	Pso2p B específico de plantas [plant-specific Pso2p B (PSB)]	Desconhecida
Artemis		
	Artemis-like de fungos	Desconhecida
	Artemis/Artemis-like de metazoários	Clivagem de *hairpins* durante a recombinação V(D)J; atuação em NHEJ para reparação de DSBs induzidos por radiação ionizante; manutenção da cromatina

[a] (?) = função ainda não confirmada.

O grupo A da família Pso2p/Snm1p é formado por subgrupos filogeneticamente diversificados, estando presente em praticamente todos os filos eucarióticos conhecidos (Capítulo 2, Figura 2; Tabela 2). Esta ampla diversificação possivelmente reflete a importância do grupo para a reparação e para a manutenção da cromatina eucariótica. No grupo A estão presentes as proteínas Pso2/Snm1 de *S. cerevisiae* (subgrupo FA), de *H. sapiens* (subgrupo MEA) e de *Gallus gallus* (subgrupo MEA) (Capítulos 1 e 2). Destaca-se também neste grupo a presença de um subgrupo formado por DNA ligases vegetais, cuja função no metabolismo de DNA é incerta (Capítulo 2, Figura 2). Por outro lado, o grupo B é formado por proteínas pouco caracterizadas bioquimicamente e a sua distribuição é mais restrita filogeneticamente, não estando presente em fungos (Capítulo 2). Entretanto, o grupo B parece possuir uma função mais abrangente para a reparação de diferentes tipos de lesões em comparação com o grupo A (Ishiai *et al.,* 2004), atuando não só na reparação de ICLs, mas também de lesões geradas por radiação ionizante. O grupo Ártemis caracteriza-se por dois subgrupos: o subgrupo Ártemis de metazoários e o subgrupo Ártemis de fungos (Capítulo 3, Figura 1; Tabela 2). A função do subgrupo Ártemis de metazoários na geração de diversidade imunológica adaptativa em vertebrados mandibulados, na reparação de danos de DNA e na manutenção da cromatina é relativamente bem conhecida (Capítulos 1 e 3), enquanto que não há dados disponíveis sobre o mecanismo de atuação do subgrupo Ártemis de fungos (Capítulo 3). Por fim, o quarto grupo é formado por seqüências Pso2p/Snm1p do gênero *Plasmodium*, de função desconhecida.

È interessante notar que, ao longo da evolução da família Pso2p/Snm1p e de seus grupos e subgrupos, houve uma conservação de domínios, conforme indicam os dados de HCA (Capítulos 1 a 4). Em um primeiro momento, comparando as seqüências ortólogas de fungos, de metazoários e de plantas foi possível identificar três domínios ou regiões conservados (CRI a CRIII) que, posteriormente, foi expandido com a inclusão de um quarto domínio (CRIV) quando foram consideradas as seqüências de protozoários de diferentes

grupos da família Pso2p/Snm1p. Foi observado que estes quatro CRs estão distribuídos ao longo dos domínios metalo-β-lactamásico e β-CASP, apesar dos mesmos não terem sido identificados preliminarmente por Callebaut *et al.* (2002). É notável também o grau de conservação de estruturas primárias e secundárias observado nos quatro CRs (Capítulos 2 a 4). Este resultado reforça a idéia de que a conformação espacial dos CRs é de extrema importância para a atividade enzimática da família Pso2p/Snm1p.

Se a filogenia e a análise de seqüências, como visto nos quatro capítulos desta tese, mostram uma conservação evolutiva da família Pso2p/Snm1p, espera-se que as proteínas desta família atuem de forma semelhante, pelo menos enzimaticamente. Infelizmente, não existem dados bioquímicos sobre os possíveis substratos de DNA utilizados pelas proteínas Pso2p/Snm1p. Porém, uma possível função destas sequências na reparação de DNA pode ser extrapolada se forem consideradas as informações existentes sobre a proteína Ártemis de metazoários (Capítulos 1 e 3). Por sua importância para a recombinação V(D)J e para a geração de diversidade imunológica adaptativa, as atividades enzimáticas da Ártemis são amplamente conhecidas (Capítulos 1 e 3). Basicamente, a Ártemis é uma exonuclease com atividade 5'→3' que atua em fita simples de DNA. Quando fosforilada pelo holocomplexo DNA-PK, adquire uma função endonucleásica com alta afinidade por estruturas secundárias do tipo *hairpin*¸ tais como aqueles gerados pelas proteínas Rags durante a recombinação V(D)J (ver Introdução, item 3.2.2.). Interessantemente, uma função exonucleásica 5'→3' foi recentemente descrita para a proteína Pso2/Snm1 de *S. cerevisiae* (Li *et al.,* 2005). Contudo, os mesmos autores não realizaram ensaios de fosforilação da proteína Pso2/Snm1 a fim de confirmar se a mesma poderia apresentar atividade endonucleásica em diferentes substratos de DNA (Li *et al.,* 2005). Até o momento, não são conhecidas as cinases que fosforilam o Pso2p/Snm1p, mas uma possível candidata foi caracterizada recentemente. Utilizando ensaios de dois-híbridos de leveduras para a detecção de possíveis proteínas

interatoras com Pso2p/Snm1p de *S. cerevisiae*, Revers *et al.* (manuscrito em preparação) isolaram uma treonina-serina cinase denominada de Pak1p. A Pak1p pertence à família das cinases dependentes de cálcio/calmodulina, que inclui a proteína CaMKK-β de mamíferos (Anderson *et al.,*1998). Foi observado que estas cinases são responsáveis pela ativação de outras proteínas cinases, tais como a Snf1p de leveduras e a AMPK de mamíferos em condições de limitação nutricional (Nath *et al.,* 2003). A caracterização genética da Pak1p mostrou que tanto Pso2p quanto Pak1p atuam epistaticamente na reparação de danos do tipo ICLs. Esta interação foi posteriormente confirmada com ensaios de fosforilação *in vitro* utilizando Pak1p e Pso2p/Snm1p purificados (Revers *et al.*, manuscrito em preparação). Por outro lado, considerando que a Ártemis de metazoários é fosforilada pelo holocomplexo DNA-PK e pelas cinases ATM e ATR (Capítulos 1 e 3), poderia-se esperar que as proteínas Pso2/Snm1 também sejam fosforiladas por diferentes cinases. Neste sentido, deve ser salientado que várias cinases homólogas a DNA-PKcs podem ser identificadas a partir de uma prospecção de dados em bancos genômicos de leveduras como, por exemplo, as proteínas Tor1 e Tor2 (*target of rapamycin*) de *S. cerevisiae* (Wedaman *et al.,* 2003). As proteínas Tors são bastante conservadas em todos os filos conhecidos, atuando em diversos processos fisiológicos (Wedaman *et al.,* 2003). A função das seqüências Tors na reparação de DNA e/ou na manutenção da integridade genômica não é conhecida mas, tal como a Pak1p, as proteínas Tors são ativadas em condições de carência nutricional (Wedaman *et al.,* 2003). Levando-se em conta que a Pak1p ativa outras cinases, pode ser que esta cinase e as proteínas Tors sejam parte de uma cascata de sinalização ou de fosforilação que ativam proteínas relacionadas à reparação de DNA, como é o caso do Pso2p/Snm1p (Figura 12).

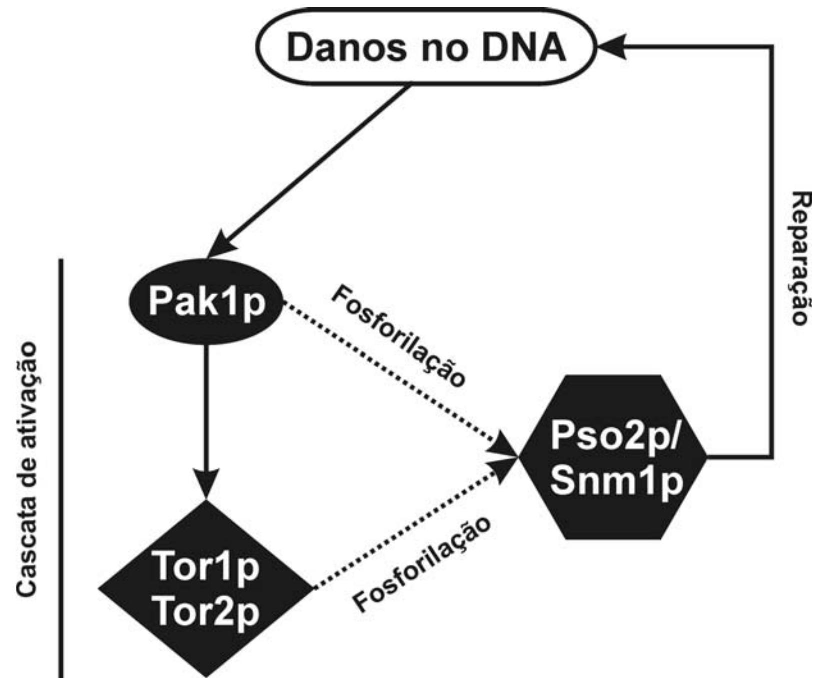


Figura 12. Mecanismo proposto de fosforilação da proteína Pso2/Snm1. Na presença de um dano no DNA, a cinase Pak1p é ativada, fosforilando diretamente as cinases Tor1 e Tor2, ou então, fosforilando Pso2p/Snm1p. Uma vez ativadas, as cinases Tor1 e Tor2 também poderiam fosforilar Pso2p/Snm1p, necessária para a reparação do dano.

Adicionalmente, os estudos utilizando linhagens mutantes de leveduras, de células de mamíferos e de linfócitos B de galinha defectivas para as proteínas Pso2p/Snm1p dos subgrupos FA e MEA mostram um aumento da sensibilidade celular para os agentes indutores de ICLs, mas não para a radiação UV e ionizante ou agentes alquilantes (Capítulo 2). Esta especificidade pode ser explicada se considerarmos que os ICLs geram mudanças conformacionais na molécula de DNA, especialmente durante o processo de replicação, gerando estruturas secundárias de DNA do tipo *hairpins* (Capítulo 1). Muitos trabalhos têm mostrado que a presença de seqüências palindrômicas no genoma de organismos eucarióticos e procarióticos induz estruturas secundárias de DNA do tipo cruciforme ou do tipo *hairpins* que, se não corretamente reparadas, são uma fonte de instabilidades

cromossômicas (Bzymek & Lovett, 2001; Farah *et al.,* 2002; Lobachev *et al.,* 2002). Os mecanismos relacionados à reparação destas estruturas ainda são desconhecidos, mas há indicações de uma possível atuação do complexo MRX e de outras proteínas da via NHEJ (Lobachev *et al.,* 2002). Outras fontes de geração de *hairpins*, como transposons, também são reparadas por vias semelhantes. Os dados obtidos com diferentes linhagens mutantes de *S. cerevisiae* transformadas com o sistema de transposição *Ac/Ds* de milho mostram que Pso2p/Snm1p e várias proteínas da via NHEJ são necessárias para a correta reparação dos *hairpins* gerados pela transposição (Yu *et al.,* 2004).

Um dos vários resultados interessantes obtidos pela análise de seqüências do grupo A diz respeito à presença de uma seqüência Pso2/Snm1 no fungo filamentoso *Aspergillus nidulans* (*Ani*Pso2p FA; Capítulo 2). Esta proteína apresenta, além dos quatro CRs característicos, dois domínios relacionados à manutenção do equilíbrio redox (domínios do tipo citocromo P450 e UbiE) e um domínio envolvido no processamento de mRNA (domínio do tipo Snp3p) (Capítulo 2). *A. nidulans* é um fungo toxígeno, que produz e acumula uma grande quantidade de esterigmatocistina, um precursor metabólico das aflatoxinas e um forte indutor de ICLs em células de mamífero (Sweeney & Dobson, 1998). A ativação da esterigmatocistina para a formação de ICLs em mamíferos depende do citocromo P450 (Sweeney & Dobson, 1998) e é esperado que o mesmo processo aconteça em *A. nidulans.* Por outro lado, a esterigmatocistina é um importante fator para a indução de esporulação em *A. nidulans* e as linhagens defectivas que não acumulam esta toxina possuem uma taxa de esporulação bastante reduzida em relação às linhagens selvagens (Calvo *et al.,* 2002). Apesar da sua função ser desconhecida, o estudo da *Ani*Pso2p FA na reparação de ICLs induzidos por esterigmatocistina poderia esclarecer como fungos produtores de aflatoxinas e seus precursores toleram altas concentrações intracelulares destes compostos.

Além da função na reparação de danos do tipo ICL, vários resultados apontam que a família Pso2p/Snm1p, especialmente as proteínas do grupo A (subgrupos FA e MEA) e do grupo Ártemis, poderia ter uma função na manutenção da estabilidade genômica eucariótica, sendo recrutada em determinadas etapas do ciclo celular a fim de identificar ou sinalizar para danos existentes na cromatina antes da célula dar prosseguimento ao seu processo de divisão (Capítulos 1 a 3). O exato mecanismo de ação do Pso2p/Snm1p MEA no controle do ciclo celular ainda não é conhecido, mas foi demonstrado que, em células humanas e em linfócitos de galinha, esta proteína co-localiza-se com a proteína ligante a p53 (53BP1) e com a γ-H2AX (Richie *et al.,* 2002). Estas duas proteínas interagem, por sua vez, com os componentes dos complexos APC (*anaphase promoting complex*)/ciclossomo e SUMO (*small ubiquitin-like modifier E3 ligase*), fundamentais para a progressão do ciclo celular e para a sinalização de danos na cromatina (Harper *et al.,* 2002; Akhter *et al.,* 2004). Uma vez que o complexo APC/ciclossomo está associado com o Pso2p/Snm1p MEA, e na presença de inibidores de fuso mitótico como o taxol, há uma parada de ciclo celular e a conseqüente reparação dos danos (Akhter *et al.,* 2004). Além disso, Akhter *et al.* (2004) demonstraram que a ausência de Pso2p/Snm1p MEA é fator decisivo para a geração de quebras cromossômicas e diminuição da viabilidade celular. Entretanto, os mesmos autores não mostraram se, na presença de ICLs, haveria também uma inibição do complexo APC/ciclossomo.

Em *S. cerevisiae* e *Schizosaccharomyces pombe* a presença de ICLs induz a uma parada de ciclo celular durante a transição G_2/M (Grossman *et al.,* 2000; Lambert *et al.,* 2003). No caso dos mutantes *pso2/snm1* de leveduras, há uma parada permanente nesta fase após o tratamento com agentes indutores de ICLs e uma conseqüente diminuição da viabilidade celular (Grossman *et al.,* 2000). Contudo, não se sabe se os inibidores de fuso mitótico poderiam induzir a uma parada de ciclo celular nos mutantes de leveduras para as proteínas Pso2/Snm1. Assim, duas hipóteses são plausíveis: (1) as proteínas Pso2/Snm1 FA

e MEA poderiam operar de forma diferenciada na manutenção da cromatina de leveduras e de organismos multicelulares, o que resultaria em respostas aparentemente antagônicas a tratamentos com agentes indutores de danos de DNA e (2) os inibidores de fuso mitótico possivelmente ativam vias de reparação e de parada de ciclo celular que são distintas daquelas observadas para os danos gerados por ICLs. É importante ressaltar que as proteínas do grupo Ártemis também possuem um papel importante na manutenção da cromatina (Capítulos 1 e 3) e as linhagens defectivas para Ártemis possuem um alto índice de translocações e quebras cromossomais (Richardson & Jasin, 2000). Tal como acontece com as proteínas Pso2p/Snm1p FA, a Ártemis é recrutada durante a transição G_2/M em resposta à radiação ionizante e UV, associando-se com as cinases ATM e ATR (Zhang *et al.*, 2004). Contudo, não há dados indicando se a Ártemis poderia ser recrutada em resposta a danos gerados por ICLs ou por inibidores de fuso mitótico ou se a mesma poderia se associar com proteínas do complexo APC/ciclossomo. Levando-se em conta a posição filogenética da Ártemis e a sua semelhança estrutural com as proteínas dos grupos A e B (Capítulos 2 e 3), não seria surpreendente a presença de uma via de manutenção genômica conservada em eucariotos onde a família Pso2p/Snm1p poderia exercer funções de reparação e/ou de sinalização de danos na cromatina.

O papel dos grupos Ártemis e *Plasmodium* na geração de diversidade biológica em metazoários, fungos e protozoários

Apesar da importância da família Pso2p/Snm1p na manutenção da cromatina eucariótica e na reparação de danos do tipo ICLs, poucos estudos têm sido realizados sobre como esta família poderia atuar na geração de diversidade biológica ou, como conhecido tecnicamente, no fenômeno da evolvabilidade. A evolvabilidade é um processo vital para todos os eucariotos, pois permite a sua adaptação em ambientes que mudam constantemente

(Barton & Charlesworth, 1998; Bürger, 1999). Há vários mecanismos moleculares que atuam na evolvabilidade dos processos bioquímicos que, em última análise, refletem no fenótipo de um organismo. Como exemplos destes mecanismos moleculares podem ser citados o sistema imunológico adaptativo de vertebrados mandibulados, a variação antigênica de protozoários e os processos recombinacionais mitóticos e meióticos, sendo que, neste último caso, caracterizado pela reprodução sexuada de grande parte dos eucariotos (Barton & Charlesworth, 1998; Bürger, 1999).

O grupo Ártemis, especialmente o subgrupo Ártemis de metazoários, tem um papel bem definido na evolvabilidade do sistema imunológico adaptativo (Capítulo 3). A função endonucleásica da Ártemis é necessária para a clivagem dos *hairpins* geradas pelas proteínas Rags e, conseqüentemente, sua atuação torna-se fundamental para a maturação de linfócitos B e T (Le Deist *et al.,* 2004). Assim, a presença de Ártemis em mamíferos e em outros vertebrados mandibulados é esperada por sua função relacionada ao sistema imunológico. Contudo, a descoberta de seqüências homólogas a Ártemis em metazoários não mandibulados e em fungos é extremamente interessante, pois sugere que as proteínas Ártemis possam estar relacionadas a outros processos moleculares (Capítulo 3). As análises filogenéticas e de seqüências claramente mostram que, tanto as proteínas Ártemis de metazoários quanto às de fungos, formam dois grupos irmãos, compartilhando um ancestral comum recente (Capítulo 3). Neste sentido, a alta conservação de estruturas primárias, secundárias e terciárias observadas entre as seqüências Ártemis atesta para a sua origem comum (Capítulo 3).

Duas hipóteses poderiam explicar a presença de Ártemis em metazoários não mandibulados e em fungos. A primeira hipótese relaciona-se à existência de uma via recombinacional ancestral que se manteve conservada ao longo da evolução eucariótica. Com a especialização tecidual, característica de organismos multicelulares, esta via recombinacional poderia ter originado as vias V(D)J e a NHEJ. A recente descoberta de

proteínas homólogas a DNA-PKcs em insetos (Dore *et al.,* 2004) corrobora esta hipótese e fundamenta a idéia de que a proteína Ártemis seja necessária para processos relacionados à reparação de DNA por vias recombinacionais em eucariotos.

A segunda hipótese que poderia explicar a presença do grupo Ártemis em eucariotos estaria relacionada com a recombinação meiótica (Capítulo 3). É fato conhecido que a recombinação meiótica é um dos principais mecanismos moleculares para a geração de evolvabilidade em eucariotos (Baarends *et al.,* 2001) e que DSBs são deliberadamente introduzidas no DNA genômico por diferentes tipos de endonucleases (Bhuiyan & Schmekel, 2004). Em termos bioquímicos, a recombinação meiótica compreende os mesmos passos básicos observados para a via HR somática, e várias das proteínas associadas com a via HR são também utilizadas pela recombinação meiótica (Baarends *et al.,* 2001). Por exemplo, o complexo MRX, necessário tanto para a via HR quanto para a via NHEJ, é recrutado durante a recombinação meiótica (Baarends *et al.,* 2001; Borde *et al.,* 2004). Interessantemente, algumas proteínas das vias V(D)J e NHEJ também são utilizadas durante a recombinação meiótica, como é o caso da DNA-PKcs e da γ-H2AX, cujas funções neste processo ainda não estão plenamente esclarecidas (Hamer *et al.,* 2003a; Hamer *et al.,* 2003b). Além destas, uma proteína homóloga à Rag-2 conhecida como Peas foi recentemente identificada em tecidos meióticos de camundongos, sendo que a presença desta proteína está relacionada ao processo de recombinação meiótica (Ohinata *et al.,* 2003). Deve ser salientado que os dados de expressão gênica da *ÁRTEMIS* de camundongos mostram uma alta expressão deste gene em tecidos meióticos (Moshous *et al.,* 2001), o que reforça a idéia de que o grupo Ártemis poderia ter uma participação nos processos de recombinação meiótica em metazoários ou em algumas espécies de fungos (Capítulo 3).

A presença de seqüências Pso2/Snm1 em protozoários foi outro achado importante deste trabalho (Capítulo 2). Por meio da filogenia e da análise de seqüências foram identificadas proteínas Pso2/Snm1 em diferentes espécies de protozoários, algumas

pertencentes ao grupo A e outras pertencentes ao grupo B (Capítulo 2). A função destas proteínas é incerta, mas considerando as suas posições filogenéticas dentro da família Pso2p/Snm1p é provável que estas seqüências estejam relacionadas à reparação de DNA e à manutenção da cromatina. Interessantemente, a filogenia das seqüências Pso2/Snm1 de *Plasmodium* mostra que estas proteínas formam um grupo evolutivamente mais antigo do que qualquer outro da família Pso2p/Snm1p. Os dados de HCA corroboram os resultados da filogenia, e mostram que a disposição das CRs e, conseqüentemente, dos domínios metalo-β-lactamásico e β-CASP, é completamente diferente do observado para as outras Pso2p/Snm1p. Estas análises, quando usadas em conjunto, reforçam o fato de que as proteínas Pso2/Snm1 de *Plasmodium* constituem um quarto grupo dentro da família Pso2/Snm1 (Capítulo 2).

O gênero *Plasmodium* (filo Apicomplexa) compreende organismos exclusivamente parasitas de metazoários, e que possuem um complexo ciclo de vida (Escalante & Ayala, 1995). Uma das principais espécies deste grupo, o *P. falciparum*, é considerado o agente etiológico da malária, responsável por milhares de mortes a cada ano em todo o mundo (Maréchal & Cesbron-Delauw, 2001). Sua capacidade de evadir o sistema imunológico de mamíferos é que torna o gênero *Plasmodium* extremamente bem-sucedido como parasita (Bhattacharyya *et al.,* 2004). A maneira como se dá à evasão se deve ao mecanismo de variação antigênica de superfície empregado pelo *Plasmodium*. Basicamente, uma família de proteínas conhecida como *var*, arranjadas em grupos ou presentes isoladamente no genoma do *Plasmodium*, são responsáveis pelo mecanismo de variação antigênica (Bhattacharyya *et al.,* 2004). Há indicações de que a família *var* sofre um rearranjo extensivo de suas seqüências, por uma via recombinacional desconhecida, durante o processo infectivo, o que leva a um aumento na variabilidade antigênica do organismo. Apesar de que os mecanismos de reparação de DNA são praticamente desconhecidos em *Plasmodium*, algumas proteínas pertencentes às via HR e NHEJ foram descritas nestes

organismos (Bhattacharyya *et al.,* 2004). Interessantemente, a presença de um grupo ancestral da família Pso2p/Snm1p em *Plasmodium*, sua posição filogenética em relação aos outros grupos e as características estruturais próprias das seqüências, abrem a possibilidade de que estas proteínas possam ter relação direta com os mecanismos recombinacionais. Seria interessante descobrir que, em *Plasmodium,* opera um mecanismo semelhante à recombinação V(D)J ou NHEJ e que este(s) mecanismo(s) seria(m) o(s) responsável(is) pela alta variabilidade clonal observada neste gênero.

As seqüências Pso2p/Snm1p de plantas

Os mecanismos de reparação de DNA em plantas são pouco conhecidos, e considerando o modo de vida das plantas (séssil e dependente da radiação solar para a realização da fotossíntese) estes são, provavelmente, mais complexos do que os mecanismos presentes em outros filos eucarióticos (Tuteja, 2001). Além disso, as plantas sofrem danos de forma constante, sejam estes gerados por agentes abióticos (tal como a presença de metais pesados no solo) ou por agentes bióticos (fungos endofíticos produtores de micotoxinas), de forma que não é surpreendente que as plantas apresentem mecanismos redundantes de reparação de DNA (Tuteja, 2001). Algumas proteínas das vias HR e NHEJ foram previamente caracterizadas em *A. thaliana*, tais como Mre11p (Bundock & Hooykaas, 2002), Rad50p (Daoudal-Cotterell *et al.,* 2002) e Ku70/Ku80 (Tamura *et al.,* 2002).

No caso da família Pso2p/Snm1p, as análises filogenéticas e de seqüências dos diferentes grupos mostram que as plantas possuem inúmeras seqüências parálogas distribuídas nos grupos A e B, cujas funções na reparação ou na manutenção genômica são praticamente desconhecidas (Capítulos 1, 2 e 4).

De uma forma geral, as seqüências Pso2p/Snm1p de plantas podem ser agrupadas filogeneticamente em quatro subgrupos: MEA, MEB, PSA e PSB (Capítulo 2; Tabela 2). As seqüências Pso2p/Snm1p de plantas, presentes no subgrupo MEA, foram caracterizadas recentemente pela alta sensibilidade que as linhagens mutantes de *A. thaliana* e *O. sativa* apresentavam a agentes indutores de EROs (Molinier *et al.,*2004; Kimura *et al.,* 2005). Para os mutantes *pso2/snm1* MEA de *O. sativa* também foi observado um aumento de sensibilidade para mitomicina C (Kimura *et al.,* 2005), provavelmente devido à geração de EROs por este agente químico. Independente da espécie de planta, a sensibilidade apresentada pelos mutantes as EROs levou os autores a concluir que estas proteínas são específicas para algum mecanismo de reparação de bases oxidadas (Molinier *et al.,*2004; Kimura *et al.,* 2005). Interessantemente, observou-se que estas seqüências agrupam fortemente com as seqüências MEA de metazoários, sendo que os dados de HCA apontam uma alta similaridade estrutural entre as mesmas (Capítulo 2). Como discutido anteriormente, as células de mamíferos defectivas para as seqüências do subgrupo MEA são altamente sensíveis à mitomicina C e à cisplatina. Entretanto, ambos os compostos são fortes indutores de EROs (Korkina *et al.,* 2000), sendo que os ICLs correspondem a apenas uma pequena fração dos danos totais formados. Uma hipótese interessante a ser testada é a de que as seqüências Pso2p/Snm1p MEA de metazoários poderiam também estar relacionadas à reparação de bases oxidadas, enquanto que os outros grupos da família Pso2p/Snm1p seriam recrutados para a reparação de danos do tipo ICL ou gerados por radiação ionizante.

Infelizmente, as seqüências de plantas pertencentes aos subgrupos MEB e PSB não possuem função definida e, sem maiores informações bioquímicas e genéticas a respeito da atuação destes subgrupos na reparação e/ou na manutenção da cromatina, torna-se muito difícil inferir qualquer função. Contudo, as seqüências Pso2p/Snm1p do subgrupo PSA são extremamente interessantes pelo fato de que uma nova família de DNA ligases dependentes

de ATP, exclusiva de plantas, está presente neste subgrupo. Por suas características filogenéticas e estruturais, esta família foi denominada de Lig6p (Capítulo 2 e 4).

As análises filogenéticas da superfamília das nucleotidiltransferases, que engloba as DNA ligases I, III, IV e as enzimas que adicionam o *5'-cap* no mRNA, indicaram que a família Lig6p divergiu de um ramo ancestral que também originou as DNA ligases I (Capítulo 4). As análises de HCA mostraram que dois domínios conservados estão presentes nas seqüências Lig6p: um domínio N-terminal, contendo as quatro CRs características das proteínas Pso2/Snm1, e um domínio C-terminal, homólogo ao da família das DNA ligases I (Capítulo 4). Estes dados de HCA foram posteriormente corroborados com uma modelagem tri-dimensional das regiões N- e C- terminais (Capítulo 4). Interessantemente, uma análise de microssintenia indicou que as Lig6p estão ligadas aos locos *SLL2* e *S* de *A. thaliana* e *Brassica* sp., respectivamente (Capítulo 4). Estes locos contêm genes que codificam para diferentes proteínas relacionadas ao controle do padrão reprodutivo de angiospermas, além de atuarem na defesa contra patógenos e no processo de desenvolvimento (Dwyer *et al.,* 1994; Pastuglia *et al.,* 1997).

Em plantas, apenas as DNA ligases I e IV foram descritas até o momento. As DNA ligases I são enzimas importantes para a replicação e para a reparação de DNA, atuando nos mecanismos de excisão (Tison *et al.,* 2000) As DNA ligases IV, por sua vez, atuam preferencialmente em processos recombinacionais, especialmente na via NHEJ (Tomkinson & Mackey, 1998). Ambas as DNA ligases estão presentes na maioria dos filos eucarióticos estudados, o que denota a sua importância para o metabolismo geral de DNA. No caso das DNA ligases I e IV de plantas, os estudos de expressão de proteína indicam que os tecidos que se dividem ativamente, tais como os tecidos meristemáticos, são os que apresentam a maior atividade de DNA ligase. Entretanto, se for considerada a alta similaridade apresentada pela região C-terminal da Lig6p com a DNA ligase I, a baixa discriminação dos ensaios usados para detectar a atividade de DNA ligase e a presença do gene *LIG6* no loco

que controla a reprodução em plantas, é plausível pensar que a Lig6p tenha uma função importante na replicação de DNA de meristemas (Capítulo 4). O domínio endonucleásico Pso2p/Snm1p da Lig6p sugere que estas enzimas também podem atuar na reparação de ICLs, de bases oxidadas e/ou na manutenção da cromatina de plantas.

Deve ser salientado que as proteínas Lig6 compõem a primeira família de DNA ligases eucarióticas com uma dupla função enzimática, atuando tanto como endonucleases quanto como ligases (Capítulo 4). As DNA ligases com dupla função enzimática só foram descritas recentemente em algumas espécies de bactérias, tais como *Mycobacterium tuberculosis* (Della *et al.,* 2004) e *Pseudomonas aeruginosa* (Zhu & Shuman, 2005). Estas DNA ligases dependentes de ATP, denominadas de LigD, são necessárias para a restauração de DSBs, atuando em uma via análoga à NHEJ de eucariotos. É possível que as LigD também tenham outras funções no metabolismo do DNA procariótico.

IV

Conclusões

Conclusão geral

A família Pso2p/Snm1p compreende proteínas evolutivamente conservadas que possuem funções na reparação de DNA, na manutenção da cromatina eucariótica e na geração de diversidade biológica. As análises filogenéticas e de seqüências mostraram que quatro grupos principais compõem a família Pso2p/Snm1p, sendo que cada grupo é formado por diversos subgrupos distintos. Adicionalmente, os dados de HCA permitiram identificar quatro regiões conservadas na proteína Pso2/Snm1, tanto em termos de estrutura primária quanto secundária, permitindo ampliar os conhecimentos estruturais sobre esta família. Mais ainda, a técnica de análise de seqüências associada à filogenia mostrou-se uma ferramenta poderosa para a identificação de novas proteínas da família Pso2p/Snm1p, cuja função ou classificação não era previamente conhecida.

Conclusões específicas

- Quatro grupos distintos compõem a família Pso2p/Snm1p, conforme os resultados obtidos com a análise filogenética da família β-CASP. São eles: grupo A, grupo B, grupo Ártemis e grupo *Plasmodium*. Cada grupo, com exceção do *Plasmodium*, é formado por subgrupos também distintos que podem ter diferentes funções na reparação de DNA;
- A análise de HCA permitiu mapear quatro regiões conservadas (CRs) presentes em todos os grupos da família Pso2p/Snm1p. Estas quatro CRs possuem uma alta similaridade, tanto de estrutura primária quanto de estrutura secundária.
- Levando-se em conta as informações bioquímicas obtidas com as proteínas do grupo Ártemis e de sua função endonucleásica na clivagem de substratos

do tipo *hairpins*, assim como as mais recentes informações sobre a indução de estruturas secundárias de DNA por eventos de transposição ou pelo colapso da forquilha de replicação de DNA, é possível que a presença de ICLs durante a replicação também origine estruturas do tipo cruciforme ou *hairpins* que, por sua vez, serviriam de substrato para as proteínas Pso2/Snm1.

- O grupo A é formado por seqüências Pso2/Snm1 amplamente distribuídas na maioria das espécies eucarióticas estudadas. Quatro subgrupos compõem o grupo A e as funções deste grupo estão centradas na reparação de danos do tipo ICL e na manutenção da integridade da cromatina. Uma seqüência Pso2p/Snm1p de *A. nidulans,* identificada pela análise de seqüências e que contém três domínios adicionais envolvidos na manutenção do equilíbrio redox e no processamento de mRNA, poderia ter um papel fundamental na reparação de danos induzidos por aflatoxinas ou por esterigmatocistina.
- O grupo B apresenta uma distribuição filogenética mais restrita, não encontrado em fungos. Tal como observado para o grupo A, este grupo é formado por diferentes subgrupos. A função do grupo B na reparação de DNA não é conhecida, apesar de haverem indícios de que estas proteínas são recrutadas para a reparação de danos gerados por agentes bifuncionais e por radiação ionizante.
- O grupo Ártemis é formado por dois subgrupos: a Ártemis de metazoários e a Ártemis de fungos. Por sua função na manutenção da cromatina, na recombinação V(D)J e na reparação de DNA, atuando em conjunto com a via NHEJ, o subgrupo Ártemis de metazoários é o que possui a melhor caracterização genética e bioquímica de todas as proteínas da família Pso2p/Snm1p. Contudo, a presença de proteínas semelhantes à Ártemis em

fungos, possivelmente indica que estas sequências tenham outras funções, seja na reparação ou na evolvabilidade.

- As seqüências Pso2p/Snm1p de *Plasmodium* constituem o grupo filogeneticamente mais antigo dentro da família Pso2p/Snm1p. Estruturalmente, os dados de HCA mostram que as proteínas Pso2/Snm1 de *Plasmodium* são diferentes daquelas observadas em outros grupos. Estas seqüências, apesar de não terem uma função conhecida, poderiam estar associadas à reparação de DNA, à manutenção da cromatina ou mesmo participarem de algum mecanismo análogo à recombinação V(D)J para a geração de diversidade antigênica em *Plasmodium.*
- Os dados de filogenia e de seqüências indicam que as proteínas Pso2/Snm1 de plantas podem ser divididas em quatro subgrupos: MEA, MEB, PSA e PSB. As funções específicas de cada subgrupo não são conhecidas, mas a presença de duas seqüências MEA em *A. thaliana* e *O. sativa*, cuja disrupção promove um aumento de sensibilidade a agentes indutores de EROs, é um indicativo de que as outras seqüências deste subgrupo, tais como as de metazoários, poderiam ter uma função na reparação de bases oxidadas.
- A prospecção de dados genômicos de plantas, seguida da análise filogenética e de seqüências, permitiu identificar uma nova família de DNA ligases específicas de plantas, as quais contêm um domínio Pso2p/Snm1p na região N-terminal e um domínio do tipo DNA ligase I na região C-terminal. Esta nova família de DNA ligases (chamada de Lig6p), pode ter importantes funções no metabolismo de DNA em plantas. A presença de um domínio Pso2p/Snm1p a caracteriza como a primeira DNA ligase eucariótica com função exo/endonucleásica.

V

Perspectivas

Perspectivas

A prospecção de dados genômicos, seguidas de uma intensa análise filogenética e de seqüências, revelou que a família Pso2p/Snm1p é formada por inúmeros grupos e subgrupos cujas funções no metabolismo do DNA genômico são praticamente desconhecidas. Os dados bioquímicos e genéticos obtidos com a Ártemis de vertebrados e também com algumas proteínas Pso2/Snm1 dos subgrupos FA e MEA apontam para funções que incluem desde a reparação de danos gerados por agentes bifuncionais até a manutenção da cromatina e a geração de diversidade biológica. Assim, propõem-se alguns estudos dentre os vários outros que poderiam ser feitos para esclarecer as funções desta família de proteínas no metabolismo de DNA:

- Verificar a indução *in vivo* de *hairpins* ou outras estruturas secundárias de DNA por agentes bifuncionais, tais como 8-MOP mais UVA, em linhagens selvagens e mutantes de *S. cerevisiae* para o gene *PSO2/SNM1*. Os dados gerados por estes experimentos poderiam ser confrontados com os resultados já existentes para os mutantes de *S. cerevisiae* transformados com o sistema de transposição de milho *Ac/Ds*, que também gera *hairpins in vivo.*
- Sabendo que diferentes tipos de cinases (DNA-PKcs, ATM e ATR) atuam na fosforilação de proteínas do subgrupo Ártemis de metazoários, ativando sua função endonucleolítica, o mesmo estudo poderia ser realizado com os mutantes de *S. cerevisiae* disruptados para as cinases *TOR1, TOR2* (ambas homólogas a DNA-PKcs) e *TEL1* (homóloga à ATM). A habilidade destes mutantes para a reparação de *hairpins* poderia ser analisada utilizando o sistema *Ac/Ds* transformado em leveduras.
- Analisar, *in vitro,* a capacidade das cinases Tor1p, Tor2p e Tel1p de fosforilar o Pso2p/Snm1p de *S. cerevisiae*. Estes resultados poderiam ser

comparados com os dados já obtidos por nosso grupo a respeito da fosforilação do Pso2p/Snm1p pela cinase Pak1p.

- A capacidade do Pso2p/Snm1p fosforilado de clivar *hairpins* também poderia ser analisada *in vitro*. Este resultado esclareceria se outros grupos da família Pso2p/Snm1p, além da Ártemis, teriam esta capacidade.
- Sabendo que as células mutantes de mamíferos para o Pso2p/Snm1p do subgrupo MEA são sensíveis à ação de inibidores do fuso mitótico, seria importante verificar se as linhagens *pso2/snm1* de *S. cerevisiae* também apresentam a mesma sensibilidade.
- A análise das seqüências semelhantes à Ártemis de fungos, em especial a da levedura *Debaryomyces hansenii*, seria de grande interesse para as pesquisas relacionadas a este grupo de proteínas. Visto que inúmeras ferramentas moleculares para a manipulação de leveduras estão disponíveis e, levando-se em conta a facilidade para induzir mutações ou disrupções sítio-específicas nestes organismos, os estudos destas seqüências poderiam esclarecer alguns pontos obscuros referentes à atuação da Ártemis na estabilidade genômica e, principalmente, na reparação de DNA.
- As proteínas Pso2/Snm1 do grupo *Plasmodium*, por suas características estruturais e por suas prováveis funções na geração de diversidade biológica, na reparação de DNA e na manutenção da cromatina, merecem um destaque especial nos estudos relacionados à família Pso2p/Snm1p. O uso da tecnologia de RNAi para o silenciamento gênico do *PSO2/SNM1* em *Plasmodium* permitiria o estudo da função destas proteínas *in vivo*.
- A presença de inúmeras seqüências Pso2/Snm1 parálogas em plantas é intrigante, e um estudo mais pormenorizado se faz necessário a fim de esclarecer suas funções no metabolismo de DNA. Poderia-se testar a

sensibilidade dos mutantes de *A. thaliana* para estes subgrupos para tratamentos com diferentes agentes genotóxicos.

- Sabendo que a disrupção da seqüência *PSO2/SNM1* MEA de *A. thaliana* e de *O. sativa* induz um aumento de sensibilidade a EROs, o mesmo fenômeno poderia ser estudos em linhagens mutantes de mamíferos e fungos para o gene *PSO2/SNM1* (subgrupos MEA e FA). Além disso, a combinação de mutações para as vias associadas à reparação de bases oxidadas com *pso2/snm1* de *S. cerevisiae* poderia trazer informações importantes a respeito de uma possível função na reparação de bases modificadas.
- A análise da atividade enzimática da família Lig6p de plantas é de fundamental interesse para todas as pesquisas relacionadas ao metabolismo de DNA em plantas. Por se tratar de uma DNA ligase com dupla função, vários ensaios *in vitro* e *in vivo* poderiam ser realizados a fim de provar a presença de uma atividade endonucleásica e de DNA ligase nesta sequência. Além disso, a construção de linhagens *lig6* de *A. thaliana* permitiria avaliar a importância desta proteína para a replicação, para a reparação e para a manutenção da integridade da cromatina.

VI

Referências

Referências bibliográficas

1. AKHTER, S., RICHIE, C. T., DENG, J. M., BREY, E., ZHANG, X., PATRICK, C., BEHRINGER, R. R. & LEGERSKI, R. J. Deficiency in SNM1 abolishes an early mitotic checkpoint induced by spindle stress. *Molecular and Cellular Biology.* 24, 10448-10455, 2004.

2. ALLARD, S., MASSON, J. –Y. & CÔTÉ, J. Chromatin remodeling and the maintenance of genome integrity. *Biochimica et Biophysica Acta.* 1677, 158-164, 2004.

3. ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 25, 3389-3402, 1997.

4. ANDERSON, K. A., MEANS, R. L., HUANG, Q. H., KEMP, B. E., GOLDSTEIN, E. G., SELBERT, M. A., EDELMAN, A. M., FREMEAU, R. T. & MEANS, A. R. Components of a calmodulin-dependent protein kinase cascade: molecular cloning, functional characterization and cellular localization of Ca^{2+}/calmodulin- dependent protein kinase kinase beta. *The Journal of Biological Chemistry.* 273, 31880-31889, 1998.

5. ARAVIND, L. An evolutionary classification of the metallo-β-lactamase fold proteins. *In Silico Biology.* 1, 69-91, 1999.

6. BAARENDS, W. M., VAN DER LAAN, R. & GROOTEGOED, J. A. DNA repair mechanisms and gametogenesis. *Reproduction.* 121, 31-39, 2001.

7. BARTON, N. H. & CHARLESWORTH, B. Why sex and recombination? *Science.* 281, 1986-1990, 1998.

8. BASU, A. K., LOECHLER, E. L., LEADON, S. A. & ESSIGMANN, J. M. Genetic effects of thymine glycol: site-specific mutagenesis and molecular modelling studies. *Proceedings of the National Academy of Sciences of the United States of America.* 86, 7677-7681, 1989.

9. BAXEVANIS, A. D. & QUELLETTE, B. F. F. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley & Sons, Nova Iorque, Estados Unidos da América. 2001.

10. BETHEA, D., FULLMER, B., SYED, S., SELTZER, G., TIANO, J., RISCHKO, C., GILLESPIE, L., BROWN, D. & GASPARRO, F. P. Psoralen photobiology and photochemotherapy: 50 years of science and medicine. *Journal of Dermatological Science.* 19, 78-88, 1999.

11. BEUKERS, R., YLSTRA, J. & BERENDS, W. The effect of the ultraviolet light sensitivity of deoxyribonucleic acid. *Recueil des travaux chimiques des Pays-Bas* 79, 101-104, 1960.

12. BHATTACHARYYA, M. K., NORRIS, D. E. & KUMAR, N. Molecular players of homologous recombination in protozoan parasites: implications for generating antigenic variation. *Infection, Genetics and Evolution.*4, 91-98, 2004.

13. BHUIYAN, H. & SCHMEKEL, K. Meiotic chromosome synapsis in yeast can occur without Spo11-induced DNA double-strand breaks. *Genetics.* 168, 775-783, 2004

14. BOITEUX, S. & GUILLET, MARIE. Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae. DNA Repair.* 3, 1-12, 2004.

15. BORDE, V., LIN, W., NOVIKOV, E., PETRINI, J. H., LICHTEN, M. & NICOLAS, A. Association of Mre11p with double-strand break sites during yeast meiosis. *Molecular Cell*. 13, 389-401, 2004.

16. BOSL, G. J. & MOTZER, R.J. Testicular germ-cell cancer. *The New England Journal of Medicine*. 337, 242-254, 1997.

17. BRENDEL, M. & HENRIQUES, J.A.P. The *pso* mutants of *Saccharomyces cerevisiae* comprise two groups: one deficient in DNA repair and another with altered mutagen metabolism. *Mutation Research.* 48, 79-96. 2001.

18. BRENDEL, M., BONATTO, D., STRAUSS, M., REVERS, L. F., PUNGARTNIK, C., SAFFI, J. & HENRIQUES, J. A. P. Role of *PSO* genes in repair of DNA damage of *Saccharomyces cerevisiae. Mutation Research.* 544, 179-193, 2003.

19. BUCKLEY, R. H. Molecular defects in human severe combined immunideficiency and approaches to immune reconstitution. *Annual Review of Immunology.* 22, 625-655, 2004.

20. BUNDOCK, P. & HOOYKAAS, P. Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in *Arabidopsis MRE11* mutants. *The Plant Cell.* 14, 2451-2462, 2002.

21. BÜRGER, R. Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics.* 153, 1055-1069, 1999.

22. BZYMEK, M. & LOVETT, S. T. Evidence for two mechanisms of palindrome-stimulated deletion in *Escherichia coli:* single-strand annealing and replication slipped mispairing. *Genetics.* 158, 527-540, 2001.

23. CALDECOTT, K. W. DNA single-strand break repair and spinocerebellar ataxia. *Cell.* 112, 7-10, 2003.

24. CALDECOTT, K. W. DNA single-strand breaks and neurodegeneration. *DNA Repair.* 3, 875-882, 2004.

25. CALDECOTT, K. W. Mammalian DNA single-strand break repair: an X-ra(y)ted affair. *BioEssays.* 23, 447-455, 2001.

26. CALLEBAUT, I., LABESSE, G., DURAND, P., POUPON, A., CANARD, L., CHOMILIER, J., HENRISSAT, B. & MORNON, J. P. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and Molecular Life Sciences.* 53, 621-645, 1997.

27. CALLEBAUT, I., MOSHOUS, D., MORNON, J. –P., & DE VILLARTAY, J. –P. Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. *Nucleic Acids Research.* 30, 3592-3601, 2002.

28. CALVO, A. M., WILSON, R. A., BOK, J. W. & KELLER, N. P. Relationship between secondary metabolism and fungal development. *Microbiology and Molecular Biology Reviews.* 66, 447-459, 2002.

29. CHENG, K. C., CAHILL, D. S., KASAI, H., NISHIMURA, S. & LOEB L. A. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G–T and A–C substitutions. *The Journal of Biological Chemistry.* 267, 166-172, 1992.

30. CHENNA, R., SUGAWARA, H., KOIKE, T., LOPEZ, R., GIBSON, T. J., HIGGINS, D. G. & THOMPSON, J. D. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research.* 31, 3497-3500, 2003.

31. CHRISTMANN, M., TOMICIC, M. T., ROOS, W. P. & KAINA, B. Mechanisms of human DNA repair: an update. *Toxicology.* 193, 3-34, 2003.

32. CONNELLY, J. C., DE LEAU, E. S., OKELY. E. A. & LEACH, D. R. Overexpression, purification, and characterization of the SbcCD protein from *Escherichia coli. The Journal of Biological Chemistry.* 272, 19819-19826, 1997.

33. COSTA, R. M. A., CHIGANÇAS, V., GALHARDO, R. S., CARVALHO, H. & MENCK, C. F. M. The eukaryotic nucleotide excision repair pathway. *Biochimie.* 85, 1083-1099, 2003.

34. CUMMINGS, J., SPANSWICK, V. J., TOMASZ, M. & SMYTH, J. F. Enzymology of mitomycin C metabolic activation in tumour tissue: implications for enzyme-directed bioreductive drug development. *Biochemical Pharmacology.* 56, 405-414, 1998.

35. D'INCALCI, M., ERBA, E., DAMIA, G., GALLIERA, E., CARRASSA, L., MARCHINI, S., MANTOVANI, R., TOGNON, G., FRUSCIO, R., JIMENO, J. & FAIRCLOTH, G. T. Unique features of the mode of action of ET-743. *The Oncologist.*7, 210-216, 2002.

36. DAOUDAL-COTTERELL, S., GALLEGO, M. E. & WHITE, C. I. The plant Rad50-Mre11 protein complex. *FEBS Letters.* 516, 164-166, 2002.

37. DE SILVA, I. U., MCHUGH, P. J., CLINGEN, P. H. & HARTLEY, J. A. Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Molecular and Celular Biology.* 20, 7980-7990, 2000.

38. DEFAZIO, L. G., STANSEL, R. M., GRIFFITH, J. D. & CHU, G. Synapsis of DNA ends by DNA-dependent protein kinase. *The EMBO Journal.* 21, 3192-3200, 2002.

39. DELLA, M., PALMBOS, P. L., TSENG, H. M., TONKIN, L. M., DALEY, J. M., TOPPER, L. M., PITCHER, R. S., TOMKINSON, A. E., WILSON, T. E. & DOHERTY, A. J. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science.* 306, 683-685, 2004.

40. DORE, A. S., DRAKE, A. C., BREWERTON, S. C. & BLUNDELL, T. L. Identification of DNA-PK in the arthropods. Evidence for the ancient ancestry of vertebrate non-homologous end-joining. *DNA Repair.* 3, 33-41, 2004.

41. DOUKI, T., REYNAUD-ANGELIN, A., CADET, J. & SAGE, E. Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation. *Biochemistry.* 42, 9221-9226, 2003.

42. DRONKERT, M. L., DE WIT, J., BOEVE, M., VASCONCELOS, M. L., VAN STEEG, H., TAN, T. L., HOEIJMAKERS, J. H. & KANAAR, R. Disruption of mouse *SNM1* causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. *Molecular and Celular Biology.* 20, 4553-4561, 2000.

43. DRONKERT, M.L. & KANAAR, R. Repair of DNA interstrand cross-links. *Mutation Research.* 486, 217-247, 2001.

44. DUDÁŠ A. & CHOVANEC, M. DNA double-strand break repair by homologous recombination. *Mutation Research.* 566, 131-167, 2004.

45. DWYER, K. G., KANDASAMY, M. K., MAHOSKY, D. I., ACCIAI, J., KUDISH, B. I., MILLER, J. E., NASRALLAH, M. E. & NASRALLAH, J. B. A superfamily of *S* locus-related sequences in *Arabidopsis:* diverse structures and expression patterns. *The Plant Cell.* 6, 1829-1843, 1994.

46. EISEN, J. A. & HANAWALT, P. C. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutation Research.* 435, 171-213, 1999.

47. ELLIS, N.A. & GERMAN, J. Molecular genetics of Bloom's syndrome. *Human Molecular Genetics.* 5, 1457-1463, 1996.

48. ESCALANTE, A. A. & AYALA, F. J. Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proceedings of the National Academy of Sciences of the United States of America.* 92, 5793-5797, 1995.

49. EVANS, M. D., DIZDAROGLU, M. & COOKE, M. S. Oxidative DNA damage and disease: induction, repair and significance. *Mutation Research.* 567, 1-61, 2004.

50. FAIVRE, L., GUARDIOLA, P., LEWIS, C., DOKAL, I., EBELL, W., ZATTERALE, A., ALTAY, C., POOLE, J., STONES, D., KWEE, M. L., VAN WEEL-SIPMAN, M., HAVENGA, C., MORGAN, N., DE WINTER, J., DIGWEED, M., SAVOIA, A., PRONK, J., DE RAVEL, T., JANSEN, S., JOENJE, H., GLUCKMAN, E. &

MATHEW, C. G. Association of complementation group and mutation type with clinical outcome in Fanconi anemia. European Fanconi Anemia Research Group. *Blood.* 96, 4064-4070, 2000.

51. FARAH, J. A., HARTSUIKER, E., MIZUNO, K., OHTA, K. & SMITH, G. R. A 160-bp palindrome is a Rad50-Rad32-dependent mitotic recombination hotspot in *Schizosaccharomyces pombe. Genetics.* 161, 461-468, 2002.

52. FERNANDEZ-CAPETILLO, O., LEE, A., NUSSENZWEIG, M. & NUSSENZWEIG, A. H2AX: the histone guardian of the genome. *DNA Repair.* 3, 959-967, 2004.

53. FLORES-ROZAS, H. & KOLODNER, R. D. Links between replication, recombination and genome instability in eukaryotes. *Trends in Biochemical Sciences.* 25, 196-200, 2000.

54. FUGMANN, S., LEE, A. I., SHOCKETT, P. E., VILLEY, I. J., & SCHATZ, D. G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annual Review of Immunology.* 18, 495-527, 2000.

55. GABORIAUD, C., BISSERY, V., BENCHETRIT, T. & MORNON, J. P. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Letters.* 224, 149-155, 1987.

56. GELLERT, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annual Review of Biochemistry.* 71, 101-132, 2002.

57. GREENBERG, R. B., ALBERTI, M., HEARST, J. E., CHUA, M. A. & SAFFRAN, W. A. Recombinational and mutagenic repair of psoralen interstrand cross-links in

Saccharomyces cerevisiae. *The Journal of Biological Chemistry*. 276, 31551-31560, 2001.

58. GROSSMANN, K. F., WARD, A. M. & MOSES, R. E. *Saccharomyces cerevisiae* lacking Snm1, Rev3 or Rad51 have a normal S-phase but arrest permanently in G_2 after cisplatin treatment. *Mutation Research*. 461, 1-13, 2000.

59. GROSSMANN, K. F., WARD, A. M., MATKOVIC, M. E., FOLIAS, A. E. & MOSES, R. E. *S. cerevisiae* has three pathways for DNA interstrand crosslink repair. *Mutation Research*. 487, 73-83, 2001.

60. GUO, L. Q. & YAMAZOE, Y. Inhibition of cytochrome P450 by furanocoumarins in grapefruit juice and herbal medicines. *Acta Pharmacologica Sinica*. 25, 129-136, 2004.

61. GUPTA, R., BECK, J. L., SHEIL, M. M. & RALPH, S. F. Identification of bifunctional GA and AG intrastrand crosslinks formed between cisplatin and DNA. *Journal of Inorganic Biochemistry*. 99, 552-559, 2005.

62. HABER, J. E. DNA recombination: the replication connection. *Trends in Biochemical Sciences*. 24, 271-275, 1999.

63. HAMER, G., ROEPERS-GAJADIEN, H. L., VAN DUYN-GOEDHART, A., GADEMAN, I. S., KAL, H. B., VAN BUUL, P. P., ASHLEY, T. & DE ROOIJ, D. G. Function of DNA-protein kinase catalytic subunit during the early meiotic prophase without Ku70 and Ku86. *Biology of Reproduction*. 68, 717-721, 2003a.

64. HAMER, G., ROEPERS-GAJADIEN, H. L., VAN DUYN-GOEDHART, A., GADEMAN, I. S., KAL, H. B., VAN BUUL, P. P. & DE ROOIJ, D. G. DNA double-

strand breaks and γ-H2AX signaling in the testis. *Biology of Reproduction.* 68, 628-634, 2003b.

65. HARPER, W., BURTON, J. L. & SOLOMON, M. J. The anaphase-promoting complex: it's not just for mitosis any more. *Genes & Development.* 16, 2179-2206, 2002.

66. HENRIQUES, J. A., BROZMANOVA, J. & BRENDEL, M. Role of *PSO* genes in the repair of photoinduced interstrand cross-links and photooxidative damage in the DNA of the yeast *Saccharomyces cerevisiae. Journal of Photochemistry and Photobiology. B, Biology*. 39,185-196, 1997.

67. HENRIQUES. J. A. & MOUSTACCHI, E. Isolation and characterization of *pso* mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae. Genetics.* 95, 273-288, 1980.

68. HEYER, W. -D. Recombination: Holliday junction resolution and crossover formation. *Current Biology.* 14, R56-R58, 2004.

69. HUANG, J. C., HSU, D. S., KAZANTSEV, A. & SANCAR, A. Substrate spectrum of human excinuclease: repair of abasic sites, methylated bases, mismatches, and bulky adducts. *Proceedings of the National Academy of Sciences of the United States of America.* 91, 12213-12217, 1994.

70. ILIAKIS. G., WANG, H., PERRAULT, A. R., BOECKER, W., ROSIDI, B., WINDHOFER, F., WU, W., GUAN, J., TERZOUDI, G. & PANTELIASC, G. Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenetic and Genome Research.* 104, 14-20, 2004.

71. ISHIAI, M., KIMURA, M., NAMIKOSHI, K., YAMAZOE, M., YAMAMOTO, K., ARAKAWA, H., AGEMATSU, K., MATSUSHITA, N., TAKEDA, S., BUERSTEDDE, J. –M. & TAKATA M. DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Molecular and Celular Biology.* 24, 10733-10741, 2004.

72. JACKSON, S. P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis.* 23, 687-689, 2002.

73. JEGGO, P. & O'NEILL, P. The Greek goddess, Artemis, reveals the secrets of her cleavage. *DNA Repair.* 1, 771-777, 2002.

74. JEGGO, P. A. Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. *Radiation Research.* 150 (1998) S80–S91.

75. JONES, D. T. & SWINDELLS, M. B. Getting the most from PSI-BLAST. *Trends in Biochemical Sciences.* 27, 161-164, 2002.

76. KAINA, B. DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling. *Biochemical Pharmacology.* 66, 1547-1554, 2003

77. KANAAR, R., HOEIJMAKERS, J. H. & VAN GENT, D. C. Molecular mechanisms of DNA double-strand break repair. *Trends in Cell Biology.* 8, 483-489, 1998.

78. KIMURA S., SAOTOME, A., UCHIYAMA, Y., MORI, Y., TAHIRA, Y. & SAKAGUCHI, K. The expression of the rice (*Oryza sativa* L.) homologue of Snm1 is

induced by DNA damages. *Biochemical and Biophysical Research Communications.* 329, 668-672, 2005.

79. KORKINA, L. G., DEEVA, I. B., DE BIASE, A., IACCARINO, M., ORAL, R., WARNAU, M. & PAGANO, G. Redox-dependent toxicity of diepoxybutane and mitomycin C in sea urchin embryogenesis. *Carcinogenesis* 21, 213-220, 2000.

80. KOUNDRIOUKOFF, S., POLO, S. & ALMOUZNI, G. Interplay between chromatin and cell cycle checkpoints in the context of ATR/ATM-dependent checkpoints. *DNA Repair.* 3, 969-978, 2004.

81. KROGH, B. O. & SYMINGTON, L. S. Recombination proteins in yeast. *Annual Review of Genetics.* 38, 233-271, 2004.

82. KUMAR, G. S., LIPMAN, R., CUMMINGS, J. & TOMASZ, M. Mitomycin C-DNA adducts generated by DT-diaphorase: revised mechanism of the enzymatic reductive activation of mitomycin C. *Biochemistry.* 36, 14128-14136, 1997.

83. LABHART, P. Nonhomologous DNA end joining in cell-free systems. *European Journal of Biochemistry.* 265, 849-861, 1999.

84. LAMBERT, S., MASON, S. J., BARBER, L. J., HARTLEY, J. A., PEARCE, J. A., CARR, A. M. & MCHUGH, P. J. *Schizosaccharomyces pombe* checkpoint response to DNA inter-strand cross-links. *Molecular and Cellular Biology.* 23, 4728-4737, 2003.

85. LAWLEY, P. D. & PHILLIPS, D. H. DNA adducts from chemotherapeutic agents. *Mutation Research.* 355, 13-40, 1996.

86. LE DEIST, F., POINSIGNON, C., MOSHOUS, D., FISCHER, A., & DE VILLARTAY, J. -P. Artemis sheds new light on V(D)J recombination. *Immunological Reviews.* 200, 142-155, 2004.

87. LEADON, S. A. & AVRUTSKAYA, A.V. Differential involvement of the human mismatch repair proteins, hMLH1 and hMSH2, in transcription-coupled repair. *Cancer Research.* 57, 3784-3791, 1997.

88. LEES-MILLER, S. P. & MEEK, K. Repair of DNA double strand breaks by non-homologous end joining. *Biochimie.* 85, 1161-1173, 2003.

89. LI, L., PETERSON, C. A., LU, X., WEI, P. & LEGERSKI, R. J. Interstrand cross-links induce DNA synthesis in damaged and undamaged plasmids in mammalian cell extracts. *Molecular and Celular Biology.* 19, 5619-5630, 1999.

90. LI, W. -H. Molecular Evolution. Sinauer Associates Publishers, Inc., Massachusets, Estados Unidos da América. 1997.

91. LI, X., HEJNA, J. & MOSES, R. E. The yeast Snm1 protein is a DNA 5'-exonuclease. *DNA Repair.*4, 163-170, 2005.

92. LIEBER, M. R. The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. *Genes to Cells.* 4, 77-85, 1999.

93. LIEBER, M. R., MA, Y., PANNICKE, U. & SCHWARZ, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nature Reviews. Molecular Cell Biology.* 4, 712-720, 2003.

94. LINDAHL, T. & WOOD R.D. Quality control by DNA repair. *Science*. 286, 1897-1905, 1999.

95. LOBACHEV, K. S., GORDENIN, D. A. & RESNICK, M. A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell*. 108, 183-193, 2002.

96. LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389, 251-260, 1997.

97. MAGAÑA-SCHWENCKE, N., HENRIQUES, J. A. P., CHANET, R. & MOUSTACCHI, E. 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 of the United States of America*. 79, 1722-1726, 1982.

98. MALINGE, J. M., GIRAUD-PANIS, M. J. & LENG, M. Interstrand cross-links of cisplatin induce striking distortions in DNA. *Journal of Inorganic Biochemistry*. 77, 23-29, 1999.

99. MARÉCHAL, E. & CESBRON-DELAUW, M. -F. The apicoplast: a new member of the plastid family. *Trends in Plant Science*. 6, 200-205, 2001.

100. MCHUGH, P. J., SONES, W. R. & HARTLEY, J. A. Repair of intermediate structures produced at DNA interstrand cross-links in *Saccharomyces cerevisiae*. *Molecular and Celular Biology*. 20, 3425-3433, 2000.

101. MCNULTY, J. M., JERKOVIC, B., BOLTON, P.H. & BASU, A. K. Replication inhibition and miscoding properties of DNA templates containing a site-specific cis-thymine glycol or urea residue. *Chemical Research in Toxicology.* 11, 666-673, 1998.

102. MEMISOGLU, A. & SAMSON, L. Base excision repair in yeast and mammals. *Mutation Research.* 451, 39-51, 2000.

103. MICHEL, B., GROMPONE, G., FLORÈS, M. –J. & BIDNENKO, V. Multiple pathways process stalled replication forks. *Proceedings of the National Academy of Sciences of the United States of America.* 101, 12783-12788, 2004.

104. MITCHELL, J. R., HOEIJMAKERS, J. H. & NIEDERNHOFER, L. J. Divide and conquer: nucleotide excision repair battles cancer and ageing. *Current Opinion in Cell Biology.* 15, 232-240, 2003.

105. MOLINIER, J., STAMM, M. E. & HOHN, B. SNM-dependent recombinational repair of oxidatively induced DNA damage in *Arabidopsis thaliana. EMBO Reports.* 5, 994-999, 2004.

106. MOORE, J. D. & KREBS, J. E. Histone modifications and DNA double-strand break repair. *Biochemistry and Cell Biology.* 82, 446-452, 2004.

107. MORISON, W. L. Psoralen ultraviolet A therapy in 2004. *Photodermatology, Photoimmunology & Photomedicine.* 20, 315-320, 2004.

108. MOSHOUS, D., CALLEBAUT, I., CHASSEVAL, R., CORNEO, B., CAVAZZANA-CALVO, M., LE DEIST, F., TEZCAN, I., SANAL, O., BERTRAND, Y., PHILIPPE, N., FISCHER, A., & DE VILLARTAY, J. -P. Artemis, a novel DNA double-strand

break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell.* 105, 177-186, 2001.

109. MOUNT, D. W. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, Nova Iorque, Estados Unidos da América. 2001.

110. NATH, N., MCCARTNEY, R. R. & SCHMIDT, M. C. Yeast Pak1 kinase associates with and activates Snf1. *Molecular and Cellular Biology.* 23, 3909-3917, 2003.

111. NELSON, D. L. & COX, M. M. Lehninger Principles of Biochemistry. Worth Publishers, Nova Iorque, Estados Unidos da América. 2000.

112. NIEDERNHOFER, L. J., ODIJK, H., BUDZOWSKA, M., VAN DRUNEN, E., MAAS, A., THEIL, A. F., DE WIT, J., JASPERS, N. G., BEVERLOO, H. B., HOEIJMAKERS, J. H. & KANAAR, R. The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Molecular and Celular Biology.* 24, 5776-5787, 2004.

113. OHINATA, Y., SUTOU, S. & MITSUI, Y. A novel testis-specific RAG2-like protein, Peas: its expression in pachytene spermatocyte cytoplasm and meiotic chromatin. *FEBS Letters.* 537, 1-5, 2003.

114. OSBORNE, M. R., WILMAN, D. E. V. & LAWLEY, P. D. Alkylation of DNA by the nitrogen mustard bis(2-chloroethyl)methylamine. *Chemical Research in Toxicology.* 2, 316-320, 1995.

115. PANASCI, L., ZHI-YUAN, X. & ALOYZ, R. Role of non homologous end-joining and recombinational DNA repair in resistance to nitrogen mustard and DNA

crosslinking agents. Em *Cancer Drug Discovery and Development: DNA Repair in Cancer Therapy* Editado por Panasci, L. C. & Alaoui-Jamali, M. A. Humana Press Inc., Totowa, Estados Unidos da América. 2004.

116. PÂQUES, F. & HABER, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews.* 63, 349-404, 1999.

117. PASTINK, A. & LOHMAN, P. H. M. Repair and consequences of double-strand breaks in DNA. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis.* 428, 141-156, 1999.

118. PASTINK, A., EEKEN, J. C. J. & LOHMAN, P. H. M. Genomic integrity and the repair of double-strand DNA breaks. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis.* 480, 37-50, 2001.

119. PASTUGLIA, M., ROBY, D., DUMAS, C. & COCK, J. M. Rapid induction by wounding and bacterial infection of an *S* gene family receptor like kinase gene in *Brassica oleracea. The Plant Cell.* 9, 49-60, 1997.

120. PETERSON, C. L. Chromatin remodeling: nucleosomes bulging at the seams. *Current Biology.* 12, R245-R247, 2002.

121. PHILLIPS, A., JANIES, D. & WHEELER, W. Multiple sequence alignment in phylogenetic analysis. *Molecular Phylogenetics and Evolution.* 16, 317-330, 2000.

122. PRAKASH, S. & PRAKASH, L. Nucleotide excision repair in yeast. *Mutation Research.* 451, 13-24, 2000.

123. PURDY, A. & SU, T. T. Telomeres: not all breaks are equal. *Current Biology.* 14, R613-R614, 2004.

124. RASSOOL, F. V. DNA double strand breaks (DSB) and non-homologous end joining (NHEJ) pathways in human leukemia. *Cancer Letters.* 193, 1-9, 2003.

125. RAVANAT, J. -L., DOUKI, T. & CADET, J. Direct and indirect effects of UV radiation on DNA and its components. *Journal of Photochemistry and Photobiology B: Biology.* 63, 88-102, 2001.

126. REARDON, J. T., BESSHO, T., KUN, H. C., BOLTON, P. H. & SANCAR A. *In vitro* repair of oxidative DNA damage by human nucleotide excision repair system: possible explanation for neurodegeneration of xeroderma pigmentosum patients. *Proceedings of the National Academy of Sciences of the United States of America.* 94, 9463-9468, 1997.

127. RICHARDSON, C. & JASIN, M. Coupled homologous and non-homologous repair of a double-strand break preserves genomic integrity in mammalian cells. *Molecular and Cellular Biology.* 20, 9068-9075, 2000.

128. RICHIE, C. T., PETERSON, C., LU, T., HITTELMAN, W. N., CARPENTER, P. B. & LEGERSKI, R. J. hSnm1 colocalizes and physically associates with 53BP1 before and after DNA damage. *Molecular and Celular Biology.* 22, 8635-8647, 2002.

129. RINK, S. M. & HOPKINS, P. B. A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry.* 34, 1439-1445, 1995.

130. ROONEY, S., ALT, F. W., LOMBARD, D., WHITLOW, S., ECKERSDORFF, M., FLEMING, J., FUGMANN, S., FERGUSON, D. O., SCHATZ, D. G. & SEKIGUCHI J. Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. *The Journal of Experimental Medicine.* 197, 553-565, 2003.

131. ROUSE, J. & JACKSON, S. P. Interfaces between the detection, signaling, and repair of DNA damage. *Science.* 297, 547-551, 2002.

132. SANDERSON, B. J. S. & SHIELD, A. J. Mutagenic damage to mammalian cells by therapeutic alkylating agents. *Mutation Research.* 355, 41-57, 1996.

133. SHÄRER, O. D. DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem.* 6, 27-32, 2005.

134. SHARPLES, G. J. & LEACH, D. R. Structural and functional similarities between the SbcCD proteins of *Escherichia coli* and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. *Molecular Microbiology.* 17, 1215-1217, 1995.

135. SLUPPHAUG, G., KAVLI, B. & KROKAN, H. E. The interacting pathways for prevention and repair of oxidative DNA damage. *Mutation Research.* 531, 231-251, 2003.

136. SOULIOTIS, V. L., DIMOPOULOS, M. A. & SFIKAKIS, P. P. Gene-specific formation and repair of DNA monoadducts and interstrand cross-links after therapeutic exposure to nitrogen mustards. *Clinical Cancer Research.* 9, 4465-4474, 2003.

137. SWEENEY, M. J. & DOBSON, A. D. W. Mycotoxin production by *Aspergillus, Fusarium* and *Penicillium* species. *International Journal of Food Microbiology.* 43, 141-158, 1998.

138. TAKATA, M., SASAKI, M. S., SONODA, E., MORRISON, C., HASHIMOTO, M., UTSUMI, H., YAMAGUCHI-IWAI, Y., SHINOHARA, A. & TAKEDA, S. 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, 1998.

139. TAMURA, K., ADACHI, Y., CHIBA, K., OGUCHI, K. & TAKAHASHI, H. Identification of Ku70 and Ku80 homologues in *Arabidopsis thaliana:* evidence for a role in the repair of DNA double strand breaks. *The Plant Journal* 29, 771-781, 2002.

140. THOMPSON, C. L. & SANCAR, A. Photolyase/cryptochrome blue-light photoreceptors use photon energy to repair DNA and reset the circadian clock. *Oncogene.* 21, 9043-9056, 2002.

141. THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.*, 22, 4673-4680, 1994.

142. TIMSON, D. J., SINGLETON, M. R. & WIGLEY, D. B. DNA ligases in the repair and replication of DNA. *Mutation Research.* 460, 301-318, 2000.

143. TOMASZ, M. Mitomycin C: small, fast and deadly (but very selective). *Chemistry & Biology.* 2, 575-579, 1995.

144. TOMKINSON, A. E. & MACKEY, Z. B. Structure and function of mammalian DNA ligases. *Mutation Research.* 407, 1-9, 1998.

145. TURNER, B. M. Cellular memory and the histone code. *Cell.* 111, 285-291, 2002.

146. TUTEJA, N., SINGH, M. B., MISRA, M. K., BHALLA, P. L. & TUTEJA, R. Molecular mechanisms of DNA damage and repair: progress in plants. *Critical Reviews in Biochemistry and Molecular Biology.* 36, 337-397, 2001.

147. VAN DEN BOSCH, M., LOHMAN, P. H. M. & PASTINK, A. DNA double-strand break repair by homologous recombination. *Biological Chemistry.* 383 (2002) 873–892.

148. VINK, A. A. & ROZA, L. Biological consequences of cyclobutane pyrimidine dimers. *Journal of Photochemistry and Photobiology B: Biology.* 65, 101-104, 2001.

149. WARREN, A. J., MACCUBBIN, A. E. & HAMILTON, J.W. Detection of mitomycin C-DNA adducts in vivo by ^{32}P-post-labeling: time-course for formation and removal of adducts and biochemical modulation. *Cancer Research.* 58, 453-461, 1998.

150. WEDAMAN, K. P., REINKE, A., ANDERSON, S., YATES, J. 3RD., MCCAFFERY, J. M. & POWERS, T. Tor kinases are in distinct membrane-associated protein complexes in *Saccharomyces cerevisiae. Molecular Biology of the Cell.* 14, 1204-1220, 2003.

151. WETERINGS, E. & VAN GENT, D. C. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair.* 3, 1425-1435, 2004.

152. WILBORN, F. & BRENDEL, M. 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, 1989.

153. WOLFFE, A. P. Histone H1. *The International Journal of Biochemistry & Cell Biology.* 29, 1463-1466, 1997.

154. WOOD, R. D., MITCHELL, M., SGOUROS, J. & LINDAHL, T. Human DNA repair genes. *Science.* 291, 1284-1289, 2001.

155. YU, J., MARSHALL, K., YAMAGUCHI, M., HABER, J. E. & WEIL, C. F. Microhomology-dependent end joining and repair of transposon induced DNA hairpins by host factors in *Saccharomyces cerevisiae. Molecular and Cellular Biology.* 24, 1351-1364, 2004.

156. ZHANG, X., SUCCI, J. FENG, Z., PRITHIVIRAJSINGH, S. STORY, M. D. & LEGERSKI, R. J. Artemis is a phosphorylation target of ATM and ATR and is involved in the G_2/M DNA damage checkpoint response. *Molecular and Cellular Biology.* 24, 9207-9220, 2004.

157. ZHOU, B. B. & ELLEDGE, S. J. The DNA damage response: putting checkpoints in perspective. *Nature.* 408, 433-439, 2000.

158. ZHU, H. & SHUMAN, S. A primer-dependent polymerase function of *Pseudomonas aeruginosa* ATP-dependent DNA ligase (LigD). *The Journal of Biological Chemistry.* 280, 418-427, 2004.

VII

Anexos

ANEXO I

TABELA PARA A NOMENCLATURA DE GENES E VIAS DE REPARAÇÃO DE DNA

Adaptado dos Anais do Simpósio: "*DNA Repair and Mutagenesis: From Molecular Structures to Biological Consequences*". Southampton, Bermudas, 2004.

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	*Human*	Activity
Base excision repair (BER)						
DNA glycosylases						Major altered base released:
	ung+	*UNG1*	*ung1+*	–	*UNG*	U
	–	–	–	*CG5285*	*SMUG1*	U, hydroxymethyl U
	–	–	–	–	*MBD4 (MED1)*	U or T opposite G at CpG sequences
	mug+	–	*thp1+*	*Thd1*	*TDG*	U, T or ethenoC opposite G
	fpg+ (mutM+)	*OGG1*		*Ogg1*	*OGG1*	8-oxoG opposite C
	mutY+	–	*myh1⁺*	–	*MYH*	A opposite 8-oxoG
	nth+	*NTG1, NTG2*	*nth1⁺*	*CG9272*	*NTH1 (NTHL1)*	Ring-saturated or fragmented pyrimidines
	alkA+, tagA+	*MAG1*	*mag1+, SPBC23G7.11*		*MPG (MAG, AAG)*	3-meA, ethenoA, hypoxanthine
	nei+	–	–	–	*NEIL1*	Removes thymine glycol
		–	–	–	*NEIL2*	Removes oxidative products of C, U
		–	–	–	*NEIL3*	Removes fragmented/oxidized pyrimidines

Pathways	E. coli	S. cerevisiae	S. pombe	Drosophila	Human	Activity
Other BER factors						
	xthA+	*APN2 (ETH1)*	*apn2+*	*Rrp1*	*APEX1 (HAP1, APE1, REF1)*	AP endonuclease
	–	–	–	*ApII*	*APEX2 (APE2)*	AP endonuclease
	nfo+	*APN1*	*apn1+*		–	AP endonuclease
	–	–	–	*CG17227*	*LIG3*	DNA ligase
		–	–	*XRCC1*	*XRCC1*	Accessory factor for *LIG3* & BER
	–	–	–	*Parp*	*PARP1 (ADPRT)*	Poly(ADP-ribose) polymerase
	–	–	–		*PARP2 (ADPRTL2)*	ADPRT-like enzyme
Direct reversal of damage						
	phrA+	*PHR1*	–	*phr*	–	CPD photolyase
	–	–	–	*phr6-4*	–	(6-4) photolyase
	–	–	*uve1*$^{+}$ *(uvde*$^{+}$*)*	–	–	UV damage endonuclease
	ada+, ogt+	*MGT1*	*SPAC1250.04c*	*agt*	*MGMT (AGT)*	O^6-meG alkyltransferase
	alkB+				*ABH2*	Reversal of alkylation damage (1-meA and 3-meC)
					ABH3 (DEPC-1)	Reversal of alkylation damage (1-meA and 3-meC)
Repair of DNA-protein cross-links						
		TDP1	*SPCP31B10.05*	*Tdp1*	*TDP1*	Removes covalently bound TopoI-DNA complexes
Mismatch excision repair (MMR)						
	mutS+	*MSH2*	*swi8+*	*spel1 (spellchecker1)*	*MSH2*	Mismatch and loop recognition
		MSH3	*swi4*	–	*MSH3*	
		MSH6	–	*CG7003*	*MSH6*	

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	Human	Activity
					MSH4	MutS homologs specialized for meiosis
					MSH5	
	mutL+			*mlh1*	*MLH1*	MutL homologs, forming dimer
		PMS1	*mlh1+*	*pms2*	*PMS2*	
					PMS1	MutL homolog
					MLH3	MutL homologs of unknown function
					PMS2L3	MutL homologs of unknown function
					PMS2L4 (PMS6)	MutL homologs of unknown function
	mutH+					GATC recognition
	uvrD+ (mutU+)					helicase aiding excision in MMR and NER
Nucleotide excision repair (NER)						
		RAD4	*rhp41+, rhp42+*	*mus210*	*XPC*	Binds distorted DNA as complex
		RAD23	*rhp23+*	*Rad23*	*RAD23B (HR23B)*	
					RAD23A (HR23A)	*RAD23B* paralog
		RAD14	*rhp14+*	*Xpac*	*XPA*	Binds DNA & proteins in pre-incision complex
	uvrA+	–	–	–	–	Binds damaged DNA in complex with UvrB
	uvrB+	–	–	–	–	Catalyzes unwinding in pre-incision complex
TFIIH subunits						Catalyzes unwinding in pre-incision complex
		SSL2 (RAD25)	*ercc3sp+*	*hay (haywire)*	*XPB (ERCC3)*	3' to 5' DNA helicase TFIIH subunit
		RAD3	*rad15+ (rad5+)*	*Xpd*	*XPD (ERCC2)*	5' to 3' DNA helicase TFIIH subunit

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	*Human*	Activity
		TFB1	*tfb1+*	*Tfb1*	*GTF2H1*	TFIIH subunit p62
		SSL1	*ssl1+*	*Ssl1*	*GTF2H2*	TFIIH subunit p44
		TFB4	*tfb4+*	*Tfb4*	*GTF2H3*	TFIIH subunit p34
		TFB2	*tfb2+*	*Tfb2*	*GTF2H4*	TFIIH subunit p52
		TFB5		*CG31917*	*GTF2H5 (TTDA)*	TFIIH subunit p8
		KIN28	–	*Cdk7*	*CDK7*	Kinase subunits of TFIIH
		CCL1	–	*CycH*	*CCNH*	
		TFB3	–	*Mat1*	*MNAT1 (MAT1)*	TFIIH subunit
NER nucleases						
	uvrC+, cho+					3' and 5' incision nuclease
		RAD2	*rad13+*	*mus201*	*XPG (ERCC5)*	3' incision nuclease
		RAD10	*swi10+*	*Ercc1*	*ERCC1*	5' incision nuclease subunits
		RAD1	*rad16+*	*mei9*	*XPF (ERCC4)*	
		RAD28	–	–	*CSA (CKN1, ERCC8)*	Cockayne syndrome; needed for TC-NER
	mfd+	*RAD26*	*rhp26+*	–	*CSB (ERCC6)*	Cockayne syndrome; needed for TC-NER
	–	–	*ddb1+*	*Ddb1*	*DDB1*	p127 subunit of DDB
	–	–	–	–	*DDB2 (XPE)*	p48 subunit of DDB, defective in XP-E
		RAD7	*rhp7+*		–	E3 ubiquitin ligase and damage binding
		RAD16	*rhp16+*		–	
		MMS19		*Mms19*	*MMS19L (MMS19)*	Transcription and NER
DNA Ligase I	*ligA+*	*CDC9*	*cdc17+*	*DNA-ligI*	*LIG1*	DNA joining
Single-stranded DNA binding protein	*ssb+*	*RFA1*	*ssb1+*	*RpA-70*	*RPA1*	Binds ssDNA intermediates in recombination, NER & gap-filling pathways
		RFA2	*ssb2+*	*RpA-30*	*RPA2*	
		RFA3	*ssb3+*	*RpA-8*	*RPA3*	
Homologous recombination (HR)						
	recA+	*RAD51*	*rhp51+*	*Rad51 (spn-A)*	*RAD51*	Formation of protein filament to mediate homologous pairing
					RAD51L1 (RAD51B)	Rad51p paralog

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	*Human*	Activity
				spn-D	*RAD51C*	Rad51p paralog
					RAD51L3 (RAD51D)	Rad51p paralog
		DMC1	*dmc1+*		*DMC1*	Rad51p paralog for meiosis
				Rad51D	*XRCC2*	DNA break and cross-link repair
	recB+, recC+, recD+					Generation of ssDNA to allow formation of RecA filament
				spn-B	*XRCC3*	DNA break and cross-link repair
	recF+, recO+, recR+	*RAD52*	*rad22+ (rad22a+), rti1+ (rad22b+)*		*RAD52*	Accessory factor for recombination
		RAD54	*rhp54+*	*okra*	*RAD54L*	Accessory factor for recombination
					RAD54B	Accessory factor for recombination
		RAD55	*rhp55+*	–	–	Recombination mediator function
		RAD57	*rhp57+*	–	–	
	–	*RAD59*	–	–	–	
		RHC18	*rad18^{+}*			
					BRCA1	Recombination; E3 ubiquitin ligase
					BRCA2 (FANCB, FANCD1)	Cooperation with *RAD51*, essential function
	sbcC+	*RAD50*	*rad50+*	*rad50*	*RAD50*	ATPase in complex with Mre11Ap, Nbs1p
	sbcD+	*MRE11*		*mre11*	*MRE11A*	3' exonuclease
		XRS2	*nbs1+*	*nbs*	*NBS1*	Mutated in Nijmegen breakage syndrome
	ruvA, ruvB					Branch migration of Holliday junctions

Pathways	E. coli	S. cerevisiae	S. pombe	Drosophila	Human	Activity
	ruvC		*(mus81-eme1)+?*			Nuclease to cleave Holliday junctions
Non-homologous end-joining (NHEJ)						
		HDF1	*pku70+*	*Irbp*	*Ku70 (G22P1)*	DNA end binding
		HDF2	*pku80+*	*Ku80*	*Ku80 (XRCC5)*	DNA end binding
		–	–		*PRKDC (DNA-PKcs, XRCC7)*	DNA-dependent protein kinase catalytic subunit
		LIG4		*ligase4*	*LIG4*	Ligase
		LIF4			*XRCC4*	Ligase accessory factor
					Artemis (SNM1C)	Nuclease
Modulation of nucleotide pools						
	MutT+	–	–	*CG10898*	*MTH1 (NUDT1)*	8-oxoGTPase
	Dut+			*dUTPase*	*DUT*	dUTPase
					p53R2	p53-inducible ribonucleotide reductase small subunit 2
DNA polymerases (catalytic subunits)	*polB+ (dinA, pol II)*					Damage responses
					POLB (β, beta)	BER in nuclear DNA
		MIP1	*SPCC24B10.22+*	*tam (tamas)*	*POLG (γ, gamma)*	Replication & BER in mitochondrial DNA
		CDC2 (POL3)	*cdc6+*	*DNA-polδ*	*POLD1 (δ, delta)*	NER and MMR
		POL2	*cdc20+*	*DNA-polε*	*POLE1 (ε, epsilon)*	NER and MMR
		REV3	*rev3+*	*mus205*	*REV3L (ζ, zeta) (PSO1)*	DNA pol zeta catalytic subunit
		REV7		*rev7*	*REV7 (MAD2L2)*	DNA pol zeta subunit
		REV1		*Rev1*	*REV1L (REV1)*	dCMP transferase & other roles in TLS
	umuC+					Catalytic subunit of Pol V for lesion bypass

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	*Human*	Activity
		RAD30	*eso1+*	*DNA-pol*η	*POLH* (η, *eta*)	Bypass of CPD, defective in XP-V
					POLI (ι, *iota; RAD30B*)	Lesion bypass
	–	–	–	*mus308*	*POLQ* (θ, *theta*)	Lesion bypass; DNA cross-link repair?
	dinB+ (*Pol IV*)				*POLK* (κ, *kappa; DINB1*)	Lesion bypass
		POL4	*SPAC2F7.06c*		*POLL* (λ, *lambda*)	Meiotic function?
					POLM (μ, *mu*)	Gap filling during non-homologous end-joining
	–	–	–	–	*POLN* (ν, *nu; POL4P*)	DNA cross-link repair?
		POL5	*pol5+*			
DNA polymerase (accessory factors)						
	dnaN+	*POL30*	*pcn1+*	*mus209*	*PCNA*	Sliding clamp
	dnaX+	*CDC44*	*rfc1+*	*Gnf1*	*RFC1*	Clamp loader, large subunit
Processing nucleases						
		MUS81	*mus81+*	*mus81*	*MUS81*	Structure-specific nuclease subunits
		MMS4	*eme1+*	*CG12936-PA*	*MMS4*	
	polA+ (*5' to 3'exo*)	*RAD27 (RTH1)*	*rad2+*	*l(3)04108*	*FEN1 (DNase IV)*	5' nuclease
					TREX1 (DNase III)	3' exonuclease
					TREX2	3' exonuclease
	recJ+, Exo1+	*EXO1*	*exo1+*	*tos (tosca)*	*EXO1 (HEX1)*	Exonuclease for MMR and other pathways
		SPO11		*meiW-68*	*SPO11*	Recombination endonuclease
	nfi+ (EndoV+)	–	*SPAC1F12.06c*	–	*ENDOV (FLJ35220)*	Incision 3' of hypoxanthine & uracil

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	*Human*	Activity
Rad6p pathway						
		RAD6		*UbcD6*	*UBE2A (RAD6A)*	E2 ubiquitin-conjugating enzyme
					UBE2B (RAD6B)	E2 ubiquitin-conjugating enzyme
		RAD18	*rhp18+*		*RAD18*	Assists repair or replication of damaged DNA
		HPR5 (SRS2, RADH)	*srs2+*			
		MMS2			*UBE2V2 (MMS2)*	Ubiquitin-conjugating complex
					UBE2N (UBC13, BTG1)	Ubiquitin-conjugating complex
Genes defective in diseases associated with sensitivity to DNA damaging agents						
	recQ+	*SGS1*	*rqh1+ (hus1+, rad12+)*	*mus309*	*BLM*	Bloom syndrome helicase
				mus304	*WRN*	Werner syndrome helicase / 3' exonuclease
					RECQL4	Rothmund-Thompson syndrome
		TEL1	*rad3+, tel1+*	*atm*	*ATM*	Ataxia telangiectasia
		HNT3	*SPCC18.09c*	*CG5316*	*APTX*	Ataxia-oculomotor apraxia syndrome (Aprataxin; interaction with *XRCC1, XRCC4*)
Fanconi anemia						
				–	*FANCA*	Tolerance or repair of DNA cross-links
				–	*FANCB*	
				–	*FANCC*	
	–	–	–	*fancd2*	*FANCD2*	
	–	–	–	–	*FANCE*	

Pathways	*E. coli*	*S. cerevisiae*	*S. pombe*	*Drosophila*	*Human*	Activity
	–	–	–	–	*FANCF*	
	–	–	–	–	*FANCG (XRCC9)*	
Other genes related to DNA repair						
		PSO2 (SNM1)		*mus322*	*DCLRE1A (PSO2, SNM1)*	DNA cross-link repair nuclease
		RAD5 (SNM2)				Post-replication repair ATPase
					SNM1B (DCLRE1B)	Related to *SNM1*
					PNKP (PNK)	Converts some DNA breaks to ligatable ends
				mus301 (spn-C)	*HEL308*	Similar to helicase domain of Mus308
Other conserved DNA damage response genes						
	–	*H2A*	*hta1+, hta2+*	*His2av*	*H2AFX (H2AX)*	Histone, phosphorylated after DNA damage
	–	–	–	*p53*	*p53 (TP53)*	Transcription factor & DNA binding
		MEC1	*rad3+, tel1+*	*mei-41*	*ATR*	ATM- and PI-3K-like essential kinase
		DDC2	*rad26+*	*mus304*	*ATRIP*	ATR interacting
		RAD17	*rad1+*	*rad1*	*RAD1*	PCNA-like DNA damage sensor (9-1-1 complex)
		DDC1	*rad9+*	*rad9*	*RAD9*	
		MEC3	*hus1+*	*Hus1-like*	*HUS1*	
		RAD24	*rad17+*	*Rad17*	*RAD17*	RFC1-like DNA damage sensor
		RAD9	*crb2+ (rhp9+)*			Checkpoint function
		CHK1	*chk1^{+} (rad27^{+})*	*grp (grapes)*	*CHEK1*	Effector kinase
		RAD53	*cds1+*	*lok (loki)*	*CHK2 (CHEK2)*	Effector kinase

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Mutation Research 544 (2003) 179–193

Reviews in Mutation Research

www.elsevier.com/locate/reviewsmr
Community address: www.elsevier.com/locate/mutres

ANEXO II

Role of PSO genes in repair of DNA damage of *Saccharomyces cerevisiae*

Martin Brendel [a], Diego Bonatto [a], Martin Strauss [b], Luis Fernando Revers [a], Cristina Pungartnik [a], Jenifer Saffi [a], João Antonio Pegas Henriques [a,*]

[a] *Departamento de Biofísica, Centro de Biotecnologia, UFRGS, Av. Bento Gonçalves 9500, 91507-970 Porto Alegre, RS, Brazil*
[b] *Institut für Mikrobiologie der J.W. Goethe Universität, Frankfurt/Main, Germany*

Received 5 May 2003; received in revised form 26 June 2003; accepted 30 June 2003

Abstract

Photoactivated psoralens used in treatment of skin diseases like Psoriasis and Vitiligo cause DNA damage, the repair of which may lead to mutations and thus to higher risk to have skin cancer. The simple eukaryote *Saccharomyces cerevisiae* was chosen to investigate the cells' genetic endowment with repair mechanisms for this type of DNA damage and to study the genetic consequences of such repair. Genetic studies on yeast mutants sensitive to photoactivated psoralens, named *pso* mutants, showed their allocation to 10 distinct loci. Cloning and molecular characterization allowed their grouping into three functional classes: (I) the largest group comprises seven PSO genes that are either generally or speci cally involved in error-prone DNA repair and thus affect induced mutability and recombination; (II) one PSO gene that represents error-free excision repair, and (III) two PSO genes encoding proteins not in uencing DNA repair but physiological processes unrelated to nucleic acid metabolism. Of the seven DNA repair genes involved in induced mutagenesis three PSO loci [*PSO1/REV3*, *PSO8/RAD6*, *PSO9/MEC3*] were allelic to already known repair genes, whereas three, *PSO2/SNM1*, *PSO3/RNR4*, and *PSO4/PRP19* represent new genes involved in DNA repair and nucleic acid metabolism in *S. cerevisiae*. Gene *PSO2* encodes a protein indispensable for repair of interstrand cross-link (ICL) that are produced in DNA by a variety of bi- and polyfunctional mutagens and that appears to be important for a likewise repair function in humans as well. In silico analysis predicts a putative endonucleolytic activity for Pso2p/Snm1p in removing hairpins generated as repair intermediates. The absence of induced mutation in *pso3/rnr4* mutants indicates an important role of this subunit of ribonucleotide reductase (RNR) in regulation of translesion polymerase ζ in error-prone repair. Prp19p/Pso4p in uences ef cienc y of DNA repair via splicing of pre-mRNAs of intron-containing repair genes but also may function in the stability of the nuclear scaffold that might in uence DNA repair capacity. The seventh gene, *PSO10* which controls an unknown step in induced mutagenesis is not yet cloned. Two genes, *PSO6/ERG3* and *PSO7/COX11*, are responsible for structural elements of the membrane and for

Abbreviations: WT, wild type; UVC, irradiation with 254 nm UV-light; 8-MOP + UVA, 3-CPs + UVA, pre-treatment with, respectively, 8-methoxypsoralen or 3-carbethoxypsoralen and irradiation with 365 nm UV-light; HN2, nitrogen mustard; HN1, nitrogen half mustard; MNNG, *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine; 4NQO, 4-nitroquinoline *N*-oxide; 8HQ, 8-hydroxyquinolineoxide; NDEA, *N*-nitrosodiethylamine; cis-DDP, *cis*Platin; ICL, interstrand cross-link; DSB, DNA double-strand breaks; NER, nucleotide excision repair; aa, amino acid

* Corresponding author. Tel.: +55-51-3316-6069; fax: +55-51-3316-6084.
E-mail address: pegas@dna.cbiot.ufrgs.br (J.A.P. Henriques).

1383-5742/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrrev.2003.06.018

a functional respiratory chain (RC), respectively, and their function thus indirectly in uences sensitivity to photoactivated psoralens.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Psoralen sensitivity; Yeast; DNA repair; Mutagenesis; Cross-link repair; pre-mRNA splicing; Ribonucleotide reductase; Ergosterol; Cytochrome c oxidase; Oxidative stress

1. Introduction

Photoactivated psoralens are already used for decades in the phototherapy (PUVA) of diverse skin diseases, e.g. Psoriasis and Vitiligo [1]. For some skin disorders this PUVA seems to exert its bene cial role through the production of DNA lesions [2,3], slowing down cell proliferation. Repair of these DNA lesions, especially the highly toxic ICL that are formed by PUVA with bi-functional psoralens [4,5], may however lead to errors, mutations [3] and thus to a higher risk for skin cancer. For a closer study of the basic mechanisms of repair of PUVA-induced DNA lesions that may lead to genetic changes and for the isolation of the responsible genes Henriques and Moustaccchi chose the unicellular eukaryote *Saccharomyces cerevisiae* [6] and isolated mutants sensitive to mono- and bi-functional photoactivated psoralens, the so-called *pso* mutants. So far, 10 *pso* mutants have been phenotypically studied and nine have been molecularly cloned and characterized. Though the last review on the *pso* mutants was published only 2 years ago [7], we wish to update this information as two new PSO genes have been cloned and substantial progress has been made in molecular and biochemical characterization of these and some others. It is for the latter reason that the depth of presentation will not be even and we, therefore, refer the reader interested in phenotypical and genetic details to our previous reviews on the *pso* mutants [8,9].

Based on all collected data, we can now divide the PSO genes into three distinct groups: with seven members group I includes the majority of the PSO genes that, either directly or indirectly, play a role in the mechanisms of error-prone repair of induced DNA damage. Two *pso* mutants were found sensitive to PUVA because they lacked protection against ROS or had an altered energy metabolism (group II), while one PSO locus could be associated with error-free NER (group III). Thus, while eight *pso* mutants are

Table 1
The 10 PSO loci: allelism, protein size and function, and phenotypes

Gene	Protein (kDa); function	Phenotype of mutant
PSO1/REV3	173; catalytic subunit of DNA polymerase ζ	Sensitive to radiation and chemical mutagens; low mutability
PSO2/SNM1	72; endonucleolytic cleavage of hairpins in ICL repair?	Sensitive to all ICL-inducing treatments; low mutability only with ICL mutagens
PSO3/RNR4	40; small subunit (Y4) of ribonucleotide reductase	Low induced mutability & recombination sensitive to low temperature growth, to ROS
PSO4/PRP19 (essential)	56.7; spliceosome associated protein	Mutagen sensitive, no mitotic recombination, low mutability, no sporulation at 30 °C; *pso4-1* thermoconditional mutant allele
PSO5/RAD16	91.3; DNA helicase of global NER	UVC-, oxidative damage-sensitive; involved in control of ageing; normal mutability
PSO6/ERG3	43; ergosterol desaturase	Sensitive to ROS, to nystatin, to calco uor white; chitin overproduction and maldistribution
PSO7/COX11	28; cytochrome c oxidase	Growing cells sensitive to 4NQO, 8HQ, NDEA, *PSO7-1* leaky mutant allele
PSO8/RAD6/UBC2	19.6; ubiquitin conjugating DNA repair	Low mutability, variable sporulation slow growth at 16 °C
PSO9/MEC3	53.2; cell cycle control regulation of DNA repair?	UVC sensitive, normal sporulation low mutability
PSO10	Unknown error-prone repair	Sensitive to PUVA, MNNG, UVC low induced mutability

interfering with the function of repair of DNA lesions, two are not (Table 1). We will adhere to this division of the 10 known PSO genes into three groups in the following compact presentation of their role in survival of mutagen-treated yeast.

2. PSO genes involved in error-prone repair

2.1. Genes PSO1, PSO8, and PSO2 encode proteins directly involved in DNA repair

Within this large group of seven genes we nd three loci encoding proteins participating in DNA repair processes. *PSO1* is allelic to *REV3* [10] that encodes the catalytic subunit of polymerase ζ (Rev3p/Rev7p dimer; polζ) [11], a translesion polymerase that can extend a mismatched primer or, generally, mismatched DNA by a few nucleotides beyond the blocking DNA lesion [12–14]. Polζ has been thoroughly studied in yeast and Rev3p homologs exist in mouse (Rev3L) [15] and two in human cells (hREV3) [16–18] and REV3l [17]. Human *REV7* and *REV1* homologs also have been found [19,20] and thus it seems likely that the three components of polymerase ζ (as known from yeast) may have the same function in humans [20].

Mutant *pso8-1* is complemented by the *RAD6* gene and sequencing of the mutant allele showed it to contain a C → T missense mutation in triplet 64, leading to a mutant protein Rad6-P64L [21]. The functions of Rad6p in ubiquitin conjugation and in DNA repair have been thoroughly investigated [22,23]. Despite a large number of molecularly characterized *rad6* mutant alleles, no Rad6-P64L mutant protein has yet been described [24]. Since homoallelic *pso8-1/pso8-1* diploids show nearly WT-like sporulation [21], *pso8-1* belongs to a group of *rad6* alleles that still retain some functionality.

Mutant *pso2-1* was found to be especially sensitive to the bi-functional 8-MOP + UVA treatment and to HN2 [6,25]. Its allelism to the *snm1* mutant [10] (*s*ensitive to *n*itrogen *m*ustard [26]) showed that *pso2/snm1* mutants had a novel phenotype: speci c sensitivity to highly cytotoxic mutagens that are able to produce, apart from mono-functional lesions, DNA ICL [26]. Speci c sensitivity of the *pso2/snm1* mutants is mainly independent of the type of bases involved in formation of such ICL and the chemical make-up of the bi- or polyfunctional mutagen [7]. Stability of ICL, however, may vary greatly [9] and this determines toxicity of the respective mutagen. The most stable ICL described is the dithymidyl psoralen cross-link that is formed by photoactivated 8-MOP. Molecular dosimetry of this ICL shows that between 16 and 20 of these lesions de ne one lethal hit (LD_{37}) for a haploid WT yeast cell [27].

Thermoconditional mutant *snm1-2ts* [28], now called *pso2-12ts* [7] carries two silent point and a missense mutation that replaces glycine with arginine at aa position 256 [29], thereby altering the hydrophilic domain of the protein. Temperature-shift experiments show a complete Pso2p function within 5–6 h after introduction of ICL by HN2 treatment [28]. Pso2p has a nuclear localisation signal and indeed is localized there [29]. With approximately 0.3 transcripts/cell *PSO2* is poorly transcribed [30] but may be induced about four-fold by ICL-producing mutagens, including UVC, but not by the monofunctional alkylating agent MNNG or by the UV-mimeticum 4NQO [31]. Inducibility appears to depend on the presence of a damage response element (DRE)-like motif in the *PSO2* promoter [30] while a downstream silencer within the ORF of neighbour-gene *CIN4* (formerly *GTP1*) [32] is responsible for its low level of constitutive expression.

DNA repair in 8-MOP + UVA-, HN2-, and *cis*DDP-treated *pso2-1* and *pso2-11* mutants showed that incision near ICL, and also (partial) excision of the DNA damage proceeds in a WT fashion but that a later step, reconstituting high molecular weight DNA from low molecular weight DNA generated by early incision/excision events, is failing [33,34]. Since incision of cross-linked DNA by enzymes of the NER is normal in *snm1/pso2* mutants, Snm1p/Pso2p is thought to be speci cally involved in a post-incision step [34] of ICL repair. Biochemical and genetic evidence suggested that two modes of DNA repair, *RAD3*- and *RAD51*-like, are necessary to remove ICL [35] and recent studies suggest that, depending on the physiological stage of the yeast cells (growing versus resting) recombinational, non-homologous end-joining, and error-prone repair was involved in repair of HN2- and *cis*DDP-induced ICL [36].

The β-lactamase motif in the C-terminus of Pso2p allows allocation of Pso2p/Snm1p to the family of β-CASP proteins (motif named after

the metallo-β-lactamase associated *C*PSF *A*rtemis *S*NM1/*P*SO2) that have nucleic acid substrates in common [37]. This family comprises eukaryotic DNA repair enzymes (Snm1p/Pso2p, Artemis) and RNA processing proteins (cleavage and the poly-adenylation speci city factor (CPSF) 73 kDa subunit). The β-CASP family of proteins have a conserved sequence in the C-terminal part of the domain [37]. Starting from these ndings the phylogenetic relationship amongst homologous proteins with functional groups comprising endo- and exo-nucleolytic, phosphotransferase or ligase function have been researched and a phylogram for Snm1p/Pso2p has been established by in silico studies. The results clearly indicate a close phylogenetic relationship of Pso2p/Snm1p to the RAG-1/RAG-2 proteins that are known to have a function in V(J)D recombination [38] and that can be biochemically de ned as proteins with endonucleolytic and phosphotransferase activities. This nding then points to a likewise role of Pso2p in DNA repair, most probably in cleavage of hairpin structures formed when a replication fork will be stalled in the neighbourhood of a stable ICL lesion (Fig. 1). This ICL removal, though leading to loss of some nucleotides and hence mutagenic, might be the critical step in overcoming the absolute block in DNA replication that is constituted by a single stable ICL lesion. While this model is clearly speculative, it certainly leads to new biochemical approaches to understand ICL repair.

Genes *PSO3*, *PSO4*, and *PSO9* encode proteins that indirectly in uence the cell's mutability, i.e. that affect physiological processes regulating the ef cienc y of error-prone DNA repair

Mutant *pso3-1* had the weakest PUVA sensitivity phenotype of the three initially isolated *pso* mutants [6] and showed nearly blocked PUVA-induced mutagenesis and mitotic gene conversion [25,39,40]. The *pso3-1* mutant is only weakly sensitive to alkylating agents, to UVC and to γ-radiation [6,39], but still exhibits defective reverse and forward mutation [39] and induced mitotic recombination by specifically reducing induced gene conversion [40] after treatment with these mutagens. The *pso3-1* mutant was also found sensitive to superoxide anion-generating paraquat [7,8,41], as well as to H_2O_2, to cadmium chloride, and to formaldehyde, suggesting an impaired repair of oxidative stress-related DNA-lesions. The original *pso3-1* mutant's higher-than-WT resistance to MNNG and high number of spontaneously generated "petites" is based on its low pools in glutathione conferred by a *gsh1*-leaky allele. After substitution of the *ghs1*-leaky by a WT allele the resulting *pso3-1 GSH1* mutant lost some phenotypes, e.g. MNNG resistance, high "petite" induction, formaldehyde and cadmium sensitivity [41]. Double mutants containing *pso3-1* and selected *rad* mutant alleles revealed epistasis of *rad3-12* and *pso2-1* mutant alleles for sensitivity to 3-CPs and 8-MOP photoaddition [42], placing the *PSO3* gene into the RAD3 pathway (NER).

Several attempts at molecular cloning *PSO3* via complementation of one of the sensitivity or non-mutability phenotypes yielded some suppressor genes that could only complement or partially complement the sensitivity phenotype of *pso3-1* while WT-like induced mutability or mitotic gene conversion was never fully restored in the respective transformants. Using the recently discovered cold sensitivity phenotype conferred by *pso3-1* the WT allele could be cloned. All *pso3-1* related phenotypes could be complemented by the *RNR4* locus that encodes the second small subunit of yeast ribonucleotide reductase (RNR) [43,44]. Molecular characterization of the *pso3-1* mutant allele showed it to encode an Rnr4-G119R missense protein with residual functionality. Whereas *rnr4Δ* mutants show neither signi cant induced forward nor reverse mutation [45] they exhibit weakly induced mutation at low mutagen doses when transformed with a single vector-contained *pso3-1* mutant allele, with the same induction kinetics as a *pso3-1* mutant strain [25].

RNR is a highly regulated enzyme complex. As a downstream member of the pathway that regulates the DNA damage response in yeast Dun1p has two functions in that it induces the degradation of Sml1p, a suppressor of RNR activity, and concomitantly activates transcription of the RNR genes [46]. Thus, the very low concentration of Rnr3p found under non-damage conditions is elevated, and Rnr1Rnr3 heterodimers with high catalytic activity are formed [47]. This, and the overall higher number of RNR enzyme complexes strongly elevate dNDP production, that is still enhanced by the Dun1p-mediated inactivation of the RNR suppressor Sml1p [46]. Chabes et al. [48] showed that a relaxed feedback inhibition which is induced by DNA damage (Sml1p destruction

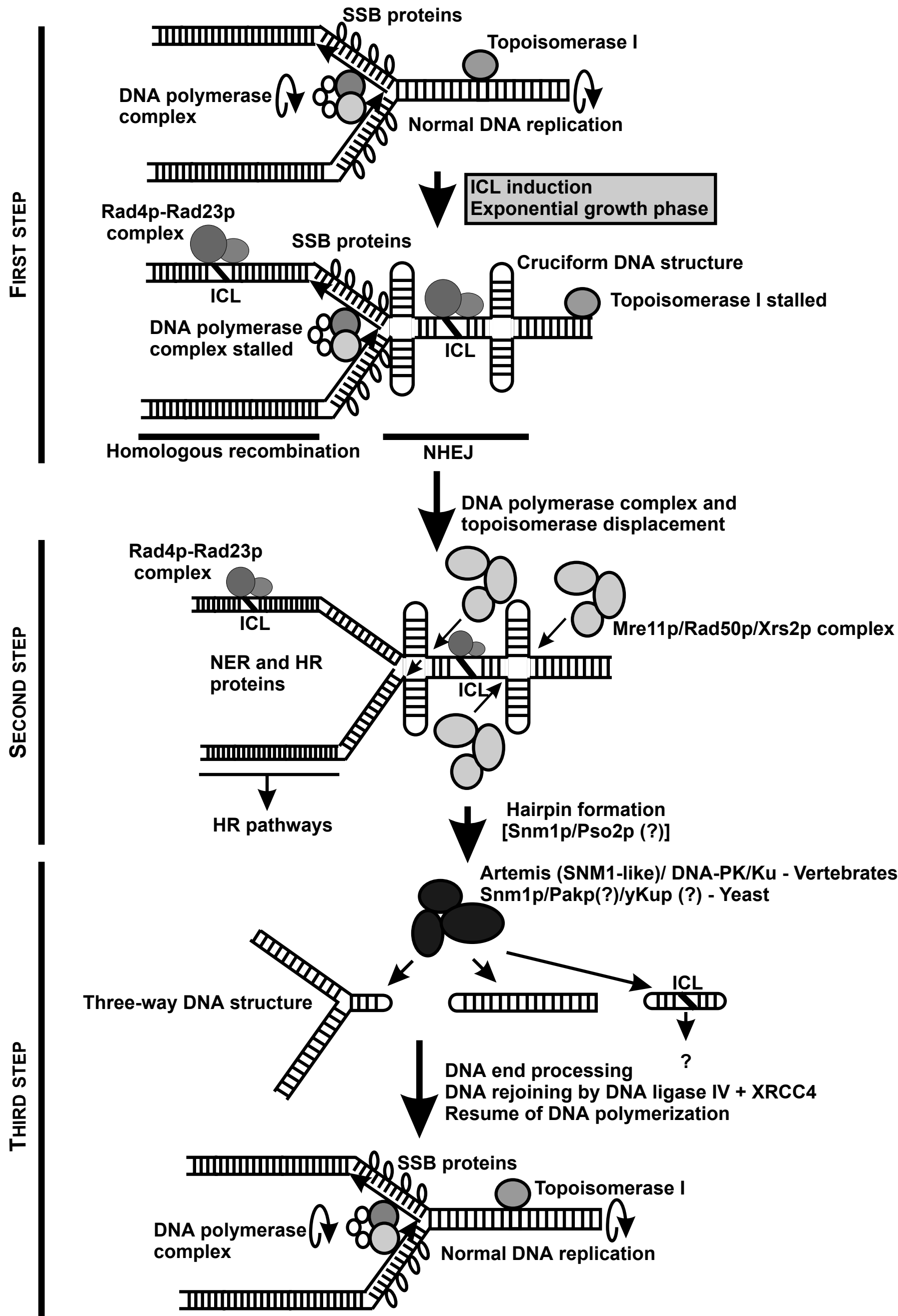


Fig. 1. Three-step model of Snm1p/Pso2p-mediated NHEJ recombination. The DNA polymerase complex and topoisomerase I in growing cells are blocked by an ICL in DNA. In the presence of a sister strand, DNA repair may proceed via homologous recombination, mediated by NER system Rad4p-Rad23p. Alternatively, stalled polymerase and topoisomerase I induce the formation of cruciform DNA structures in the vicinity of ICL, especially when palindromic sequences are present. Cruciform structures are recognized by the Mre11p-Rad50p-Xrs2p complex that cuts the single strand DNA regions and induces the formation of hairpins at the extremities of DNA with (or without) the help of Snm1p. In the third step, both three-way DNA structures, DNA capped hairpin extremities, and ICL-containing capped hairpins are cleaved by the Artemis (SNM1-like)/DNA-PK/Ku protein complex in vertebrates or by Snm1p/Pakp/Kup in yeast. DNA ligase IV and Xrcc4 then join the fragment ends and DNA replication is resumed by DNA polymerase. ICL-containing DNA is degraded.

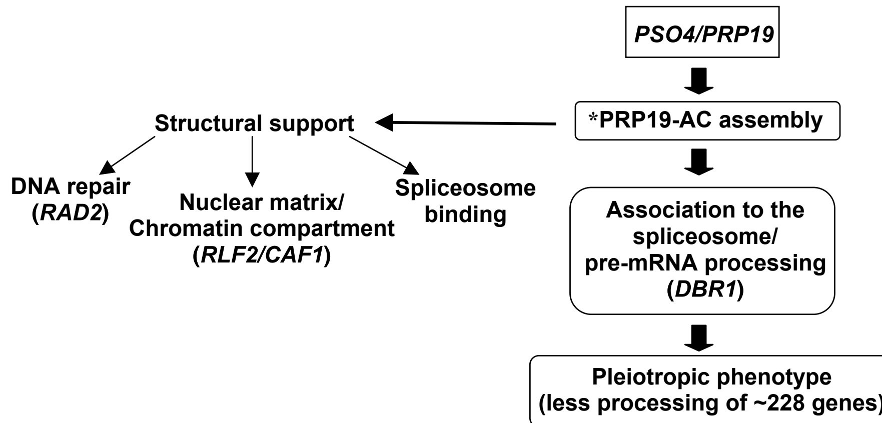


Fig. 2. Possible functions of Prp19p/Pso4p in yeast. Large arrows: known function; thin arrows: possible other function. The asterisk shows the process impaired by the mutant Pso4-1p, which thereby blocks all further steps in splicing.

by Dun1p?) leads to signi cantly higher dNTP pools that in turn allow a higher rate of induced mutation, most likely by the enhanced lesion bypass activity of a better dNTP-alimented polζ (Fig. 2). Arguing in the opposite direction we may then deduce that, in absence of a functional Rnr4p in *rnr4Δ* mutants, or in presence of a missense mutant protein Pso3-1p (Rnr4-G119R) the RNR complex functions only poorly and the resulting small dNTP pools are preferentially consumed by replicative polymerases with an at least 10-fold higher binding speci city , leaving polζ without dNTP supply (Fig. 2).

Mutant *pso4-1*, a slightly X-ray sensitive haploid mutant formerly named *xs9* [49], is highly sensitive to photo-activated psoralens [50], has a pleiotropic repair-defect phenotype [9] and is a thermo-conditional mutant with no viability at 36 °C [51]. Molecular cloning showed its allelism to *PRP19* [51], an essential yeast locus encoding a spliceosomal complex-associated protein [52,53].

This suggested two reasons for the pleiotropic phenotype of *pso4-1*: (a) the Pso4p/Prp19p has more than one function, and one of these would directly affect DNA repair and recombination, or (b) non-effective pre-mRNA splicing of one to several of the 238 known intron-containing pre-mRNAs of yeast [54] at permissible temperature would lead to partially impaired cell physiology. Thus, mutagen sensitivity, lower mutability, recombination, and sporulation found for *pso4-1* and its homozygous diploid could all be the result of non-splicing (or partially non-splicing) of the respective pre-mRNAs. This second hypothesis was tested by a simple experiment employing the intron containing *RAD14* repair gene. If faulty processing of *RAD14*-transcribed pre-mRNA were the only reason that would lead to the observed UVC sensitivity in *pso4-1*, a double mutant combining *pso4-1* and *rad14Δ* should show the same UVC sensitivity. This was clearly not the case, the *pso4-1 rad14Δ* double mutant always displaying a higher UVC sensitivity than *rad14Δ* alone (synergistic interaction), even at the permissive temperature, thus pointing to contribution of further gene(s) to the UVC-sensitivity phenotype of *pso4-1* [55].

Several interaction trap screenings yielded 32 isolates that could be allocated to 13 ORFs, of which v e are with as yet unknown function. The remaining eight interacting proteins can be roughly grouped into four functional classes: (a) DNA repair (one ORF), (b) growth and cell cycle regulation (four ORFs), (c) chromatin structure and chromosome dynamics (one ORF), and (d) pre-mRNA splicing (two ORFs) [55]. Eight proteins interact with each other to form the Prp19p-associated complex but not all seem essential for the splicing process. It is tempting to speculate that perhaps one or more of the Pso4p interactors with as yet unknown function might be members of this associated complex or might themselves react (bind to) proteins from the Prp19-associated complex.

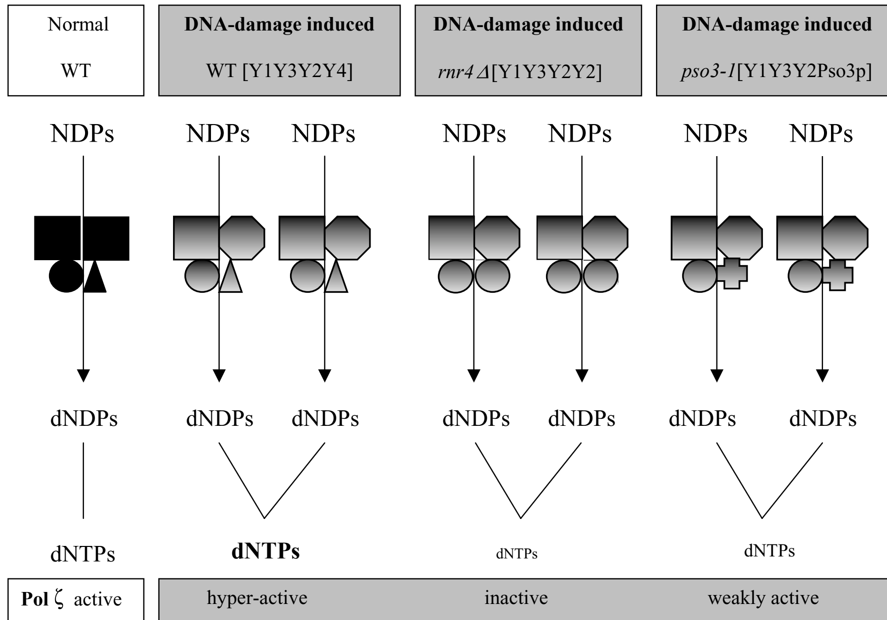


Fig. 3. Putative correlation of error-prone DNA repair with ef cienc y of NDP reduction by yeast ribonucleotide reductase. From left to right: normal RNR enzyme complex in WT; DNA damage-induced ampli cation (symbolized by two enzyme complexes) of RNR in WT with highly ef cient Y1Y3 (square–octogon) heterodimer large subunit; same in mutant *rnr4Δ* where RNR contains (low-ef cienc y) small subunit Y2Y2 (circle–circle) homodimer; same in mutant *pso3-1* where RNR has (low-ef cienc y) small subunit heterodimer Y2Pso3-1p (circle–cross). Whereas above-normal concentration of dNTPs leads to increased mutability [48], lower-than-normal dNTP pools presumably prevent activity of polζ.

Recently the human hNMP200, an ortholog to Pso4p/Prp19p, has been found to be part of the nuclear matrix [56]. This allows us to discuss an additional and different function for Prp19p. Evidence suggests that the nuclear matrix is involved in various nuclear processes and in genome maintenance, e.g. in repair and replication of DNA, in transcription and RNA processing [57–59]. Like hNMP200, Prp19p could be a yeast nuclear matrix protein. As a member of such non-chromatin scaffold it could provide structural support for the machinery processing RNA and/or for protein complexes engaged in replication, repair, and transcription (Fig. 3).

Mutant *pso9-1* is not only sensitive to photo-activated mono- and bi-functional psoralens but also to UVC and MNNG and has impaired mutability after treatment of stationary phase cells with these mutagens. Mutant *pso9-1* is not sensitive to H_2O_2 and paraquat-induced oxidative damage. Reverse mutation was signi cantly lower than in WT but not as much suppressed as in *pso8-1* and *pso10-1* mutants. Forward mutation in the *CAN1* locus was severely inhibited as compared to the WT for all four tested mutagens. These phenotypic characteristics justify its association with the repair group of the *pso* mutants. Molecular cloning showed that *pso9-1* is a mutant allele of *MEC3*. Forming a complex with Ddc1p and Rad17p [60], Mec3p is involved in the regulatory cascade that controls cell cycle response to DNA damage and that eventually induces transcription of repair-relevant genes. Amongst other gene products, *Mec3p* is also required for UVC-induced mutagenesis [61] thus explaining the low UV-mutability phenotype of the *pso9-1* mutant.

As the seventh of *pso* mutants that impair error-prone repair *pso10-1* has not yet been molecularly cloned. Its high sensitivity to photo-activated psoralens and to UVC and its low induced mutability suggest *pso10-1* to lack a functional protein that either regulates error-prone DNA repair or is itself an active repair enzyme. Homoallelic *pso10-1/pso10-1* diploids do not sporulate which might be seen as

further indication that *PSO10* may be a member of the group of genes allocated to the RAD6 epistasis group.

3. One PSO gene involved in nucleotide excision repair

3.1. PSO5 is engaged in global NER

Stationary phase cells harbouring mutant allele *pso5-1* are moderately sensitive to UVC, to the UV-mimeticum 4NQO and to the radiomimetica HN1 and HN2 [62]; they are also cross-sensitive to 3-CPS + UVA, which produces either 4′–5′-furan-side pyrimidine mono-adducts in DNA or singlet oxygen [63,64]. Furthermore, they are sensitive to other oxidative stress-enhancing chemicals, such as H_2O_2 and paraquat [62,65]. By molecular analysis *pso5-1* was found to be a mutant allele of *RAD16* [65]. Rad16p functions in global genome repair, a sub-pathway of NER that preferentially repairs CPD in either the HMLa or HMLα loci [66], in silent regions of DNA, and in the non-transcribed strands of active genes [66–68].

Comparative studies in expression of β-galactosidase from DNA damage-inducible *RNR2–lacZ* and *RNR3–lacZ* fusion constructs in WT and *pso5/rad16* transformants have shown that the DNA damage-induced expression of *RNR2* and *RNR3* not only depends on a functional Pso5p/Rad16p but also on the type of DNA damage [69]. While UVC, 4NQO, and H_2O_2 induced *RNR2* and *RNR3* via DNA damage, the oxidative stressors *t*BOOH and paraquat could not. Thus, the latter two agents may form DNA lesions unable to initiate the signal cascade of inducible NER [70] or the signals are not addressing the Dun1 kinase controlled pathway speci c for DRE motif-inducible genes *RNR2* and *RNR3* [71,72]. For some mutagens, however, Pso5p/Rad16p seems to function in the signal transducing pathway controlling DNA damage-inducible components of NER and associated genes.

Using the two-hybrid system, Rad16p/Pso5p was shown to interact with the *SGS1*-encoded protein Sgs1p [73] that in vitro displays 3′ → 5′ helicase as well as DNA-dependent ATPase activity [74,75]. Sgs1p also interacts with topoisomerases II and III [76,77] and is involved in premature ageing in yeast [78]. It is homologous to the human Wrn, Blm, and RTS proteins, which are, respectively, responsible for the Werner, Bloom, and Rothmund–Thomson syndromes that are related to premature ageing and to cancer proneness [76,79,80]. A *rad16–sgs1* double mutant displays epistasis after treatment with several mutagens. The *sgs1* mutant's sensitivity to MMS, UVC, 4-NQO, and H_2O_2 [73] and the sensitivity to UVC and γ-ray irradiation of *sgs1*Δ [81] point to a function of Sgs1p in DNA repair. It appears that the helicase activity of Sgs1p is responsible for most elements of the *sgs1* mutant phenotype, including its sensitivity to hydroxyurea [82].

Deletion of a functional Rad16p signi cantly reduces the average life span of the mutants [73]. The *rad16*Δ–*sgs1*Δ double mutant displayed a life span comparable to that of the *sgs1*Δ single mutant. This epistatic interaction of *rad16* and *sgs1* mutations suggests that *RAD16*/*PSO5* and *SGS1* share functions in a common pathway of life span control. The life span reduction observed in *pso5*/*rad16* mutants suggests that the RAD52 DNA repair pathway (Rad50p, Rad51p, Rad52p, Rad57p) [83], involved in DSB and homologous recombination repair, is not the only one responsible for life span control in yeast.

4. PSO genes that are involved in mutagen metabolism and not in DNA repair

The two mutants *pso6-1* and *pso7-1* were only slightly sensitive to 8-MOP + UVA but signi cantly so to treatment with the mono-functional 3-CPs + UVA [62]. In addition to 3-CPs-thymine DNA mono-adducts this treatment also generates singlet oxygen, a reactive oxygen species (ROS) that has been shown to intensify oxidative DNA damage [63]. This latter activity may be suppressed in the presence of the singlet oxygen quencher sodium azide [62] and the resulting near WT resistance indicates that the enhanced sensitivity of *pso6-1* and *pso7-1* may be the result of increased DNA damage induced by an activated oxygen species that is generated during photo-activation of 3-CPs. Lack of repair of 3-CPs + UVA induced DNA mono-adducts, therefore, is probably not the reason for the mutants' sensitivity. Since exposure to other mutagens mostly lead to WT-like survival and WT-like induced reverse

mutation, these two mutants clearly differed from the above described other seven. The pronounced sensitivity to other ROS generating, e.g. to paraquat and to H_2O_2, clearly set *pso6-1* apart from *pso7-1* which displayed WT-resistance to these chemicals and was only sensitive to 3-CPs + UVA in stationary phase cells [62].

The *PSO6* gene was molecularly cloned via complementation of the paraquat sensitivity of *pso6-1* and found to be allelic to the *ERG3* locus [84] that encodes the enzyme sterol Δ^5-desaturase. In yeast membranes ergosterol is the most prominent sterol, in contrast to higher eukaryotic cells, where cholesterol is the main sterol. Due to their hydrophobic structure both sterols can contribute to membrane stability. Anchoring amongst the polar long chain of fatty acids, they are thought to influence membrane fluidity, thereby regulating flow, permeability, and enzyme activity, and as a consequence, also cell growth [85,86].

Lipid peroxidation (LP) is known to be one of the most toxic events related to oxidative stress. ROS, especially $OH^{\bullet}$ and $HOO^{\bullet}$, can pull (extract) a bi-allelic hydrogen atom of unsaturated fatty acid (LH) to form lipid alkyl radical ($L^{\bullet}$), which can be oxidized to a lipid peroxil radical ($LOO^{\bullet}$). This may attack adjacent LH and propagate the radical chain reaction [87,88]. Ergosterol is able to inhibit LP [89] and it was suggested [90] that this is due to the sterols of endoperoxide and hydroperoxide formed instead (only from ergosterol and not from episterol) which can protect membrane integrity. Ergosterol appears to play an important role in mediating the cytotoxic effects of singlet oxygen [85,91,92]. Subnormal content of membrane ergosterol in *pso6/erg3* mutants would, therefore, explain their low efficiency of transformation (altered permeability to Li^+), lack of protection from ROS generated in respiratory metabolism of non-fermentable substrates like ethanol and glycerol (uncoupling of oxidative phosphorylation), poor mating (altered fusion ability) and maldistribution of chitin (episterol in place of ergosterol leads to over-expression of the enzyme Chs3, responsible for chitin membrane deposition) [84].

The WT-like resistance to HN1, HN2, 4NQO, and UVC of *pso6/erg3* suggests normal function of DNA repair in this mutant. Oxidative stress, however, might be enhanced by mutagens that themselves generate ROS and this might ultimately lead to a higher number of oxidative base damages in DNA. This is suggested by the enhanced mutability of *pso6/erg3* after treatment with 3-CPs + UVA (and to a lesser extent, after 8-MOP + UVA), whereas induced mutation by UVC, HN1, and HN2 is WT-like [62]. Thus, the *PSO6/ERG3*-encoded sterol Δ^5-desaturase is most probably not involved in any kind of DNA repair, but through its enzyme activity will contribute to a final product that might have a protective function in preventing (or lowering) oxidative stress in yeast cells.

When in exponential phase of growth *pso7-1* mutant cells are highly sensitive to 4NQO and this fact has been exploited to molecularly clone *PSO7* via complementation [93]. The *pso7-1* mutation in yeast gene *COX11* that encodes a protein indispensable for the assembly of a functional cytochrome c oxidase [94], which located in the inner mitochondrial membrane, is the final electron acceptor of the respiratory chain (RC) responsible for reducing O_2 to H_2O. The *pso7-1* mutant allele is leaky as the mutant still contains about 5% of the WT activity of cytochrome c oxidase whereas the *cox11Δ* mutant has no detectable enzyme activity [93]. This allows *pso7-1* to still grow, though very slowly, on non-fermentable substrates [93], while *cox11Δ* strains are petite and only grow in presence of fermentable carbon sources [94]. The RC in *S. cerevisiae* grown on non-fermentable carbon sources behaves as one unit, implying that the different respiratory complexes physically interact [95], i.e. there is a coupling between the steady-state levels of the different complexes [96]. Therefore, the absence of a functional cytochrome c oxidase would not permit the RC to act as a single unit in *cox11Δ*, whereas the leaky *pso7-1* allele might still permit the RC to act as a super-molecular entity with a control coefficient for respiration of one [96].

The moderate to higher sensitivity to oxidative stress-generating treatments like 3-CPs + UVA [63] or 4NQO [97] has been explained by disturbed electron flows in the *pso7/cox11* mutants resulting in a higher rate of toxic LP [98] and to genotoxicity via a higher rate of oxidative DNA damage [63,99]. Also, in *pso7/cox11* altered metabolization of certain mutagens (e.g. 4NQO, that is a pro-mutagen and carcinogen which undergoes a four-electron reduction to become a nitro-radical anion [100]) might lead to a higher-than-normal production of metabolites able to

generate elevated intracellular oxidative stress with a higher DNA damaging potential.

Repair-proficient *pso7/cox11* mutants were found highly sensitive to the mutagens NDEA, an alkylating chemical that is metabolized via redox cycling to yield hydroxylamine radicals, ROS and LP [88,101,102] and to 8-hydroxyquinoline (8HQ), which may also be activated via altered oxygen metabolism [103] and possibly form diol-epoxide derivatives [104,105]. It should be noticed that *pso6/erg3* mutants were also sensitive to NDEA, but not to 8HQ, most probably due to the inability of 8HQ to cause LP [103].

We may state that the sensitivity response of the two non-repair mutants *pso6-1* and *pso7-1* is strictly due to altered metabolism of some mutagens that is caused by alterations of membrane lipids and of complex IV of the RC which ultimately results in enhanced cell inactivation. Thus, although not participating in any role in DNA repair, ergosterol and cytochrome c oxidase may be considered important factors in modulation of intracellular oxidative stress responses. Actually, these two essential metabolic components may be functionally closely related to each other. A perturbation of mitochondrial electron flow can indeed arise from a decrease in ergosterol in the inner mitochondrial membrane (extended oxidative stress if late stages of ergosterol biosynthesis are inhibited) or from a direct interaction between the applied mutagen and the mitochondrial enzyme complexes [106]. This would subsequently enhance permeability of membranes to, e.g. photo-sensitizers that in turn could lead to cellular damage, impairment of mitochondrial function, and cell inactivation [90].

One could, therefore, assume that a *pso6 pso7* double mutant had a much higher sensitivity to predominantly oxidative stress-causing agents. And also, that these two mutants alleles, in conjunction with other specific DNA repair mutant alleles, could be employed as eukaryotic models for the typing of unknown mutagens since they would detect and report, by changes in biological endpoints, specific and different reaction mechanisms of the chemicals: those capable of

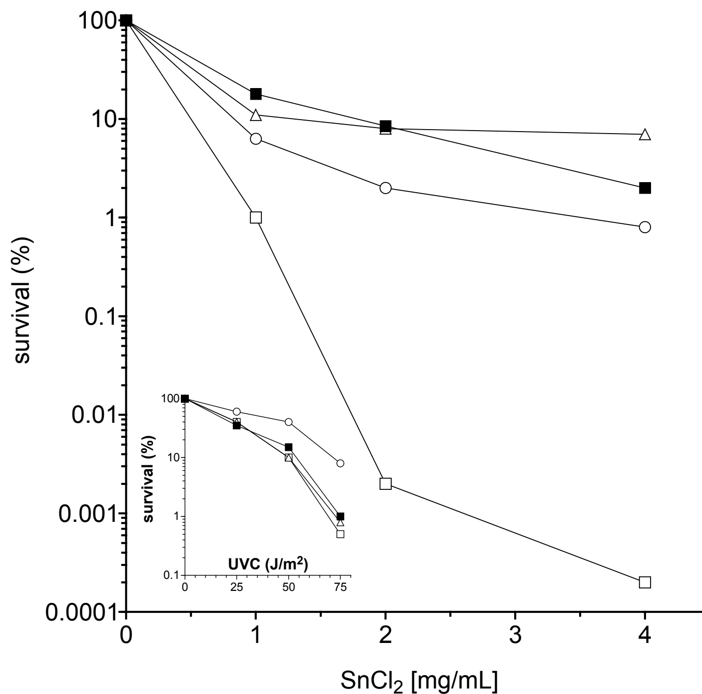


Fig. 4. Genotoxic effect of $SnCl_2$ exposure in four haploid strains of *Saccharomyces cerevisiae* derived from a tetratype ascus with all combinations of *ERG3* and *COX11* WT and mutant alleles: (■) WT; (△) *erg3Δ*, (○) *cox11Δ*, (□) *erg3Δcox11Δ*. Stationary cells were exposed to $SnCl_2$ for 20 min before diluting and plating for survival. Inset shows survival response of the four strains to UVC treatment.

inducing direct DNA damage versus others inducing mainly LP or, when activated via redox cycle, enhancing intracellular oxidative stress. Our rst assumption could be con rmed by showing that a *pso6 pso7* double mutant had a much higher sensitivity to the oxidative stressor $SnCl_2$ [107] than either single mutant (Fig. 4). This clear response of the double mutant (synergistic effect) may then be exploited for the classi cation of putative oxidative stress-causing agents.

5. Concluding remarks

Since 30 years, PUVA is applied in effective treatment of skin disorders. However, this photo-chemotherapeutical method produces, amongst other effects, DNA damage and thus evokes the function of DNA repair mechanisms whose enzymatic reactions might eliminate the DNA lesions either in an error-free or error-prone fashion. The latter process may lead to mutations that sometimes may constitute the rst step in a cascade of malfunctions of cellular physiology, ultimately leading to carcinogenesis. Clinical follow-up studies of patients having received PUVA treatment show that they indeed have an increased risk for skin cancer, especially squamous cell carcinoma [108], while the risk of inducing melanoma is still controversially discussed [109,110]. Molecular epidemiology revealed that more than half of the tumors had at least one mutation in p53 and about half of those could be attributed to PUVA treatment [111]. Also a high frequency in Ha-ras mutations was detected in squamous cell carcinoma from PUVA-treated Psoriasis patients [112]. Interestingly, DNA repair capacity of the treated patients played an important role in the development of PUVA-induced skin cancer as individuals with low repair capacity had a six-fold higher skin-cancer risk, and developed skin cancer at an earlier age, than those with high DNA repair capacity [113]. These ndings clearly show the interrelation of PUVA photo-chemotherapy and DNA repair and suggest that a stimulation of error-free repair processes or, alternatively, a suppression of error-prone repair, in PUVA-treated patients might help to minimize the collateral genetical damage of this treatment.

The universality of many DNA repair mechanisms from microorganisms to mammalian cells [114] suggests that our ndings on the genetical damage induced by PUVA treatment in the unicellular eukaryote *S. cerevisiae* and especially in the *pso* mutants, can contribute to the understanding on PUVA DNA lesion processing in mammalian cells as well. The isolation of *pso* mutants from a mutagenized culture of WT strain N123 seems to have favored selection of mutant alleles quite different from the "mainstream" *rad* and *rev* mutants, as only a third was found allelic to that category of yeast genes (*pso1-1/REV3*, *pso5-1/RAD16*, *pso8-1/RAD6*). While the putative role of Pso2p/Snm1p in ICL repair might indicate a fourth *pso* gene coding for a DNA repair enzyme, all other mutants sensitive to PUVA contained de ciency alleles of loci not directly involved in DNA repair. It is tempting to attribute induction and selection of these *pso* mutants to two special features: (1) PUVA treatment used for mutant selection produces a different spectrum of DNA lesions (psoralen mono- and di-adducts, and, via singlet oxygen, oxidized bases) and (2) the WT N123 from which the *pso* mutants were derived was a *ghs1* leaky mutant, i.e. had a lower than normal glutathione pool. The combination of these peculiarities might have led to the observed spectrum of PSO genes that, together with many RAD loci, are responsible for PUVA resistance and PUVA mutagenesis of the WT, be it by protective measures (*PSO6/ERG3*, *PSO7/COX11*) or by metabolic steps preceding the proper repair of the PUVA-induced DNA lesions (*PSO3/RNR4*, *PSO4/PRP19*, *PSO9/MEC3*).

Without isolation of the *pso* mutants, we would not have had the interesting result that this sensitivity-to-PUVA phenotype had shown us genes as varied as *PSO3/RNR4*, *PSO4/PRP19* and *PSO9/MEC3* to be involved in error-prone DNA repair. Also most certainly, *erg3* and *cox11* mutants would not have been our rst choice when thinking of mutagen sensitivity. It is this uncertainty and the surprises linked to it that makes research both a challenge and a joy.

Acknowledgements

Research was supported by CNPq, CAPES, GENOTOX, and by the DFG. M.S. was fellow of the Hessische Graduiertenförderung. DAAD and CNPq furnished travel grants to J.A.P.H. and M.B.

References

[1] T. Horio, Indications and action mechanisms of phototherapy, J. Dermatol. Sci. 23 (Suppl. 1) (2000) S17–S21.
[2] L. Musajo, G. Rodighiero, F. Dall'Aqua, Evidences of a photoreaction of the photosensitizing furocoumarins with DNA and with pyrimidine nucleosides, Experientia 21 (1965) 24–25.
[3] B.R. Scott, M.A. Pathak, G.R. Mohn, Molecular and genetic basis of furocoumarin reactions, Mutat. Res. 39 (1976) 29–74.
[4] L. Musajo, G. Rodighiero, Mode of photosensitizing action of furocoumarins, Photophysiology 7 (1972) 115–147.
[5] F. Dall'Acqua, S. Marciani, D. Vevaldi, G. Rodighiero, Skin photosensitization and crosslinking formation in native DNA by furocoumarins, Z. Naturforsch. 290 (1974) 635–636.
[6] J.A.P. Henriques, E. Moustacchi, Isolation and characterization of *pso* mutants sensitive to photoaddition of psoralen derivatives in *Saccharomyces cerevisiae*, Genetics 95 (1980) 273–288.
[7] M. Brendel, J.A.P. Henriques, The *pso* mutants of *Saccharomyces cerevisiae* comprise two groups: one de cient in DNA repair and another with altered mutagen metabolism, Mutat. Res. 48 (2001) 79–96.
[8] J.A.P. Henriques, M. Brendel, The role of PSO and SNM genes in dark repair of the yeast *Saccharomyces cerevisiae*, Curr. Genet. 18 (1990) 387–393.
[9] J.A.P. Henriques, J. Brozmanova, M. Brendel, Role of PSO genes in the repair of photoinduced interstrand cross-links and photooxidative damage in the DNA of the yeast *Saccharomyces cerevisiae*, J. Photochem. Photobiol. B 39 (1997) 185–196.
[10] C. Cassier-Chauvat, E. Moustacchi, Allelism between *pso1-1* and *rev3-1* mutants and between *pso2-1* and *snm1* mutants in *Saccharomyces cerevisiae*, Curr. Genet. 13 (1988) 37–40.
[11] C.W. Lawrence, D.C. Hinkle, DNA polymerase zeta and the control of DNA damage induced mutagenesis in eukaryotes, Cancer Surv. 28 (1996) 21–31.
[12] J.R. Nelson, C.W. Lawrence, D.C. Hinkle, Thymine–thymine dimer bypass by yeast polymerase zeta, Science 272 (1996) 1646–1649.
[13] C.W. Lawrence, Cellular roles of DNA polymerase ζ and Rev1 protein, DNA Repair 1 (2002) 425–435.
[14] C.W. Lawrence, P.E.M. Gibbs, R.S. Murante, X.-D. Wang, Z. Li, T.P. McManus, W.G. McGregor, J.R. Nelson, D.C. Hinkle, V.M. Maher, Roles of DNA polymerase ζ and Rev1 protein in eukaryotic mutagenesis and translesion replication, Cold Spring Harbor Sympm. Quant. Biol. 65 (2000) 61–69.
[15] W. Xiao, T. Lechler, B.L. Chow, T. Fontanie, M. Agustus, K.C. Carter, Y.F. Wei, Identi cation, chromosomal mapping and tissue-speci c expression of hREV3 encoding a putative human DNA polymerase zeta, Carcinogenesis 19 (1998) 945–949.
[16] W. Lin, X. Wu, Z. Wang, A full-length cDNA of hREV3 is predicted to encode DNA polymerase zeta for damage-induced mutagenesis in humans, Mutat. Res. 433 (1999) 89–98.
[17] C. Morelli, A.J. Mungall, M. Negrini, G. Barbanti-Brodano, C.M. Croce, Alternative splicing, genomic structure, and ne chromosome localization on REV31, Cytogenet. Cell Genet. 83 (1998) 18–20.
[18] Y. Murakumo, T. Roth, H. Ishii, D. Rasio, S. Numata, C.M. Croce, R. Fishel, A human REV7 homolog that interacts with the polymerase zeta catalytic subunit hREV3 and the spindle assembly checkpoint protein hMAD2, J. Biol. Chem. 275 (2000) 4391–4397.
[19] P.E. Gibbs, X.D. Wang, Z. Li, T.P. McManus, W.G. McGregor, C.W. Lawrence, V.M. Maher, The function of the human homolog of *Saccharomyces cerevisiae REV1* is required for mutagenesis induced by UV light, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 4186–4191.
[20] K. Kawamura, J.O. Wang, R. Bahar, N. Koshikawa, T. Shishikura, A. Nakagawara, S. Sakiyama, K. Kajiwara, M. Kimura, M. Tagawa, The error-prone DNA polymerase zeta catalytic subunit (*Rev3*) gene is ubiquitously expressed in normal and malignant human tissues, Int. J. Oncol. 18 (2001) 97–103.
[21] H. Rolla, M. Grey, C.L. Schmidt, E. Niegemann, M. Brendel, J.A.P. Henriques, Mutant *pso8*-1 of *Saccharomyces cerevisiae*, sensitive to photoactivated psoralens, UV radiation, and chemical mutagens, contains a *rad6* missense mutant allele, Curr. Genet. 41 (2002) 217–223.
[22] S. Sung, S. Prakash, L. Prakash, The RAD6 protein of *Saccharomyces cerevisiae* polyubiquitinates histones, and its acidic domain mediates this activity, Genes Dev. 2 (1988) 1476–1485.
[23] S. Prakash, P. Sung, L. Prakash, DNA repair genes and proteins of *Saccharomyces cerevisiae*, Annu. Rev. Genet. 27 (1993) 33–70.
[24] G. Freiberg, A.D. Mesecar, H. Huang, J.Y. Hong, S.W. Liebman, Characterization of novel *rad6/ubc2* ubiquitin-conjugating enzyme mutants in yeast, Curr. Genet. 37 (2000) 221–233.
[25] C. Cassier, R. Chanet, J.A.P. Henriques, E. Moustacchi, The effects of the three PSO genes on induced mutagenesis: a novel class of mutationally defective yeast, Genetics 96 (1980) 841–857.
[26] M. Kircher, M. Brendel, DNA alkylation by mustard gas in yeast strains of different repair capacity, Chem. Biol. Interact. 44 (1983) 27–39.
[27] M. Bankmann, M. Brendel, Molecular dosimetry of 8-MOP + UVA-induced DNA photoadducts in *Saccharomyces cerevisiae*: correlation of lesion number with genotoxic potential, J. Photochem. Photobiol. B 4 (1989) 57–74.
[28] W. Siede, M. Brendel, Isolation and characterization of yeast mutants with thermoconditional sensitivity to the bifunctional alkylating agent nitrogen mustard, Curr. Genet. 4 (1981) 145–149.
[29] E. Niegemann, M. Brendel, A single amino acid change in SNM1-encoded protein leads to thermoconditional de cienc y for DNA cross-link repair in *Saccharomyces cerevisiae*, Mutat. Res. 315 (1994) 275–279.
[30] D. Richter, E. Niegemann, M. Brendel, Molecular structure of the DNA cross-link repair gene *SNM1* (*PSO2*) of the yeast

Saccharomyces cerevisiae, Mol. Gen. Genet. 231 (1992) 194–200.

[31] R. Wolter, W. Siede, M. Brendel, Regulation of *SNM1*, an inducible *Saccharomyces cerevisiae* gene required for repair of DNA cross-links, Mol. Gen. Genet. 250 (1996) 162–168.

[32] R. Wolter, D. Richter, E. Niegemann, M. Brendel, Molecular characterization of *GTP1*, a *Saccharomyces cerevisiae* gene encoding a small GTP-binding protein, Curr. Genet. 26 (1994) 564–566.

[33] F. Wilborn, M. Brendel, Formation and stability of interstrand cross-links induced by *cis*- and *trans*-diamminedichloroplatinum (II) in the DNA of *Saccharomyces cerevisiae* strains differing in repair capacity, Curr. Genet. 16 (1989) 331–338.

[34] N. Magaña-Schwencke, J.A.P. Henriques, R. Chanet, E. Moustacchi, The fate of 8-methoxypsoralen photoinduced cross-links in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair de cient strains, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 1722–1726.

[35] W.J. Jachymczyk, R.C. von Borstel, M.R.A. Mowat, P.J. Hastings, Repair of interstrand cross-links in DNA of *Saccharomyces cerevisiae* requires two systems of DNA repair: the *RAD3* system and the *RAD51* system, Mol. Gen. Genet. 182 (1981) 196–205.

[36] P.J. McHugh, W.R. Sones, J.A. Hartley, Repair of intermediate structures produced at DNA interstrand cross-links in *Saccharomyces cerevisiae*, Mol. Cell Biol. 20 (2000) 3425–3433.

[37] I. Callebaut, D. Moshous, J.-P. Mornon, J.-P. de Villartay, Mettalo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family, Nucleic Acids Res. 30 (2002) 3592–3601.

[38] Y. Ma, U. Pannicke, K. Schwarz, M.R. Lieber, Hairpin opening and overhang processing by an Artemis/DANN-dependent protein kinase complex in nonhomologos end joining and V(J)D recombination, Cell 106 (2002) 781–794.

[39] C. Cassier, E. Moustacchi, Mutagenesis induced by mono- and bi-functional alkylating agents in yeast mutants sensitive to photo-addition of furocoumarins (*pso*), Mutat. Res. 84 (1981) 37–47.

[40] H.H.R. de Andrade, E. Moustacchi, J.A.P. Henriques, The *PSO3* gene is involved in error-prone intragenic recombinational DNA repair in *Saccharomyces cerevisiae*, Mol. Gen. Genet. 219 (1989) 75–80.

[41] M. Brendel, M. Grey, A.F. Maris, J. Hietkamp, F. Fesus, C.T. Pich, L. Dafré, M. Schmidt, F. Eckardt-Schupp, J.A.P. Henriques, Low glutathione pools in the original *pso3* mutant of *Saccharomyces cerevisiae* are responsible for its pleiotropic sensitivity phenotype, Curr. Genet. 33 (1998) 4–9.

[42] M.S. Benfato, M. Brendel, J.A.P. Henriques, The DNA repair gene *PSO3* of *Saccharomyces cerevisiae* belongs to the *RAD3* epistasis group, Curr. Genet. 21 (1992) 85–90.

[43] P.J. Wang, A. Chabes, R. Casagrande, X.C. Tian, L. Thelander, T. Huffaker, Rnr4p, a novel ribonucleotide reductase small-subunit protein, Mol. Cell. Biol. 17 (1997) 6114–6121.

[44] M. Huang, S.J. Elledge, Identi cation of *RNR4*, encoding a second essential small subunit of ribonucleotide reductase in *Saccharomyces cerevisiae*, Mol. Cell. Biol. 17 (1997) 6105–6113.

[45] M. Strauss, M. Grey, J.A.P. Henriques, M. Brendel, *RNR4* mutant alleles *rnr4Δ* and *pso3-1* block induced forward and reverse mutation, in preparation.

[46] X. Zhao, R. Rothstein, The Dun1 checkpoint kinase phosphorylates and regulates the ribonucleotide reductase inhibitor Sml1, Proc. Natl. Acad. Sci U.S.A. 99 (2002) 3746–3751.

[47] V. Domkin, L. Thelander, A. Chabes, Yeast DNA damage-inducible Rnr3 has a very low catalytic activity strongly stimulated after the formation of a cross-talking Rnr1/Rnr3 complex, J. Biol. Chem. 277 (2002) 18574–18578.

[48] A. Chabes, B. Georgieva, V. Domkin, X. Zhao, R. Rothstein, L. Thelander, Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase, Cell 112 (2003) 391–401.

[49] A. Benaten, C.A. Beam, The genetic control of X-ray resistance of budding yeast cells, Radiat. Res. 69 (1977) 91–116.

[50] J.A.P. Henriques, E.J. Vicente, K.V.C.L. daSilva, A.C.G. Schenberg, *PSO4*: a novel gene involved in error-prone repair in *Saccharomyces cerevisiae*, Mutat. Res. 218 (1989) 111–124.

[51] M. Grey, A. Düsterhöft, J.A.P. Henriques, M. Brendel, Allelism of *PSO4* and *PRP19* links pre-mRNA processing with recombination and error-prone DNA-repair in *Saccharomyces cerevisiae*, Nucleic Acids Res. 24 (1996) 4009–4014.

[52] W.-Y. Tarn, K.R. Lee, S.-C. Cheng, The yeast Prp19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-mRNA and prior to formation of the functional spliceosome, Mol. Cell Biol. 13 (1993) 1883–1891.

[53] S.C. Cheng, W.-Y. Tarn, T.Y. Tsao, J. Abelson, Prp19: a novel spliceosomal component, Mol. Cell Biol. 13 (1993) 1876–1882.

[54] H.W. Mewes, D. Frishman, C. Gruber, B. Geier, D. Haase, A. Kaps, K. Lemcke, G. Mannhaupt, F. Pfeiffer, C. Schuller, S. Stocker, B. Weil, MIPS: a database for genomes and protein sequences, Nucleic Acids Res. 28 (2000) 37–40.

[55] L.F. Revers, J.M. Cardone, H. Feldmann, M. Grey, M. Brendel, J.A.P. Henriques, Thermoconditional modulation of the pleiotropic sensitivity phenotype by the *Saccharomyces cerevisiae* PRP19 mutant allele *pso4-1*, Nucleic Acids Res. 30 (2002) 4993–5003.

[56] T. Erdemir, B. Bilican, D. Oncel, C.R. Goding, U. Yavuzer, DNA damage-dependent interaction of the nuclear matrix protein C1D with translin-associated factor X (TRAX), J. Cell Sci. 115 (2002).

[57] J. Gotzmann, C. Gerner, M. Meissner, K. Holzmann, R. Grimm, W. Mikulits, G. Sauermann, HNMP200: a novel human common nuclear matrix protein combining structural and regulatory functions, Exp. Cell Res. 261 (2000) 166–179.

[58] T. Pederson, Half a century of 'the nuclear matrix', Mol. Biol. Cell 11 (2000) 799–804.
[59] S. Kamiuchi, M. Saijo, E. Citterio, M. deJager, J.H. Hoeijmakers, K. Tanaka, Translocation of Cockayne syndrome group A protein to the nuclear matrix: possible relevance to transcription-coupled DNA repair, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 201–206.
[60] T. Kondo, K. Matsumoto, K. Sugimoto, Role of complex containing Rad17, Mec3, and Ddc1 in the yeast DNA damage checkpoint pathway, Mol. Cell. Biol. 19 (1999) 1136–1143.
[61] A.G. Paulovich, C.D. Armour, L.H. Hartwell, The *Saccharomyces cerevisiae RAD9*, *RAD17*, *RAD24* and *MEC3* genes are required for tolerating irreparable, ultraviolet-induced DNA damage, Genetics 150 (1998) 75–93.
[62] C.B. Querol, S.O. Paesi-Toresan, L.B. Meira, M. Brendel, J.AP. Henriques, Isolation and characterization of three mutants with increased sensitivity to photoactivated 3-carbethoxypsoralen in *Saccharomyces cerevisiae*, Curr. Genet. 25 (1994) 407–411.
[63] E. Sage, T.L. Doan, V. Boyer, D.E. Helland, L. Kittler, C. Helene, E. Moustacchi, Oxidative DNA damage photoinduced by 3-carbethoxypsoralen and other furocoumarins: mechanisms of photo-oxidation and recognition by repair enzymes, J. Mol. Biol. 209 (1989) 297–314.
[64] D. Averbeck, M. Dardalhon, N. Magaña-Schwencke, L.B. Meira, V. Meniel, New aspects of the repair and genotoxicity of psoralen photoinduced lesions in DNA, J. Photochem. Photobiol. B Biol. 14 (1992) 47–63.
[65] S.O. Paesi-Toresan, C.T. Pich, M. Grey, D. Keszenman, M. Brendel, J.A.P. Henriques, Gene *PSO5* of *Saccharomyces cerevisiae*, involved in repair of oxidative damage, is allelic to *RAD16*, Curr. Genet. 27 (1995) 493–495.
[66] D. Schild, B.J. Glasner, R.K. Mortimer, M. Carlson, B.C. Laurent, Identi cation of *RAD16*, a yeast excision repair gene homologous to the recombinational repair gene *RAD54* and to the *SNF2* gene involved in transcriptional activation, Yeast 8 (1992) 385–395.
[67] D.D. Bang, R. Verhage, N. Goosen, J. Brouwer, P. van de Putte, Cloning of *RAD16*, a gene involved in differential repair in *Saccharomyces cerevisiae*, Nucleic Acids Res. 20 (1992) 3925–3931.
[68] R. Verhage, A.M. Zeeman, N. de Groot, F. Gleig, D.D. Bang, P. van de Putte, J. Brouwer, The *RAD7* and *RAD16* genes, which are essential for pyrimidine dimer removal from the silent mating type loci, are also required for repair of nontranscribed strand of an active gene in *Saccharomyces cerevisiae*, Mol. Cell Biol. 14 (1994) 6135–6142.
[69] S.O. Paesi-Toresan, A.F. Maris, M. Brendel, J.A.P. Henriques, The *Saccharomyces cerevisiae* gene *PSO5/RAD16* is involved in the regulation of DNA damage-inducible genes *RNR2* and *RNR3*, Curr. Genet. 34 (1998) 124–127.
[70] T.A. Navas, Y. Sanchez, S.J. Elledge, *RAD9* and DNA polymerase epsilon form parallel sensory branches for tranducing the DNA damage checkpoint signal in *Saccharomyces cerevisiae*, Genes Dev. 10 (1996) 2632–2643.
[71] Z. Zhou, S.J. Elledge, *DUN1* encodes a protein kinase that controls the DNA damage response in yeast, Cell 75 (1993) 1119–1127.
[72] W. Siede, E.C. Friedberg, Regulation of the yeast *RAD2* gene: DNA damage-dependent induction correlates with protein binding to regulatory sequences and their deletion in uences survival, Mol. Gen. Genet. 232 (1992) 247–256.
[73] J. Saf , H. Feldmann, E.L. Winnacker, J.A.P. Henriques, Interaction of the yeast Pso5/Rad16 and Sgs1 proteins: in uences on DNA repair and aging, Mutat. Res. 486 (2001) 195–206.
[74] J. Lu, J.R. Mullen, S.J. Brill, S. Kleff, M. Romeo, R. Sternglanz, Human homologues of yeast helicase, Nature 383 (1996) 678–679.
[75] R.J. Benett, J.A. Sharp, J.C. Wang, Puri cation and characterization of the Sgs1 DNA helicase activity of *Saccharomyces cerevisiae*, J. Biol. Chem. 273 (1998) 9644–9650.
[76] P. Watt, E. Louis, R. Borts, I. Hickson, Sgs1: a eukaryotic homolog of *E. coli* RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation, Cell 81 (1995) 253–260.
[77] S. Gangloff, J. McDonald, C. Bendixen, L. Arthur, R. Rothstein, The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase, Mol. Cell Biol. 14 (1994) 8391–8398.
[78] L. Guarente, Do changes in chromosomes cause aging? Cell 86 (1996) 9–12.
[79] P.W. Watt, I.D. Hickson, R.H. Borts, E.J. Louis, *SGS1*, a homologue of the Bloom's and Werner's Syndrome genes, is required for maintenance of genome stability in *Saccharomyces cerevisiae*, Genetics 144 (1996) 935–945.
[80] F.B. Johnson, D.A. Sinclair, L. Guarente, Molecular biology of aging, Cell (1999) 297–302.
[81] J. Saf , V.R. Pereira, J.A.P. Henriques, Importance of the Sgs1 helicase activity in DNA repair of *Saccharomyces cerevisiae*, Curr. Genet. 37 (2000) 75–78.
[82] J.R. Mullen, V. Kaliraman, S.J. Brill, Bipartite structure of the SGS1 DNA helicase in *Saccharomyces cerevisiae*, Genetics 154 (2000) 1101–1114.
[83] P.U. Park, P. A Defossez, L. Guarente, Effects of mutations in DNA repair genes on formation of ribosomal DNA circles and life span in *Saccharomyces cerevisiae*, Mol. Cell Biol. 19 (1999) 3848–3856.
[84] C.L. Schmidt, M. Grey, M. Schmidt, M. Brendel, J.A.P. Henriques, Allelism of *Saccharomyces cerevisiae* genes *PSO6*, involved in survival after 3-CPs + UVA induced damage, and *ERG3*, encoding the enzyme sterol C-5 desaturase, Yeast 15 (1999) 1503–1510.
[85] S.J. Smith, L.W. Parks, The *ERG3* gene in *Saccharomyces cerevisiae* is required for the utilization of respiratory substrates and in heme-de cient cells, Yeast 9 (1993) 1177–1187.
[86] M.E. van der Rest, A.H. Kamminga, A. Nakano, Y. Anraku, B. Poolman, W.N. Konings, The plasma membrane of *Saccharomyces cerevisiae*: structure, function and biogenesis, Microbiol. Rev. 59 (1995) 304–322.

[87] H. Kappus, H. Sies, Toxic drug-effects associated with oxygen metabolism: redox cycling and lipid peroxidation, Experientia 37 (1981) 1233–1241.
[88] M. Ahotupa, V. Bussacchini-Griot, J.C. Bereziat, A.M. Camus, H. Bartsch, Rapid oxidative stress induced by *N*-nitrosamines, Biochem. Biophys. Res. Commun. 146 (1987) 1047–1054.
[89] H. Wiseman, M. Cannon, H.R.V. Arnstein, B. Halliwell, Enhancement by tamoxigfen of the membrane antioxidant action of the yeast membrane sterol ergosterol: relevance to the antiyeast and anticancer action of tamoxifen, Biochem. Biophys. Acta 1181 (1993) 201–206.
[90] P.W. Albro, P. Bislki, J.T. Corbett, J.L. Schroeder, C.F. Chignell, Photochemical reaction and phototoxicity of sterols: novel self-perpetuation mechanism for lipid photooxidation, Photochem. Photobiol. 66 (1997) 316–325.
[91] W.M. Casey, L.W. Parks, A role for sterols in the porphyrin mediated photosensitization of yeast, Photochem. Photobiol. 50 (1989) 553–556.
[92] T. Bocking, K.D. Barrow, A.G. Netting, T.C. Chilcott, H.G. Coster, M. Hofer, Effects of singlet oxygen on membrane sterols in the yeast *Saccharomyces cerevisiae*, Eur. J. Biochem. 267 (2000) 1607–1618.
[93] C. Pungartnik, M. Brendel, J.A.P. Henriques, Mutant allele *pso7-1*, that sensitizes *Saccharomyces cerevisiae* to photoactivated psoralen, is allelic with *COX11*, encoding a protein indispensable for a functional cytochrome c oxidase, Curr. Genet. 36 (1999) 124–129.
[94] A. Tzagoloff, N. Capitanio, M.P. Nobrega, D. Gatti, Cytochrome oxidase assembly in yeast requires the product of *COX11* a homolog of the *P. denitri cans* protein encoded by ORF3, EMBO J. 9 (1990) 2759–2764.
[95] H. Boumans H, L.A. Grivell, J.A. Berden, The respiratory chain in yeast behaves as a single functional unit, J. Biol. Chem. 273 (1998) 4872–4877.
[96] H.H. Boumans, J.A. Berden, L.A. Grivell, K. van Dam, Metabolic control analysis of the bc_1 complex of *Saccharomyces cerevisiae*: effect on cytochome c oxidase, respiration and growth rate, Biochem. J. 331 (1998) 877–883.
[97] T. Nunoshiba, B. Demple, Potent intracellular oxidative stress exerted by the carcinogen 4-nitroquinoline-*N*-oxide, Cancer Res. 53 (1993) 3250–3252.
[98] A. Boveris, Biochemistry of free radicals: from electrons to tissues, Medicina (Buenos Aires) 58 (1998) 350–356.
[99] A. Martinez, A. Urios, M. Blanco, Mutagenicity of 80 chemicals in *Escherichia coli* tester strains IC203, de cient in OxyR, and its oxyR(+) parent WP2 uvrA/pKM101: detection of 31 oxidative mutagens, Mutat. Res. 467 (2000) 41–53.
[100] T. Yano, S. Takahashi, T. Ichikawa, Active oxygen generated in the process of carcinogen metabolism can induce oxidative damage in nuclei, Res. Commun. Mol. Pathol. Pharmacol. 87 (1995) 367–370.
[101] H. Bartsch, E. Hietanem, C. Malavielle, Carcinogenic nitrosamines: free radical aspects of their action, Free Radic. Biol. Med. 7 (1989) 637–644.
[102] J. Fridovich, Oxygen toxicity: a radical explanation, J. Exp. Biol. (1998) 1203–1209.
[103] C. Pungartnik, J. Picada, M. Brendel, J.A.P. Henriques, Further phenotypic characterisation of *pso* mutants of *Saccharomyces cerevisiae* with respect to DNA repair and response to oxidative stress, Genet. Mol. Res. (2002) 79–89.
[104] E.H. Vock, W.K. Lutz, O. Ilinskaya, S. Vamvakas, Discrimination between genotoxicity and cytotoxicity for the induction of DNA double-strand breaks in cells treated with aldehydes and diepoxides, Mutat. Res. 441 (1999) 85–93.
[105] J.N. Picada, N.N. Khromov-Borisov, J.A.P. Henriques, Deletogenic activity of 1,2:7,8-diepoxyoctane in the *Salmonella typhimurium* tester strain TA102, Mutat. Res. 437 (1999) 165–173.
[106] G.F. Bammert, J.M. Fostel, Genome wide expression patterns in *Saccharomyces cerevisiae*: comparison of drug treatments and genetic alterations affecting biosynthesis of ergosterol, Antimicrob. Agents Chemot. 44 (2000) 1255–1265.
[107] J.C.P. Mattos, F.J.S. Dantas, R.J.A.C. Bezerra, M. Bernardo-Filho, J.B. Cabral-Neto, C. Lage, A.C. Leitão, A. Caldeira-de-Araújo, Damage induced by stannous chloride in plasmid DNA, Toxicol. Lett. 116 (2000) 159–163.
[108] A. Hannuskela-Svahn, E. Pukkala, E. Laara, K. Poikolainen, J. Karvonen, Psoriasis, its treatment, and cancer in a cohort of Finnish patients, J. Invest. Dermatol. 114 (2000) 587–590.
[109] R.S. Stern, The risk of melanoma in association with long-term exposure to PUVA, J. Am. Acad. Dermatol. 44 (2001) 755–761.
[110] B. Lindelof, B. Sigurgeirsson, E. Tegner, O. Larko, A. Johannesson, B. Berne, B. Ljunggren, T. Anersson, L. Molin, E. Nylander-Lundqvist, L. Emtestam, PUVA and cancer risk: the Swedish follow-up study, Br. J. Dermatol. 141 (1999) 108–112.
[111] R.S. Stern, S. Bolshakov, A.J. Nataraj, H.N. Ananthaswamy, p53 mutation in nonmelanoma skin cancers occurring in psoralen ultraviolet A-treated patients: evidence for heterogeneity and eld cancerization, J. Invest. Dermatol. 119 (2002) 522–526.
[112] H. Kreimer-Erlacher, H. Seidl, B. Back, H. Kerl, P. Wolf, High mutation frequency at Ha-ras exons 1–4 in squamous cell carcinomas from PUVA-treated Psoriasis patients, Photochem. Photobiol. 74 (2001) 323–330.
[113] M. Dybdahl, G. Frentz, U. Vogel, H. Wallin, B.A. Nexo, Low DNA repair is a risk factor in skin carcinogenesis: a study of basal cell carcinoma in Psoriasis patients, Mutat. Res. 433 (1999) 15–22.
[114] J.A. Eisen, P.C. Hanawalt, A phylogenomic study of DNA repair genes, proteins, and processes, Mutat. Res. 435 (1999) 171–213.

VIII

Curriculum vitae

Diego Bonatto

CURRICULUM VITAE

PORTO ALEGRE
2005

CURRICULUM VITAE

Abril, 2005

1 DADOS PESSOAIS

Nome: Diego Bonatto
Nome em citações bibliográficas: BONATTO, D.
Sexo: masculino
Filiação: Victor Thadeu Bonatto e Margareth Bonatto
Nascimento: 20/11/1975, Porto Alegre/RS - Brasil
Carteira de identidade: 1057300699 / SJS / RS / 11/02/1998
CPF: 80301690006

Endereço profissional: Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia.
AV. BENTO GONCALVES , 9500 - PREDIO 43421 - Lab. 219
AGRONOMIA
91501-970 PORTO ALEGRE, RS - Brasil - Caixa Postal: 15005
Telefone: (51) 33167608 Ramal: 7608 Fax: 33166084
E-mail: DIEGOBONATTO@HOTMAIL.COM

Endereço residencial: RUA SÃO CARLOS, 244/05
FLORESTA
90220-120 PORTO ALEGRE, RS - Brasil
Telefone: (51) 32285990 Fax: 33166084
E-mail: diegobonatto@gmail.com

2 FORMAÇÃO ACADÊMICA/TITULAÇÃO

1998 - 2000 Mestrado em Biologia Celular e Molecular.
Universidade Federal do Rio Grande do Sul, UFRGS, Rio Grande do Sul, Brasil.
Título: Isolamento e caracterização de novas linhagens eubacterianas produtoras de ácidos poli(hidroxialcanóicos). Ano de obtenção: 2000.
Orientador: João Antonio Pêgas Henriques.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: Ecologia microbiana; Biotecnologia; PHAs.
Áreas do conhecimento: Microbiologia Industrial e de Fermentação; Bacterologia; Biologia Molecular.
Setores de aplicação: Produtos e processos biotecnológicos; Desenvolvimento de novos materiais.

1994 - 1998 Graduação em Ciências Biológicas.
Universidade Federal do Rio Grande do Sul, UFRGS, Rio Grande do Sul, Brasil.
Título: Indução do gene de reparo por excisão de nucleotídeos PSO5 (RAD16) de Saccharomyces cerevisiae em diferentes linhagens desta levedura e em diferentes condições de estresse e cultivo.
Orientador: João Antonio Pêgas Henriques.
Bolsista do(a): Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPQ, Brasil.

2000 Doutorado em Biologia Celular e Molecular.
Universidade Federal do Rio Grande do Sul, UFRGS, Rio Grande do Sul, Brasil.
Título: A proteína Pso2p/Snm1p de Saccharomyces cerevisiae e suas implicações no reparo de DNA.
Orientador: João Antonio Pêgas Henriques.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: Interstrand cross-link; PSO2; Saccharomyces cerevisiae; NHEJ; HMG-1/2; Filogenia de proteínas.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Genética Evolutiva; Biologia e Fisiologia dos Microorganismos.

3 FORMAÇÃO COMPLEMENTAR

2003 - 2003 IV Escola Franco Brasileira de Biofísica.
Universidade Federal do Rio de Janeiro, UFRJ, Rio de Janeiro, Brasil.
2003 - 2003 Métodos de Análise Filogenética. (Carga horária: 3h)
Sociedade Brasileira de Genética, SBG, São Paulo, Brasil.
2001 - 2001 Bioprocessos Industriais. (Carga horária: 4h)
Sociedade Brasileira de Microbiologia, SBM, São Paulo, Brasil.
2001 - 2001 Estrutura Molecular das Bactérias Lácticas. (Carga horária: 4h)
Sociedade Brasileira de Microbiologia, SBM, São Paulo, Brasil.
1999 - 1999 Aplicando Conceitos de Qualidade num Laboratório. (Carga horária: 3h)
Sociedade Brasileira de Genética, SBG, São Paulo, Brasil.
1999 - 1999 História (Humanizada) da Biologia Molecular. (Carga horária: 3h)
Sociedade Brasileira de Genética, SBG, São Paulo, Brasil.
1998 - 1998 IV Encontro Gaúcho de Imunologia.
Sociedade Brasileira de Imunologia - Regional Sul, SBI, Rio Grande do Sul, Brasil.
1998 - 1998 Marcadores Citológicos e Moleculares em Mutagênese. (Carga horária: 3h)
Sociedade Brasileira de Genética, SBG, São Paulo, Brasil.
1998 - 1998 Expressão Gênica em Fungos. (Carga horária: 3h)
Sociedade Brasileira de Genética, SBG, São Paulo, Brasil.
1998 - 1998 Extensão universitária em I Conferência Científica da UFRGS. (Carga horária: 4h)
Universidade Federal do Rio Grande do Sul, UFRGS, Rio Grande do Sul, Brasil.
1997 - 1997 Curso de Introdução à Terapia Gênica. (Carga horária: 15h)
Hospital de Clínicas de Porto Alegre, HCPA, Rio Grande do Sul, Brasil.
1995 - 1995 10 Anos do PCR: Métodos e Aplicações em Genética. (Carga horária: 3h)
Sociedade Brasileira de Genética, SBG, São Paulo, Brasil.

4 ATUAÇÃO PROFISSIONAL

Universidade Federal do Rio Grande do Sul - UFRGS

Vínculo institucional

1998 - 2000 Vínculo: Outro, Enquadramento funcional: Estudante de Pós-graduação - Mestrado.
1994 - 1998 Vínculo: Outro, Enquadramento funcional: Estudante de Graduação - Ciências Biológicas, Regime: Dedicação exclusiva.
2000 - Atual Vínculo: Outro, Enquadramento funcional: Estudante de Pós-graduação - Doutorado.

Atividades

11/1998 - 7/2000

Linhas de pesquisa
1. Isolamento e Caracterização de Novos Microrganismos Produtores de PHAs.

3/1994 - 10/1998

Linhas de pesquisa
1. Reparação de DNA em Saccharomyces cerevisiae.

11/2000 - Atual

Linhas de pesquisa
1. Reparação de DNA em Saccharomyces cerevisiae.

Universidad de Buenos Aires - U.B.A.

Vínculo institucional

1999 - 1999 Vínculo: Outro, Enquadramento funcional: Estágio não remunerado.

Atividades

2/1999 - 2/1999

Estágios realizados
1. Técnicas de identificação de microrganismos produtores de PHAs.

5 LINHAS DE PESQUISA

1 Isolamento e Caracterização de Novos Microrganismos Produtores de PHAs.

2 Reparação de DNA em Saccharomyces cerevisiae.

3 Reparação de DNA em Saccharomyces cerevisiae.

6 ÁREAS DE ATUAÇÃO

1 Genética, Genética Molecular e de Microorganismos.
2 Genética, Genética Evolutiva.
3 Biofísica, Biofísica Molecular.
4 Microbiologia Aplicada, Microbiologia Industrial e de Fermentação.
5 Bioquímica, Biologia Molecular.
6 Bioquímica, Bioquímica dos Microorganismos.

7 IDIOMAS

Compreende: Espanhol (Bem), Inglês (Bem).
Fala: Espanhol (Razoavelmente), Inglês (Bem).
Lê: Espanhol (Bem), Inglês (Bem).
Escreve: Espanhol (Razoavelmente), Inglês (Bem).

8 PRÊMIOS E TÍTULOS

2003 Destaque XV Salão de Iniciação Científica, Universidade Federal do Rio Grande do Sul.

9 PRODUÇÃO CIENTÍFICA, TECNOLÓGICA E ARTÍSTICA/CULTURAL

9.1 PRODUÇÃO BIBLIOGRÁFICA

9.1.1 Trabalhos completos em anais de eventos

1 MATIAS, Fernanda; LISBÔA, Marcia Pagno; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. Analysis of cellular disposition of PHAs granules in the actinomycetes bacterial group using optical microscopy. In: XVIII CONGRESSO DA SOCIEDADE BRASILEIRA DE MICROSCOPIA E MICROANÁLISE, 2001, Águas de Lindóia. 2001.
Palavras-chave: PHAs; Grânulos intracelulares; Microscopia óptica; Microscopia de interferência; Actinomicetes.
Áreas do conhecimento: Citologia e Biologia Celular; Biologia e Fisiologia dos Microorganismos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Inglês; Meio de divulgação: Impresso.

9.1.2 Resumos simples em anais de eventos

1 BONATTO, Diego; BRENDEL, Martin; HENRIQUES, João Antonio Pêgas. A new group of plant specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis and three-dimensional modeling. In: DNA REPAIR AND MUTAGENESIS: FROM MOLECULAR STRUCTURES TO BIOLOGICAL CONSEQUENCES, 2004, Southampton. **DNA Repair and Mutagenesis: From Molecular Structures to Biological Consequences.** 2004. p. 76-76.
Palavras-chave: Filogenia de proteínas; Reparação de DNA; PSO2; DNA ligases; Modelagem de proteínas; Mecanismos de

reparação de DNA em plantas.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Genética Evolutiva; Genética Vegetal; Genética Evolutiva; Genética Molecular e de Microorganismos; Genética Evolutiva.
Referências adicionais: Classificação do evento: Internacional; Bermudas/Inglês; Meio de divulgação: Impresso.

2 POLETTO, Nadine Paese; SOARES, Daniele Grazziotin; BONATTO, Diego; SALVADOR, Mirian; SAFFI, Jenifer; SCHWARTSMANN, Gilberto; HENRIQUES, João Antonio Pêgas. Efeitos citotóxico, mutagênico e recombinogênico do alcalóide ecteinascidin 743 na levedura Saccharomyces cerevisiae. In: XIV ENCONTRO DE GENETICISTAS DO RIO GRANDE DO SUL, 2004, Canoas, RS. Canoas, RS: Editora da ULBRA, 2004.
Palavras-chave: ET-743; Saccharomyces cerevisiae; Reparação de DNA; Drogas anticâncer; Mutagênese; Recombinogênese.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Genética Evolutiva; Farmacologia Bioquímica e Molecular; Bioquímica dos Microorganismos.
Referências adicionais: Classificação do evento: Regional; Brasil/Português; Meio de divulgação: Impresso.

3 BONATTO, Diego; REVERS, Luis Fernando; BRENDEL, Martin; HENRIQUES, João Antonio Pêgas. Molecular Evolution of the Beta-CASP Family Suggest Functional Homology with RAG1/RAG2 Proteins: A New Model for DNA Interstrand Crosslink Repair, Nonhomologous End-Joining and V(D)J Recombination in Yeast and Vertebrates. In: 49° CONGRESSO BRASILEIRO DE GENÉTICA - SOCIEDADE BRASILEIRA DE GENÉTICA, 2003, Águas de Lindóia - SP. **Resumos do 49° Congresso Brasileiro de Genética.** 2003.
Palavras-chave: Genética evolutiva; Filogenia de proteínas; NHEJ; PSO2; Saccharomyces cerevisiae.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Genética Evolutiva.
Referências adicionais: Classificação do evento: Nacional; Brasil/Inglês; Meio de divulgação: Impresso.

4 MATIAS, Fernanda; LISBÔA, Marcia Pagno; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. Análise por MET de grânulos de PHAs em Streptomyces sp. isolado de solo do Estado do RS. In: XI CONGRESSO BRASILEIRO DE BIOLOGIA CELULAR, 2002, Porto Alegre. **Anais do XI Congresso Brasileiro de Biologia Celular.** 2002. p. 167.
Palavras-chave: Actinomicetes; Grânulos intracelulares; PHAs; Microscopia eletrônica de transmissão.
Áreas do conhecimento: Microbiologia de Solos; Biologia e Fisiologia dos Microorganismos; Microbiologia Aplicada.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

5 MATIAS, Fernanda; LISBÔA, Marcia Pagno; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. Relação de Coliformes Fecais e Totais com Actinomicetos Produtores de PHAs na Areia do Litoral Norte Gaúcho. In: 8° ENCONTRO NACIONAL DE MICROBIOLOGIA AMBIENTAL, 2002, Rio de Janeiro - RJ. **Resumos do 8° Encontro Nacional de Microbiologia Ambiental.** 2002.
Palavras-chave: Actinomicetes; Coliformes fecais; PHAs; Ecologia microbiana.
Áreas do conhecimento: Microbiologia de Solos; Microbiologia Aplicada; Biologia e Fisiologia dos Microorganismos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

6 MATIAS, Fernanda; LISBÔA, Marcia Pagno; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. A presença de actinomicetes acumuladores de PHAs como um bioindicador de excesso de contaminação ambiental com carbono orgânico. In: V CONGRESSO DE ECOLOGIA DO BRASIL, 2001, Porto Alegre. 2001.
Palavras-chave: Actinomicetes; Ecologia microbiana; Coliformes fecais; PHAs.
Áreas do conhecimento: Microbiologia de Solos; Biologia e Fisiologia dos Microorganismos; Ecologia Microbiana.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

7 MATIAS, Fernanda; LISBÔA, Marcia Pagno; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. Analysis and characterization of new PHA-producing actinomycetes strains from different types of soils in Rio Grande do Sul, Brazil. In: XXI CONGRESSO BRASILEIRO DE MICROBIOLOGIA, 2001, Foz do Iguaçu. **Anais do XXI Congresso Brasileiro de Microbiologia.** 2001.
Palavras-chave: Actinomicetes; PHAs; Ecologia microbiana; Microbiologia de solos; Microbiologia industrial.
Áreas do conhecimento: Ecologia Microbiana; Biologia e Fisiologia dos Microorganismos; Microbiologia de Solos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Inglês; Meio de divulgação: Impresso.

8 LISBÔA, Marcia Pagno; MATIAS, Fernanda; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. Isolation and characterization of a PHA-producing Lactobacillus strain from industrial yogurt. In: XXI CONGRESSO BRASILEIRO DE MICROBIOLOGIA, 2001, Foz do Iguaçu. **Anais do XXI Congresso Brasileiro de Microbiologia.** 2001.
Palavras-chave: Lactobacillus; PHAs; Ecologia microbiana; iogurte.
Áreas do conhecimento: Ecologia Microbiana; Bioquímica dos Microorganismos; Microbiologia de Alimentos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Inglês; Meio de divulgação: Impresso.

9 BONATTO,Diego; MATIAS, Fernanda; LISBÔA, Marcia Pagno; BOGDAWA, Heique Marlis; HENRIQUES, João Antonio Pêgas. Production of PHB by a Pseudomonas sp. strain isolated from soil in culture medium containing high concentrations of sugar cane molasses. In: XXI CONGRESSO BRASILEIRO DE MICROBIOLOGIA, 2001, Foz do Iguaçu. **Anais do XXI Congresso Brasileiro de Microbiologia.** 2001.

Palavras-chave: Pseudomonas sp; PHAs; sugar cane molasses; Ecologia microbiana; Metabolismo de carboidratos.
Áreas do conhecimento: Microbiologia Industrial e de Fermentação; Ecologia Microbiana; Bioquímica dos Microorganismos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Inglês; Meio de divulgação: Impresso.

10 SILVEIRA, Isabel Cristina Telles; BONATTO, Diego; HENRIQUES, João Antonio Pêgas; MONTEGGIA, Luiz Olinto. MONITORAMENTO DE BIOMASSA ANAERÓBIA PRESENTE EM REATORES DE BAIXA CARGA: TÉCNICAS CONVENCIONAIS x TÉCNICAS DA BIOLOGIA MOLECULAR. In: XXVII CONGRESSO INTERAMERICANO DE ENGENHARIA SANITÁRIA E AMBIENTAL, 2000, Porto Alegre - RS. 2000.

Palavras-chave: Biologia Molecular; Microrganismos acetoclásticos; Técnicas de Genética Microbiana; Reatores UASB.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Bioquímica dos Microorganismos; Biologia Molecular.
Setores de aplicação: Captação, tratamento e distribuição de água, limpeza urbana, esgoto e atividades conexas.
Referências adicionais: Classificação do evento: Internacional; Brasil/Português; Meio de divulgação: Impresso.

11 BONATTO, Diego; SILVEIRA, Isabel Cristina Telles; MONTEGGIA, Luiz Olinto; HENRIQUES, João Antonio Pêgas. Acompanhamento da dinâmica populacional de microrganismos acetoclásticos usando como marcador o gene para a subunidade alfa da enzima monóxido de carbono desidrogenase/acetil-CoA sintetase. In: 45 CONGRESSO NACIONAL DE GENÉTICA - SOCIEDADE BRASILEIRA DE GENÉTICA, 1999, Gramado - RS. 1999.

Palavras-chave: Ecologia microbiana; Genética Microbiana; Técnicas de Genética Microbiana; Microrganismos acetoclásticos.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biologia Molecular; Bioquímica dos Microorganismos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

12 HAMDAN, Ana Lúcia; BONATTO, Diego; HENRIQUES, João Antonio Pêgas. Evidências fenotípicas e genotípicas de um evento ancestral de duplicação gênica em Saccharomyces cerevisiae envolvendo os genes ALR1 e ALR2. In: 45 CONGRESSO NACIONAL DE GENÉTICA - SOCIEDADE BRASILEIRA DE GENÉTICA, 1999, Gramado. 1999.

Palavras-chave: Biologia Molecular; Genética evolutiva; Saccharomyces cerevisiae; Resistência a aluínio.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biologia Molecular; Genética Evolutiva.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

13 MARIS, Angélica Francesca; ASSUMPÇÃO, André; BONATTO, Diego; BOCCARDI, Fabiane; HENRIQUES, João Antonio Pêgas. Intrinsic resistance of non-fermenting yeast against hydroperoxides does not depend on mitochondrial functions. In: 45 CONGRESSO NACIONAL DE GENÉTICA - SOCIEDADE BRASILEIRA DE GENÉTICA, 1999, Gramado. 1999.

Palavras-chave: Genética Microbiana; Saccharomyces cerevisiae; Metabolismo de carboidratos.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biologia Molecular; Bioquímica dos Microorganismos.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

14 BONATTO, Diego; TORESAN, Suelen Paesi; MARIS, Angélica Francesca; HENRIQUES, João Antonio Pêgas. Induction of PSO5(RAD16), an excision repair gene of Saccharomyces cerevisiae, under hydrogen peroxide induced stress. In: SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR - XXVI REUNIÃO ANUAL, 1997, Caxambu. 1997.

Palavras-chave: Biologia Molecular; Saccharomyces cerevisiae; NER; PSO5.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biologia Molecular.
Referências adicionais: Classificação do evento: Nacional; Brasil/Português; Meio de divulgação: Impresso.

9.1.3 Artigos completos publicados em periódicos

1 BONATTO, Diego; BRENDEL, Martin; HENRIQUES, João Antonio Pêgas. A new group of plant specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis, and three-dimensional modeling. **Functional Plant Biology,** v. 32, n. 2, p. 161-174, 2005.

Palavras-chave: DNA ligases; Filogenia de proteínas; Hydrophobic cluster analysis; Mecanismos de reparação de DNA em plantas; PSO2; Genética evolutiva.
Áreas do conhecimento: Genética Vegetal; Genética Evolutiva; Microbiologia Aplicada; Genética Evolutiva; Química de Macromoléculas.
Referências adicionais: Austrália/Inglês; Meio de divulgação: Impresso; ISSN/ISBN: 14454408.
No prelo

2 SOARES, Daniele Grazziotin; POLETTO,Nadine Paese; BONATTO,Diego; SALVADOR, Mirian; SCHWARTSMANN,Gilberto; HENRIQUES,João Antonio Pêgas. Low cytotoxicity of Ecteinascidin 743 in yeast lacking the major endonucleolytic enzymes of base and nucleotide excision repair pathways. **BiochemicalPharmacology,** 2005.

Palavras-chave:ET-743; BER; NER; Saccharomyces cerevisiae; Apn1p; Rad1p.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Farmacologia Bioquímica e Molecular; Mutagenese.
Referências adicionais: Estados Unidos/Inglês; Meio de divulgação: Impresso; ISSN/ISBN: 00062952.
Aceito para a publicação

3 KERN, Ana Lúcia; BONATTO,Diego; DIAS, Johnny Ferraz; YONEAMA, Maria-lucia; BRENDEL, Martin; HENRIQUES,João Antonio Pêgas. The Function of Alr1p of Saccharomyces cerevisiae in Cadmium Detoxification: Insights from Phylogenetic Studies and Particle-Induced X-ray Emission. **Biometals,** v. 18, n. 1, p. 31-41, 2005.

Palavras-chave:Saccharomyces cerevisiae; Filogenia de proteínas; Proteínas transmembrana; PIXE; Hydrophobic cluster analysis; metal uptake.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Metabolismo e Bioenergética; Genética Evolutiva;Física Atômica e Molecular.
Referências adicionais: Holanda/Inglês;Meio de divulgação: Impresso; ISSN/ISBN: 09660844.
Aceito para publicação

4 KERN, Ana Lúcia; BONATTO,Diego; DIAS, Johnny Ferraz; YONEAMA, Maria-lucia; BRENDEL, Martin; HENRIQUES,João Antonio Pêgas. The importance of yeast Alr proteins in cadmium detoxification as indicated by Particle-Induced X-ray Emission and Phenotypic Analyses. **X-Ray Spectrometry,** 2005.

Palavras-chave: Proteínas transmembrana; PIXE; Saccharomyces cerevisiae; Tolerância a metais; Destoxificação de cádmio; Resistência a alumínio.
Áreas do conhecimento: Métodos Experimentais e Instrumentação para Partículas Elementares e Física Nuclear;Genética Evolutiva;Espectroscopia; Genética Molecular e de Microorganismos; Genética Evolutiva.
Referências adicionais: Inglaterra/Inglês;Meio de divulgação: Impresso; ISSN/ISBN: 00498246.
Aceito para publicação

5 BONATTO,Diego; MATIAS, Fernanda; LISBÔA, Marcia Pagno; BOGDAWA, Heique Marlis; HENRIQUES, João Antonio Pêgas. Production of short side chain-poly[hydroxyalkanoate] by a newly isolated Ralstonia pickettii strain. **World Journal of Microbiology & Biotechnology,** Holanda, v. 20, n. 4, p. 395-403, 2004.

Palavras-chave:PHAs; Microbiologia de solos; Ralstonia pickettii; Sugarcane molasses; Bioprospecção microbiana; Sacarose.
Áreas do conhecimento: Microbiologia Aplicada; Microbiologia de Solos; Genética Molecular e de Microorganismos.
Referências adicionais: Holanda/Inglês;Meio de divulgação: Impresso; ISSN/ISBN: 09593993.

6 BONATTO,Diego; REVERS, Luis Fernando; BRENDEL, Martin; HENRIQUES,João Antonio Pêgas. The eukaryotic Pso2/Snm1/Artemis proteins and their function as genomic and cellular caretakers. **Brazilian Journal of Medical and Biological Research,** 2004.

Palavras-chave:Filogenia de proteínas; NHEJ; DSB; PSO2; Artemis; Saccharomyces cerevisiae.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Genética Evolutiva; Biologia Molecular; GenéticaEvolutiva; Bioquímica dos Microorganismos.
Referências adicionais: Brasil/Inglês; Meio de divulgação: Impresso; ISSN/ISBN: 0100879X.
No prelo

7 BRENDEL, Martin; BONATTO,Diego; STRAUSS, Martin; REVERS, Luis Fernando; PUNGARTNIK,Cristina; SAFFI, Jenifer; HENRIQUES,João Antonio Pêgas. Role of PSO genes in repair of DNA damage of Saccharomyces cerevisiae. **Mutation Research Reviews In Mutation Research,** v. 544, n. 2-3, p. 179-193, 2003.

Palavras-chave:Saccharomyces cerevisiae; oxidative stress; NER; PSO2; PSO3; Interstrand cross-link.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biofísica Molecular; Genética Evolutiva.
Referências adicionais: Estados Unidos/Inglês; Meio de divulgação: Impresso; ISSN/ISBN: 13835742.

8 REVERS, Luis Fernando; CARDONE, Jaqueline Moraes; BONATTO,Diego; SAFFI, Jenifer; GREY, Martin; FELDMANN, Heidy; BRENDEL, Martin; HENRIQUES,João Antonio Pêgas. Thermoconditional modulation of the pleiotropic sensitivity phenotype by the Saccharomyces cerevisiae PRP19 mutant allele pso4-1. **Nucleic Acid Research,** Inglaterra, v. 30, n. 22, p. 4993-5003, 2002.

Palavras-chave:Saccharomyces cerevisiae; NER; PSO4; Splicing; mutantes termocondicionais.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Bioquímica dos Microorganismos; Biologia e Fisiologia dos Microorganismos.
Referências adicionais: Inglaterra/Inglês;Meio de divulgação: Impresso; ISSN/ISBN: 0305048.

9 MARIS, Angel F; ASSUMPÇÃO, Andre L K; BONATTO, Diego; BRENDEL, Martin; HENRIQUES, João Antonio Pêgas. Diauxic shift-induced stress resistance against hydro-peroxides in Saccharomyces cerevisiae is not an adaptative stress response and does not depend on functional mitochondria. **Current Genetics,** Estados Unidos, v. 39, n. 3, p. 137-149, 2001.

Palavras-chave: Saccharomyces cerevisiae; oxidative stress; mitochondria; diauxic shift; hydro-peroxides.
Áreas do conhecimento: Bioquímica dos Microorganismos; Genética Molecular e de Microorganismos; Biologia Molecular.
Referências adicionais: Estados Unidos/Inglês; Meio de divulgação: Impresso; ISSN/ISBN: 01728083.

9.1.4 Capítulos de livros publicados

1 BONATTO, Diego; ROSA, Renato Moreira; SAFFI, Jenifer; HENRIQUES, João Antonio Pêgas. Estresse oxidativo e envelhecimento. In: HENRIQUES, João Antonio Pêgas; SALVADOR, Mirian. (Org.). **Radicais Livres e Defesas Antioxidantes no Estresse Oxidativo.** Canoas - RS, 2004.

Palavras-chave: Estresse oxidativo; Envelhecimento; EROs; Saccharomyces cerevisiae; Mamíferos; Pássaros.
Áreas do conhecimento: Fisiologia de Órgãos e Sistemas; Microbiologia de Solos; Metabolismo e Bioenergética; Biofísica de Processos e Sistemas.
Referências adicionais: Brasil/Português; Meio de divulgação: Impresso.

2 BONATTO, Diego; MATIAS, Fernanda; LISBÔA, Marcia Pagno; BOGDAWA, Heique Marlis; HENRIQUES, João Antonio Pêgas. Production of PHB by a Pseudomonas sp. strain grown in sugarcane molasses. In: JONAS, Rainer; PANDEY, Ashok; THARUN, Günter. (Org.). **Biotechnological Advances and Applications in Bioconversion of Renewable Raw Materials.** Braunschweig, 2004, p. 97-101.

Palavras-chave: PHAs; Bioprospecção microbiana; Microbiologia industrial; Sugarcane molasses; Sacarose; Microbiologia de solos.
Áreas do conhecimento: Genética Evolutiva; Biologia e Fisiologia dos Microorganismos; Microbiologia Aplicada; Genética Evolutiva.
Referências adicionais: Alemanha/Inglês; Meio de divulgação: Impresso; ISBN: 3925268250.

9.1.5 Textos em jornais de notícias

1 BONATTO, Diego; BOGDAWA, Heique Marlis. Doutorandos da UFRGS pesquisam produção de plástico biodegradável. **Jornal da Universidade,** Porto Alegre - RS, p. 13.27 ago. 2002.

Palavras-chave: Bioplástico; Microbiologia industrial; Transgenia; PHAs; sugar cane molasses.
Áreas do conhecimento: Microbiologia Industrial e de Fermentação; Bioquímica dos Microorganismos; Genética Vegetal.
Referências adicionais: Brasil/Português; Meio de divulgação: Impresso; Data de publicação: 27/08/2002.

2 BONATTO, Diego; BOGDAWA, Heique Marlis. O Plástico que não dura. **Zero Hora,** Porto Alegre - RS, p. 7-7, 12 ago. 2002.

Palavras-chave: Bioplástico; Ecologia microbiana; Microbiologia industrial; PHAs.
Áreas do conhecimento: Microbiologia Industrial e de Fermentação; Bioquímica dos Microorganismos.
Referências adicionais: Brasil/Português; Meio de divulgação: Impresso; Data de publicação: 12/08/2002.

3 BONATTO, Diego; BOGDAWA, Heique Marlis; MATIAS, Fernanda; LISBÔA, Marcia Pagno; PASQUALI, Giancarlo; HENRIQUES, João Antonio Pêgas. Pesquisa desenvolve bioplástico no Estado. **Jornal do Comércio,** Porto Alegre, p. 10.13 ago. 2002.

Palavras-chave: Biotecnologia; Transgenia; Técnicas de Genética Microbiana; PHAs; Bioplástico.
Áreas do conhecimento: Microbiologia Aplicada.
Setores de aplicação: Fabricação de produtos de borracha e plástico.
Referências adicionais: Brasil/Português; Meio de divulgação: Impresso; Data de publicação: 13/08/2002.

4 BONATTO, Diego; BOGDAWA, Heique Marlis; PASQUALI, Giancarlo; HENRIQUES, João Antonio Pêgas. UFRGS faz plástico à base de melaço. **Correio do Povo,** Porto Alegre - RS, p. 14-14, 14 ago. 2002.

Palavras-chave: Bioplástico; Biotecnologia; Microbiologia industrial; sugar cane molasses; Transgenia; Biotecnologia de plantas.
Áreas do conhecimento: Microbiologia Aplicada; Genética Molecular e de Microorganismos; Genética Vegetal.
Referências adicionais: Brasil/Português; Meio de divulgação: Impresso; Data de publicação: 14/08/2002.

9.2 PRODUÇÃO TÉCNICA

9.2.1 Demais tipos de produção técnica

1 BONATTO, Diego. **Análise de Bactérias Produtoras de Poliéster Biológico.** 2002. (Apresentação de trabalho/Comunicação).

Palavras-chave: Bioplástico; Ecologia microbiana; PHAs; Microbiologia industrial.
Áreas do conhecimento: Microbiologia Aplicada; Biologia e Fisiologia dos Microorganismos.
Referências adicionais: Brasil/Português; Meio de divulgação: Outro; Local: Instituto de Biociências - UFRGS; Cidade: Porto Alegre -

RS; Evento: Seminários II do Curso de Ciências Biológicas; Inst. promotora/financiadora:Instituto de Biociências - UFRGS.

2 BONATTO,Diego. **Embalagens: tendências e futuros.** 2000. (Apresentação de trabalho/Seminário).

Palavras-chave:PHAs; Técnicas de Genética Microbiana; Biotecnologia.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biologia Molecular; Bioquímica dos Microorganismos.
Setores de aplicação: Produtos e processos biotecnológicos; Produtos e serviços voltados para a defesa e proteção do meio ambiente, incluindo o desenvolvimento sustentado; Desenvolvimento de novos materiais.
Referências adicionais: Brasil/Português; Local: FIERGS; Cidade: Porto Alelgre; Evento: Embalagens: tendências e futuros; Inst. promotora/financiadora:SIA/RS; SINPLAST/RS e ARDEF.

3 BONATTO,Diego; BOGDAWA, Heique Marlis; HENRIQUES,João Antonio Pêgas. **Produção de PHAs: uma nova perspectiva industrial.** 1999. (Apresentação de trabalho/Congresso).

Palavras-chave:Biotecnologia; PHAs.
Áreas do conhecimento: Bacterologia; Microbiologia Industrial e de Fermentação; Bioquímica dos Microorganismos.
Setores de aplicação: Produtos e serviços voltados para a defesa e proteção do meio ambiente, incluindo o desenvolvimento sustentado; Fabricação de produtos químicos; Desenvolvimento de novos materiais.
Referências adicionais: Brasil/Português; Local: Belo Horizonte; Cidade: Minas Gerais; Evento: Congresso Internacional sobre meio ambiente e oportunidades de negócios; Inst. promotora/financiadora:FIEMG.

4 BONATTO,Diego. **Tecnologia de DNA Recombinante.** 1999. (Apresentação de trabalho/Outra).

Palavras-chave:Biotecnologia; Genética Microbiana; Biologia Molecular; Transgenia.
Áreas do conhecimento: Genética Molecular e de Microorganismos; Biologia Molecular; Microbiologia Aplicada.
Setores de aplicação: Produtos e processos biotecnológicos.
Referências adicionais: Brasil/Português; Local: Faculdade de Química; Cidade: Porto Alegre; Evento: Palestra para a Faaculdade de Química; Inst. promotora/financiadora:PUC - RS.

9.3 DEMAIS TRABALHOS

1 BONATTO,Diego; HENRIQUES,João Antonio Pêgas. **Disciplina de Biofísica III - Curso de Ciências Biológicas.** 2000. (Atividades Didáticas).

Palavras-chave:Biofísica das radiações; Reparação de DNA; Saccharomyces cerevisiae; Fotobiologia e Radiobiologia.
Áreas do conhecimento: Radiologia e Fotobiologia; Biofísica Celular; Biofísica Molecular.
Referências adicionais: Brasil/Português; Meio de divulgação: Vários; Finalidade: Atividades Didáticas; Local do evento: Insituto de Biociências - Departamento de Biofísica - UFRGS.
Disciplina ministrada no primeiro semestre do ano 2000, totalizando 15 horas-aula. Atividade curricular obrigatória do Programa de Pós-graduação em Biologia Celular e Molecular (PPGBCM) da UFRGS

2 BONATTO,Diego; BRENDEL, Martin. **Microbial Genetics.** 1997. (Monitoria).

Palavras-chave:Saccharomyces cerevisiae; Bacillus; Genética Microbiana; Técnicas de Genética Microbiana.
Áreas do conhecimento: Bacterologia; Micologia; Virologia.
Referências adicionais: Brasil/Inglês; Meio de divulgação: Vários; Finalidade: Curso ministrado pelo Prof.Martin Brendel(Frankfurt, Alemanha) sobre Genética Microbiana; Local do evento: Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul.
Monitoria exercida durante todo o curso "Microbial Genetics", totalizando cerca de 160 horas-aula.

10 DADOS COMPLEMENTARES

10.1 PARTICIPAÇÃO EM BANCAS EXAMINADORAS

10.1.1 Trabalhos de Conclusão de Curso de Graduação

1 BONATTO,Diego; CORDELLINI, Valeriano Antônio; HENRIQUES,João Antonio Pêgas. Participação em banca de Fernanda Matias. **Isolamento e caracterização de novas linhagens de actinomicetos produtores de PHAs de solos do Rio Grande do Sul.** 2003. Trabalho de Conclusão de Curso (Graduação em Ciências Biológicas) - Universidade Federal do Rio Grande do Sul.

Palavras-chave: PHAs; Bioplástico; Actinomicetes; Microbiologia de solos; Microscopia eletrônica de transmissão; Microscopia óptica.
Áreas do conhecimento: Microbiologia de Solos; Biologia e Fisiologia dos Microorganismos; Citologia e Biologia Celular.
Referências adicionais: Brasil/Português.

2 BONATTO,Diego; SAFFI, Jenifer; RAMOS, Ana Lígia Lia de Paula. Participação em banca de Renato Moreira Rosa. **Reparo de danos oxidativos no mutante pso3-1 de Saccharomyces cerevisiae.** 2002. Trabalho de Conclusão de Curso (Graduação em Faculdade de Farmácia) - Universidade Federal do Rio Grande do Sul.

Palavras-chave:Saccharomyces cerevisiae; oxidative stress; PSO3.
Áreas do conhecimento: Bioquímica dos Microorganismos; Biologia e Fisiologia dos Microorganismos.

Referências adicionais: Brasil/Português.

10.2 PARTICIPAÇÃO EM EVENTOS

1 **49° Congresso Nacional de Genética.** 2003. (Participação em eventos/Congresso).
Áreas do conhecimento: Genética.
Referências adicionais: Brasil; Meio de divulgação: Digital; Nome do evento: 49° Congresso Nacional de Genética; Nome da instituição promotora: Sociedade Brasileira de Genética; Cidade: Águas de Lindóia - SP.

2 **XXI Congresso Brasileiro de Microbiologia.** 2001. (Participação em eventos/Congresso).
Palavras-chave: Microbiologia industrial; Microbiologia de solos; Genética Microbiana; Técnicas de Genética Microbiana.
Áreas do conhecimento: Microbiologia; Microbiologia Aplicada; Genética Molecular e de Microorganismos.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: XXI Congresso Brasileiro de Microbiologia; Nome da instituição promotora: Sociedade Brasileira de Microbiologia; Local: Rafain Palace Hotel; Cidade: Foz do Iguaçu - PR.

3 **More Quality of Life by Means of Biotechnology.** 2000. (Participação em eventos/Simpósio).
Áreas do conhecimento: Genética; Bioquímica; Microbiologia.
Referências adicionais: Alemanha; Meio de divulgação: Impresso; Nome do evento: International Symposium on the Bioconversion of Renewable Raw Materials; Nome da instituição promotora: Gesellschaft für Biotechnologische Forschung (GBF); Local: Gesellschaft für Biotechnologische Forschung (GBF); Cidade: Braunschweig - Alemanha.

4 **45° Congresso Nacional de Genética.** 1999. (Participação em eventos/Congresso).
Palavras-chave: Genética Microbiana; Saccharomyces cerevisiae; Técnicas de Genética Microbiana.
Áreas do conhecimento: Genética.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: 45° Congresso Nacional de Genética; Nome da instituição promotora: Sociedade Brasileira de Genética; Cidade: Gramado - RS.

5 **Congresso Internacional: Meio Ambiente - Oportunidades de Negócios.** 1999. (Participação em eventos/Congresso).
Áreas do conhecimento: Microbiologia; Administração; Economia.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: Congresso Internacional: Meio Ambiente - Oportunidades de Negócios; Nome da instituição promotora: Confederação Nacional das Indústrias e Federação das Indústrias do Estado de Minas Gerais; Local: Minas Trade Center; Cidade: Belo Horizonte - MG.

6 **44° Congresso Nacional de Genética.** 1998. (Participação em eventos/Congresso).
Áreas do conhecimento: Genética.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: 44° Congresso Nacional de Genética; Nome da instituição promotora: Sociedade Brasileira de Genética; Local: Hotel Monte Real Resort; Cidade: Águas de Lindóia - SP.

7 **IV Encontro Gaúcho de Imunologia.** 1998. (Participação em eventos/Encontro).
Áreas do conhecimento: Genética; Imunologia.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: IV Encontro Gaúcho de Imunologia; Nome da instituição promotora: Sociedade Brasileira de Imunologia - Regional Sul; Local: Pontifícia Universidade Católica do Rio Grande do Sul; Cidade: Porto Alegre - RS.

8 **XXVI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular.** 1997. (Participação em eventos/Congresso).
Áreas do conhecimento: Bioquímica.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: XXVI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular; Nome da instituição promotora: Sociedade Brasileira de Bioquímica e Biologia Molecular; Cidade: Caxambu - MG.

9 **41° Congresso Nacional de Genética.** 1995. (Participação em eventos/Congresso).
Áreas do conhecimento: Genética.
Referências adicionais: Brasil; Meio de divulgação: Impresso; Nome do evento: 41° Congresso Nacional de Genética; Nome da instituição promotora: Sociedade Brasileira de Genética; Local: Caxambu; Cidade: Caxambu - MG.

11 INDICADORES DE PRODUÇÃO

Produção bibliográfica

Artigos publicados em periódicos - 9
Completos - 9

Trabalhos em eventos - 15

Completos - 1
Resumos - 14

Livros e capítulos - 2
Capítulos de livros publicados - 2

Textos em jornais ou revistas (magazines) - 4
Jornais de notícias - 4

Produção técnica

Demais tipos de produção técnica - 4

Demais trabalhos

Dados complementares

Participação em bancas examinadoras - 2

Participação em eventos - 9