UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

DILAINE ROSE SILVA SCHNEIDER

"EXPRESSÃO, PURIFICAÇÃO E CARACTERIZAÇÃO PARCIAL DE PROTEÍNAS RELACIONADAS À PATOGENICIDADE DE *Magnaporthe grisea*"

Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Genética e Biologia Molecular na área de Genética de Microorganismos

Orientadora: Profa. Dra. Anete Pereira de Souza

Campinas, 2010

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

Sch57e

Schneider, Dilaine Rose Silva
Expressão, purificação e caracterização parcial de proteínas relacionadas à patogenicidade de *Magnaporthe grisea* / Dilaine Rose Silva Schneider. – Campinas, SP: [s.n.], 2010.

Orientadora: Anete Pereira de Souza.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. *Magnaporthe grisea*. 2. Proteínas intrinsicamente desordenadas. 3. Proteínas efetoras. 4. Patogenicidade. 5. Xilanases. I. Souza, Anete Pereira de. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

Título em inglês: Expression, purification and partial characterization of proteins related to the pathogenicity of *Magnaporthe grisea*.
Palavras-chave em inglês: *Magnaporthe grisea*; Intrinsically disordered proteins; Effector proteins; Pathogenicity; Xylanases.
Área de concentração: Genética de Microorganismos.
Titulação: Doutora em Genética e Biologia Molecular.
Banca examinadora: Anete Pereira de Souza, Michel G. A . Vincentz, Olívia Márcia Nagy Arantes, Flávio Henrique da Silva, Aline Aparecida Pizzirani-Kleiner.
Data da defesa: 29/07/2010.
Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 29 de Julho de 2010.

BANCA EXAMINADORA

Profa. Dra. **Anete Pereira de Souza** (Orientadora)

Prof. Dr. **Michel G. A. Vincentz**

Profa. Dra. **Olivia Márcia Nagy Arantes**

Prof. Dr. **Flávio Henrique da Silva**

Profa. Dra. **Aline Aparecida Pizzirani-Kleiner**

Prof. Dr. **Márcio de Castro Silva Filho**

Prof. Dr. **Nilson Ivo Zanchin**

Profa. Dra. **Alessandra de Souza**

Dedicatória

Para meus pais, Dimas e Eclair,

que me presentearam com o seu imenso e acalentador amor altruísta.

Agradecimentos

Tudo o que sou, de algum modo, procede de alguém.

Assim, agradeço:

Ao único Deus verdadeiro, pela dádiva preciosa da vida possibilitando-me perceber e analisar a sua grandiosa criação;

Aos meus pais, meus irmãos e minhas irmãs que sempre me asseguraram de sua amizade incondicional - em especial ao Dilson, pelo encorajamento e pela companhia;

Aos meus amados filhos (Elisa, Tomás, Hugo e Álvaro), pelo amparo de seu terno amor e porque, apesar de jovens, souberam aguardar pacientemente o "fim da tese da mãe";

Ao Homero, marido e companheiro, pelo apoio e pela compreensão;

À Anete, que, além de excelente orientadora neste trabalho e referência profissional de liderança, não se esquivou de me apoiar em assuntos pessoais. Você será sempre uma grata lembrança para mim, Anete!

Aos meus colegas e amigos do LAGM/CBMEG/UNICAMP: Alexandre, Carol, Clelton, Eliane, Karen, Lilian, Letícia, Luciana, Marianna, Roberta e Susy pelo conhecimento compartilhado, pelo auxílio nos experimentos e pelo ambiente de trabalho que cada um proporcionou dentro de sua capacidade; em especial, ao Adriano, ao Antonio, à Camília, à Cleide, ao Marcelo e ao Zé Sérgio que também me permitiram dividir, com eles, anseios e dificuldades pessoais;

Aos meus colegas e amigos do Barracão da Genética/UNICAMP pelo auxílio profissional e pelos momentos agradáveis (como nas festinhas e nos churrascos);

Ao Juverlande e à Rosana: meus sempre grandes amigos;

À Zilda, com seu sorriso fácil e sua boa vontade, por ter providenciado prontamente o material de laboratório adequadamente limpo e esterilizado e pelo seu café sempre revigorante;

À Tânia, à Sandra e à Lourdes pela eficiência no cumprimento de suas funções;

Aos membros da pré-banca e da banca de defesa de tese por sua pronta disposição em utilizar seus conhecimentos no julgamento desta tese;

À CAPES e à FAPESP pelo apoio financeiro e pelas avaliações dos relatórios;

À UNICAMP, por oferecer a estrutura física e a estrutura administrativa adequadas para o desenvolvimento deste trabalho.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

PREFÁCIO

Os resultados aqui apresentados foram obtidos no estudo de nove diferentes genes (e de alguns de seus domínios funcionais) relacionados à patogenicidade do ascomiceto fitopatogênico *Magnaporthe grisea*. Estes resultados serão apresentados na forma de dois artigos científicos (um deles publicado e o outro submetido à publicação) e de um capítulo de resultados complementares.

No artigo intitulado "Overexpression and purification of PWL2D, a mutant of the effector protein PWL2 from *Magnaporthe grisea*" encontra-se o componente principal do trabalho desenvolvido nesta tese. Nele descrevem-se os procedimentos que culminaram com a proposta de um modelo para explicar a susceptibilidade e a resistência do hospedeiro à proteína efetora de avirulência PWL2D. Tal manuscrito foi publicado no periódico *Protein Expression and Purification*.

Em outro artigo, intitulado "Cloning, overexpression, purification and partial characterization of the XYL5 xylanase from *Magnaporthe grisea* in *Escherichia coli*", descreve-se a obtenção de uma nova xilanase e também de seu domínio catalítico. No capítulo Resultados Complementares são apresentados resultados adicionais no estudo das proteínas PWL2D, XYL5 e de outros genes e proteínas analisados durante a condução deste trabalho.

Em anexo encontram-se dois artigos publicados relacionados a trabalhos desenvolvidos em assuntos não detalhados nessa tese. Ambos os artigos envolveram o grupo de pesquisa onde esta tese foi desenvolvida. O primeiro artigo, publicado em 2009, intitula-se "Molecular and cytogenetic characterization of na AT-rich satellite DNA family in *Urvillea chacoensis* Hunz. (Paullinieae, Sapindaceae)", tendo sido publicado no periódico Genetica, 136: 171-177; neste artigo sequências de satDNA da família *U. chacoensis* foram caracterizadas, mapeadas e comparadas com 45S rDNA para a obtenção de um padrão de heterocromatina. O segundo artigo, também publicado em 2009, intitula-se "Functional and small-angle X-ray scattering studies of a new stationary phase survival protein E (SurE) from *Xylella fastidiosa* evidence of allosteric behavior", tendo sido publicado no periódico FEBS Journal 276(22):6751-62; neste artigo foi feita a caracterização funcional de uma nucleotidase da bactéria fitopatogênica *Xylella fastidiosa.*

Na apresentação dos resultados são feitas as discussões pertinentes, e, ao final, apresentam-se as conclusões. Finalmente, apresentam-se as perspectivas de pesquisas com as proteínas estudadas neste trabalho.

RESUMO

A brusone do arroz (*rice blast disease*) causada pelo ascomiceto fitopatógeno *Magnaporthe grisea* continua a ter um enorme impacto nas culturas de arroz (*Oryza sativa*) no Brasil e no mundo. PWL2, uma proteína efetora, é um conhecido produto de um gene AVR (avirulência). O gene *PWL2* impede que o fungo infecte *weeping lovegrass* (*Eragrostis curvula*). Neste trabalho nós identificamos em uma linhagem de *M. grisea* um gene que produz uma proteína diferente de PWL2, denominada PWL2D. A seqüência do gene *PWL2D* tem duas bases que diferem do gene *PWL2*, as quais produzem alterações nos resíduos de 90 e 142 da proteína. A alteração do resíduo 90 (de D90 para N90) é fundamental para a avirulência. Neste trabalho foram efetuadas a clonagem do gene PWL2D no vetor pET32-Xa/LIC, a expressão em *Escherichia coli* e a avaliação da estrutura de PWL2D por técnicas espectroscópicas. A proteína PWL2D fusionada à cauda TRX é propensa a agregação, e sua solubilidade é melhorada quando super-expressa sem o seu peptídeo-sinal original. Os resultados estruturais obtidos indicam que a proteína PWL2D possivelmente é intrinsecamente desordenada. Foi elaborado um modelo para a resistência/susceptibilidade do hospedeiro à *M. grisea* considerando a atuação de PWL2D como uma proteína intrinsecamente desordenada. Os resultados obtidos deverão facilitar a análise estrutural de PWL2D e podem contribuir para a compreensão da função do gene nas interações fungo / planta.

Oito diferentes genes de *M. grisea*, além de *PWL2D,* foram também estudados neste trabalho. Dentre estes, destacam-se o gene que produz a xilanase XYL5 e seu domínio catalítico, o gene que codifica a chaperona ABC1 e seus dois domínios funcionais, e o gene que codifica a trealase PTH9, sendo todos estes relacionados à patogenicidade do fungo *M. grisea*. A xilanase XYL5 (EC 3.2.1.8) e seu domínio catalítico conservado (XYL5/DOM) foram fusionados à Maltose Binding Protein (MBP) ou à tiorredoxina

(TRX) e expressas em *E. coli*. A produção de proteína solúvel e ativa foi influenciada pelo tipo de fusão. Os extratos solúveis contendo as proteínas de fusão MBP-XYL5 e MBP-XYL5/DOM apresentaram atividade xilanolítica em relação ao controle. Entretanto, durante o processo de purificação, a atividade foi perdida. Assim, obteve-se pela primeira vez o gene de patogenicidade *XYL5* de *M. grisea* expresso com sucesso em *E. coli* e sua atividade enzimática xilanolítica foi demonstrada. Não foi possível expressar a chaperona ABC1 na forma solúvel nos sistemas de expressão utilizados, e a sequência gênica referente à trealase PTH9 – por mostrar a presença de introns após o seqüenciamento do gene amplificado na linhagem de *M. grisea* em estudo, mostrou-se inadequado para a sua expressão protéica no sistema de expressão procariótico utilizado durante a realização deste trabalho.

ABSTRACT

The rice blast disease caused by the ascomycete phytopathogen *Magnaporthe grisea* continues to have a huge impact on crops of rice (Oryza sativa) in Brazil and worldwide. PWL2, an effector protein, is a product of an AVR (avirulence) gene . The gene PWL2 prevents fungus from infecting weeping lovegrass (Eragrostis curvula). In this work we identified in a strain of *M. grisea* a gene that produces a protein different from PWL2, called PWL2D. The gene sequence PWL2D has two bases that differ from PWL2 gene, which produce changes in residues 90 and 142 of the protein. The change of residue 90 (from D90 to N90) is critical to avirulence. In this work it was realized the cloning of the gene in the vector PWL2D pET32-Xa/LIC, the expression in *Escherichia coli* and the assessment of PWL2D structure by spectroscopic techniques. The protein fused to the tag PWL2D TRX is prone to aggregation, and its solubility is improved when overexpressed without its original signal peptide. The structural results obtained indicate that possibly the protein PWL2D is intrinsically disordered. A model for the resistance/susceptibility of the host to *M. grisea* was developed considering the performance of PWL2D as an intrinsically disordered protein. The results should facilitate structural analysis of PWL2D and may contribute to the understanding of gene function in the interactions fungus/plant.

Eight different genes of *M. grisea*, besides *PWL2D*, were also studied in this work. Among these, stands out the gene that produces xylanase XYL5 and its catalytic domain, the gene that codify the chaperone ABC1 and its two functional domains, and the gene that codify the trehalase PTH9, all them being related to the pathogenicity of the fungus *M. grisea*. The xylanase XYL5 (EC 3.2.1.8) and its retained catalytic domain (XYL5/DOM) were fused to the solubilizing proteins (MBP) or thioredoxin (TRX) and expressed into *E. coli*. The production of soluble and active protein was influenced by the type of fusion. The soluble extracts containing the fusion proteins MBP-XYL5 and MBP-XYL5/DOM showed xylanolytic activity compared to the control. However,

during the purification process, the activity was lost. Thus, we obtained for the first time the gene pathogenicity XYL5 *M. grisea* expressed successfully in *E. coli* and its enzymatic xylanolytic activity was demonstrated. It was not possible to express the chaperone ABC1 in soluble form in the expression systems used, and the gene sequence related to trehalase PTH9 - by showing the presence of introns after the sequencing of the gene amplified in the strain of *M. grisea* under study, rendered inadequate for its protein expression in the prokaryotic system used during the realization of this work.

1. INTRODUÇÃO

Muitas doenças importantes de humanos e outros animais são causadas por bactérias e vírus. Em contraste, os vegetais são mais comumente afetados por fungos e vírus, o que se deve, em parte, às diferenças entre plantas e animais como habitats para crescimento microbiano (Lucas, 1998a). A vasta maioria dos fungos são estritamente saprófitas. De milhares de espécies de fungos conhecidas menos que 10% são capazes de colonizar plantas, e uma fração ainda menor é capaz de causar doença; isto talvez se dê em virtude de a doença ser uma exceção, não a regra. No entanto, entre os agentes causais de doenças em plantas, os fungos fitopatógenos têm um papel predominante tanto por causar devastações epidêmicas como pela perda significativa de plantações anuais. Esta evidência tem feito destes microrganismos um sério fator econômico, atraindo a atenção de agricultores, cultivadores e cientistas (Knogge, 1996; Laugé; De Wit, 1998; Kershaw; Talbot, 2009).

Devido ao valor econômico e social de diferentes plantas, comestíveis ou não, o conhecimento no nível genético e no nível molecular da patogenicidade e da virulência de fungos patogênicos mostra-se de suma importância. No Brasil e em diferentes países do mundo, doenças significativas acometem inúmeras variedades de plantas. A severidade destas doenças deve-se ao fato de os fungos fitopatogênicos poderem afetar diferentes partes dos vegetais em qualquer época de desenvolvimento. A descoberta de mecanismos de patogenicidade, bem como de genes de patogenicidade e de avirulência, conduzem ao entendimento mais sofisticado da interação entre

planta-fungo (Kahmann; Basse, 2001). Diferentes e diversas técnicas têm sido desenvolvidas para analisar este processo no nível molecular (Skiner *et al.*, 2001; Sweigard; Ebbole, 2001).

O fungo ascomiceto *Magnaporthe grisea* causa a brusone, uma das mais devastadoras doenças em arroz cultivado em todo o mundo, utilizando mecanismos sofisticados, como a formação de estruturas de infecção (Sesma; Osbourn, 2004). Estudos da patogênese neste microrganismo identificaram diversos genes neste ascomiceto envolvidos no desenvolvimento da doença (Tucker *et al*., 2004; Fourrey; Talbot, 2005). *M. grisea* tem despontado como prospectivo modelo para o estudo de interações patógeno-hospedeiro. Uma potente combinação de genética clássica, biologia molecular, citologia e biologia celular permite, agora, a análise de elementos críticos destas interações. Tais elementos podem ser resumidos como se segue: mecanismos de patogenicidade, especificidade para espécies-hopedeiras, especificidade para cultivar-hospedeiro, mecanismos de variação genética espontânea.

Dos diferentes grupos de proteínas (enzimas, proteínas redox, carreadoras e de armazenamento, hormônios, e anticorpos), muitos produtos gênicos envolvidos na patogenicidade de fungos são enzimas. A atividade biológica das proteínas depende da estrutura tridimensional, obtida através do processo de enovelamento protéico, sendo que as informações necessárias para o enovelamento correto estão na própria seqüência primária de aminoácidos (McPherson, 2004). Portanto, a partir do

conhecimento sobre a estrutura de uma proteína pode ser possível inferir algo sobre sua função (Barth, 2004). A Biologia Molecular tem dado suporte para a clonagem de genes, a seleção de transformantes, a ruptura de genes, e a caracterização de sequências de DNA repetitivo, caracterizando assim linhagens de *M. grisea* patogênicas de arroz (Valent; Chumley, 1991; Kadotani *et al.,* 2003). A elucidação do genoma de *M. grisea* possibilitou o estudo do envolvimento de diferentes genes na interação planta-hospedeiro (Dean, 2005; Kershaw; Talbot, 2009). *M. grisea* apresenta um genoma que compreende 40Mb organizadas em 7 cromossomos e seu genoma foi seqüenciado por um consórcio entre diferentes organizações (The *Magnaporthe grisea* Genome Project, http://www.ncbi.nlm.nih.gov).

No presente trabalho, diferentes genes de patogenicidade e genes de avirulência de linhagens de *M. grisea* patogênicas de arroz foram escolhidos, por meio de critérios específicos, para a avaliação de seus produtos gênicos estrutural ou funcionalmente.

1.1. *Magnaporthe grisea* E A BRUSONE (*BLAST DISEASE*)

M. grisea é um microrganismo patogênico altamente virulento associado a um vasto espectro de plantas hospedeiras, incluindo mais de 50 espécies de gramíneas. Diferentes linhagens podem infectar cevada, trigo, arroz ou milho (Talbot *et al.*, 1993; Howard; Valent, 1996). Economicamente, este é um dos mais importantes fungos fitopatogênicos do mundo, podendo devastar plantações que alimentariam milhões de

pessoas por ano. Em virtude disto, ele tem sido extensivamente estudado genéticamente, molecular e ultraestruturalmente. *M. grisea* é de fácil manipulação genética (clássica e molecular), constituindo-se em um sistema experimental conveniente para elucidação de numerosos aspectos da patogênese, incluindo a morfogênese relacionada com a infecção, a especificidade de cultivares e espécies-hospedeiros, bem como de vias de sinalização celular. Diferentes genes de patogenicidade desta espécie têm sido identificados. Muitos deles estão associados com a produção de estruturas de infecção, resposta ao ambiente interno do hospedeiro e sinalização para responder a mudanças no ambiente (Idnurm; Howlett, 2001).

M. grisea produz conidióforos cinzentos, geralmente solitários, com conídios acrógenos. Micélio cinza, geralmente pouco desenvolvido. Conídios piriformes medindo 17 a 23 por 8 a 11 μm, a maioria com 2 septos. O ciclo de vida do fungo pode ser descrito tanto por meio de seu desenvolvimento sexual como assexual. O inóculo primário pode ser disperso pelo vento, água ou sementes infectadas. O conídio entra em contato com a superfície foliar do hospedeiro e adere se a ela com a ajuda de uma mucilagem preexistente em uma bolsa localizada na extremidade do conídio. Estudos mostram que esta mucilagem é adesiva devido à sua composição de carboidratos e glicoproteínas (rico em resíduos de a- glucosil e/ou manosil) (Howard & Valent, 1996). Para a penetração e colonização do tecido foliar é necessária a formação de uma célula especializada (*apressório)*. A colonização de plantas hospedeiras suscetíveis desencadeia uma desorganização do metabolismo celular, resultando na

formação de lesões típicas. Do tecido colonizado são produzidos esporos assexuais que darão início a um novo ciclo. O ciclo de infecção completa-se em cinco dias.

A brusone (queimadura do arroz), causada por *M. grisea,* foi provavelmente reconhecida pela primeira vez como "doença da febre do arroz" na China em 1637. Mais tarde foi descrita como *byo-imochi* no Japão em 1704, e como "brusone" na Itália em 1828. O fungo está atualmente presente em pelo menos 85 países, incluindo o Brasil. Em 1996, foi encontrada brusone no arroz na Califórnia, e desde então tem sido encontrada nas gramíneas em campos de golfe no meio-oeste dos Estados Unidos. Linhagens do fungo podem infectar diferentes gramíneas, tais como cevada, trigo, centeio, milheto e gramados, além de arroz. Assim, mesmo quando as colheitas são queimadas para destruir a infecção por fungos, plantas daninhas podem atuar como um reservatório da doença. A doença pode receber diferentes denominações dependendo da cultura infectada: brusone do arroz, brusone do trigo, brusone da cevada, e assim por diante (Cereal Knowledge Bank, 2010).

A brusone do arroz é a mais severa e mais abrangente doença de arroz cultivado, prevalecente tanto nos trópicos como em zonas temperadas. A brusone ocorre em todo o território brasileiro, do Rio Grande do Sul ao Amazonas; os prejuízos são variáveis, sendo maiores em arroz de terras altas. Em condições favoráveis à doença, na região Centro-Oeste brasileira, as perdas podem chegar a até 100% (Embrapa, 2010). A doença pode atingir, com diferentes graus de severidade, diferentes partes da planta em diferentes estágios de desenvolvimento (Figuras 1 e 2).

Figura 1. Escala para diferentes graus de severidade da brusone em folha de *Oryza sativa*: 0 Sem sintomas; **0 a 3**: plantas resistentes (R) ou incompatíveis; **4 a 9**: reações susceptíveis (S) ou compatíveis. *www.cnpaf.embrapa.br/*

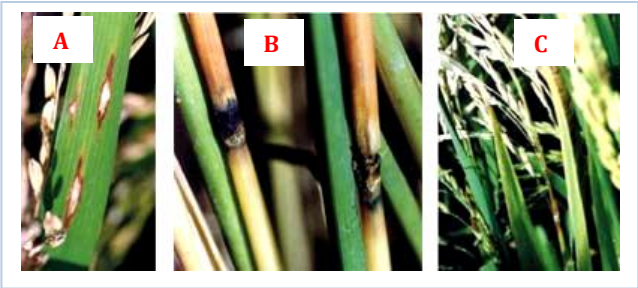


Figura 2. Sintoma de brusone em diversas partes da planta *Oryza sativa*: (A) na folha, (B) no nó dos colmos e (C) na panícula. *sistemasdeproducao.cnptia.embrapa.br*

A busca para a explicação da quebra rápida da resistência ao patógeno tem movido debates na Patologia de Plantas. O debate fica centrado em torno da origem da diversidade de patótipos, ou raças, detectados no campo. Em um extremo, o patógeno tem sido descrito como "hipervariável", com a capacidade de colonizar novos patótipos a partir de um único esporo assexual. Esta possibilidade impele a inúmeras

análises e testes de combinações envolvendo patógeno-hospedeiro a fim de se obter uma raça de planta de resistência durável (Zeigler, 1998).

A severidade da brusone depende de uma série de condições relacionadas à resistência do hospedeiro, à presença de raças do patógeno e à prevalência de fatores do ambiente favoráveis ou não à doença. Aspectos relacionados ao controle de ervas daninhas e à colheita também têm sua importância. A cultura deve ser mantida no limpo para impedir que estas plantas atuem como hospedeiros intermediários do fungo ou mesmo tornem o microclima da cultura mais favorável ao patógeno. A colheita tardia pode favorecer a infecção dos grãos por fungos saprófitas ou parasitas. Recomenda-se que os grãos sejam colhidos com 22% de umidade ou quando a panícula apresentar 2/3 dos grãos maduros (Santos, 2010).

O controle da brusone, tanto no sistema irrigado como no sistema de sequeiro, envolve também o emprego de produtos químicos aplicados como tratamento de sementes e em pulverização da parte aérea. Vários produtos têm sido utilizados. A escolha dos mesmos pode ser feita de acordo com a eficiência do fungicida, sua disponibilidade no mercado e economicidade. Dentre os produtos comumente recomendados estão o benomyl, blasticidin-S, carbendazin, carboxin, ediferiphos, kasugamicina, kitazin, maneb, mancozeb, thiabendazol, triciclazol e pyroquilon (Santos, 2010).

Atualmente, a estratégia mais estudada para controlar a doença na cultura do arroz é baseada, principalmente, no desenvolvimento de cultivares resistentes. No entanto, a "quebra" ou "perda" da resistência das cultivares depois de 3 a 4 anos de seu cultivo intenso e contínuo tem minimizado a eficiência desta alternativa de controle. Os aspectos que envolvem a resistência à brusone do arroz também são bastante conhecidos, tendo sido descritos atualmente em torno de 30 genes de resistência, sendo cerca de 10 destes associados com resistência parcial e 20 com resistência completa (Yoshida *et al*., 2009)

A análise genética clássica de *M. grisea* tem permitido o desenvolvimento de linhagens férteis em laboratório, o isolamento de mutantes, a caracterização da fase diplóide vegetativa e o uso de marcadores genéticos. A infertilidade de linhagens de *M. grisea* que infectam diferentes espécies de gramíneas tem permitido análise genética da especificidade no nível de espécies hospedeiras (Kang *et al*., 1995).

Diversos genes foram identificados em *M. grisea* envolvidos no desenvolvimento da doença que pode causar devastações em plantações de diferentes de cereais e gramíneas (Fourrey; Talbot, 2005; Collemare *et al.*, 2008). Estudos da patogênese neste microrganismo possibilitaram a identificação de diversos genes neste ascomiceto envolvidos no desenvolvimento da doença (Tucker *et al*., 2004; Fourrey; Talbot, 2005).

1.2 OS MECANISMOS DE PATOGENICIDADE DE *Magnaporthe grisea*

A invasão das células da planta por um microrganismo patogênico é um processo dinâmico e contínuo, mas certos estágios-chave podem ser identificados. Tais estágios incluem o contato com o hospedeiro, a adesão à superfície do deste, a penetração e a entrada, a colonização de tecidos, a supressão ou a sobrepujança aos mecanismos de defesa do hospedeiro, a reprodução do patógeno, e a dispersão do patógeno a partir do hospedeiro (Lucas, 1998b). Para cada um destes estágios o fungo pode se mostrar mais eficiente, dependendo da estratégia por ele utilizada.

Um processo de infecção eficiente deve ser aquele que efetivamente causa a doença. O fungo fitopatogênico deve reconhecer e interagir com seu hospedeiro durante os variados estágios do ciclo da doença. A descoberta de mecanismos de patogenicidade bem como de genes de patogenicidade e de avirulência conduzem ao entendimento mais sofisticado da interação entre planta-fungo (Kahmann; Basse, 2001). Diferentes técnicas vêm sendo desenvolvidas para analisar este processo no nível molecular (Skiner *et al.*, 2001; Sweigard; Ebbole, 2001; Caracuel-Rios *et al*., 2007; Mosquera *et al*., 2009).

Os mecanismos pelos quais diferentes patógenos de plantas penetram em seus hospedeiros têm sido o objeto de estudo de inúmeros pesquisadores já por muitos anos. Devido ao impacto danoso nos cultivares de arroz (*Oryza sativa*), *M. grisea* (comumente referido como “rice blast fungus” ou fungo causador da “brusone”) tem

sido alvo de muita atenção. Descrições detalhadas de seu ciclo de vida e de muitos processos biológicos complexos envolvidos tem sido o tópico de vários artigos de revisão (Howard *et al*., 1996; Talbot, 2003; Goriely *et al*., 2006).

Patógenos desenvolvem diferentes estratégias para sobrepujar diversas barreiras que podem encontrar durante o processo de infecção de seus hospedeiros. *M. grisea,* após a sua penetração na célula hospedeira, diferencia hifas invasivas que preenchem o lúmen celular vegetal e estão em contato direto com a membrana da célula infectada. A célula vegetal infectada é viva, exibindo considerável acúmulo de vesículas próximas ao fungo, o que é consistente com o estabelecimento de uma fase biotrófica neste estágio da infecção. A colonização dos tecidos do hospedeiro pelo fungo ocorre através da perfuração da parede celular das células adjacentes, provavelmente usando plasmodesmos como pontos de ruptura, ou através do crescimento de hifas no apoplasma. Depois de alguns dias de crescimento biotrófico dentro dos tecidos da planta, o fungo muda para uma fase necrotrófica-*like* associada ao início da esporulação, levando à lesões visíveis. (Ribot, *et al*. 2008).

M. grisea pode atacar todas as partes da planta, mas tipicamente investe contra folhas e panículas, produzindo aí as lesões mais sérias (Figura 2). Por muitos anos pesquisadores focaram suas buscas na identificação de raças patogênicas do fungo e na incorporação de genes de resistência correspondentes nas variedades de arroz (*Oryza sativa*). Entretanto, resistência durável ao patógeno tem frustrado plantadores e fitopatologistas a despeito de esforço sério e concentrado de ambas as partes.

Tipicamente, uma variedade de arroz assumida como resistente à brusone mostra sinais de suscetibilidade após poucas estações de plantio em ambientes livres do patógeno (Sesma; Osbourn, 2004).

Para infectar tecidos da planta hospedeira, *M. grisea* pode apresentar tanto a reprodução sexuada como a assexuada para produzir estruturas de infecção especializadas conhecidas como *apressório* (Figura 3).

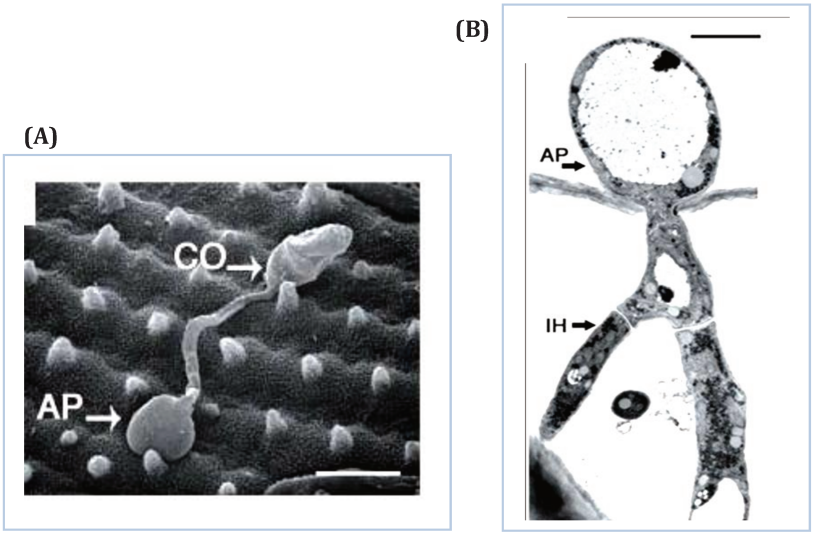


Figura 3- Micrografia por transmissão eletrônica de invasão na cutícula da folha de *Oryza sativa*: (**A**) Conídio (CO) e apressorium (AP). Barra de escala: 10 µm; (**B**) apressório (AP) mostrando uma hifa biotrófica invasiva penetrando na folha de *Oryza sativa*. Barra de escala: 5 µm . (Dean *et al*., 2005)

Assim, uma vez depositado na superfície da planta e na presença de água livre, o conídio germina, produzindo um tubo germinativo e o apressório. A penetração é feita diretamente através da cutícula, raramente pelos estômatos (Saunders *et al*. 2010). A

colonização dos tecidos é facilitada por toxinas (p.ex., pyriculariol), que provocam a morte de células, e por hifas que se desenvolvem no tecido morto (Jo *et al*., 2007).

O modo de infecção da raiz é a mesma de muitos fungos, ou seja, uma longa hifa cresce formando uma estrutura de infecção para penetrar no interior da raiz. Alcançando o interior da raiz o fungo pode iniciar o processo de colonização. Pode haver também invasão do sistema vascular da planta, com *M. grisea* se desenvolvendo dentro do xilema e do floema e bloqueando o transporte de nutrientes e água das raízes, e assim produzir lesões na parte aérea (Gupta; Chattoo, 2007). A infecção de raiz e o bloqueio dos tecidos vasculares tem potencial para matar a planta.

M. grisea é capaz de selecionar resistência aos tratamentos químicos e ao melhoramento genético desenvolvido pelos melhoristas de plantas em alguns tipos de arroz. Pensa-se que o fungo pode atingir a alteração genética referente à resistência por meio de mutação (Jo *et al*., 2007). Com o sequenciamento do genoma de *M. grisea* espera-se que os mecanismos de interação planta-patógeno fúngico, tanto na parte aérea como na infecção da raiz sejam revelados e também que o desenvolvimento de métodos de controle efetivo seja possível (Dean *et al*.,2005; Gupta; Chattoo, 2007).

1.3. GENES DE PATOGENICIDADE E GENES DE AVIRULÊNCIA

A compreensão dos mecanismos de patogenicidade empregados por *M. grisea* é, indubitavelmente, de suma importância. Genes de patogenicidade e genes de

avirulência têm sido identificados neste microrganismo (Idnurm; Howlett, 2001). Os produtos dos genes envolvidos na patogenicidade podem indicar o modo de ação utilizado por determinado microrganismo para causar a doença. Muitos destes produtos gênicos são polipeptídeos cuja função ainda não foi determinada ou não está plenamente caracterizada. Sabe-se que as diversas funções biológicas de moléculas complexas são determinadas por e dependentes de suas estruturas tridimensionais e, também, da habilidade destas estruturas responderem a outras moléculas por mudanças conformacionais (Blundell; Johnson, 1976).

Os tipos de genes essenciais para a patogenicidade dependem do processo de infecção adotado por um determinado fungo. Tem-se arbitrariamente dividido estes genes de patogenicidade de fungos em diferentes grupos, embora alguns genes possam ser classificados em mais de um grupo. Assim, seis grupos de genes de patogenicidade podem ser listados: (1) produção de estrutura de infecção, (2) degradação de parede celular e de cutícula, (3) resposta ao ambiente interno no hospedeiro, (4) produção de toxinas fúngicas, (5) produção de componentes de sinalização em cascata, e (6) genes de patogenicidade que manifestam um novo modo de patogenicidade ainda não descrito (Idnurm; Howlett, 2001).

Existem diversas estratégias que podem ser empregadas por fungos fitopatogênicos para infectar seus hospedeiros (Figura 4). Muitas das interações entre fungo e plantas são do tipo “não-hospedeiro”, em que o fungo não é patogênico, provavelmente porque o microrganismo necessite dos fatores básicos de patogenicidade (Laugé; De

Wit, 1998). As interações do tipo "hospedeiro" podem ser divididas em "compatível" e "incompatível". A herdabilidade de virulência (capacidade de causar doença) ou avirulência (incapacidade de causar doença em cultivares particulares) tem sido investigada, sendo que a primeira parece ser recessiva e a segunda, dominante (Laugé; De Wit, 1998). Esta constatação permitiu o conceito "gene-a-gene", o qual postula que para todo gene dominante determinando resistência no hospedeiro existe um gene dominante para avirulência, relacionado a ele, no patógeno. Assim o gene de avirulência (Avr) de um patógeno deve existir em virtude de um gene de resistência (R) na planta-hospedeiro. Para um elicitador ser um candidato a um produto de um gene de avirulênca ele deve mostrar atividade em plantas contendo o gene complementar R, e ser inativo em plantas carecendo deste gene (Lucas, 1998b).

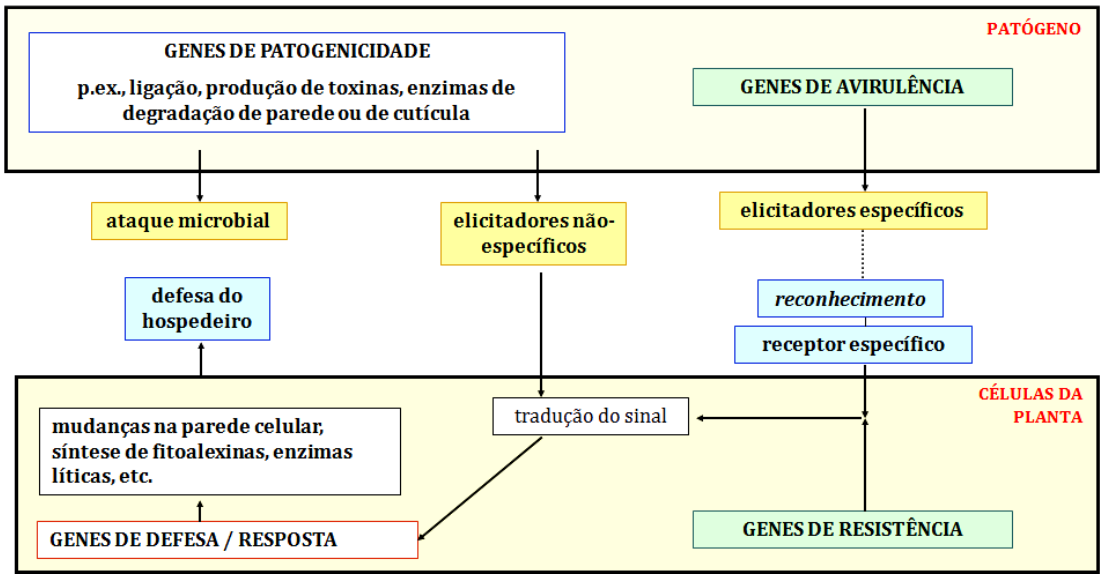


Figura 4- Exemplo de interações estratégicas no mecanismo de infecção por fungos nas células da planta

Elucidar as bases moleculares das interações do fungo com o hospedeiro levará à compreensão das relações patógeno-hospedeiro. O reconhecimento do que não lhe é próprio (*non-self*) e a subseqüente ativação de defesa contra o ataque do patógeno são conhecidos em todos os organismos multicelulares. Os hospedeiros expressam *receptores de reconhecimento de padrões* (PRR) que reconhecem os chamados *padrões moleculares associados ao micróbio ou patógeno* (M/PAMP). Estes padrões são estruturas invariáveis que são indispensáveis ao microrganismo, não existem no hospedeiro e, assim, permitem ao hospedeiro reconhecê-los como não-próprio afastando patógenos invasores. A anteriormente denominada *resistência basal não-cultivar específica* - agora designada *imunidade disparada por PAMP* (PTI) - é suficiente para restringir o crescimento do patógeno (Postel; Kemmerling, 2009).

Para crescer de modo eficiente e proliferar em seu hospedeiro, patógenos virulentos precisam sobrepujar a primeira linha de defesa. Esses patógenos injetam proteínas efetoras na célula de planta que pode suprimir a PTI. *Efetores* são definidos como proteínas ou pequenas moléculas do patógeno que alteram a função e a estrutura da célula do hospedeiro. Estas alterações facilitam a infecção (toxinas ou fatores de virulência), ou disparam respostas de defesa (elicitadores ou fatores de avirulência), ou ambos (Ellis *et al*., 2009).

Patógenos de plantas suprimem a resposta basal de defesa através de efetores que são capazes de causar a doença (De Wit, 2007). Com efetores adicionais e fazendo uso de

nutrientes do hospedeiro os patógenos podem sobreviver e completar seu ciclo de vida. Este fenômeno é chamado de *susceptibilidade disparada por efetores* (ETS). Não obstante, visto não possuírem uma adaptação do sistema imunológico, as plantas desenvolveram uma segunda linha de defesa com moléculas de detecção específicas, as chamadas *proteínas de resistência* (R). Efetores que capacitam ao patógeno sobrepujar a PTI são reconhecidos pelos genes específicos de resistência à doença. O efetor reconhecido é denominado de *proteína de avirulência* (Avr) (Jones; Dangle, 2006). Assim, fatores de virulência são transformados em fatores de avirulência, que permitem com que a planta especificamente detecte o patógeno com eficiência. A percepção da presença destes fatores de avirulência conduz a uma resposta drástica e rápida de hipersensibilidade que restringe o crescimento do agressor (Postel; Kemmerling, 2009).

Fatores de virulência são todos os mecanismos que permitem a invasão pelo hóspede (como estruturas, produtos ou estratégias que o invasor utiliza para sobrepujar o sistema de defesa do hospedeiro) e causa a infecção (Trabulsi *et al.*, 2008). No momento da invasão, *M. grisea* secreta efetores que modificam as defesas e os processos celulares do hospedeiro à medida que ele sucessivamente invade células de arroz. Em plantas resistentes, os efetores são direta ou indiretamente reconhecidos por proteínas de resistência que residem no interior da célula vegetal ou nas membranas plasmáticas. Vários efetores secretados funcionam dentro da célula hospedeira, mas o mecanismo de absorção não é ainda conhecido (Mosquera *et al.*, 2009).

A variação observada entre os efetores de fungos mostra dois tipos de seleção que aparecem para se relacionar com o fato de eles interagirem direta ou indiretamente com suas proteínas de resistência. Em alguns casos, pode ocorrer o reconhecimento direto entre o produto do gene de resistência (R) e o produto do gene de avirulência (Avr). Interações diretas parecem favorecer mutações pontuais nos genes efetores, levando a substituições de aminoácidos (Stergiopoulos; De Wit, 2009). Em outros casos não tem sido observados os pares R/Avr. Esta observação conduziu à formação da hipótese da guarda, em que o modelo proposto é tal que as proteínas R interagem, ou guardam, uma proteína conhecida como *guardee* (a qual é o alvo do produto do gene Avr); quando há a detecção da interferência com a proteína *guardee*, há a ativação da resistência (Dodds *et al.*, 2009; De Wit *et al*., 2009).

1.4. A PROTEÍNA PWL2 DE *Magnaporthe grisea*

Ferramentas de genética molecular em anos recentes têm permitido a identificação e a análise detalhada de genes envolvidos nos mecanismos de interação entre patógeno e seus hospedeiros. Para subverter as defesas do hospedeiro, *M. grisea* produz uma bateria de moléculas efetoras, incluindo alguns com atividade de avirulência (Avr), que são reconhecidos pelas proteínas de resistência do hospedeiro (R), resultando em ativação rápida e eficaz da imunidade inata (Yoshida *et al.,* 2009).

Investigações conduziram à descoberta de uma família multigênica *PWL* (*pathogenicity toward weeping lovegrass*) altamente diversificada, composta de 4 genes que definem 3 subgrupos. Alguns destes genes são determinantes de especificidade e outros interagem com genes de resistência em plantas. Estes genes são encontrados em muitas linhagens de *M. grisea* patogênicas de arroz. Esta família gênica encontra-se no cromossomo 2, numa região de DNA repetitivo (Sweigard *et al.*, 1995).

A análise genética de especificidade de hospedeiro em *M. grisea* permitiu a identificação de um gene, *PWL2*, cujo efeito mais significativo reside na habilidade do fungo infectar *weeping lovegrass* (*Eragrostis curvula*). O gene *PWL2* foi inicialmente clonado e verificou-se que tem propriedades análogas à classe de genes de avirulência, cuja função previne a infecção de certos cultivares de espécies particulares de hospedeiros (Sweigard *et al.*, 1995).

Hifas biotróficas invasivas de *M. grisea* (Figura 3) secretam proteínas efetoras que modificam a defesa do hospedeiro e os processos celulares à medida que ocorre a invasão das células de arroz (Mosquera *et al*., 2009). Poucos genes que são especificamente expressos nestas hifas e muitos genes que codificam efetores secretados têm sido identificados, incluindo o gene de avirulência *PWL2 (*Kang *et al.,* 1995; Sweigard *et al*., 1995; Talbot *et al*., 2003; Donofrio *et al*., 2006; Mosquera *et al*., 2009)

O gene *PWL2,* presente nas linhagens que são patogênicas de arroz, codifica a proteína PWL2, de 16,16 kDa, a qual não tem função conhecida, apresenta uma possível região de ligação transmembrana para os 25 primeiros aminoácidos (possível sequência-sinal) e um conteúdo de 19% de glicina. O loco *PWL2* é altamente polimórfico entre patógenos de arroz proveniente de diversas localizações geográficas. Uma simples substituição de uma única base resulta na perda da avirulência (Kang *et al.,* 1995; Sweigard *et al.*, 1995).

1.5. PROTEÍNAS INTRINSECAMENTE DESORDENADAS

Diferentes doenças em organismos superiores e drogas potenciais eficientes em impedir interações proteína-proteína têm sido associados à desordem estrutural das proteínas. Regiões desordenadas (DR) são proteínas inteiras ou regiões de proteínas que carecem de uma estrutura terciária fixa. Tais regiões desordenadas podem estar parcialmente ou totalmente não-dobradas. Essas regiões desordenadas têm demonstrado estarem envolvidas em uma variedade de funções, incluindo reconhecimento de DNA, modulação de especificidade / afinidade da ligação entre proteínas, e no controle de tempo de vida de proteínas. Embora essas DR careçam de uma estrutura definida em 3-D em seus estados nativos, elas freqüentemente sofrem transições desordem-para-ordem em ligação com os seus pares moleculares. Sabe-se que a seqüência determina a estrutura, mas alguns autores têm assumido que a seqüência deveria determinar a falta de estrutura também (Li *et al*.,1999; Romero *et al*., 2001; Romero *et al*., 2004).

O conceito predominante de que a estrutura 3D ordenada da proteína determina a sua função está passando por uma reavaliação. O interesse em proteínas intrinsecamente não estruturadas está aumentando por causa do reconhecimento de que funções biológicas derivam tanto de estrutura 3D ordenada como da falta de estrutura específica. Na verdade, algumas proteínas exigem a ausência de prévia estrutura 3D para realizar suas funções. Muitas proteínas confirmam que a estrutura inicial em 3D não é necessária para o reconhecimento biomolecular que envolve as proteínas intrinsecamente desordenadas, sendo que as regiões desordenadas podem se ligar a seus alvos com alta especificidade e baixa afinidade; além disso, a desordem intrínseca promove a diversidade de ligação permitindo a interação de proteínas com numerosos pares moleculares (Iakoucheva *et al.*, 2002).

Vários métodos têm sido utilizados para detectar e caracterizar a desordem intrínseca em proteínas, cada um com seus próprios pontos fortes e fracos: cristalografia de raios-X, dicroísmo circular (CD), proteólise limitada, ressonância magnética nuclear (NMR), e outros (Dunker *et al.,* 2002). Na determinação das estruturas das proteínas por cristalografia de raios-X a desordem conduz à perda de densidade eletrônica, o que poderia facilitar a determinação dos locais onde a desordem se apresenta. Entretanto, a maior incerteza na informação proveniente da cristalografia de raios-X é que sem experimentos adicionais não fica claro se a região de perda de densidade eletrônica é um domínio oscilante, ou é intrinsecamente desordenado ou ainda se é o resultado de dificuldades técnicas.

A informação sobre a estrutura para proteínas em solução também é fornecida por dicroísmo circular. *Far-UV CD spectra* fornece uma estimativa da estrutura secundária e distingue *random coils*. No entanto, este método é apenas semi-quantitativo e carece de informações sobre resíduos específicos e, portanto, não fornece dados claros para as proteínas que contêm tanto regiões ordenadas quanto regiões desordenadas (Dunker *et al.,* 2001).

Quanto ao método de proteólise limitada, até o final dos anos 1940 já era reconhecido que a digestão por protease fornece uma visão mais detalhada sobre a a estrutura das proteínas. Estudos mais recentes forneceram evidências convincentes de que a flexibilidade, não a simples exposição da superfície da proteína, é o principal determinante para a digestão de possíveis sítios de corte por proteases (Dunker *et al.,* 2001). Assim, a hipersensibilidade às proteases é uma evidência segura da desordem da proteína. A digestão por proteases dá informação específica sobre posição. No entanto, a exigência de resíduos sensíveis à protease limita a demarcação de ordem/desordem por este método. A digestão por protease é especialmente útil quando usada em combinação com outros métodos (Dunker *et al.,* 2002). Por exemplo, digestão por protease é útil quando combinada com espectros de CD, que carece de informação sobre a posição específica. Finalmente, a combinação da proteólise e espectrometria de massas para a identificação dos fragmentos mostra a promessa especial para indicar a presença de desordem intrínseca (Uversky, 2002).

No caso de NMR, estruturas 3D de proteínas podem ser determinadas em solução. A ausência do requisito de cristalização significa que NMR fornece uma estimativa menos tendenciosa da desordem em comparação com cristalografia, mas domínios de proteínas do tipo *molten globule* serão subestimados na caracterização por NMR.

As relações entre sequência e estrutura indicam que a desordem é uma propriedade codificada (Dunker *et al.,* 2008). Predições em 29 genomas indicam que as proteínas de eucariotos têm muita desordem intrínseca (Dunker *et al.,* 2001). Muitas moléculas de interesse (como drogas potenciais) têm se mostrado atuantes no bloqueio das interações proteína-proteína, baseado na desordem intrínseca de um de seus pares (Iakoucheva *et al.,* 2002); este fato pode incrementar a busca de soluções para inúmeras patogenicidades, incluindo aquelas provocadas por fungos fitopatogênicos, soluções estas fundamentadas em interações envolvendo proteínas distinguidas pela desordem intrínseca de um de seus pares.

1.6. XILANASES

Celulose, hemicelulose (que contem xilana) e lignina são os principais componentes das paredes das células vegetais. A xilana é um heteroglicano com um elevado potencial de conversão para os produtos finais úteis. A conversão completa da hemicelulose requer a ação de várias enzimas que clivam a cadeia principal e a cadeia lateral: endoxilanase (endo-1,4-β-xilanase, EC 3.2.1.8) β xilosidase (1,4-β-xilosidase, EC3.2.1.37), α- glucuronidase (α-glucosiduronase, EC3.2.1.139), α-

arabinofuranosidase (α-L arabinofuranosidase, EC3.2.1.55) acetylxylan e esterase (CE 3.1.1.72). Endo-β-1,4-xilanases catalisam a hidrólise de xilana na cadeia principal para produzir xilooligosacarídeos que, por sua vez, podem ser convertidos em xilose por xilosidase-β (Zhang *et al*., 2007).

Com base em homologias de seqüência de aminoácidos e análise de *clusters*, as xilanases foram agrupadas principalmente em duas famílias de glicosil hidrolases: família F ou GH10 e família G ou GH11 (Jeffries 1996; Zhou *et al*., 2008). No entanto, outras famílias glicosídeo hidrolase (GH5, GH7, GH8 e GH43) também foram encontrados contendo diferentes domínios com atividade endo-1,4-β-xilanase (Collins *et al*., 2005).

A natureza é repleta de bactérias e fungos que podem produzir enzimas de degradação da parede celular. A maioria dos microrganismos isolados de resíduos do solo/compostagem é capaz de produzir um espectro de enzimas de degradação da parede celular (Badhan *et al*., 2007). Os fungos filamentosos têm sido utilizados por mais de 50 anos na produção de enzimas industriais. Eles são amplamente utilizados como produtores de xilanase e são geralmente considerados como mais potentes do que bactérias e leveduras (Polizeli *et al*., 2005; Pedersen *et al*., 2007). Gêneros de fungos que são conhecidos por produzir xilanases incluem *Aspergillus, Disporotrichum, Penicillium, Neurospora, Fusarium, Chaetomium, Trichoderma*, etc (Kulkarni *et al*., 1999; Mach; Zeilinger, 2003; Saleem *et al*., 2008). As xilanases de fungos têm muitos usos comerciais tais como na confecção de papel, na ração animal,

na confecção de pães, na indústria produtora de sucos e vinho, na indústria têxtil, na indústria farmacêutica, etc. (Polizeli *et al.*, 2005).

Sabendo que patógenos de plantas penetram em seus hospedeiros por diferentes mecanismos incluindo pela degradação de parede celular, muito tem sido descrito a respeito de seu ciclo de vida e de muitos processos biológicos complexos envolvidos no desenvolvimento da patogenicidade de *M. grisea* (Kwon *et al.*, 2008; Yara *et al.*, 2008; Talbot, 2003; Goriely *et al*., 2006). Um dos mecanismos que *M. grisea* utiliza para penetrar o hospedeiro é através de um complexo enzimático xilanolítico de degradação de parede celular da planta. Diferentes xilanases têm sido identificadas em *M. grisea* e estão possivelmente envolvidas nos mecanismos de infecção das plantas hospedeiras (Shengcheng *et al*. 2001). Atualmente, a amplificação e clonagem de três genes putativos de xilanases, XYL3, XYL4 e XYL5 (respectivos números de acesso no GenBank: **AY144348** a **144.350**) foram relatados (Wu *et al*., 2006). Dentre o complexo xilanolítico, a enzima mais importante é a endoxilanase (endo-1,4-β-xilanase, EC 3.2.1.8) (Zhou, 2008). A proteína XYL5 é uma putativa endo-1,4-β-xilanase que contém um domínio conservado da família 10 de Glicosídeo Hidrolases GH10 (Berrin; Juge, 2008).

Xilanases têm sido isoladas de microrganismos de vários gêneros e expressos em *Escherichia coli*. Entretanto, em bactérias, as xilanases são produzidas em níveis muito menores e restritas às frações periplasmática e intracellular. Além disso, as enzimas expressas em bactéria não estão sujeitas às modificações pós traducionais (Oliveira,

2006; Polizeli *et al.*, 2005). Assim, investimentos diversos têm sido feitos no sentido de utilizar ferramentas de genética molecular objetivando-se entender os mecanismos complexos que estão envolvidos na interação entre patógenos e seus hospedeiros. Para tal, diferentes estratégias de clonagem, expressão e purificação de proteínas patogênicas têm sido desenvolvidas visando aperfeiçoar cada etapa do processo que tem como meta a caracterização estrutural e funcional de tais proteínas recombinantes. Deste modo, otimiza-se o uso de vetores plasmidiais diversos para facilitar os procedimentos que possibilitem aumentar o produto protéico purificado, com base na especificidade para a ligação em resinas utilizadas em cromatografia de afinidade. Otimiza-se também a utilização de linhagens de expressão que aumentem a solubilidade de proteínas recombinantes (Nallamsetty, 2007; Thiede *et al*., 2005; Nishihara, 1998; Mortimer *et al*., 1997). Neste trabalho, nós clonamos, expressamos e avaliamos a atividade da enzima xilanolítica XYL5 íntegra e de seu domínio GH10 como proteínas de fusão com tiorredoxina (TRX) ou maltose binding protein (MBP).

2. OBJETIVOS

2.1. OBJETIVOS GERAIS

- Contribuir para a compreensão do mecanismo utilizado pelo fitopatógeno *M. grisea* para infectar e causar doença (brusone ou *blast disease*) na planta hospedeira
- Auxiliar no entendimento da função gênica na patogenicidade
- Fornecer subsídios para a seleção de plantas resistentes ao patógeno, visando à aplicação prática de cunho econômico

2.2. OBJETIVOS ESPECÍFICOS

- Clonagem e expressão, em *E. coli*, de genes relacionados à patogenicidade de *M. grisea* visando à purificação de suas proteínas relacionadas
- Avaliar aspectos estruturais e funcionais das proteínas PWL2D e XYL5 que fossem relevantes para a compreensão das interações do fungo com seu hospedeiro

3. METODOLOGIA

As amplificações das seqüências-alvo para a obtenção e a purificação dos insertos a serem clonados foram feitas de acordo com os manuscrito 1 e 2. A transformação das células de *E. coli* com PEG (polietilenoglicol 10%) foi feita de acordo com Sambrook (Sambrook; Russell, 2001). Os procedimentos de clonagem, as condições de cultivo e de indução de expressão, as purificações por cromatografia de afinidade e a análise por SDS-PAGE das proteínas foram realizadas de acordo com o manuscrito 1 para todos os genes de *M. grisea* aqui estudados. Em caso de outro método usado para os procedimentos anteriores, será especificado no texto.

3.1. AVALIAÇÃO DA ATIVIDADE XILANOLÍTICA PELO MÉTODO DO DNS

A quantificação dos açúcares redutores totais presentes no extrato enzimático foi feita pelo método do ácido 3,5-dinitrosalicílico (DNS) conforme Miller (Miller, 1959) utilizando solução de xilose como padrão. De acordo com este método, o DNS sofre redução de um de seus grupos nitro ao reagir com os carboidratos, formando um composto que apresenta forte absorção em 540nm. Na reação ocorre uma redução do ácido 3,5-dinitrosalicílico para 3-amino-5-nitrosálico, que é um produto de cor laranja, enquanto o grupo aldeído dos açúcares redutores é oxidado para carbonila.

A solução com o reagente DNS foi preparada (dissolvendo-se 10g DNS, 2g de fenol, 0,5g de sulfito de sódio e 10g de NaOH em 1000ml de água destilada), e armazenada em frasco plástico protegido da luz à temperatura ambiente.

A solução de tartarato duplo de sódio e potássio foi preparada (4 g do sal em um volume final de 100 ml de água destilada), armazenada protegida da luz e em temperatura ambiente.

A solução de xilose para a curva-padrão foi preparada a 1 mg/ml. A curva-padrão foi feita com a leitura da absorbância para as concentrações de xilose a 1,0 mg/ml, 0,5 mg/ml, 0,25 mg/ml, 0,12 mg/ml, 0,06 mg/ml e 0 mg/ml.

A solução de xilana (*birchwood*, Sigma) a ser usada como substrato foi preparada a 1%. O conteúdo de proteínas totais nas amostras foi padronizado por meio da leitura em 280nm espectrofotômetro. Um volume de 665 μl do substrato xilana em solução foi adicionado a 335 μl se solução contendo as amostras padronizadas, levado a banho de 50°C por 20 min. Um volume de 300 μl desta mistura foi adicionado a tubos contendo 300 μl da solução de DNS. Os tubos, fechados, foram levados a banho de 90° C por 15 min. Os tubos foram retirados do banho, e 100 μl da solução de tartarato foi adicionado. A leitura foi feita em 540 nm, sendo que cada leitura foi realizada em triplicata proveniente de medidas independentes.

Para detectar as unidades redutoras não provenientes da hidrólise enzimática da xilana foi utilizado como controle uma mistura da solução de xilana, tampão de extração e o DNS. Uma unidade de atividade da enzima foi definida como a quantidade de enzima requerida para gerar 1 μmol de xilose equivalente por minuto a 50°C, e a atividade específica foi definida como unidades por mg de proteína.

O procedimento anterior foi também utilizado para as frações provenientes da purificação por cromatografia de afinidade à amilose ou ao níquel: as frações eluídas contendo as proteínas de interesse foram dializadas contra tampão Tris (50 mM Tris-HCl, pH 7.0), por 8-12 horas, sendo posteriormente quantificadas e submetidas à análise da atividade xilanolítica. Detalhes sobre a avaliação da atividade xilanolítica utilizando o método DNS para as proteínas XYL5 e XYL5/DOM e para as proteínas de fusão MBP-XYL5 e MBP-XYL5/DOM encontram-se no manuscrito 2.

3.2. AVALIAÇÃO DA ATIVIDADE XILANOLÍTICA PELO MÉTODO DA COLORAÇÃO DE MEIO EM PLACAS

O método de avaliação da atividade xilanolítica em placas foi feito com uma adaptação de Teather (Teather; Wood, 1982) e de Kasana (Kasana *et al.*, 2008). Foi utilizada uma solução de Congo Red 2%. Este corante (3,3'-(4,4'-Bifenildiilbisazo) bis- (4-amino-1-naftalinsulfonato)- de dissódio) tem uma forte afinidade por fibras de celulose, mas esta afinidade parece não ser por ligação covalente (Howie *et al*., 2008). A agregação

do corante tende ao vermelho no espectro de absorção, onde a ligação à fibras de celulose ou à fibras de amilóides tem o efeito oposto, ou seja, a ausência da coloração vermelho-alaranjado (Teather; Wood, 1982; Li *et al*., 2010).

Placas de Petri contendo 1% de agarose e 1% de xilana (w/v) foram preparadas; poços com cerca de 3,0mm de diâmetro foram feitos no meio solidificado. Nestas placas foram inoculados 20 μL de extrato enzimático contendo as proteínas de interesse (este extrato foi proveniente da fração solúvel obtida de 3 mg de massa de bactérias contendo MBP-XYL5 ou MBP-XYL5/DOM). Como controle negativo foi utilizado o mesmo procedimento para com o massa de bactérias contendo o vetor pSV282 sem inserção dos genes das proteínas de interesse. As placas inoculadas foram incubadas por 22h a 25°C. Aproximadamente 10 ml de solução de Congo Red foi cuidadosamente depositada sobre o meio nas placas, seguido de incubação à temperatura ambiente por 15 min. A solução de Congo Red foi removida e 10 ml de solução salina (1M NaCl) foi adicionada. Após 30 min a solução salina foi removida e uma solução de 50 g/l de ácido acético foi adicionada para permitir a visualização dos halos de degradação da xilana.

O procedimento anterior foi também utilizado para as frações provenientes da purificação por cromatografia de afinidade ao níquel ou à amilose: as frações eluídas contendo as proteínas de interesse foram dializadas contra tampão Tris (50 mM Tris-HCl, pH 7.0), por 8-12 horas, sendo posteriormente quantificadas e submetidas à análise da atividade xilanolítica em placa. Como controle negativo foi utilizado o

mesmo procedimento com a fração solúvel proveniente da massa de bactérias contendo o vetor pSV282 sem inserção dos genes das proteínas de interesse.

Detalhes sobre a avaliação da atividade xilanolítica utilizando o método de coloração com Congo Red para as proteínas de fusão MBP-XYL5 e MBP-XYL5/DOM encontram-se no manuscrito 2.

4. MANUSCRITOS E INFORMAÇÕES ADICIONAIS

4.1. MANUSCRITO 1

Este manuscrito foi publicado no periódico *Protein Expression and Purification* sob o código de referência DOI information 10.1016/j.pep. 2010.04.020.

Overexpression and purification of PWL2D, a mutant of the effector protein PWL2 from *Magnaporthe grisea*

Schneider, DRS; Saraiva, AM; Azzoni, AR; Miranda, HRCAN; de Toledo, MAS; Pelloso , AC; de Souza, AP

ARTICLE IN PRESS

Protein Expression and Purification xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Overexpression and purification of PWL2D, a mutant of the effector protein PWL2 from *Magnaporthe grisea*

D.R.S. Schneider [a], A.M. Saraiva [a], A.R. Azzoni [a], H.R.C.A.N. Miranda [a], M.A.S. de Toledo [a], A.C. Pelloso [a], A.P. Souza [a,b,*]

[a] Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas, CP 6010, CEP 13083-970 Campinas, SP, Brazil
[b] Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, CEP. 13083-970 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 25 January 2010
and in revised form 22 April 2010
Available online xxxx

Keywords:
PWL2D
Magnaporthe grisea
Escherichia coli
Effector protein
Intrinsically disordered protein
Thioredoxin tag

ABSTRACT

The rice blast disease caused by the ascomycete *Magnaporthe grisea* continues to cause a tremendous impact in rice (*Oryza sativa*) cultures around the world. Elucidating the molecular basis of the fungus interactions with its host might help increase the general understanding of the pathogen–host relationship. At the moment of invasion, the fungus secretes effectors that modify host defenses and cellular processes as they successively invade living rice cells. PWL2, an effector protein, is a known AVR (avirulence) gene product. The *PWL2* gene prevents the fungus from infecting weeping lovegrass (*Eragrostis curvula*). In this study, we identified a *PWL2* allele gene (which we termed *PWL2D*) in a strain of *M. grisea*. The sequence of *PWL2D* has only two bases different from that of *PWL2*, producing alterations in residue 90 and residue 142. However, the alteration of residue 90 (from D_{90} to N_{90}) is critical to gene function. Here, we cloned the gene *PWL2D* in a pET System vector, expressed the gene product in *Escherichia coli* and evaluated by spectroscopic techniques some aspects of the PWL2D structure. While TRX-tagged PWL2D is prone to aggregation, the solubility of PWL2D is improved when it is overexpressed without its original signal peptide. Expression and purification procedures for these constructs are described. Finally, we found out that the protein seems to be an intrinsically disordered protein. Results from these studies will facilitate structural analysis of PWL2D and might contribute to understanding the gene's function and of fungal/plant interactions.

Introduction

Magnaporthe grisea is the causal agent of one of the most devastating rice (*Oryza sativa*) diseases in the world. The fungus uses sophisticated mechanisms such as the formation of infection structures for its development through colonization, leading to the emergence of disease in plant [1–3]. This microorganism is known to be a "cereal killer" pathogen that targets many species of grasses [4,5] such as *Eragrostis curvula* ("weeping lovegrass") and cause devastation in many different grain crops, such as *Eleusine coracana*, *Pennisetum glaucum*, barley (*Hordeum vulgare*), wheat (*Triticum aestivum*) and maize (*Zea mays*).

The genome of *M. grisea* has already been sequenced and includes approximately 40 MB packaged in seven chromosomes [6]. A classical genetic analysis has allowed the development of fertile laboratory strains, the isolation of mutants, the characterization of diploid vegetative stages and the use of genetic markers. Investigations led to the discovery of a gene family (PWL – Pathogenicity toward Weeping Lovegrass) which is highly diversified, consisting of four genes (*PWL1*, *PWL2*, *PWL3 and PWL4*) that define three subgroups. Some of these genes are determinants of host specificity [5] and their products interact with other resistance gene products in plants. These genes are found in many strains of *M. grisea* that are rice pathogens [7].

PWL2, a member of the family PWL, exerts a major effect on the ability of *M. grisea* to infect weeping lovegrass [5,7]. Biotrophic invasive hyphae (IH)[1] of the blast fungus *Magnaporthe oryzae* secrete effectors to modify host defenses and cellular processes as they successively invade living rice cells [8]. A few genes that are specifically expressed in blast IH [4,9] and several genes encoding secreted effectors, including the known AVR (avirulence) gene product PWL2, have been identified [7,8].

In the family PWL, multiple homologues have been identified with varying degrees of sequence homology. Studies indicate that PWL genes may have other functions besides eliciting host defense responses [5].

* Corresponding author at: Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas, CP 6010, CEP 13083-970 Campinas, SP, Brazil. Fax: +55 19 3521 1089.
E-mail address: anete@unicamp.br (A.P. Souza).

[1] *Abbreviations used:* IH, invasive hyphae; IMAC, metal ion affinity chromatography.

1046-5928/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2010.04.020

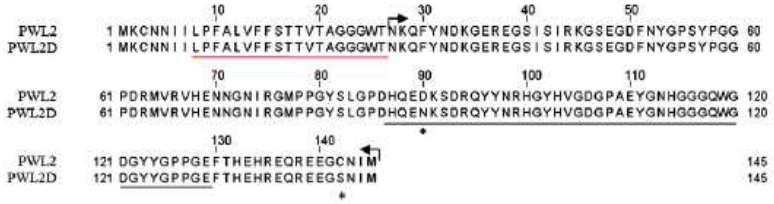


Fig. 1. Alignment of PWL2 and PWL2D protein sequences from *M. grisea*. The Accession Number from GenBank, a protein sequence database, for PWL2 is AAA91019. The sequence alignment was generated using Clustal W2. Arrows indicate the start and the end of the region *sh*PWL2D (PWL2D without signal peptide). The black underline indicates the start and the end of the region *or*PWL2D (predicted ordered region of PWL2D). The red underline indicates the signal peptide of PWL2D. The asterisks indicate the different residues between PWL2 and PWL2D.

In this study, we identified a gene allele to PWL2 (we termed it *PWL2D*, an allele of the gene PWL2 obtained from *M. grisea* strain D). The sequence of *PWL2D* differs by two bases from that of *PWL2* (GenBank database under the Accession No. U26313.1). These two different bases produce alterations in residues 90 and 142 (Fig. 1). The avirulence gene *PWL2* encodes a secreted effector protein [8]. The alteration in residue 90 (from D_{90} to N_{90}) is critical for gene avirulence function and this alteration results in the loss of recognition of the pathogen by the host plants [7]. The alteration in residue 142 (from C_{142} to S_{142}) eliminates one cysteine in the protein sequence. We cloned the gene *PWL2D* in a pET System vector, expressed the gene product in *Escherichia coli* and used spectroscopic and NMR techniques to evaluate some aspects of the structure of PWL2D that are relevant to the understanding of fungal/plant interactions.

Many proteins have been classified as intrinsically unstructured. Their functional roles are in crucial areas such as transcriptional regulation, modulation of specificity/affinity of protein binding, translation and cellular signal transduction, and these roles have only recently been recognized, as a consequence of new paradigms in biochemical methodology [10]. The possibility that PWL2D is an intrinsically disordered protein (and the consequences of this in light of its role as an effector protein) is also discussed.

Materials and methods

Materials

The strain Py416-L1-3 (termed "D" strain in this study) of the filamentous fungus *M. grisea* was obtained from the National Center for Research on Beans and Rice (CNPAF, GO, Brazil). The strain was used as a source of genomic DNA and to obtain the gene *PWL2D*. Strains of *E. coli* DH5α [F2 lacZD M15 hsdR17 (r2 m2) gyr A36] and BL21 (DE3) (Novagen, Madison, WI, USA) were used as the host strains for subcloning and gene expression, respectively. The plasmid vector pET32-Xa/LIC (*Novagen*, Madison, WI, USA) was used for cloning and gene amplification, as well as for protein expression.

The modification enzymes T4 DNA ligase and Taq polymerase were obtained from Invitrogen (CA, USA). The Ni–NTA Superflow Resin (*QIAexpress System Qiagen*, Hilden, Germany) was used for purification by nickel affinity chromatography. The HiTrap ANX-FF Chromatography Column (*GE Healthcare Inc.*, Uppsala, Sweden) was used for purification by anion-exchange chromatography. The proteases factor Xa (*Novagen*, USA) or thrombin (*GE-Pharmacia Biosciences*, UK) were used to cleave the solubility tag. Fragments amplified by polymerase chain reaction (PCR) were purified using QIAquick PCR Purification Kit (*Qiagen*, Hilden, Germany). Plasmid DNA was purified using a QIAprep Spin Miniprep Kit (*Qiagen*, Hilden, Germany). All chemicals were obtained from *Sigma* (St. Louis, MO, USA) unless otherwise specified.

Genomic DNA extraction and purification

The method for genomic DNA extraction and purification from *M. grisea* was adapted from a method described previously [11]. Petri dishes of 9.0 cm in diameter, with solid medium PDA, were covered with sterile and permeable cellophane disks. Seven "plugs" from plates with mycelium were placed on this cellophane. After 4–6 days of incubation in the incubator at 28 °C, the mycelium was scraped from the cellophane with a properly sterilized spatula. The material was transferred to 1.5 mL microcentrifuge tubes and 600 μL of extraction buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA, 100 mM NaCl, 2% SDS) was added to each tube. The following cycle was repeated three times: the tubes were vortexed for 30 s, frozen in liquid nitrogen for 30 s, and incubated in a dry incubator for 30 s (70 °C) until thawing was complete. DNA was purified using the classic phenol–chloroform–isoamyl alcohol method. The DNA was precipitated with ethanol and sodium acetate. The pellets of DNA were resuspended in 20 μL TE buffer [12].

Construction of expression vectors

Three vectors were constructed to express PWL2D (full sequence, 16.16 kDa), *sh*PWL2D (*short sequence* of PWL2D, PWL2D without the predicted signal peptide: residues 27–145, 13.38 kDa) and *or*PWL2D (*ordered region* from PWL2D, predicted ordered region from PWL2D: residues 86–129, 4.92 kDa) as TRX-fusion proteins. All proteins were expressed as TRX-fusion proteins from the cloning vector in pET32-Xa/LIC.

The target gene amplification for cloning into pET32-Xa/LIC was done in two steps by polymerase chain reaction. First, the DNA fragments encoding the sequences *PWL2D*, *sh*PWL2D and *or*PWL2D were amplified from purified genomic DNA and primer set A from Table 1. These primers amplify around 300 bp upstream and 300 bp downstream of the *PWL2D* gene. The amplification protocol consisted of a 3 min denaturation at 94 °C followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at adequate temperature for the primers for 1 min and 30 s, extension at 72 °C for 2 min, and final extension at 72 °C for 7 min. Second, this first amplification was used as a template for the specific amplification to clone into pET32-Xa/LIC: *PWL2D*, *shPWL2D* and *orPWL2D* were amplified with primer sets B–D from Table 1, respectively. The primers were designed to generate products with vector cohesive overhangs (underlined bases on the Table 1). In this second amplification, we used the same amplification protocol, except for the annealing temperature that was 55 °C. The amplified DNA fragments were analyzed on 1% agarose gel. The blunt-ended PCR products were purified, treated and cloned in the linearized pET32-Xa/LIC vector according to the vector supplier's instructions. After cloning in the expression vector, the constructs pET32.*PWL2D*, pET32.*shPWL2D* and pET32.*orPWL2D* were introduced by chemical transformation into *E. coli* DH5α [12]. As a negative control of expression, vectors without the inserts of interest

Table 1
Sequences of oligonucleotide primers used for amplification and cloning.

Primer sets	Sequence	Gene	Cloning and expression plasmidial vector
A	5′ – CGC GGA CGG GAC GAG TAA AAA – 3′ 5′ – GGT GGG TCT GGG GGC GTG AT – 3′	*PWL2D*, *shPWL2D* and *orPWL2D*	-
B	5′ – GGT ATT GAG GGT CGC *ATG* AAA TGC AAC AAC ATC – 3′ (Start codon; Specific sequence for cloning in pET32-XaLIC) 5′ – AGA GGA GAG TTA GAG CCT *TAC* ATA ATA TTG GAG CC – 3′ (Stop codon; Specific sequence for cloning in pET32-XaLIC)	*PWL2D*	pET32-XaLIC
C	5′ – GGT ATT GAG GGT CGC *ATG* AAC AAA CAA TTT TAC – 3′ (Start codon; Specific sequence for cloning in pET32-XaLIC) 5′ – AGA GGA GAG TTA GAG CCT *TAC* ATA ATA TTG GAG CC – 3′ (Stop codon; Specific sequence for cloning in pET32-XaLIC)	*shPWL2D*	pET32-XaLIC
D	5′ – GGT ATT GAG GGT CGC *ATG* GAT CAT CAG GAA AAT – 3′ (Start codon; Specific sequence for cloning in pET32-XaLIC) 5′ – AGA GGA GAG TTA GAG CCT *TTA* CTC GCC TGG CGG – 3′ (Stop codon; Specific sequence for cloning in pET32-XaLIC)	*orPWL2D*	pET32-XaLIC

were also introduced into *E. coli* DH5α. All sequences were verified by automated DNA sequencing.

Expression of fusion proteins in E. coli

PWL2D, *sh*PWL2D and *or*PWL2D were expressed in *E. coli*. The constructs pET32.*PWL2D*, pET32.*shPWL2D* and pET32.*orPWL2D* were introduced into *E. coli* BL21(DE3) by chemical transformation. Vectors without inserts were also used as expression controls. After incubation and selection of positive clones, cultures were grown overnight at 37 °C and 300 rpm and diluted 1:100 in fresh LB medium [12] supplemented with ampicillin 100 μg/mL. Cultures were then incubated at 37 °C and 300 rpm until optical density (600 nm) reach 0.8. For the induction of the target protein expression, lactose was introduced into the 1.0 L bacteria shake flasks to a final concentration of 5.6 mM. The cells were then incubated at 37 °C and 300 rpm for 4 h. The expression of fusion proteins in *E. coli* was monitored by SDS–PAGE.

Extraction and purification of TRX–PWL2D, TRX–shPWL2D and TRX–orPWL2D proteins from E. coli

For the *E. coli* pellet obtained from the centrifugation of 500 ml of culture medium, 20 ml of extraction buffer (50 mM Tris–HCl, 300 mM NaCl, pH 7.5) was added. After efficient sonication, two centrifugations were performed (23,000*g*, 20 min, 4 °C). The precipitate (insoluble material) was stored at −20 °C, and the supernatant fraction (soluble fraction) was used for purification. The recombinant proteins were purified by metal ion affinity chromatography (IMAC) using Ni–NTA or by ANX-FF anion-exchange chromatography.

In the affinity chromatography, the soluble fractions obtained from the protein extraction step were loaded into the IMAC column (1 mL of resin) pre-equilibrated with extraction buffer. The column was washed with extraction buffer, and bound protein was eluted using an imidazole gradient from 5 to 500 mM on an AKTA FPLC chromatography system (*GE Healthcare Inc.*, Uppsala, Sweden) or a gravity chromatography system. In the gravity chromatography system, the soluble fraction obtained was carried through a column containing the resin Ni–NTA, with increasing concentrations of the eluent imidazole (5, 10, 50, 100, 200, 250 and 500 mM). Protein elution was monitored at 280 nm, and the resulting fractions were analyzed by SDS electrophoresis on Gradient 8–25 PhastGelTM (*GE Healthcare Inc.*, Uppsala, Sweden). All material that passed through the column in each step was collected.

To perform anion-exchange chromatography the soluble fractions obtained from the protein extraction step were dialyzed against buffer 1 (50 mM Tris–HCl, 100 mM NaCl, pH 7.5) overnight at 10 °C. After this step, the soluble fractions were dialyzed against buffer 2 (50 mM Tris–HCl, 50 mM NaCl, pH 7.5) overnight at 10 °C. The samples were applied to an ANX-FF anion-exchange chromatography column pre-equilibrated with buffer 2. Elution was performed using a gradient of up to 500 mM NaCl. Protein elution was monitored at 280 nm, and the resulting fractions were analyzed by SDS–PAGE.

Cleavage of the fusion protein

To test protein solubility in absence of the TRX tag, purified proteins were dialyzed against protease cleavage buffer, and cleaved using recombinant proteases factor Xa or thrombin. The resulting samples were loaded on Ni–NTA resin or on an anion-exchange chromatography column (ANX-FF) to optimize the purification and the characterization.

Peptide mass fingerprint (PMF) by MALDI-TOF/MS

Sample preparation and analysis were performed as previously described [13]. Briefly, each protein sample was subjected to electrophoresis (SDS–PAGE), and then the band of interest was cut from the gel. The gel fraction was treated for the removal of the dye. Analyses of all samples obtained by tryptic digestion were prepared using the dried droplet method. Trypsin was added for digestion, keeping the material at 37 °C for 12 h. The peptides obtained were extracted and taken to dry in a vacuum microcentrifuge. The sample was acidified by the addition of two sample volumes of 0.1% (v/v) trifluoroacetic acid (TFA) and was left at room temperature for a few minutes, to reduce the droplet volume through evaporation. The matrix [1% (w/v) α-cyano-4-hydroxycinnamic acid in 1:1 (v/v) H_2O/ACN solution containing 0.1% (v/v) TFA] was added, and the sample was allowed to dry at room temperature. All measurements were performed on a Matrix-Assisted Laser/Desorption Ionization combined with Time-of-Flight mass spectrometer (MALDI-TOF–mass spectrometer, *Waters-Micromass*,

UK) in a previously cleaned microplate. Analyses were performed exclusively for positively charged ions, in reflectron mode. For subsequent data evaluation, including peptide identification, the software package Masslynx (*Waters Corporation*, MA, USA) was used.

Circular dichroism spectroscopy

Far-UV Circular dichroism (CD) measurements were performed using a Jasco J-715 spectropolarimeter (*Jasco*, Tokyo, Japan). After purification TRX-PWL2D, TRX-*sh*PWL2D, TRX-*or*PWL2D, PWL2D, *sh*PWL2D and *or*PWL2D were dialyzed overnight against the CD buffer (5 mM sodium phosphate, pH 7.5). The dialyzed samples were diluted to 0.5 mg/mL or 0.25 mg/mL before data collection. Data points were collected from 190 to 260 nm at 20 °C using a 1.00-mm cuvette. Each spectrum was the average of 20 scans. The CD spectra were obtained in milli-degrees and converted to molar ellipticity.

Nuclear magnetic resonance

We used spectroscopy by nuclear magnetic resonance (1D-NMR) to obtain information on the structure and dynamics of the proteins. After cleavage of the fusion tag, the sample relating to *or*PWL2D was purified as described. The sample was dialyzed in 20 mM sodium phosphate buffer pH 6.8 plus 200 mM NaCl. The experiments were performed under a frequency field of 600 MHz, with 90° pulses in relation to the Z axis (parameter PW = 8 s) on a Varian Inova Spectrometer (11.7 T) (*Varian Inc.*, CA, USA). The data were analyzed with the use of compatible software.

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) of the fusion proteins

The soluble and insoluble fractions (inclusion bodies) of the fusion proteins were analyzed by SDS–PAGE. *E. coli* that expressed the TRX-fusion protein were disrupted by efficient sonication. The pellets containing the inclusion bodies and the supernatant with soluble protein were collected by centrifugation for 30 min at 23,000g. Samples of these were put on loading gel with buffer (0.5 M Tris–HCl, pH 6.8, 2% glycerol, 10% SDS, 0.1% bromophenol blue) and separated by 12% (w/v) SDS–PAGE. The proteins were then stained with Coomassie Brilliant Blue R-250. Total protein concentration was determined by the ratio between the molar extinction coefficient and absorbance at 280 nm.

Results and discussion

Design and construction of fusion protein expression vectors

Before the design and construction of the cloning vectors for the expression of the PWL2D protein, we considered some relevant aspects regarding this protein. Since sequence determines structure, some authors have assumed that sequence would also determine lack of structure [14–16]. In this way, a series of neural network predictors (NNPs) have been developed that use amino acid sequence data to predict disorder in a given region of a protein. This collection of predictors of natural disordered regions was termed PONDR [14]. In order to increase the chances of soluble protein expression in *E. coli*, we examined some characteristics of the PWL2D protein using prediction softwares [14–17]. Since there is evidence that signal-peptides for translocation of proteins to the cytoplasm membrane may result in the increase of the inclusion bodies formation [18], we decided to remove the putative signal peptide coding region (membrane inside and transmembrane helix segments, residues 1–26, Fig. 1 and Table 2) and this protein was named *sh*PWL2D. We also selected a region inside the full protein PWL2D that was predicted as ordered containing the residue 90 considered critic to gene function [7]. Therefore we cloned *orPWL2D*, a sequence related to ordered region (residues 86–129) (Fig. 1 and Table 2), to verify its expression and solubility.

Table 2
Comparison of predictions of natural disordered regions and transmembrane helices in proteins PWL2 and PWL2D.

Predicted Characteristics	PWL2	PWL2D
Number of residues	145	145
Number of disordered residues	51	55
Disordered segment	37–50	37–50
	61–85	61–85
	130–141	130–145
Membrane inside segment	1–6	1–6
Transmembrane helix segment	7–26	7–26
Membrane outside segment	27–145	27–145

The chosen cloning vector pET32-Xa/LIC vector contains the coding sequence of TRX (thioredoxin) protein that has been used to improve the solubility of proteins fused to it [19]. In this vector, TRX tag is followed by a sequence of 6 histidine residues (His-tag) and a sequence of 15 residues (S-tag). The protein of interest is fused after to these three regions. PWL2D (full sequence, 16.16 kDa), *sh*PWL2D (residues 27–145, 13.38 kDa) and *or*PWL2D (residues 86–129, 4.92 kDa) were cloning in the pET32-Xa/LIC plasmid vector, and three different constructions were obtained: pTRX–PWL2D, pTRX–*sh*PWL2D and pTRX–*or*PWL2D, all fused to TRX protein (17.16 kDa) and containing a His-tag for purification using IMAC.

Expression and purification of thioredoxin-tagged PWL2D, shPWL2D and orPWL2D

Table 3 presents a series of physicochemical characteristics of the different PWL2 constructs that allowed us to optimize the expression and purification of proteins either fused to TRX or without this solubility tag. The extraction and purification conditions were determined taken into account the data presented in Table 3. The instability index and the aliphatic index were also evaluated [20,21]. These indices may be regarded as positive factors for an increase in protein thermostability [21]. The GRAVY (Grand Average of Hydropathy) value for a peptide or protein is calculated as "the

Table 3
Prediction of the physicochemical parameters to the proteins.

Predicted physicochemical parameters	PWL2	TRX-PWL2	PWL2D	TRX-PWL2D	*sh*PWL2D	TRX-*sh*PWL2D	*or*PWL2D	TRX-*or*PWL2D
Residue number	145	307	145	307	119	281	44	206
Molecular weight	16.18	33.34	16.16	33.33	13.38	30.57	4.92	22.11
pI	5.76	5.87	5.94	5.95	5.76	5.86	5.34	5.76
Cysteine number	2	4	1	3	1	3	0	2
Instability index	36.36	27.36	37.25	27.79	34.78	26.30	27.28	21.02
Aliphatic index	36.28	59.45	36.28	59.45	24.54	56.62	8.86	64.95
Hydropathy	−1.110	−0.677	−1.133	−0.688	−1.548	−0.822	−1.907	−0.635

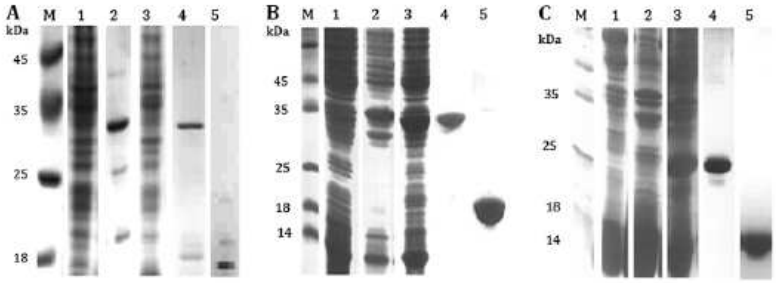


Fig. 2. SDS–PAGE expression, purification and cleavage of the TRX tag from thioredoxin-tagged proteins (A) PWL2D, (B) *sh*PWL2D and (C) *or*PWL2D. M-LWM Marker (*GE, Pharmacia*). (1) Soluble fraction without induction. (2) Insoluble fraction after induction of 4 h with 5.6 mM lactose. (3) Soluble fraction after induction of 4 h with 5.6 mM lactose. (4) Purification by nickel affinity chromatography after elution with 200 mM imidazole (A) or 100 mM imidazole (B and C). (5) Purification by nickel affinity after cleavage of the TRX tag with 0.02 U thrombin/10 μg protein chromatography (A and B) or by anion-exchange chromatography after cleavage of the TRX tag with 0.5 U factor Xa/10 μg protein (C).

sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. This is very important in protein structure; hydrophobic amino acids tend to be internal (with regard to the 3D shape of the protein) while hydrophilic amino acids are more commonly found towards the protein surface" [22]. Data in Table 3 show that the TRX solubility tag might improve the instability and aliphatic indices.

The expression of TRX-tagged proteins in *E. coli* was monitored by SDS–PAGE. The results showed that optimal expression for TRX-PWL2D was obtained using *E. coli* BL21(DE3) cells grown 4 h at 37 °C after induction with 5.6 mM lactose (Fig. 2).

While some of the recombinant PWL2D expressed in this manner was soluble, a large fraction of the protein was present in the insoluble cell pellet (Fig. 2). To improve the solubility of TRX-PWL2D, we tested several different strains such as BL21(DE3)-pLysS, BL21(DE3)Rosetta, BL21(DE3)GroE and BL21(DE3)slyD⁻. We also tested a reduced concentration for the expression inductor (0.56 mM lactose) and different incubation temperatures (22 °C and 28 °C). In all these cases, there was no improvement in the solubility of TRX-PWL2D (data not shown). The expression of *PWL2D* using different expression vectors (pET28a(+), pET29a(+), and pET-SUMO/Pro) was also performed but again no improvement in protein solubility was detected (data not shown).

Our results indicated that optimal expression for the truncated proteins TRX-*sh*PWL2D and TRX-*or*PWL2D was also obtained using BL21(DE3) cells grown for 4 h at 37 °C after induction with 5.6 mM lactose (Fig. 2). However, unlike the TRX-PWL2D expression, large fractions of the proteins TRX-*sh*PWL2D and TRX-*or*PWL2D were found in the soluble fraction, despite some of the recombinant proteins expressed in this manner were insoluble (Fig. 2).

The soluble fractions obtained from the proteins extraction steps under optimal expression conditions were then loaded into the IMAC column for purification. TRX-PWL2D was mainly eluted at 200 mM imidazole concentration, while TRX-*sh*PWL2D and TRX-*or*PWL2D were mainly eluted using 100 mM imidazole (Fig. 2). The TRX-PWL2D, TRX-*sh*PWL2D and TRX-*or*PWL2D protein yield was approximately 0.8, 1.7 and 2.0 mg/g wet cell pellet, respectively. Therefore, it seems the TRX fusion tag was less effective at improving solubility than the removal of the signal peptide.

To cleave the TRX tag, the purified TRX-PWL2D and TRX-*sh*PWL2D were dialyzed against cleavage buffer and cleaved using the recombinant protease thrombin. The purified TRX-*or*PWL2D was dialyzed against cleavage buffer and cleaved using recombinant protease factor Xa. The cleavage of the fused tag was made. Each 10 μg of TRX–PWL2D or TRX–*sh*PWL2D was cleaved over a period of 4 h using 0.02 U thrombin in 5 μL of cleavage buffer, and each 10 μg of TRX–*or*PWL2D was cleaved overnight using 0.5 U factor Xa in 5 μL of cleavage buffer. Mass spectrometry confirmed that thrombin cleaved TRX-PWL2D correctly. The resulting cleaved samples were loaded again on Ni–NTA resin for final purification.

In parallel, different proteases were also tested (thermolysin, subtilisin, trypsin, chymotrypsin, papain) for different lengths of time (15 min, 30 min, 60 min, 2 h and 4 h) to detect a stable core

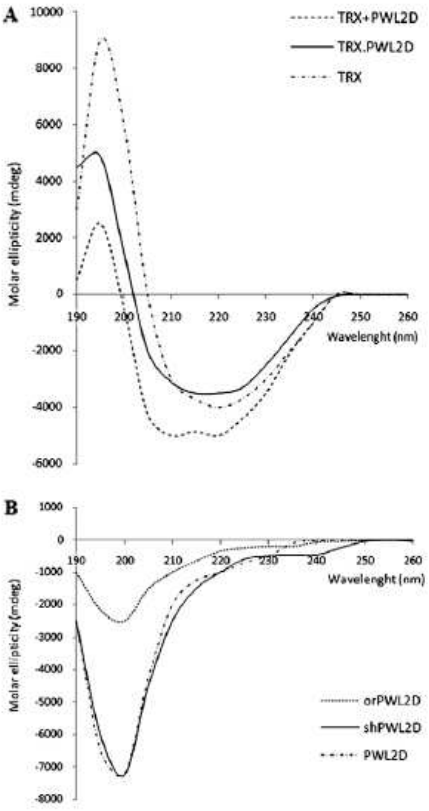


Fig. 3. Circular dichroism (CD) spectra of (A) TRX-fusioned PWL2D (TRX.PWL2D), TRX and PWL2D after cleavage (TRX + PWL2D), and TRX alone; (B) After the cleavage of the TRX tag: PWL2D, *sh*PWL2D and *or*PWL2D, all unstructured.

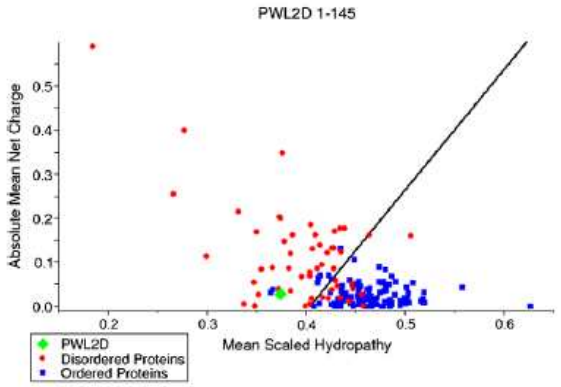


Fig. 4. Prediction of disordered regions in the PWL2D protein by use of the software predictor of naturally disordered regions (PONDR). The green point shows the localization of the PWL2D protein among the disordered proteins.

in the recombinant protein, but no satisfactory results were obtained (data not shown). This result confirmed the prediction of the bioinformatics tool *Peptide Cutter* (http://us.expasy.org).

Circular dichroism and NMR studies

After the separation of the tag and the protein purification, circular dichroism (CD) spectroscopy was used to monitor proper folding of PWL2D, *sh*PWL2D and *or*PWL2D. The CD spectra for TRX-PWL2D, TRX-*sh*PWL2D and TRX-*or*PWL2D constructs indicate well-folded proteins with a mixture of alpha-helix and beta-sheet structural elements [17]. The CD spectra in the Fig. 3A show that the presence of TRX contributes to the structuring of PWL2D. However, after the removal of TRX tag, CD spectra of the proteins indicated mainly random coils (Fig. 3B). Far-UV CD spectra have been used to characterize unfolded polypeptide chains, which display an intensive minimum in the vicinity of 200 nm and an ellipticity

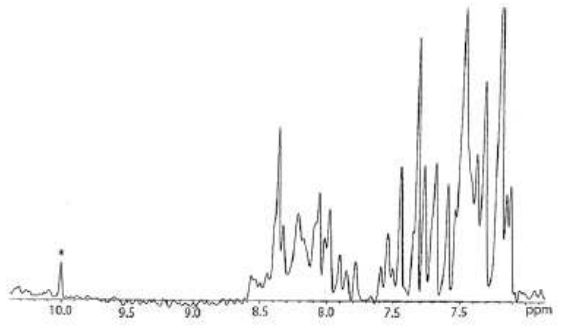


Fig. 5. 1D-NMR spectrum of the peptide *or*PWL2D (4.92 kDa): the structure presented in random coil, in agreement with the CD. Spectrum is showing the chemical shifts between 11.0 and 6.5 ppm. Asterisk indicates the peak to the tryptophan exists in the protein sequence.

close to zero in the vicinity of 222 nm [28]. The spectra of PWL2D (Fig. 3B) show an intensive minimum at around 198 nm ($[\theta] = -7500$) and an ellipticity close to zero around 222 nm ($[\theta] = -1000$), like many natively unfolded proteins [28]. Furthermore, when we plotted Absolute Mean Net Charge versus Mean Scaled Hydropathy [14–16] the initial prediction of natural disordered regions (Table 2) pointed to PWL2D being a naturally disordered protein (Fig. 4). Finally, 1D-NMR studies using *or*PWL2D confirmed that this sequence was unstructured (Fig. 5). Therefore, these results seem to confirm what was predicted for the structure of these protein sequences (Table 4) [14–16,23–25]. All together, our results seem to indicate that PWL2D, and consequently *sh*PWL2D and *or*PWL2D, are naturally unstructured proteins.

Also significant is the fact that PWL2D has two residues which differ from PWL2 (Fig. 1). It is known that the residue $Aspartate_{90}$ is crucial to avirulence functionality [7]. We assessed whether the change at residue 90 (from $Aspartate_{90}$ to $Asparagine_{90}$) could lead to alterations in possible post-translational modifications. Using the site prediction programs NetPhos 2.0 Server and YinOYang 1.2 [26,27], the possibility for phosphorylation of serine 92 of PWL2D decreased with the change at residue 90 (from $Aspartate_{90}$ to $Asparagine_{90}$). Also, it is predicted a possibility of 53% for the emergence of a glycosylation site at the modified residue 90. Moreover, the alpha-helix before modification at residue 90 was predicted be formed by five residues (residue 89–residue 93), and after modification at residue 90, the alpha-helix was predicted be formed by only two residues (residue 89 and residue 90). Moreover, the $Serine_{142}$, instead of $Cysteine_{142}$, prevents the formation of disulfide bonds (Fig. 1 and Table 3) and increases the disorder in the final region of PWL2D (Table 2). Despite not sufficient to explain the lack of an ordered structure, all these modifications can cause alterations in structural conformation that may explain the loss of function.

The plot showed in Fig. 4 indicates that PWL2D is located among naturally unstructured proteins or at least has in its structure a significant part of disordered regions (DRs). These regions are characterized by the lack of a fixed tertiary structure; in essence they are partially or fully unfolded. These disordered regions have been shown to be involved in a range of functions, including DNA recognition, modulation of specificity/affinity of protein binding, molecular threading, activation by cleavage and control of protein lifetime [14–16].

The functional importance of being disordered has been intensively analyzed. Many proteins have been classified as "intrinsically unstructured". The functional role of these proteins in crucial areas such as transcriptional regulation, translation and cellular signal transduction has only recently been recognized, as a consequence of the use of new paradigms in biochemical methodology [10]. Molecular recognition is important to the infection process of the disease caused by *M. grisea* and to host–pathogen interactions mediated by effector proteins [29–31]. Increased intrinsic plasticity represents an important prerequisite for effective molecular recognition. The majority of intrinsically disordered proteins undergo disorder-to-order transitions upon binding to their partners [28,32,33].

Because the DRs lack a defined 3-D structure in their native states, it has been suggested that intrinsic disorder and inherent flexibility in these regions is probably essential for their function.

Table 4
Prediction of the percentage of structural elements and random coil.

Predicted structure characteristic	PWL2	TRX-PWL2	PWL2D	TRX-PWL2D	*sh*PWL2D	TRX-*sh*PWL2D	*or*PWL2D	TRX-*or*PWL2D
Structural elements (%)	46.21	57.33	45.52	57.00	34.21	54.35	22.73	52.43
Random coil (%)	53.79	42.67	54.48	43.00	65.79	45.65	77.27	47.57

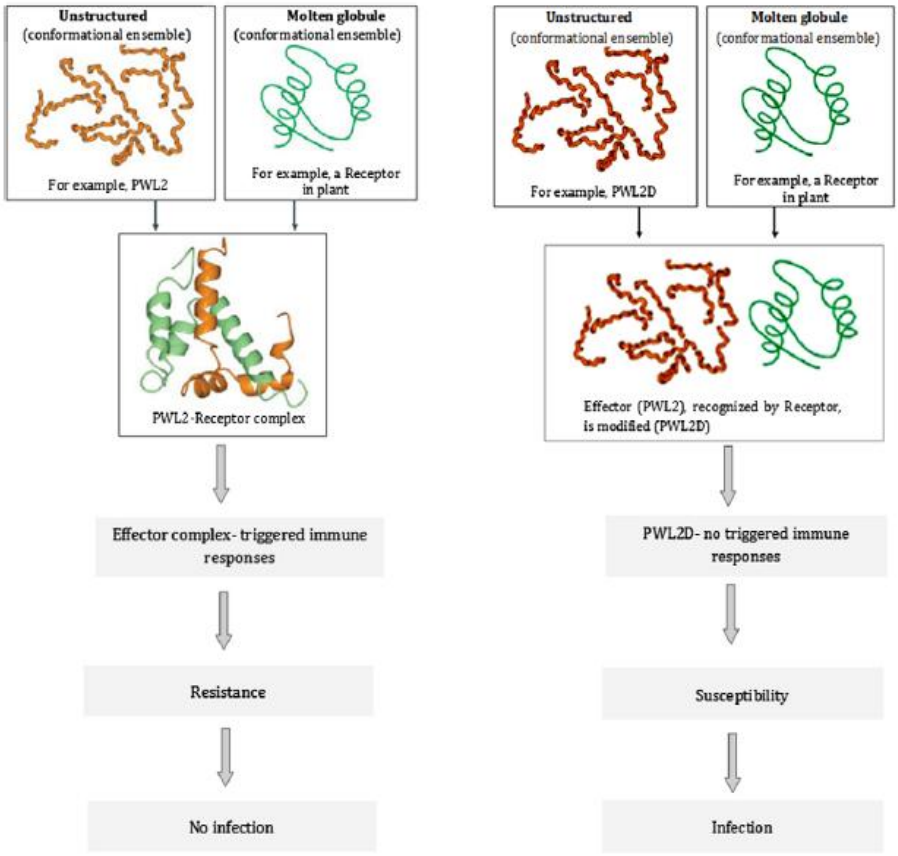


Fig. 6. Model proposed for the action of an intrinsically disordered effector protein. *Left*: PWL2 made the disorder-to-order transitions and bound to its partners in the host plant. The effector complex triggered an immune response, and no infection occurred there (resistance of the host). *Right*: Because of a modification at residue 90, critical to gene function, PWL2D made no disorder-to-order transitions, and binding to its partners in the host plant did not happen. Also, there was no triggered immune response, and infection occurred there (susceptibility of the host). The representative protein drawings are from reference [10] with permission from the authors.

Flexibility can be necessary to allow for adaptation to conformational requirements [10]. This quality is well suited to the role of an effector protein. In fact, to be unstructured does not mean to be without function. It has been suggested that the natively unfolded protein can benefit from the advantages of a flexible structure during disorder–order transitions, in comparison with rigid proteins [33]. By taking this into account, we propose here a model of the action of an intrinsically unstructured effector protein (Fig. 6). As we already discussed, the modification in the critical residue 90 of PWL2D decreases the number of residues in the alpha-helix, which perhaps alters the structural flexibility. Therefore, this modification could prevent disorder-to-order transitions upon PWL2D binding to its partners in the host plant. Thus PWL2D, possibly an intrinsically unstructured protein, would not function as an effector protein because of modifications to important residues (Fig. 6).

Conclusions

PWL2D was expressed in *E. coli* BL21(DE3) in the soluble fraction when it was fused with a solubility tag (Thioredoxin, or TRX). However, when the TRX tag is removed using protease thrombin, PWL2D tends to aggregate. We removed the signal peptide of PWL2D (producing a peptide sequence, *sh*PWL2D) to improve solubility and obtained soluble TRX-*sh*PWL2D as well *sh*PWL2D. We also expressed an ordered region from PWL2D (producing a peptide sequence, *or*PWL2D), and obtained soluble TRX-*or*PWL2D as well *or*PWL2D.

Far-UV CD showed unstructured PWL2D, *sh*PWL2D and *or*PWL2D, and NMR also confirmed that *or*PWL2D was unstructured. Furthermore, initial prediction of natural disordered regions pointed to PWL2D among disordered proteins when we plotted Absolute Mean Net Charge versus Mean Scaled Hydropathy [14–16]. Therefore, PWL2D might be an intrinsically unstructured protein. Flexibility can be necessary to allow adaptation to conformational requirements [10], and this functionality is appropriate to the role of an effector protein.

The wide spectra of pathways that involve effector proteins have also evoked considerable interest, such as the development of drugs that can modulate their actions [34]. Results described here for PWL2D may facilitate detailed structure–function studies of this protein in the near future. PWL2D may play an important role in the avirulence of *M. grisea*, and it might be involved in a number of physiological functions. Finally, the studies of PWL2D may also increase the current knowledge on the mechanism of fungal pathogenicity and on the role of diseases caused by unstructured proteins.

Acknowledgments

The authors are grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Brazil) for financial support (Project No. 2003/08963-0) and graduate fellowships to A.M.S. and A.C.P., graduate fellowships to H.R.C.A.N.M. and M.A.S.T., and research fellowship to A.R.A.; to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, São Paulo, Brazil) for a graduate fellowship to D.R.S.S., and to Conselho Nacional de Pesquisa (CNPq) for a research fellowship to A.P. Souza. We would like to thank Dr. M. Sforzza and Dra. A.C. Neri from the National Synchrotron Light Laboratory (LNLS) for the NMR experiments; Dr. H.J. Dyson and Dr. P.E. Wright (Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA) who allowed us to use some of the drawings from your article, and the Centro Nacional de Pesquisa em Arroz e Feijão (CNPAF, Embrapa, Goiás, GO, Brazil) for the supply of the *M. grisea* strain.

References

[1] A. Sesma, A.E. Osbourn, The rice leaf blast pathogen undergoes developmental process typical of root-infecting fungi, Nature 43 (2004) 516–517.
[2] S.L. Tucker, C.R. Thornton, K. Tasker, C. Jacob, G. Giles, M. Egan, N.J. Talbot, A fungal metallothionein is required for pathogenicity of *Magnaporthe grisea*, Plant Cell 16 (2004) 1575–1588.
[3] C. Veneault-Fourrey, N.J. Talbot, Moving toward a systems biology approach to the study of fungal pathogenesis in the rice blast fungus *Magnaporthe grisea*, Adv. Appl. Microbiol. 57 (2005) 177–215.
[4] N.J. Talbot, On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*, Annu. Rev. Microbiol. 57 (2003) 177–202.
[5] S. Kang, J.A. Sweigard, B. Valent, The *PWL* host specificity gene family in the blast fungus *Magnaporthe grisea*, Mol. Plant Microbe Interact. 8 (1995) 939–948.
[6] R.A. Dean, N.J. Talbot, D.J. Ebbole, M.L. Farman, T.K. Mitchell, M.J. Orbach, M. Thon, R. Kulkarni, J.R. Xu, H. Pan, N.D. Read, Y.H. Lee, I. Carbone, D. Brown, Y.Y. Oh, N. Donofrio, J.S. Jeong, D.M. Soanes, S. Djonovic, E. Kolomiets, C. Rehmeyer, W. Li, M. Harding, S. Kim, M.H. Lebrun, H. Bohnert, S. Coughlan, J. Butler, S. Calvo, L.J. Ma, R. Nicol, S. Purcell, C. Nusbaum, J.E. Galagan, B.W. Birren, The genome sequence of the rice blast fungus *Magnaporthe grisea*, Nature 434 (2005) 980–986.
[7] J. Sweigard, A.M. Caroll, S. Kang, F.G. Chumley, B. Valent, Identification, cloning and characterization of *pwl* 2, a gene for host-species specificity in the rice blast fungus, Plant Cell 7 (1995) 1221–1233.
[8] G. Mosquera, M.C. Giraldo, C.H. Khang, S. Coughlan, B. Valent, Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1–4 as biotrophy-associated secreted proteins in rice blast disease, Plant Cell 21 (2009) 1273–1290.
[9] N.M. Donofrio, Y. Oh, R. Lundy, H. Pan, D.E. Brown, J.S. Jeong, S. Coughlan, T.K. Mitchell, R.A. Dean, Global gene expression during nitrogen starvation in the rice blast fungus, *Magnaporthe grisea*, Fungal Genet. Biol. 4 (2006) 605–617.
[10] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat. Rev. Mol. Cell. Biol. 6 (2005) 197–208.
[11] G. Lecellier, P. Silar, Rapid methods for nucleic acids extraction from Petri dish-grown mycelia, Curr. Genet. 25 (1994) 122–123.
[12] J. Sambrook, D.W. Russell, Plasmids and their usefulness in molecular cloning, in: J. Sambrook, D.W. Russel (Eds.), Molecular Cloning, a Laboratory Manual, third ed., CSHL Press, New York, 2001, pp. 132–134.
[13] A. Riffel, A. Brandelli, C.M. Bellato, G.H. Souza, M.N. Eberlin, F.C. Tavares, Purification and characterization of a keratinolytic metalloprotease from Chryseobacterium sp. kr6, J. Biotechnol. 128 (2007) 693–703.
[14] X. Li, P. Romero, M. Rani, A.K. Dunker, Z. Obradovic, Predicting protein disorder for N-, C-, and internal regions, Genome Inform. 10 (1999) 30–40.
[15] P. Romero, Z. Obradovic, X. Li, E. Garner, C. Brown, A.K. Dunker, Sequence complexity of disordered protein, Proteins: Struct. Funct. Gen. 42 (2001) 38–48.
[16] P. Romero, Z. Obradovic, X. Li, E. Garner, C. Brown, A.K. Dunker, Natively disordered proteins: functions and predictions, Appl. Bioinform. 3 (2004) 105–113.
[17] A. Krogh, B. Larsson, G. von Heijne, E.L.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[18] QIAGEN, The Qiaexpressionist: A handbook for high-level expression and purification of 6×His-tagged proteins, fifth ed., Hilden, Germany, 2003, pp. 102. Available at www.qiagen.com.
[19] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Protein Sci. 8 (1999) 1668–1674.
[20] K. Guruprasad, B.V.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Eng. 4 (1990) 155–161.
[21] A.J. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.
[22] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[23] N. Sapay, Y. Guermeur, G. Deleage, Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier, BMC Bioinform. 7 (2006) 255.
[24] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[25] A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences, Science 252 (1991) 1162–1164.
[26] N. Blom, S. Gammeltoft, S. Brunak, Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362.
[27] R. Gupta, S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function, Pacific Symp. Biocomp. 7 (2002) 310–322.
[28] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci. 11 (2002) 739–756.
[29] Q.H. Chen, Y.C. Wang, A.N. Li, Z.G. Zhang, X.B. Zheng, Molecular mapping of two cultivar-specific avirulence genes in the rice blast fungus *Magnaporthe grisea*, Mol. Genet. Genomics 277 (2007) 139–148.
[30] B.J. Lugtenberg, T.F. Chin-A-Woeng, G.V. Bloemberg, Microbe–plant interactions: principles and mechanisms, Antonie Van Leeuwenhoek 81 (2002) 373–383.
[31] R. Kahmann, C. Basse, Fungal gene expression during pathogenesis-related development and host plant colonization, Curr. Opin. Microbiol. 4 (2001) 374–380.
[32] A.K. Dunker, C.J. Oldfield, J. Meng, P. Romero, J.Y. Yang, J.W. Chen, V. Vacic, Z. Obradovic, V.N. Uversky, The unfoldomics decade: an update on intrinsically disordered proteins, BMC Genomics 9 (suppl. 2) (2008) S1.
[33] J. Gsponer, M.M. Babu, The rules of disorder or why disorder rules, Prog. Biophys. Mol. Biol. 99 (2009) 94–103.
[34] A.F. Bent, D. Mackey, Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions, Annu. Rev. Phytopathol. 45 (2007) 399–436.

Please cite this article in press as: D.R.S. Schneider et al., Overexpression and purification of PWL2D, a mutant of the effector protein PWL2 from *Magnaporthe grisea*, Protein Expr. Purif. (2010), doi:10.1016/j.pep.2010.04.020

4.2. INFORMAÇÕES ADICIONAIS SOBRE A FAMÍLIA PWL

4.2.1. Genes e proteínas íntegras da família PWL: clonagem, expressão e purificação

Empregando o DNA genômico como molde, os genes de interesse foram inicialmente amplificados com o uso de *primers* flanqueadores ao gene e posteriormente clonados com o uso de *primers* específicos para os vetores de clonagem. Seguiu-se com o seqüenciamento a fim de se comparar as seqüências genômicas das linhagens do laboratório com aquelas presentes no GenBank.

O fragmento correspondente ao gene *PWL1* foi amplificado apresentando-se em uma banda eletroforética correspondente a 2,5 vezes o tamanho estimado no banco de dados. O genoma de *M. grisea* pode conter arranjos cromossômicos instáveis, e genes que pertencem às famílias gênicas como *PWL* ou que estão em posição telomérica (como o gene *PWL1*) podem ter a sua amplificação dificultada (McPherson *et al.,* 1995; Valent; Chumley, 1991), o que poderia, a princípio, justificar o resultado obtido.

Quanto ao gene *PWL2*, o seqüenciamento dos fragmentos amplificados a partir do DNA genômico das linhagens A e E mostrou uma seqüência de nucleotídeos, nestas linhagens, idêntica àquela do banco de dados (Figura 5).

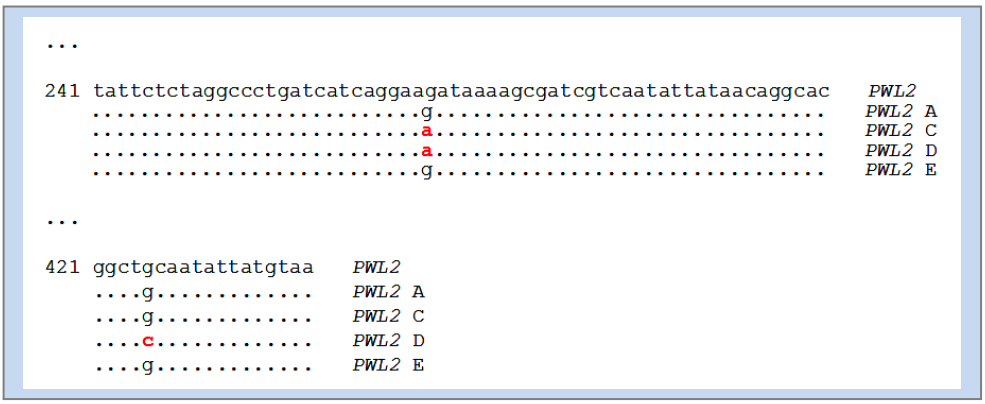


Figura 5. Alinhamento das seqüências dos genes *PWL2* (GenBanK: **U26313.1**), *PWL2A*, *PWL2C*, *PWL2D* e *PWL2E*.

A linhagem C apresentou uma diferença na seqüência de nucleotídeos que conduziu a uma variação na seqüência de proteína, a saber, asparagina no lugar de aspartato na posição 90 da proteína (Figura 6).

```
CLUSTALW 2.0.12 multiple sequence alignment


PWL2/A/E        MKCNNIILPFALVFFSTTVTAGGGWTNKQFYNDKGEREGSISIRKGSEGDFNYGPSYPGG 60
PWL2C           MKCNNIILPFALVFFSTTVTAGGGWTNKQFYNDKGEREGSISIRKGSEGDFNYGPSYPGG 60
PWL2D           MKCNNIILPFALVFFSTTVTAGGGWTNKQFYNDKGEREGSISIRKGSEGDFNYGPSYPGG 60
                ************************************************************
                                          90
                                          |
PWL2/A/E        PDRMVRVHENNGNIRGMPPGYSLGPDHQEDKSDRQYYNRHGYHVGDGPAEYGNHGGGQWG 120
PWL2C           PDRMVRVHENNGNIRGMPPGYSLGPDHQENKSDRQYYNRHGYHVGDGPAEYGNHGGGQWG 120
PWL2D           PDRMVRVHENNGNIRGMPPGYSLGPDHQENKSDRQYYNRHGYHVGDGPAEYGNHGGGQWG 120
                ****************************:***************************** 
                                    142
                                     |
PWL2/A/E        DGYYGPPGEFTHEHREQREEGCNIM 145
PWL2C           DGYYGPPGEFTHEHREQREEGCNIM 145
PWL2D           DGYYGPPGEFTHEHREQREEGSNIM 145
                ********************.***
```

Figura 6. Alinhamento das seqüências das proteínas PWL2 (GenBanK: **AAA91019.1**), PWL2A, PWL2C, PWL2D e PWL2E utilizando o software ClustalW2.0 (www.ebi.ac.uk/Tools/clustalw2/index.html). PWL2C apresenta asparagina no lugar de aspartato na posição 90 da proteína, enquanto PWL2D apresenta asparagina no lugar de aspartato na posição 90 da proteína e serina no lugar de cisteína na posição 142 da proteína, em relação à PWL2.

O fragmento amplificado correspondente ao gene *PWL2* proveniente da linhagem D apresentou duas diferenças na seqüência de nucleotídeos em relação à seqüência

disponível no GenBank (Figura 5); tais diferenças na linhagem D conduzem a duas alterações na seqüência de aminoácidos (asparagina no lugar de aspartato, na posição 90 da proteína, e serina no lugar de cisteína, na posição 142 da proteína).

Destaca-se a diferença na posição 90 da proteína: já foi relatado que esta diferença é suficiente para cessar a função de avirulência do gene *PWL*2, conduzindo à susceptibilidade do hospedeiro (Sweigard *et al.,* 1995). Os alelos do gene *PWL2* foram denominados *PWL2A*, *PWL2C, PWL2D e PWL2E* (ou seja, os alelos do gene PWL2 nas linhagens A, C, D e E, respectivamente). O manuscrito 1 avalia as implicações destas alterações dos resíduos quanto às possíveis mudanças nas interações do produto do gene de avirulência *PWL2D* com outras proteínas bem como o reconhecimento do patógeno pelo hospedeiro em relação a estas mudanças.

O fragmento amplificado correspondente ao gene *PWL3* apresentou diferenças na seqüência de nucleotídeos da linhagem em relação à seqüência disponível no banco de dados nas linhagens utilizadas: uma deleção (T) na posição 301, uma transição (G > A) na posição 363 e uma transversão (A > C) na posição 391. A proteína PWL3 foi expressa na fração insolúvel, tanto na linhagem de *E. coli* BL21(DE3) como na linhagem BL21(DE3)ptGroE. O gene *PWL3* e sua proteína relacionada integraram-se a outro projeto de pesquisa.

O gene *PWL4* não foi amplificado a despeito dos *primers* utilizados, fossem eles específicos para a seqüência gênica ou fossem flanqueadores a ela. Devido ao gene

PWL4 ser um alelo do gene *PWL3*, o qual foi amplificado a partir do genoma das linhagens utilizadas, a existência de um dos genes no genoma haplóide de *M. grisea* implica na exclusão do outro (Kang *et al*., 1995), podendo justificar o resultado obtido.

4.2.2. Regiões intragênicas no gene *PWL2*

Utilizando o software de predição TMHMM Server v. 2.0 (disponível em www.cbs.dtu.dk/services/TMHMM-2.0/) para a avaliação da existência de regiões transmembranas, puderam-se identificar as seqüências referentes ao peptídeo-sinal nas proteínas traduzidas (Figura 7) dentro da seqüência dos genes de interesse (*PWL2A, PWL2C, PWL2D e PWL2E*).

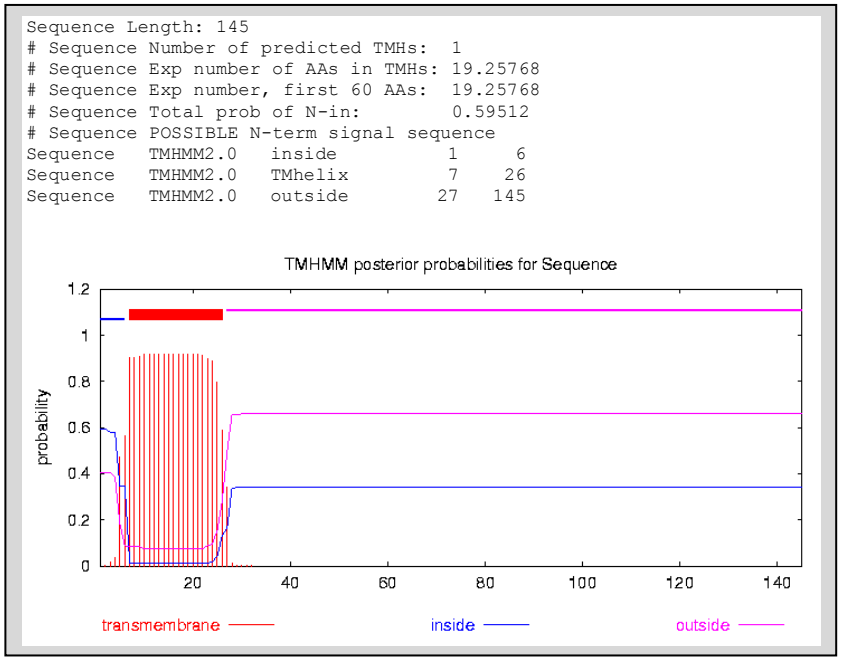


Figura 7. Predição de peptídeo sinal na proteína PWL2, por intermédio do *software* TMHMM2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/)

Empregando o software *Predictors of Natural Disordered Regions* - PONDR (Li *et al.*, 1999; Romero *et al.*, 2001; Romero *et al.*, 2004) para a predição de regiões desordenadas dentro de seqüências protéicas, identificaram-se também tanto regiões ordenadas como regiões desordenadas nas proteínas relacionadas aos genes em questão (Figura 8).

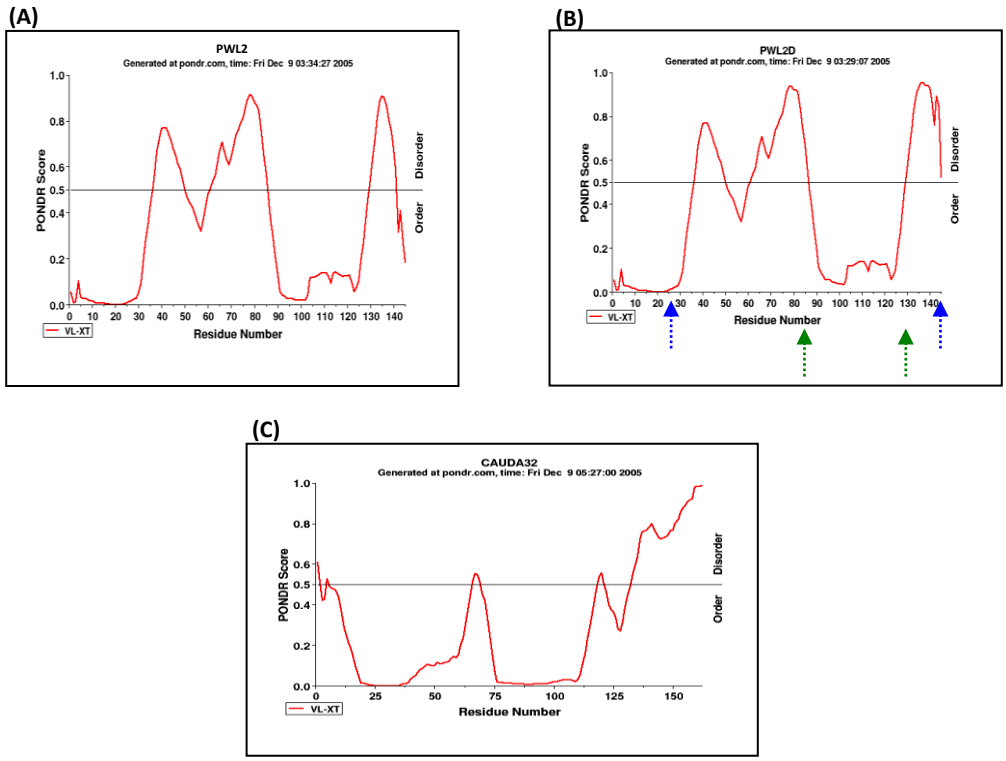


Figura 8. Predição das regiões ordenadas e desordenadas nas proteínas **(A)** PWL2; **(B)** PWL2D; **(C)** TRX tag (proveniente do pET32/XaLIC). Setas verdes: *or*PWL2D (resíduos 86-129). Setas azuis: *sh*PWL2D (resíduos 26-145)

Sabe-se que a natureza hidrofóbica de peptídeos-sinal (Figura 9) interfere na translocação de proteínas para a membrana citoplasmática, podendo resultar na formação de corpos de inclusão, inclusive no periplasma (QIAGEN Manual, 2003). Assim, a retirada do peptídeo-sinal tem sido recomendada em procedimentos de

expressão de proteínas recombinantes visando melhorar a obtenção de proteína solúvel (QIAGEN Manual, 2003). Adotou-se, deste modo, a estratégia de clonagem dos genes de interesse sem a seqüência referente ao peptídeo-sinal visando uma possível melhora na aquisição da fração protéica solúvel expressa.

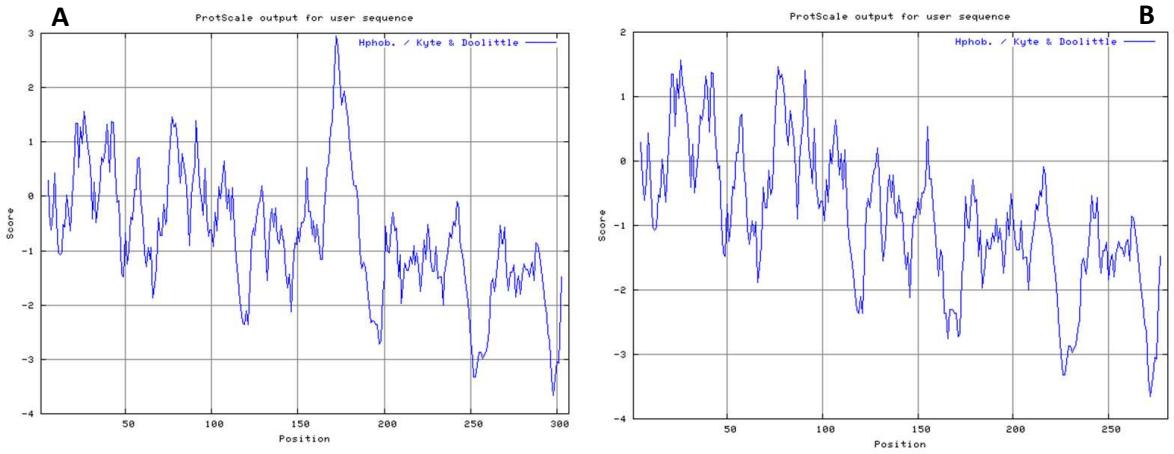


Figure 9. Predição da hidrofobicidade de **(A)** TRX-PWL2D: notar que o peptídeo-sinal (resíduos 163-188) contribui para aumentar a hidrofobicidade**. (B)** TRX-*sh*PWL2D. TRX tag (resíduos 1-162). Software disponível em www.expasy.ch/tools/protscale.html

Devido à predição do peptídeo-sinal, à predição das regiões desordenadas e à predição de uma região ordenada existentes nas proteínas traduzidas a partir dos genes *PWL*2A, *PWL*2C e *PWL*2D, entendemos que a amplificação de regiões intragênicas poderia otimizar a produção e a solubilidade das proteínas de interesse. Denominou-se *sh*PWL2_ (*short sequence of* PWL2_) à seqüencia protéica sem o peptídeo-sinal, e denominou-se *or*PWL2_ (*ordered sequence in* PWL2_) à seqüência ordenada e contendo o resíduo 90 (considerado crítico para a função de avirulência do gene, segundo Sweigard *et al.*, 1995), conforme a Figura 1 no manuscrito 1.

4.2.3. Clonagem e expressão dos alelos de *PWL2* e de suas regiões intragênicas

Distintos sistemas de clonagem e expressão foram utilizados para os alelos de *PWL2* e suas regiões intragênicas. Os genes íntegros *PWL2*A, *PWL2*C e *PWL2*D foram clonados em pET32Xa/LIC, em pETSUMO/Pro (com auxílio das endonucleases de restrição *Bam*HI e *Bsma*I), e em pET29a(+) (com auxílio das endonucleases de restrição *Nde*I *e Xho*I). As regiões intragênicas *orPWL2*A*, shPWL2*A, *orPWL2*C*, shPWL2*C, *orPWL2*D e sh*PWL2*D foram clonados em pET32Xa/LIC e em pET29a(+) (com auxílio das endonucleases de restrição *Nde*I *e Xho*I).

Juntamente com os vetores listados anteriormente no item 5.2 puderam-se compor os sistemas de expressão utilizados para as os genes íntegros [nas linhagens BL21(DE3), BL21(DE3)ptGroE, BL21(DE3)AD494, BL21(DE3)pRIL, BL21(DE3)pLysS, BL21 (DE3)Rosetta e BL21 (DE3)SlyD$^-$], e para as regiões intragênicas clonadas [nas linhagens BL21 (DE3) e BL21 (DE3)ptGroE].

Visto que a linhagem D (Py 416-L1-3) contem o gene *PWL2* com duas alterações na seqüência gênica (denominado *PWL2D,* significando alelo do gene *PWL2* na linhagem D) esta linhagem foi utilizada para a avaliação mais detalhada da proteína íntegra expressa e dos peptídeos relativos às suas regiões intragênicas (*or*PWL2D e *sh*PWL2D), conforme o manuscrito 1. O maior rendimento em termos de solubilidade e de nível de expressão foi alcançado com a combinação do sistema BL21(DE3), o vetor pET32-Xa/LIC e as regiões intragênicas *sh*PWL2 e *or*PWL2 (Figura 2 do

manuscrito 1). Este resultado reafirmou que, neste caso, a retirada do peptídeo-sinal pode melhorar a solubilidade da proteína-alvo (QIAGEN Manual, 2003). A expressão com o vetor pET29a(+)não produziu proteína solúvel.

4.2.4. Técnicas espectroscópicas

4.2.4.1. Espectrometria de massas

A espectrometria de massas (MS) por MALDI-TOF foi realizada inicialmente com o intuito de se certificar da fidelidade da clivagem feita pelas proteases e das massas moleculares das proteínas e peptídeos de interesse resultantes após a clivagem da cauda fusionada. A técnica de MS utilizada, descrita no manuscrito 1, confirmou que a trombina clivou PWL2D fusionada à cauda do pET32-Xa/LIC no sítio esperado, visto que o espectro de peptídeos tripticos gerados pela ação da tripsina sobre a proteína PWL2D clivada com trombina correspondeu ao previsto (Figura 10).

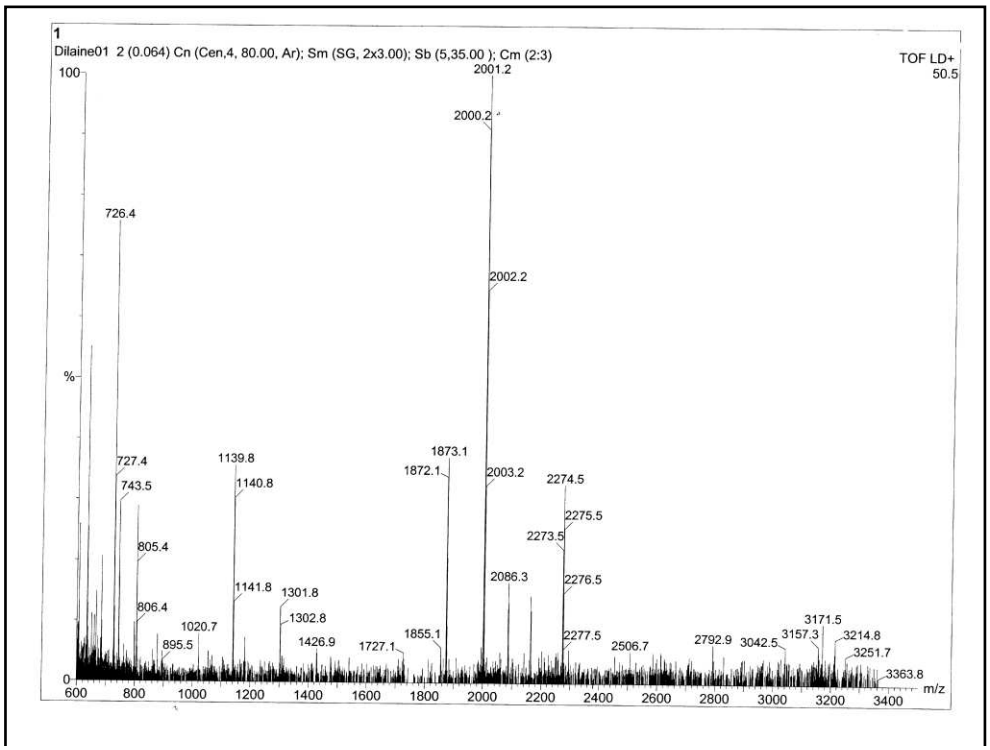


Figura 10. Espectrometria de massas. Espectro de peptídeos tripticos gerados pela ação da tripsina sobre a proteína TRX- PWL2D clivada com a proteína trombina

4.2.4.2. Dicroísmo Circular (CD)

Espectroscopia de dicroísmo circular (CD) foi usada para monitorar o enovelamento apropriado das proteínas TRX-PWL2D, TRX-*sh*PWL2D e TRX-*or*PWL2D, antes e após a clivagem da cauda TRX de solubilização. Os espectros de CD para TRX-PWL2D, PWL2D, *sh*PWL2D, *or*PWL2D e para a cauda resultante após a clivagem são mostrados e avaliados no manuscrito 1. Os espectros antes da clivagem da cauda TRX indicam proteínas bem dobradas com uma mistura de elementos estruturais alfa-hélice e beta-sheet (Krogh *et al*., 2001). No entanto, após a remoção da cauda TRX, os espectros de CD das proteínas indicaram a existência de *random coil.* A influência da estrutura da

cauda TRX na manutenção da estrutura das proteínas pode ser inicialmente avaliada pelo espectro de CD desta cauda (manuscrito 1, Figura 3), indicando que a presença de elementos estruturais, nas proteínas recombinantes fusionadas, antes da clivagem da cauda de solubilização TRX, pode ser justificada na própria estruturação da cauda, explicando, a princípio, a perda de estrutura das proteínas após a clivagem.

4.2.4.3. Espectroscopia de Ressonância Magnética Nuclear (NMR)

O espectro unidimensional das proteínas não marcadas foi coletado, estando estas em tampão fosfato de sódio 100 mM, pH6.8 a 25^0C para a análise do estado das estruturas protéicas através da dispersão dos picos. Pelos espectros gerados em NMR, confirmaram-se os resultados do dicroísmo circular: após a clivagem com trombina, *or*PWL2D apresentou-se desestruturada em *random coil*, conforme mostrado na Figura 5 do manuscrito 1.

4.3. MANUSCRITO 2

Cloning, expression, purification and partial characterization of the XYL5 xylanase from *Magnaporthe grisea*

Autores: Dilaine R.S. Schneider; Antonio M. Saraiva; Adriano R. Azzoni; Alexandre C. Pelloso; Marcelo A.S. de Toledo; Anete P. de Souza

Cloning, expression, purification and partial characterization of the XYL5 xylanase from *Magnaporthe grisea*

Dilaine R.S. Schneider[1]; Antonio M. Saraiva[1]; Adriano R. Azzoni[1]; Alexandre C. Pelloso[1]; Marcelo A.S. de Toledo[1]; Anete P. de Souza[1,2]

[1]Centro de Biologia Molecular e Engenharia Genética (CBMEG), [2]Departamento de Biologia Vegetal – Instituto de Biologia. Universidade Estadual de Campinas, SP, Brazil

Correspondence to Anete P. de Souza, Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, CEP 13083-970, Campinas, São Paulo, Brazil
Fax: +55 19 3521 1089
Tel: +55 19 3521 1132
E-mail: anete@unicamp.br

Abstract

We cloned the putative xylanase XYL5 (EC3.2.1.8) gene of *Magnaporthe grisea* and its conserved catalytic domain (XYL5/DOM). These proteins were fused to MBP and expressed in *Escherichia coli*. The soluble extract containing the fusion proteins MBP-XYL5 and MBP-XYL5/DOM showed positive activity compared to the control MBP. We report the expression of a novel xylanolytic enzyme that presents activity without any purification or excision of the solubilization tag, by the use of a heterologous system as simple as *E. coli*. These aspects are economically advantageous in terms of cost and production time for industrial processes use xylanolytic enzyme complexes. To our knowledge, this is the first time that the *M. grisea XYL5* pathogenicity gene has been successfully expressed in *E. coli* and the first time that its xylanolytic enzyme activity was demonstrated.

Keywords: Xylanase XYL5; *Magnaporthe grisea*; MBP fusion; recombinant protein; pSV282.

Abbreviations: MBP- Maltose Binding Protein;

1 Introduction

Xylan is the second most abundant renewable polysaccharide in nature after cellulose. It is one of the major polysaccharide components of secondary plant cell walls in cereals, in which it maintains cell wall integrity by attaching to cellulose, lignin, pectin and other polysaccharides (Zhou *et al*. 2008). Unlike cellulose, xylan is a complex polymer consisting of a β-1,4-linked xylopyranoside backbone substituted with side chains. The heterogeneity and complexity of xylan has resulted in an abundance of diverse xylanases that differ in their physico-chemical properties, modes of action and substrate specificities. For example, Endo-1,4-β-D-xylanases (EC 3.2.1.8) randomly cleave the xylan backbone, while β-D-xylosidases (EC 3.2.1.37) release xylose monomers from the nonreducing end of xylo-oligosaccharides and xylobiose (Berrin and Juge 2008).

Xylanolytic enzymes, from a variety of microorganisms, have attracted attention mainly because of their potential biotechnological uses in several industries, such as the food, feed, biofuel and pulp and paper industries (Jun *et al*. 2009). Today, the most effective application of xylanases is in the pre-bleaching of kraft pulp to minimize the use of harsh chemicals in the subsequent treatment stages of pulp bleaching (Beg *et al*. 2001). In the feed industry, the incorporation of xylanase into the rye-based diet of broiler chickens results in the reduction in intestinal viscosity, thereby improving both the weight gain of chicks and their feed conversion efficiency (Talbot 2003). In baking, xylanases improve the elasticity and strength of dough, increasing loaf volumes and the texture of breads (Ahmed *et al*. 2009; Polizeli *et al*. 2005). Xylanases, in combination with other enzymes, can also be used for the generation of biological fuels, such as ethanol from lignocellulosic materials; for degumming of fibers, such as flax, hemp, jute and ramie; and for the deinking of waste newspapers, which are used in the papermaking process (Beg *et al*. 2001; Jun *et al*. 2009).

Fungi, actinomycetes and bacteria are some of the most important xylanolytic enzyme producers. Many of these organisms have been shown to produce different forms of xylanases, which may have diverse physico-chemical properties, specific activities and overlapping but dissimilar specificities, increasing the efficiency and extent of hydrolysis (Ahmed *et al*. 2009). For scientific and industrial purposes, different systems have been used for cloning and expression of xylanolytic enzymes, especially from fungi. Recombinant *Escherichia coli*, yeast and fungi have been studied, and each system has advantages and disadvantages that must be evaluated in order to obtain functional recombinant enzyme (Cao*et al*. 2007; Carapito *et al*. 2009; Nevalainen *et al*. 2005; Sakaguchi *et al*. 2008). Among these systems, *E. coli* has been widely used for the cloning and heterologous expression of recombinant xylanases from fungi (Zhou *et al*. 2008; Lee *et al*. 2007). This employ is mainly due to the widespread use of simple DNA cloning techniques, the large number of commercial cloning vectors, the simplicity of overproduction of recombinant enzymes, etc. (Zhou *et al*. 2008). However, there are also important restrictions on the use of *E. coli* as a cloning host for fungal enzymes, such as the absence of post-translational modifications, which may affect the activity of the resulting recombinant xylanase (Ahmed *et al*. 2009). The lack of glycosylation is known to be an important

factor that accounts for the reduction in the affinity between the enzyme and substrate as well as the decrease in enzyme stability (Ahmed *et al.* 2009; Polizeli *et al.* 2005). Although several fungi and yeasts have been used as expression hosts for fungal xylanases, *E. coli* continues to be used as the preferred cloning host. In general, it provides a convenient way to produce significant amounts of enzyme for functional studies and structural elucidation that can guide further improvements in enzyme characteristics by point mutation and protein engineering approaches (Ahmed *et al.* 2009).

In this work, we cloned, expressed and partially characterized a putative xylanase from *Magnaporthe grisea* using *E. coli. M. grisea* is a filamentous fungus, an Ascomycete, and a pathogen of economically important cereals, such as wheat and barley, and causes the most severe disease in rice (Talbot 2003). The genome sequence of the *Magnaporthe grisea* was published recently (Dean *et al*. 2005). Different xylanases have been identified in *Magnaporthe grisea* and are possibly involved in the fungi plant infection mechanisms (Shengcheng *et al.* 2001). Currently, the amplification and cloning of three putative xylanase genes, *XYL3*, *XYL4* and *XYL5* (GenBank accession numbers **AY144348** to **144350**), have been reported (Wu *et al.* 2006). The XYL5 protein is a putative Endo-1,4-β-D-xylanase (EC 3.2.1.8) that contains a conserved domain of family 10 of Glycoside Hydrolases (GH10) (Berrin and Juge 2008). Here, we cloned, expressed in *Escherichia coli* and evaluated the xylanolytic activity of the intact XYL5 enzyme and the conserved catalytic domain of XYL5 from *M. grisea* as fusion proteins with maltose binding protein (MBP). This is the first report of the successful expression of the XYL5 xylanase in *E. coli,* a finding with biotechnological and agricultural implications.

2 Materials and Methods

2.1 Materials

The strain Py 416-L1-3 of the filamentous fungus *Magnaporthe grisea*, obtained from the National Center for Research on Beans and Rice (CNPAF - GO, Brazil), was used as the source of genomic DNA, which was used to obtain the XYL5 gene. The *Escherichia coli* DH5α [F2 lacZD M15 hsdR17 (r2 m2) gyr A36] and BL21(DE3) (*Novagen*, Madison, WI, USA) strains were used as host strains for subcloning and gene expression, respectively.

The plasmid vectors and expression strains were obtained from *Novagen* (*Novagen*, Madison, WI, USA). The plasmid vector pSV282 (from of the vector pET27-a, *Novagen*), was used for cloning, gene amplification and protein expression.

The restriction enzymes, *BamH*I, *Bpi*I and *Xho*I were obtained from *Fermentas* (MD, USA). The modification enzymes T4 DNA ligase and Taq polymerase were obtained from *Invitrogen* (Carlsbad, CA, USA).

Fragments amplified by Polymerase Chain Reaction (PCR) were purified using the QIAquick PCR Purification Kit (*Qiagen*, Hilden, Germany). Plasmid DNA was purified using the QIAprep Spin Miniprep Kit (*Qiagen*, Hilden, Germany). All chemicals were obtained from *Sigma* (St. Louis, MO, USA) unless otherwise specified.

2.2 Genomic DNA extraction and purification

Genomic DNA extraction and purification procedures from *M. grisea* were adapted using a method described previously (Lecellier and Silar 1994). Petri dishes of 9.0 cm in diameter, with solid medium PDA, were covered with a sterile, permeable cellophane disk. Seven "plugs" from plates with mycelium were placed on the disk. After 4-6 days of incubation at 28°C, the mycelium was scraped from the cellophane. The material was transferred to 1.5 ml microcentrifuge tubes and 600 μL of extraction buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl and 2% SDS) was added to each tube. The following cycle was repeated three times: the tubes were vortexed for 30 s, frozen in liquid nitrogen for 30 s and incubated in a dry incubator for 30 s (70^{o}C) until completely thawed. DNA was purified using the classical phenol-chloroform-isoamyl alcohol method. The DNA was precipitated with ethanol and sodium acetate. The DNA pellets were resuspended in 20 μL of TE buffer (Sambrook and Russell 2001).

2.3 Construction of the XYL5 and XYL5/DOM expression vectors

An alignment of XYL5 with several xylanases from different fungi is presented in Figure 1; it indicates the presence of conserved regions. To know the activity of XYL5 and compare it to the activity its conserved domain XYL5/DOM we expressed them as MBP-fusion proteins using the cloning vector pSV282. This vector, from of the vector pET27-a (*Novagen*), has the following basic structure: *His tag site* (6 histidines) – *MBP tag site* (45kDa) – *Tobacco Etch Virus* (*TEV) protease site – Bam*HI *restriction enzyme site – site to the gene cloning*.

The DNA fragment encoding the gene XYL5 was amplified from genomic DNA by polymerase chain reaction (PCR, MJ Research, Inc., USA). We initially used primer set A from Table 1, which flanks the gene sequence of interest. The fragments were purified and used as a template for PCR with specific primers for the cloning vector pSV282.

For cloning into the vector pSV282, the sequences encoding *XYL5* and *XYL5/DOM* were amplified using primer sets B and C, respectively (Table 1). The products amplified by PCR were then purified and digested

with *Xho*I and *Bpi*I. *Bpi*I is a class II restriction enzyme that was used because XYL5 and XYL5/DOM are cut by the restriction enzyme *BamH*I. Therefore, we created a recognition specific site for restriction that was similar to the site for *BamH*I to clone XYL5 and XYL5/DOM into the pSV282 vector. The pSV282 was digested with *BamH*I and *Xho*I and then ligated to the digested fragments to generate the constructs pSV282-XYL5 and pSV282-XYL5/DOM.

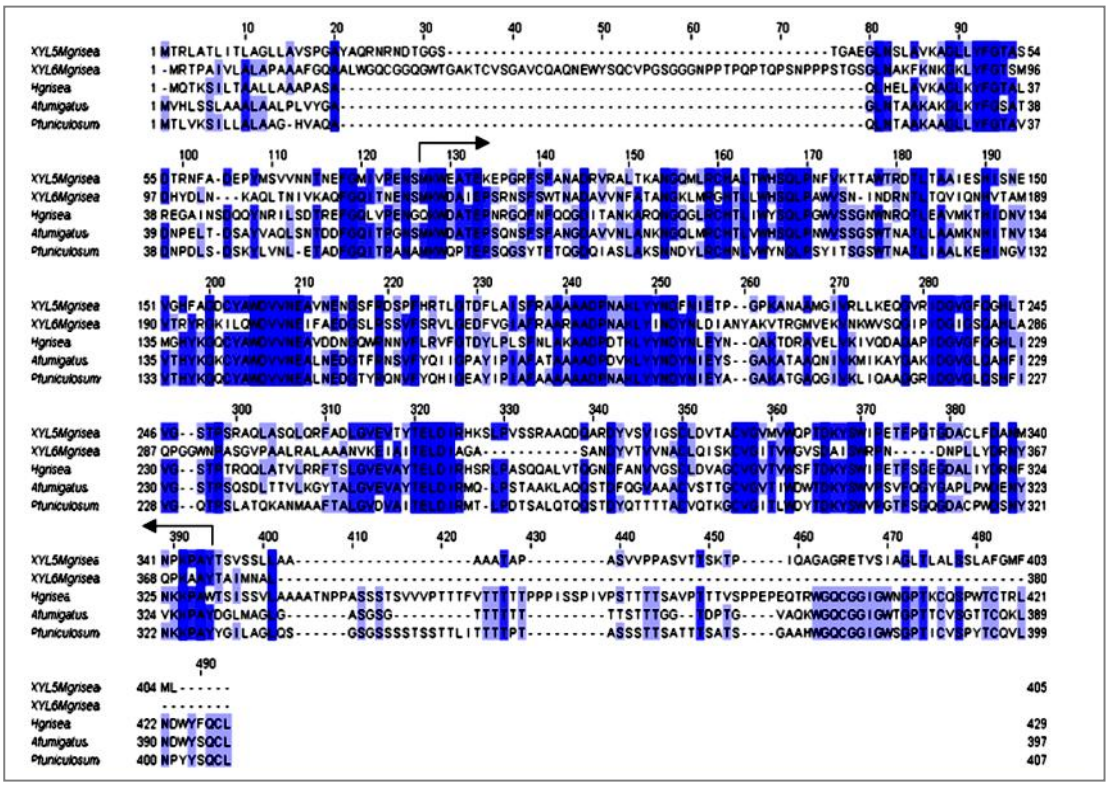


Fig. 1 - Alignment of xylanase protein sequences from selected species. Sequences are from the GeneBank, a protein sequence database. The accession numbers are **AAN60060.1** for XYL5 of *Magnaporthe grisea*, **AAM95237.1** for XYL6 of *Magnaporthe grisea,* **CAG25554.1** for *Penicillium funiculosum*, **BAA19220.1** for *Humicola grisea* and **EAL89199.1** for *Aspergillus fumigatus*. The sequence alignment was generated using Clustal W2. Arrows indicate the start and end of the XYL5/DOM region (glycosyl hydrolase domain of XYL5). The conserved and similar amino acids in the xylanases are indicated by solid and grey boxes, respectively.

After cloning into the expression vector, the constructs pSV282-XYL5 and pSV282-XYL5/DOM were introduced by chemical transformation into *E. coli* DH5α (Sambrook and Russell 2001). As a negative control for expression, the vector without inserts was also transformed into *E. coli* DH5α. All cloned fragments were verified by DNA sequencing.

Table 1
The sequences of oligonucleotide primers used to cloning and expression with the plasmid vector pSV282

Primer sets	Sequence	Gene	Cloning and expression plasmid vector
A	5'-ACCGGTCCTGACTCCCCTTGAAAC-3' 5'-GATACCCGCACGACCCCTGGATA-3'	*XYL5* and *XYL5/DOM*	-
B	5'-ATGAAGACTGGATCC*ATG*ACAAGACTAGCAACACTCA-3' *Bpi*I Recognition/restriction site; Start codon 5'-AGAACCTCGAG*TCA*GAGCATAAACATGCCA-3' *Xho*I restriction site; Stop codon	*XYL5*	pSV282
C	5'-ATGAAGACTGGATCC*ATG*AAGTGGGAGGCCACC-3' *Bpi*I Recognition/restriction site; Start codon 5'-ACTAGTCTCGAG*TCA*GTAGGCCGGCTTGGGAT-3' *Xho*I restriction site; Stop codon	*XYL5/DOM*	pSV282

2.4 Expression of the XYL5 and XYL5/DOM fusion proteins in *E. coli*

XYL5 and XYL5/DOM were expressed in *E. coli*. The constructs pSV282-XYL5 and pSV282-XYL5/DOM were introduced into *E. coli* BL21(DE3) by chemical transformation. The vectors without inserts were also introduced into *E. coli* BL21(DE3) as controls for expression. After incubation and selection of positive clones, cultures were grown overnight at 37°C and 300 rpm and then diluted 1:100 in fresh TB medium (Sambrook and Russell 2001) plus 0.2% glucose and supplemented with 30 μg/ml kanamycin. Cultures were incubated at 37°C and 300 rpm, until the optical density reached 0.8, and were then induced by 2, 4 and 6 hours with IPTG (1 mM).

2.5 Extraction and purification of MBP–XYL5 and MBP–XYL5/DOM from *E. coli*

For each bacterial pellet from 1 L of induced bacterial culture, 4 ml of extraction buffer (50 mM Tris-HCl buffer containing 300 mM NaCl at pH 7.5) was added to the suspension. After sonication, the suspensions were centrifuged twice (23,000 g, 20 minutes, 4°C). The precipitate (insoluble fraction) was stored at -20°C, and the supernatant fraction (soluble fraction) was immediately submitted to the purification procedure as follows. In the vector pSV282, the gene of the solubilization protein MBP is fused in-frame with an N-terminal hexahistidine tag for purification of the recombinant proteins via immobilized metal affinity

chromatography (IMAC System, QIAexpressionist, Qiagen, Germany). Therefore, the proteins expressed in the vector pSV282 can be purified either by nickel or amylose affinity chromatography.

The soluble fraction obtained was applied to a column containing Ni-NTA resin. The resin was equilibrated with extraction buffer, and elutions with increasing imidazole concentrations in the extraction buffer (5, 10, 50, 100, 200, 250 and 500 mM) were collected. For purification by amylose affinity chromatography, the resin was equilibrated with extraction buffer. The elution buffer was composed of extraction buffer supplemented with different concentrations of maltose (0.01, 0.05, 10 and 50 mM). The eluted fractions containing the protein of interest were submitted to five sequential dialyses. Each dialysis was performed in extraction buffer, in a volume equal to 1000 times the sample volume, at 4°C for 6 hours. The dialyzed material was quantified by spectrophotometry.

2.6 Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) of the XYL5 and XYL5/DOM fusion proteins

The soluble and insoluble fractions (inclusion bodies) of the fusion proteins XYL5 and XYL5/DOM were analyzed by SDS-PAGE. *E. coli* that expressed the MBP-fusion proteins were disrupted by sonication. The pellet containing the inclusion bodies and the supernatant with soluble protein were collected by centrifugation for 30 min at 23,000 g. Samples of these fractions were mixed with gel loading buffer (0.5 M Tris-HCl, pH 6.8, 2% glycerol, 10% SDS and 0.1% bromophenol blue) and analyzed on 12% (w/v) SDS-PAGE gels. The proteins were then stained with Coomassie Brilliant Blue R-250.

2.7 Circular dichroism spectroscopy

Circular dichroism (CD) measurements were performed using a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan). After purification, MBP-XYL5 and MBP-XYL5/DOM were dialyzed overnight against the CD buffer (5 mM sodium phosphate, pH 7.5) and diluted to 0.5 mg/ml or 0.25 mg/ml before data collection. Data points were collected at 20°C using a 1 mm path length cuvette. The spectrum is presented as an average of 20 scans recorded from 190 to 260 nm at a rate of 20 nm/min.

2.8 Biological activity of recombinant MBP-XYL5 and MBP-XYL5/DOM

Two methods were used to assess the biological activity of recombinant MBP-XYL5 and MBP-XYL5/DOM. For both, the non-purified soluble extracts containing MBP-XYL5 or MBP-XYL5/DOM and the dialyzed purified fractions from the amylose affinity chromatography were tested.

With the first method (Teather and Wood 1982; Kasana *et al*. 2008), Petri plates containing 1% agarose and 1% (w/v) birchwood xylan were prepared. Wells about 3.0 mm in diameter were made after the medium had

solidified on the plates. The plates were used to inoculate 20 μL of either of the dialyzed purified samples or non-purified soluble extracts containing MBP-XYL5 or MBP-XYL5/DOM. The inoculations were done with the protein from 3 mg of the bacterial pellet. The elution fractions from the purification containing the protein were dialyzed against Tris buffer (50 mM Tris-HCl, pH 7.0) overnight, and the dialyzed material was quantified by spectrophotometry. The procedure was also performed for the negative control using the expression vector (pSV282) without insert. Inoculated plates were then incubated for 22 h at 25°C. Approximately 10 ml of 0.5% Congo Red was carefully deposited on the medium and incubated for 15 minutes at room temperature. The solution of Congo Red was removed, and 10 ml of saline solution (1 M NaCl) was added. After 30 minutes, the saline was removed and a solution of 50 g/L acetic acid was added to allow the formation of halos, which indicate the degradation of xylan.

With the other method used (Miller, 1959; Nair *et al*. 2008), the specificities of MBP-XYL5 and MBP-XYL5/DOM against birchwood xylan, carboxymethyl cellulose (CMC), microcrystalline cellulose (Avicel) and starch were tested. Substrates (1% w/v) were prepared in buffer (50 mM Tris HCl, pH 7.5 and 300 mM NaCl). A standard amount (3 mg) of purified MBP-XYL5 or MBP-XYL5/DOM or non-purified soluble extracts from 3 mg of bacterial pellet containing the enzymes were used to test biological activity. Xylanase and cellulase activities were measured using 665 μL of substrate solutions and 335 μL of samples at 50°C for 20 min. The liberation of reducing sugars was estimated by the dinitrosalicylic acid (DNS) method (Miller 1959; Nair *et al*. 2008) using xylose as a standard. Reducing sugars were determined by measuring the absorption at 540 nm relative to a D-xylose standard curve. One unit of enzyme activity was defined as the quantity of enzyme required to generate 1 μmol of xylose equivalents per minute at 50°C, and the specific activity was defined as units per mg protein. The results presented were calculated as the average of three independent measurements.

3 Results

3.1 Design and construction of the fusion protein expression vectors

We used the expression strain *E. coli* BL21(DE3) harboring vector pSV282, which provide the possibility of increased protein solubility by fusion to MBP. Two types of expression vectors were constructed for the expression of XYL5 and XYL5/DOM fused to MBP. The MBP protein has been extensively used to improve the solubility of recombinant proteins (Kapust and Waugh 1999). However, the level of expression does not necessarily correlate with the solubility of the proteins of interest (Kapust and Waugh 1999).

For the expression of the XYL5 and XYL5/DOM fusion proteins, were inserted into the pSV282 vector, which contained the coding sequence for MBP, to generate pSV282–XYL5 and pSV282–XYL5/DOM. The

expression plasmid pSV282 is based on the pET expression system (*Novagen*). In this vectors, a His tag was fused to the N-termini of the XYL5 and XYL5/DOM, allowing purification of the fusion proteins by nickel affinity chromatography (IMAC System); MBP was also fused to the N-termini of XYL5 and XYL5/DOM to enable the purification of the encoded proteins by amylose affinity chromatography.

We also tested several expression strains [BL21(DE3), BL21(DE3)ptGroE, BL21(DE3)pRIL, BL21(DE3)Rosetta and BL21(DE3)AD494] harboring other plasmid vectors (pET32-Xa/LIC, pET28-a and pET29-a) without obtaining adequate solubility of the protein of interest (data not shown).

3.2 Optimization of MBP-XYL5 and MBP-XYL5/DOM expression and solubility in *E. coli*

The cultures were incubated at 37°C and 300 rpm, until the optical density reached 0.8, and were then induced by 2, 4 and 6 hours with IPTG (1 mM). The cultures induced by 4 hours were chosen to the purification (Fig. 2 and 3). The constructs tested for expression of XYL5 and its conserved domain were fusions with MBP (MBP-XYL5 and MBP-XYL5/DOM). The MBP-XYL5 protein (Table 2) was partially expressed in the soluble fraction in the BL21(DE3) strain. Coomassie Brilliant Blue staining revealed a protein band of 89 kDa, which corresponded to the molecular weight of MBP-XYL5, in all fractions examined (Fig. 2).

The glycosyl hydrolase domain XYL5/DOM protein was expressed as an MBP-XYL5/DOM fusion using the *E. coli* BL21(DE3) strain. Coomassie Brilliant Blue staining revealed a protein band of 74 kDa, which corresponded to the expected molecular weight of MBP-XYL5/DOM (Table 2 and Fig. 3).

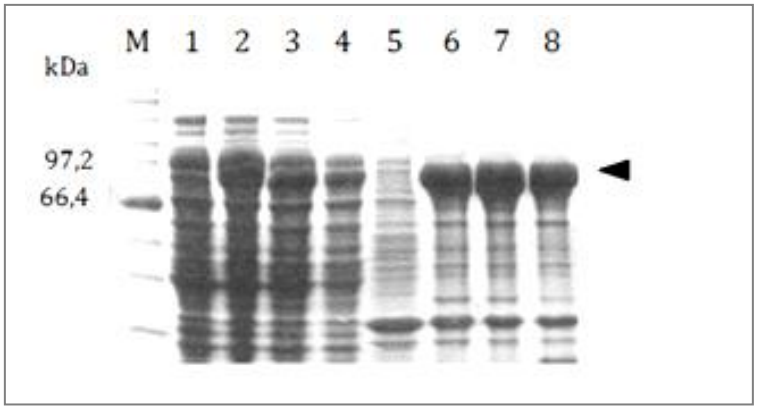


Fig. 2 - Expression of recombinant XYL5 fusion protein in *E.coli* Bl21(DE3) harboring pSV282-XYL5. The arrow indicates the bands corresponding to the molecular weight of 89kDa to MBP-XYL5. The cells were cultured at 37ºC, 300rpm at 1mM IPTG. Extraction of soluble fraction (lanes 1, 2, 3 and 4), and insoluble fraction (lanes 5, 6, 7and 8). Lane 1, No induction; lane 2, induction at 2h; lane 3, induction at 4h; lane 4, induction at 6h; Lane 5, No induction; lane 6, induction at 2h; lane 7, induction at 4h; lane 8, induction at 6h. Lane M- Protein Marker, Broad Range (2-212 kDa) from BioLabs.

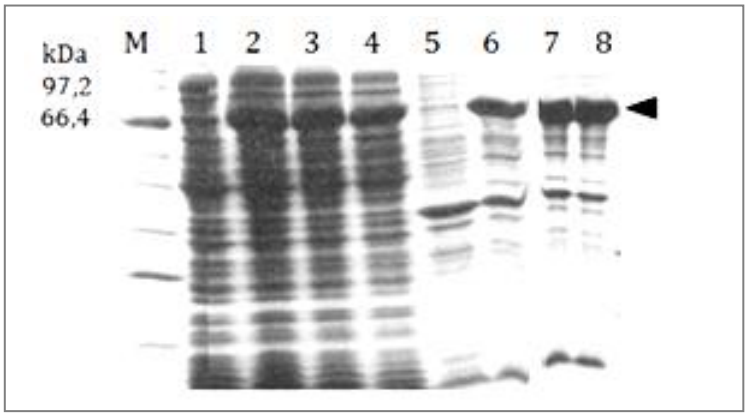


Fig. 3 - Expression of recombinant XYL5/DOM fusion protein in *E. coli* Bl21(DE3) harboring pSV282-XYL5/DOM. The arrow indicates the bands corresponding to the molecular weight of 74kDa to MBP-XYL5/DOM. The cells were cultured at 37°C, 300rpm at 1mM IPTG. Extraction of soluble (lanes 1, 2, 3 and 4) and insoluble (lanes 5, 6, 7and 8) fractions. Lane 1, No induction; lane 2, induction at 2h; lane 3, induction at 4h; lane 4, induction at 6h; Lane 5, No induction; lane 6, induction at 2h; lane 7, induction at 4h; lane 8, induction at 6h. Lane M- Protein Marker, Broad Range (2-212 kDa) from BioLabs.

Table 2
Predicted pI values, molecular weights and numbers of cysteines for XYL5, MBP-XYL5, XYL5/DOM and MBP-XYL5/DOM.

	XYL5	XYL5/DOM	MBP- XYL5	MBP- XYL5/DOM
pI	5.84	5.97	5.51	5.54
MW (kDa)	43.25	29.04	89.01	73.95
Cysteines	5	5	5	5

3.3 Purification of MBP-XYL5 and MBP-XYL5/DOM from *E. coli*

We purified MBP–XYL5 by amylose affinity chromatography. By exploiting the affinity of MBP for amylose, the MBP fusion proteins were adsorbed onto an amylose resin column (Fig. 4A). MBP-XYL5 was then eluted from the amylose-coated resin with 10 mM maltose. The eluted MBP-XYL5 was then subjected to SDS-PAGE, which revealed a protein band with a molecular weight of 89 kDa, corresponding to the expected molecular weight of MBP-XYL5. The yield of purified MBP-XYL5 was estimated to be 2.4 mg of fusion protein per gram of initial cell pellet.

As with MBP-XYL5, the MBP-XYL5/DOM protein was purified by amylose resin (Fig. 4B). SDS-PAGE of the purified fractions revealed a protein band with a molecular weight of 74 kDa, which corresponded to the

expected molecular weight of MBP-XYL5/DOM. The yield of purified MBP–XYL5/DOM was 3.8 mg of fusion protein per gram of initial cell pellet, a total higher than that obtained for MBP-XYL/DOM.

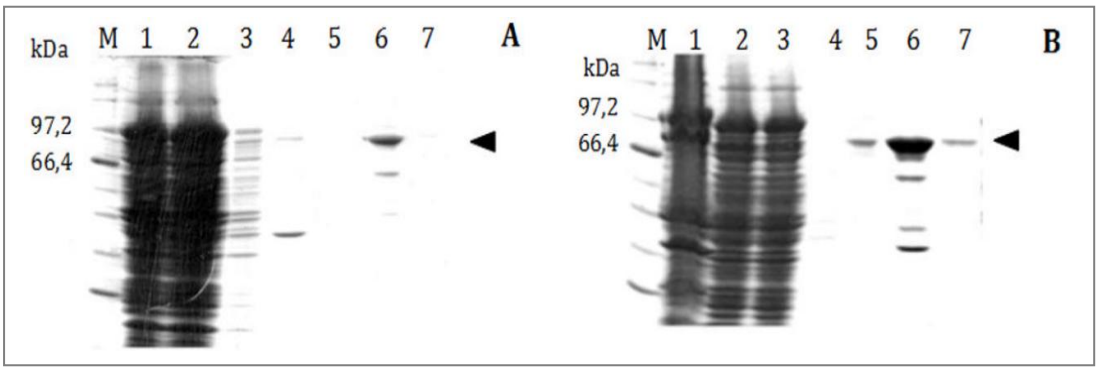


Fig. 4 (A) Purification of MBP-XYL5; the arrow indicates the bands corresponding to the molecular weight of 89kDa to MBP-XYL5. **(B)** Purification of MBP-XYL5/DOM; the arrow indicates the bands corresponding to the molecular weight of 74kDa to MBP-XYL5/DOM. MBP-XYL5 and MBP-XYL5/DOM were purified via MBP affinity chromatography using Amylose Resin. Lane 1, total soluble fraction; lane 2, flow trough; lane 3, wash fraction; lanes 4-7, elution fractions (0.01mM, 0.05mM maltose, 10 mM maltose and 50 mM maltose, respectively). Lane M- Protein Marker, Broad Range (2-212 kDa) from BioLabs .

The purification of MBP-XYL5 and MBP-XYL5/DOM by nickel affinity chromatography did not permit recovery of the proteins (data not shown).

3.4 Circular dichroism spectroscopy

We used circular dichroism (CD) spectroscopy to monitor the proper folding of MBP-XYL5 and MBP-XYL5/DOM. The spectra (Fig. 5) of both constructs were characteristic of well-folded proteins with a mixture of alpha-helix and beta-sheet structural elements (Kelly *et al*., 2005; Khrapunov, 2009). The CD results for both proteins confirmed the structures predicted with the use of specialized software (Jones 1999).

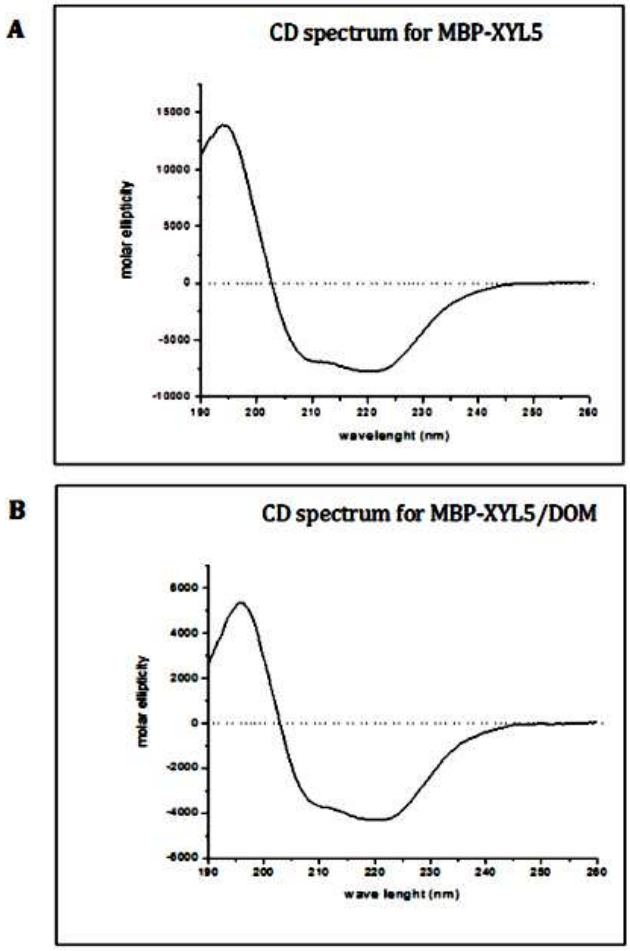


Fig. 5 - Circular dichroism of **(A)** MBP-XYL5 and **(B)** MBP-XYL5/DOM. The spectra of both constructs show well-folded proteins with a mixture of alpha-helix and beta-sheet structural elements

3.5 Biological activity of xylanase fused to MBP

The next step was to confirm the biological activity of MBP-XYL5 and MBP-XYL5/DOM by xylanolytic activity assays. MBP-XYL5 and MBP-XYL5/DOM were produced, purified and dialyzed. After standardization by protein mass, the biological activities of MBP-XYL5 and MBP-XYL5/DOM were evaluated by two different methods. The biological activity was tested using both the soluble extract and purified fractions containing the protein of interest. The substrate specificity of the enzyme was tested using birchwood xylan, CMC, microcrystalline cellulose (Avicel) and starch.

The soluble extracts containing MBP-XYL5 and MBP-XYL5/DOM showed activity only with birchwood xylan. In the assessment method using Petri dishes with Congo Red to measure halos of degradation of xylan, the extract containing MBP-XYL5/DOM presented a higher level of xylanolytic activity than the extract containing MBP-XYL5 (Fig. 6). A similar result was obtained when using the dinitrosalicylic acid (DNS)

method for detection of hydrolysis and liberation of reducing sugars; the xylanolytic activity was measured by xylose liberation from birchwood xylan on 50°C to MBP-XYL5, MBP-XYL5/DOM, and MBP: 15.56, 19.75, and 0.06 μg/ml xylose/min respectively. Reducing sugars were measured using the absorption at 540 nm relative to D-xylose standards. The soluble extract containing MBP-XYL5/DOM presented a higher level of xylanolytic activity than that obtained for the extract containing MBP-XYL5 (Fig. 6).

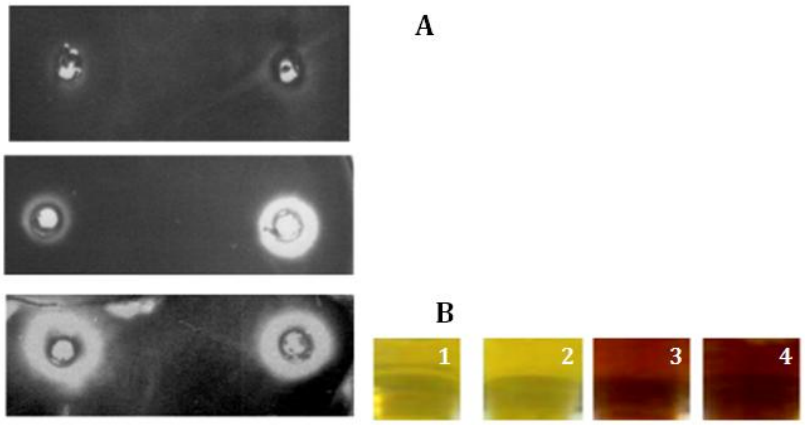


Fig. 6 (A) Assessment method in Petri dishes with Congo Red showing the halos of xylan degradation. Wells with 20 μl soluble extract containing MBP (upper); wells with 20 μl soluble extract containing MBP-XYL5 (middle); wells with 20 μl soluble extract containing MBP-XYL5/DOM (bottom). **(B)** Assessment method of estimating the liberation of reducing sugars by the dinitrosalicylic acid (DNS) using xylose as a standard: tube 1, soluble extract containing extraction buffer; tube 2, soluble extract containing MBP; tube 3, soluble extract containing MBP-XYL5/DOM; tube 4, soluble extract containing MBP-XYL5.

In all tests, a negative control of MBP was used. It did not show the presence of xylanolytic activity (Fig. 6). Surprisingly, all fractions containing MBP-XYL5 or MBP-XYL5/DOM purified by chromatography on amylose resin showed no xylanolytic activity, despite the assay method used.

4 Discussion

Our goal was to demonstrate, for the first time, the xylanolytic activity of the *XYL5* pathogenicity gene product of *Magnaporthe grisea*. A high level of expression of soluble XYL5 and XYL5/DOM was achieved with MBP. The soluble extract containing the fusion proteins MBP-XYL5 and MBP-XYL5/DOM showed xylanolytic activity compared to the control MBP. In *E. coli,* eukaryotic proteins are frequently expressed as insoluble aggregates called inclusion bodies. The protocol described in this work resulted in significant improvements in the yields of soluble XYL5 and XYL5/DOM. The proteins fused to MBP were present in both the total protein and soluble protein fractions. Our results agree with previous studies in which the use of MBP as a fusion partner was shown to improve the quantity and solubility of recombinant proteins (Cho *et al*.

2008; Gutierrez-Lugo *et al*. 2006; Korepanova *et al*. 2007; Niiranen *et al*. 2007; Tolun *et al*. 2007; Zhu *et al*. 2009).

Cloning in the pSV282 vector allowed the expression of proteins fused to a His tag and an MBP tag, which could subsequently be purified either by nickel or amylose affinity chromatography respectively. In this study, the purification of MBP-XYL5 and MBP-XYL5/DOM by nickel affinity chromatography did not permit recovery of the proteins (data not shown). This may result from structural features (folding) of the protein if the His tag was embedded within the structure, preventing access to the resin (Smyth *et al*. 2003). However, the purification of MBP-XYL5 and MBP-XYL5/DOM by affinity chromatography with amylose resin was successful.

MBP has emerged as an excellent solubilization tag, better than GST or TRX (Kapust and Waugh 1999). It has been suggested that aggregation occurs during, rather than after, the folding of proteins. Thus, although fusion to MBP has a beneficial impact on the folding of some proteins in *E. coli*, we cannot say with certainty what effect it may have on proteins that form insoluble aggregates in their native state (Kapust and Waugh 1999). In this study, we have verified the formation of insoluble aggregates despite fusion to MBP because we obtained only a portion of the total expressed protein fused to MBP as soluble (Fig. 2 and Fig. 3).

Despite the fact that the soluble extract containing the fusion proteins MBP-XYL5 and MBP-XYL5/DOM showed xylanolytic activity compared to the control MBP, the proteins lost activity during the purification procedure. We obtained expression of the MBP-XYL5 or MBP-XYL5/DOM proteins in both the soluble fraction and insoluble fraction (Fig. 2 and Fig. 3), and part of the MBP-XYL5 and MBP-XYL5/DOM proteins do not remain adsorbed on the amylose resin (Fig. 4). Thus, we formulated a hypothesis based on a model that has been proposed to explain how MBP can improve the solubility and promote the proper folding of its fusion partners (Kapust and Waugh 1999). In this model, it was suggested that MBP acts as a chaperone-like protein, promoting not only the solubility but also the folding of the protein of interest. Also in the model, the unfolded protein gives rise to structures for folding intermediates, which can have one of three different conformations or a combination of them: they may become insoluble (inclusion bodies), soluble intermediate structures that do not retain biological activity or native soluble structures with biological activity.

Thus, based on that model, we assume that the total extract contains MBP-XYL5 or MBP-XYL5/DOM in a mixture of three types of structures: insoluble (inclusion bodies), soluble intermediate structures that do not retain biological activity and soluble native structures with biological activity. We can then assume that in the soluble fraction there are both soluble intermediate structures that do not retain biological activity and native soluble structures with biological activity. We hypothesized that the soluble fraction would possess xylanolytic biological activity because this fraction contains native soluble structures with biological activity, in addition to soluble intermediate structures that do not retain biological activity. However, if we purified the

soluble fraction, the purified fraction would have less biological activity because the purified fraction would mostly contain soluble intermediate structures that do not retain biological activity.

Many other explanations can account for the loss of xylanolytic activity after the purification of the soluble MBP-XYL5 or MBP-XYL5/DOM, such as: **(1)** variations of the protein structures during the steps of the purification procedure - however, circular dichroism (Fig. 4) showed that purified MBP-XYL5 and MBP-XYL5/DOM were structured as expected. **(2)** It has been shown that many xylanases require metal ions as modulators of activity (Carmona *et al*. 2005; Fernandez-Espinar *et al*. 1994). We know that XYL5 may have hydrolase activity on glycosyl bonds and may be capable of binding cations (UniProtKB/TrEMBL entry Q8J1Y4, available on www.uniprot.org/uniprot/Q8J1Y4). Therefore, metal ions or cations could also be necessary for the activity of MBP-XYL5 and MBP-XYL5/DOM. If these metal ions were not kept together proteins during the purification procedure, this could result in the loss of activity. **(3)** Xylanases can also be inhibited by maltose (which is significant because the elution in affinity chromatography by amylose is made with maltose) (Khanna 1993; Kubackova *et al*. 1978; Thiagarajan and Gunasekaran 2006). **(4)** It is also important to note that XYL5 and XYL5/DOM have five cysteines (Table 2) and that disulfide bonds are important for their structures and activity (Buchmeier *et al*. 2006; Karagüler *et al*. 2007). These bonds may be maintained by the environment of the soluble fraction before purification; if this environment has been changed in the process of purification, the disulfides may be lost, possibly changing the xylanolytic activity. These and other factors may be, individually or jointly, responsible for the loss of enzymatic activity after purification of MBP-XYL5 and MBP-XYL5/DOM.

In conclusion, in this work we cloned and expressed an active xylanase from *Magnaporthe grisea* using *E. coli*. We obtained the enzyme (MBP-XYL5) and its catalytic domain (MBP-XYL5/DOM) as purified fusion proteins. It has been shown that the processes of purification and excision of the solubilization tag from recombinant proteins can lead to, for various reasons, a great loss of material as well as a loss of enzymatic activity (Smyth *et al*. 2003; Adikesavan *et al*. 2005). However, we obtained an active xylanolytic enzyme without any purification or excision of the solubilization tag in a heterologous system as simple as *E. coli*. These procedures are economically advantageous in terms of cost and production time for industrial processes that use xylanolytic complex because one of the main constraints on the industrial application of xylanases is the high cost of production (Ahmed *et al*. 2009; Jun *et al*. 2009; Kirk *et al*. 2002). The verification of the hypotheses for the reductions in the biological activity of MBP-XYL5 or MBP-XYL5/DOM after the purification procedure deserves further investigation. Despite of this, these results contribute to the evaluations of xylanase xyl5 the pathogenicity of *Magnaporthe grisea*.

Acknowledgments

This study was supported by grants from the Fundação de Amparo à Pequisa do Estado de São Paulo (FAPESP, Brazil) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil).

We would like to thank the Centro Nacional de Pesquisa em Arroz e Feijão (CNPAF, Embrapa, Brazil) for supplying the *Magnaporthe grisea* strain, and Dr. Laura Mizoue (Center for Structural Biology at Vanderbilt University, USA) for supplying the plasmid vector pSV282.

References

Adikesavan NV, Mahmood SS, Stanley N, Xu Z, Wu N, Thibonnierb M, Shohama M (2005) A C-terminal segment of the V1R vasopressin receptor is structured in the crystal structure of its chimera with the maltose-binding protein. Acta Cryst; F61:341–345.

Ahmed S, Riaz S, Jamil A (2009) Molecular cloning of fungal xylanases: an overview. Appl Microbiol Biotechnol 84(1):19-35.

Beg QK, Kapoor M, Mahajan L, Hoondal GS (2001) Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol 56(3-4):326-338.

Berrin JG, Juge N (2008) Factors affecting xylanase functionality in the degradation of arabinoxylans. Biotechnol Lett 30(7):1139-1150.

Buchmeier NA, Newton GL, Fahey RC (2006) A mycothiol synthase mutant of *Mycobacterium tuberculosis* has an altered thiol-disulfide content and limited tolerance to stress. J Bacteriol 188(17):6245-6252.

Cao Y, Qiao J, Li Y, Lu W (2007) De novo synthesis, constitutive expression of *Aspergillus sulphureus* β-xylanase gene in *Pichia pastoris* and partial enzymic characterization. Appl Microbiol Biotechnol 76:579–585.

Carapito R, Carapito C, Jeltsch JM, Phalip V (2009) Efficient hydrolysis of hemicellulose by a *Fusarium graminearum* xylanase blend produced at high levels in *Escherichia coli*. Bioresource Technology 100:845–850.

Carmona EC, Fialho MB, Buchgnani EB, Coelho GD, Brocheto-Braga MR, Jorge JA (2005) Production, purification and characterization of a minor form of xylanase from *Aspergillus versicolor*. Proc Biochem 40:359-364.

Cho HJ, Lee Y, Chang RS, Hahm MS, Kim MK, Kim YB, Oh YK (2008) Maltose binding protein facilitates high-level expression and functional purification of the chemokines RANTES and SDF-1alpha from *Escherichia coli*. Protein Expr Purif 60(1):37-45.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeyer C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434(7036):980-986.

Fernandez-Espinar M, Pinaga F, de Graaff L, Visser J, Ramon D, Valles S (1994) Purification, characterization and regulation of the synthesis of an *Aspergillus nidulans* acidic xylanase. Appl Microbiol Biotechnol 42:555-562.

Gutierrez-Lugo MT, Newton GL, Fahey RC, Bewley CA (2006) Cloning, expression and rapid purification of active recombinant mycothiol ligase as B1 immunoglobulin binding domain of streptococcal protein G, glutathione-S-transferase and maltose binding protein fusion proteins in *Mycobacterium smegmatis*. Protein Expr Purif 50:128-136.

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292:195-202.

Jun H, Bing Y, Keying Z, Xuemei D, Daiwen C (2009) Expression of a *Trichoderma reesei* beta-xylanase gene in *Escherichia coli* and activity of the enzyme on fiber-bound substrates. Protein Expr Purif 67(1):1-6.

Kapust RB, Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Science 8:1668–1674.

Karagüler NG, Sessions RB, Clarke AR (2007) Effects of disulphide bridges on the activity and stability of the formate dehydrogenase from *Candida methylica*. Biotechnol Lett 29(9):1375-1380.

Kasana RC, Salwan R, Dhar H, Dutt S, Gulati A (2008) A rapid and easy method for the detection of microbial cellulases on agar plates using gram's iodine. Curr Microbiol 57(5):503-507.

Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta. 1751(2):119-39.

Khanna S (1993) Regulation, purification and properties of xylanase from *Cellulomonas fimi*. Enzyme Microb Technol 15:990-995.

Khrapunov S (2009) Circular dichroism spectroscopy has intrinsic limitations for protein secondary structure analysis. Anal Biochem 389(2):174-176.

Kirk O, Borchert TV, Fuglsang CC (2002) Industrial enzyme applications. Curr Opin Biotechnol 13(4):345-51.

Korepanova A, Moore JD, Nguyen HB, Hua Y, Cross TA, Gao F (2007) Expression of membrane proteins from *Mycobacterium tuberculosis* in *Escherichia coli* as fusions with maltose binding protein. Protein Expr Purif 53(1):24-30.

Kubackova M, Karacsonyi S, Bilisics L, Toman R (1978) Some properties of an endo-1,4-beta-D-xylanase from the ligniperdous fungus *Trametes hirsute*. Folia Microbiol (Praha) 23:202-209.

Lecellier G, Silar P (1994) Rapid methods for nucleic acids extraction from Petri dish-grown mycelia. Curr Genet 25(2):122-123.

Lee TH, Lim PO, Lee YE (2007) Cloning, characterization, and expression of xylanase A gene from Paenibacillus sp. DG-22 in *Escherichia coli.* J Microbiol Biotechnol 17(1):29-36.

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Analytical Chemistry 31(30):426-428.

Nair SG, Sindhu R, Shashidhar S (2008) Purification and biochemical characterization of two xylanases from *Aspergillus sydowii* SBS 45. Appl Biochem Biotechnol 149(3):229-243.

Nevalainen KMH, Teo VSJ, Bergquist PL (2005) Heterologous protein expression in filamentous fungi. Trends in Biotechnology 23 (9):468-474.

Niiranen L, Espelid S, Karlsen CR, Mustonen M, Paulsen SM, Heikinheimo P, Willassen NP (2007) Comparative expression study to increase the solubility of cold adapted Vibrio proteins in *Escherichia coli.* Protein Expr Purif 52(1):210-218.

Polizeli ML, Rizzatti AC, Monti R, Terenzi HF, Jorge JA, Amorim DS (2005) Xylanases from fungi: properties and industrial applications. Appl Microbiol Biotechnol 67(5):577-591.

Sakaguchi M, Niimiya K, Takezawa M, Toki T, Sugahara Y, Kawakita M (2008) Construction of an expression system for aqualysin I in *Escherichia coli* that gives a markedly improved yield of the enzyme protein. Biosci Biotechnol Biochem 72(8):2012-2018.

Sambrook J, Russell, DW (2001) Plasmids and their usefulness in molecular cloning. In: Molecular cloning, a laboratory manual. 3rd edn. Vol. 1, pp 1.32 - 1.34. CSHL Press, New York.

Shengcheng W, Zhiying Z, Darvill AG, Albersheim P (2001) Xylanases of Magnaporthe grisea: differential requirement for pathogenicity. In: Host-Parasite Interactions, XXI Fungal Genetics Conference, Asilomar, California,USA.

Smyth DR, Mrozkiewicz MK, Mcgrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. Protein Science 12:1313–1322.

Talbot NJ (2003) On the trail of a cereal killer () Exploring the Biology of *Magnaporthe grisea*. Annu Rev Microbiol 57:177–202.

Teather RM, Wood PJ (1982)Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43(4):777-780.

Thiagarajan S, Jeya M, Gunasekaran P (2006) Purification and characterization of a high molecular weight endoxylanase from the solid-state culture of an alkali-tolerant *Aspergillus fumigatus* MKU1. World J Microbiol Biotechnol 22:487-492.

Tolun AA, Dickerson IM, Malhotra A (2007) Overexpression and purification of human calcitonin gene-related peptide-receptor component protein in *Escherichia coli.* Protein Expr Purif 52(1):167-174.

Wu SC, Halley JE, Luttig C, Fernekes LM, Gutiérrez-Sanchez G, Darvill AG, Albersheim P (2006) Identification of an endo-beta-1,4-D-xylanase from *Magnaporthe grisea* by gene knockout analysis, purification, and heterologous expression. Appl Environ Microbiol 72(2):986-993.

Zhou C, Bai J, Deng S, Wang J, Zhu J, Wu M, Wang W (2008) Cloning of a xylanase gene from *Aspergillus usamii* and its expression in *Escherichia coli*. Bioresour Technol 99(4):831-838.

Zhu S, Yang G, Yang X, Zhao Y, Li X, Deng P, Xie Y, Gan Z, Liu Y, Li Z, Liao J, Yu M, Liao F (2009) Soluble expression in *Escherichia coli* of active human cyclic nucleotide phosphodiesterase isoform 4B2 in fusion with maltose-binding protein. Biosci Biotechnol Biochem 73(4):968-970.

4.4. INFORMAÇÕES ADICIONAIS SOBRE A XILANASE XYL5

4.4.1. Gene *XYL5*, sua região intragênica e suas proteínas correlatas

A amplificação e a clonagem do gene íntegro *XYL5* e do seu domínio *XYL5/DOM* (Figuras 11 e 12) foi feita com o uso de diversos *primers* internos (Tabela 1 do manuscrito 2) visando ao seqüenciamento. A clonagem no vetor pSV282 foi realizada utilizando-se uma proporção de 1:3 (vetor:inserto). A eficiência da clonagem foi avaliada mediante o seqüenciamento de diferentes clones para cada gene. As condições de clonagem, cultivo, extração das proteínas, expressão e purificação das proteínas encontram-se no manuscrito 2.

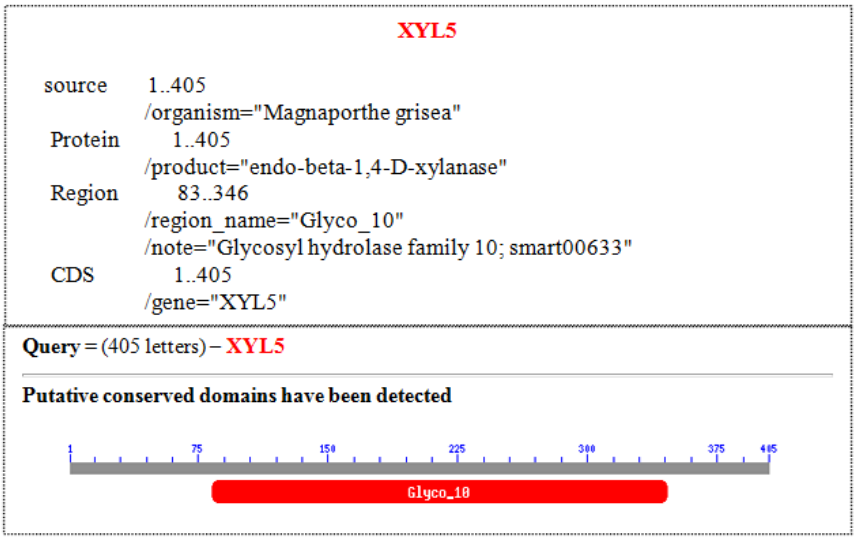


Figura 11. Predição da função da proteína XYL5 e de seu domínio catalítico (http://www.ncbi.nlm.nih.gov/).

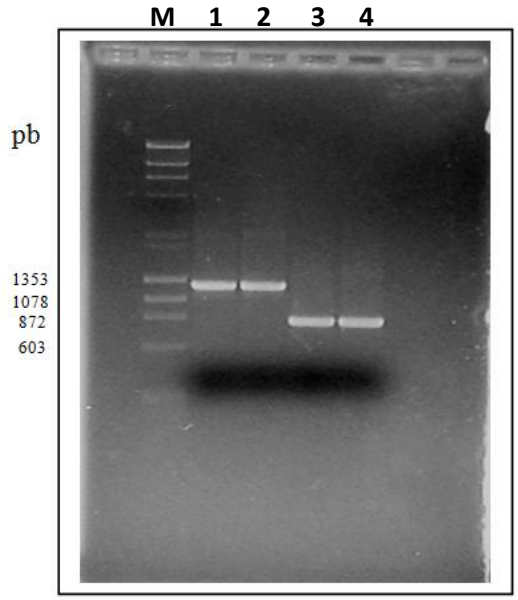


Figura 12. Gel de agarose 1%. Purificação dos insertos para a clonagem no vetor pSV282. **M**: Marcador Fermentas **1-2**: Gene *XYL5* (1218pb) **3-4**: Região intragênica *XYL5.DOM* (792pb)

A expressão dos genes de interesse em diferentes linhagens de expressão pode ser vista na tabela 2. Observa-se que, com exceção da expressão de *XYL5/DOM*.pSV282 na linhagem AD494, todas as demais combinações mostraram que a expressão protéica ocorreu. Entretanto, quanto à solubilidade, na maioria dos casos a proteína expressa foi recuperada na fração insolúvel.

Visto que a concentração de sal, a temperatura, a aeração do meio de cultura regulada pela velocidade de rotação e o pH do tampão de extração, bem como agentes caotrópicos como 4M uréia podem influir de modo considerável na solubilidade de proteínas, diferentes testes de solubilização foram realizados com as proteínas XYL5 e XYL5.DOM (Gray *et al*., 2000; Mayer, 2004; Tresaugues *et al*., 2004) . Com o uso de uréia, pôde-se observar a solubilização, mas não o *refolding* correto das proteínas,

visto que não ficaram adsorvidas na resina Ni-NTA. Entretanto, a clonagem no vetor pSV282 e a expressão na linhagem BL21(DE3) possibilitou a recuperação das proteínas de interesse na fração solúvel ainda que parcialmente (Tabela 1).

Tabela 1. Clonagens de *XYL5* e de *XYL5/DOM;* avaliação da expressão e da solubilidade. A linhagem de clonagem foi *E.coli* DH5-∝ em todos os casos.

Proteína	**Vetor de clonagem e de expressão**	**Linhagem de expressão *E. coli***	**Expressão/Solubilidade**
XYL5	pSV282	BL21(DE3)	Parcialmente solúvel
XYL5	pSV282	BL21(DE3) pRIL	Insolúvel
XYL5	pSV282	BL21(DE3)Rosetta	Insolúvel
XYL5	pET32Xa/LIC	BL21(DE3)AD494	Insolúvel
XYL5/DOM	pSV282	BL21(DE3)	Parcialmente solúvel
XYL5/DOM	pSV282	BL21(DE3) pRIL	Insolúvel
XYL5/DOM	pSV282	BL21(DE3)Rosetta	Insolúvel
XYL5/DOM	pET32Xa/LIC	BL21(DE3)AD494	(sem expressão detectável)

A obtenção da expressão também na fração insolúvel (Figura 13) pode ser inicialmente explicada pelas regiões transmembranas preditas tanto para XYL5 como

para XYL5/DOM, conforme pode ser visto na Figura 14, visto que regiões hidrofóbicas, como as regiões transmembrana o são, favorecem a insolubilidade das proteínas.

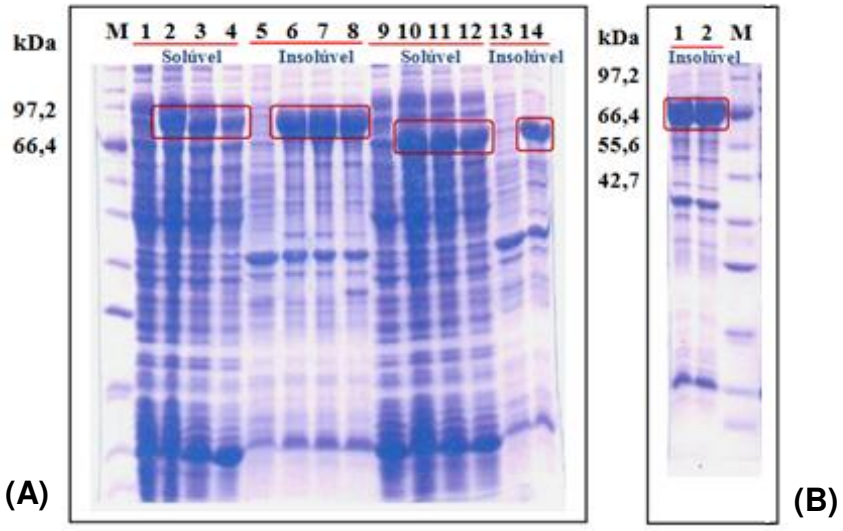


Figura 13. Extração das frações solúveis e insolúveis após diferentes tempos de indução. SDS-PAGE 12% - Amostras obtidas a partir dos genes clonados em pSV282 expressos em *E.coli* BL21(DE3). **M**- Marcador BioLabs (**A) 1-4:** MBP-XYL5 (0h, 2h, 4h, 6h de indução)**; 5-8**: MBP-XYL5 (0h, 2h, 4h, 6h de indução); **9-12:** MBP-XYL5/DOM (0h, 2h, 4h, 6h de indução)**; 13-14**: MBP-XYL5/DOM (0h, 2h de indução). **(B) 1-2**: MBP-XYL5/DOM (4h, 6h de indução). Tamanhos esperados para as bandas eletroforéticas: MBP-XYL5 (89kDa), MBP-XYL5/DOM (74kDa)

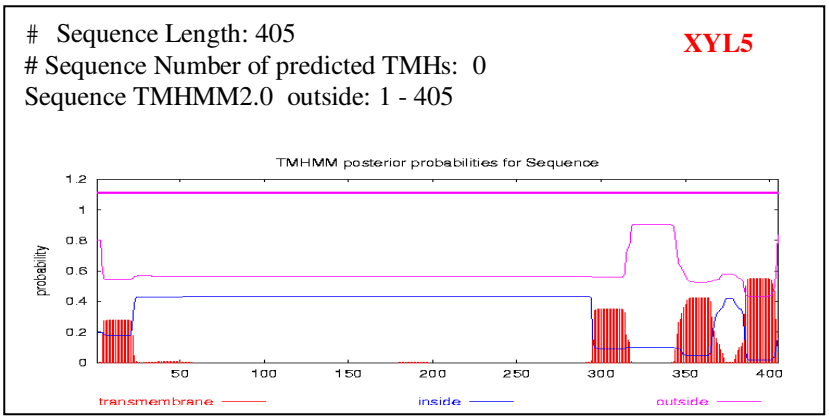


Figura 14. Predição das regiões transmembranas e de peptídeo- sinal na proteína XYL5; XYL5/DOM: resíduos 83 – 346.

Com a escolha do sistema ideal de expressão, a saber, a linhagem BL21(DE3) e o vetor pSV282, pôde-se purificar e testar a atividade das proteínas XYL5 e XYL5/DOM tanto com a cauda de solubilização MBP fusionada como sem a fusão, sendo que os detalhes e os resultados destas etapas encontram-se no manuscrito 2.

5. RESULTADOS COMPLEMENTARES E DISCUSSÃO

5.1. LINHAGENS FITOPATOGÊNICAS DE *Magnaporthe grisea*

As linhagens de *M. grisea* utilizadas foram cedidas e caracterizadas como fitopatogênicas pelo Centro Nacional de Pesquisa em Arroz e Feijão (CNPAF/ Embrapa, GO): linhagem **A** (Py 2672-L7-1), linhagem **B** (Py 2618-L1-1), linhagem **C** (Py 2588-L5-1), linhagem **D** (Py 416-L1-3) e linhagem **E** (Py 424- L8-3) – as siglas entre parêntesis referem-se aos registros destas linhagens no banco de origem, no CNPAF.

5.2. SISTEMAS DE CLONAGEM E DE EXPRESSÃO

Diferentes estratégias de clonagem, expressão e purificação de proteínas têm sido desenvolvidas com o intuito de otimizar cada etapa do processo que visa a caracterização estrutural e funcional de proteínas recombinantes (Tudzynski, *et al*. 2003; Maxwell *et al.*, 2003; Thiede *et al*., 2005). Os esforços neste sentido permitem: o uso de vetores diversos adequados para facilitar a clonagem (Studier; Moffatt, 1986; Malakov, 2004), a utilização de linhagens de *E. coli* que possibilitam aumentar o produto purificado com base na especificidade para a ligação em resinas utilizadas em cromatografia de afinidade (Roof, 1997), bem como a utilização de linhagens que aumentem a solubilidade de proteínas recombinantes (Nishihara, 1998).

Diferentes vetores foram utilizados em etapas distintas do trabalho sempre visando: clonagem eficiente, qualidade e quantidade de produto protéico expresso, facilidade e eficiência no procedimento de purificação protéica.

A clonagem no vetor pET32-Xa/LIC (*Novagen*), que não requer o uso de endonucleases de restrição, forneceu como opção o emprego de diferentes proteases para a recuperação da proteína de interesse. A estrutura deste vetor é, basicamente, a que se segue: *TRX tag (109aa) — His tag (6aa) — S tag (15aa) — Proteína-alvo*. Neste vetor, a proteína tioredoxina ou TRX (cuja função é aumentar a solubilidade da proteína recombinante) é seguida de uma seqüência de 6 resíduos de histidina (His-tag, responsável pela posterior ligação da proteína expressa fusionada à resina Ni-NTA na purificação por cromatografia de afinidade ao níquel) e de uma seqüência de 15 resíduos (S-tag, utilizados na purificação com resinas próprias para esta cauda). A proteína de interesse encontra-se fusionada posterior a estas três regiões. Neste contexto, têm-se duas possibilidades de excisão da causa fusionada: a protease trombina cliva entre a His-tag e a S-tag, e a protease fator Xa cliva após a S-tag e imediatamente antes da proteína-alvo. O fator Xa, portanto, é a primeira protease a ser empregada e sugerida pelo fabricante do vetor, visto resultar em uma proteína-alvo sem quaisquer outros resíduos indesejáveis. A trombina, que cliva entre a His-tag e a proteína-alvo, gera uma cauda resultante de cerca de 17,6 kDa e deixa permanecer, junto à proteína de interesse, uma pequena seqüência de cerca de 15 resíduos (pET System Manual, 2006).

O uso de outros vetores do sistema pET (*Novagen*) visou otimizar diferentes etapas dos processos de expressão, extração, purificação e solubilização de genes de interesse e de suas proteínas correlatas. Vetores como pET28a(+) e pET29a(+) foram inicialmente escolhidos por oferecerem possíveis vantagens, a saber: diferentes opções de sítios de clonagem (que podem determinar o tamanho da cauda fusionada à proteína de interesse), indução da expressão por lactose ou por IPTG, além de ser possível fusionar a cauda de histidina tanto no N-terminal como no C-terminal da proteína de interesse. Os *primers* para tais vetores foram desenhados considerando-se os sítios de clonagem e as enzimas de restrição mais convenientes, de acordo com a compatibilidade entre tais enzimas e a sequência gênica de interesse.

A utilização de vetor pETSUMO/Pro objetivou a melhoria da expressão e da solubilização das proteínas de interesse. Este vetor possui, basicamente, uma estrutura como a que se segue: *His tag (6aa)— [sítio de reconhecimento da SUMO protease] — Proteína-alvo.* A excisão da cauda fusionada de 12kDa é realizada por meio da protease de alta especificidade. O vetor possibilita um aumento na solubilidade da proteína de interesse devido à ação da proteína SUMO (*small ubiquitin-like modifier*) que regula interações entre proteína-proteína e proteína-DNA bem como a localização e a estabilidade da proteína-alvo (Malakhov *et al.*, 2004).

A escolha do vetor de expressão pSV282, derivado do vetor pET27 (*Novagen*), foi feita visando um aumento na solubilização das proteínas expressas. Este vetor possui uma estrutura como se segue: *His tag (6aa)— [Maltose Binding Protein tag] — [60*

resíduos] — sítio de clivagem da TEV protease —sítio de clonagem com BamHI — Proteína-alvo. A MBP (*Maltose Binding Protein*) é responsável pela solubilização de proteínas a ela fusionadas (Kim *et al.*, 2007). A clonagem no vetor pSV282 demandou o uso de enzimas de restrição, sendo que a enzima *Bam*HI é obrigatoriamente requerida para o procedimento.

Visando melhorar a expressão e a solubilização, diferentes sistemas de expressão foram utilizados para os genes íntegros clonados. A linhagem BL21(DE3), utilizada para a expressão de proteínas recombinantes, possui baixos índices de expressão de proteases que poderiam degradar a proteína de interesse. Esta linhagem possui o gene DE3 que é responsável pela síntese da enzima T7 RNA polimerase e é controlado pelo promotor *lac*UV5 que é ativado pela adição de agentes indutores como IPTG (isopropyl-β-D- 1-thiogalactopyranoside) ou lactose (pET Manual, 2010).

Utilizaram-se linhagens com outras características além das apresentadas pela BL21(DE3): BL21(DE3)ptGroE, BL21(DE3)AD494, BL21(DE3)pRIL, BL21(DE3)pLysS, BL21 (DE3)Rosetta e BL21 (DE3)SlyD$^-$.

A linhagem BL21(DE3)ptGroE, por meio da chaperona GroE, propõe promover maior produto solúvel. A linhagem BL21(DE3)AD494 promove a formação de pontes dissulfeto. A linhagem BL21(DE3)pRIL e a linhagem BL21 (DE3)Rosetta contem tRNAs para os códons raros. A linhagem BL21(DE3)pLysS possui o plasmídeo pLysS, responsável pela síntese de lisozima, que age reduzindo os níveis de produção da T7

RNA polimerase, a qual está envolvida na transcrição; o uso desta linhagem é, portanto, conveniente, quando se quer reduzir a expressão basal de proteínas sabidamente ou supostamente tóxicas à linhagem hospedeira. A linhagem BL21 (DE3)SlyD$^-$ tem o gene *slyD* deletado. Este gene codifica para uma proteína — SLYD — que, por conter muitos resíduos de histidina em sua seqüência, compete na etapa de purificação por afinidade com a cauda de histidina fusionada à proteína de interesse clonada em alguns vetores do sistema pET (Bernhardt *et al.*, 2000); esta competição é inadequada para o processo de purificação por cromatografia de afinidade ao níquel, visto que pode ocasionar uma considerável redução na quantidade de produto purificado.

5.3. GENES DE *Magnaporthe grisea* SELECIONADOS E SUAS RESPECTIVAS PROTEÍNAS

Alguns aspectos foram analisados a fim de se ter um critério para a escolha dos genes de patogenicidade ou de avirulência de *M. grisea*. Dentre tais aspectos, descreve-se:

(1) valor sócio-econômico do fungo fitopatogênico em virtude da importância da planta-hospedeira;
(2) a informação disponível sobre a identificação e a caracterização previamente feitas dos genes considerados como “genes relacionados à patogenicidade”;

(3) a informação disponível quanto à descrição da seqüência de nucleotídeos de tais genes, com a verificação da existência de introns e do número de pares de base (tendo em vista o vetor de clonagem e o sistema de clonagem a serem utilizados);

(4) a informação disponível quanto à descrição da seqüência de aminoácidos das proteínas codificadas por tais genes, com a verificação da homologia com outras seqüências (homologia igual ou maior que 40%, numa extensão maior ou igual a 70% do comprimento da seqüência de interesse) e com verificação da existência da estrutura em cristal;

(5) a massa molar e a possível função da proteína codificada.

Assim, diferentes fontes e diversos bancos de dados em Bioinformática foram utilizados com o objetivo de identificar e selecionar os genes de patogenicidade e suas respectivas proteínas. Os genes selecionados (com seqüências depositadas no GenBank - www.ncbi.nlm.nih.gov) e as proteínas por eles codificadas encontram-se na tabela 2.

Tabela 2. Genes de *Magnaporthe grisea* selecionados e suas respectivas proteínas.

GENE		**PROTEÍNA**		**Observação**
Nome (Acesso no GenBank)	**pb**	**kDa**	**Função prevista**	**Linhagens A, B, C, D**
PWL1 (**U36923.1**)	444	16	ORF (*open reading frame*)	—
PWL2 (**U26313.1**)	438	16	ORF	Proteína Íntegra e proteínas de 2 regiões intragênicas (shPWL2 e orPWL2)
PWL3 (**U36995.1**)	414	15	ORF	—
PWL4 (**U36996.1**)	417	15	ORF	—
XYL5 (**AY144350.1**)	1218	45	- Endo-beta-1,4-D-Xilanase - Domínio de 29kDa: Glicosyl Hydrolase family 10: XYL5/DOM	Proteína Íntegra e proteína de 1 região intragênica (XYL5/DOM)
AVR-PITA **(AF207841.1)**	672	25	- Metaloprotease	—
ABC1 (**AF420471.1**)	2181	80	- Chaperona - Domínio de 13kDa (ABC1 family - suprime defeito de tradução do mRNA do citocromo B) – ABC1/DOM1 - Domínio de 43kDa (Kinase não usual) – ABC1/DOM2	Proteína Íntegra e proteínas de 2 regiões intragênicas (domínios ABC1/DOM1 e ABC1/DOM2)
PTH3 (**AF027980.1**)	765	24	- Imidazole glycerol phosphate dehydratase (IGPD)	—
PTH9 (**AF027981.1**)	2211	81	- Trealase	Proteína Íntegra e proteína de 1 região intragênica

5.4. GENES *AVR-PITA, ABC1, PTH3* E *PTH9,* E SUAS PROTEÍNAS CORRELATAS

O gene *AVR*-PITA não pôde ser amplificado nem com *primers* flanqueadores ao gene nem com *primers* específicos para a clonagem nos vetores pET32-Xa/LIC, pET28a(+) e pET29a(+). Este resultado pode ser inicialmente justificado por se tratar de um gene

de avirulência, que limita sua ocorrência em alguns genótipos de *M. grisea* (Valent; Chumley, 1991). Além disso, mapeamento genético mostrou que o gene de avirulência *AVR-Pita* está intimamente ligado ao telômero do cromossomo 3 de *M. grisea* (Orbach *et al*., 2000). Ao serem introduzidos em patógenos virulentos de arroz, o gene clonado confere especificamente avirulência para cultivares de arroz que contêm Pi-ta. Freqüente perda espontânea de AVR-Pita parece ser o resultado de sua localização telomérica – o que pode justificar sua ausência nas linhagens utilizadas (Orbach *et al.,* 2000).

A amplificação do gene íntegro *ABC1* visando o seqüenciamento exigiu o uso de diversos *primers* internos em virtude do tamanho da seqüência. O seqüenciamento de diferentes clones contendo este gene mostrou que as linhagens disponíveis no laboratório apresentam disparidade na seqüência em relação ao banco de dados de referência. As clonagens nos vetores pET32-Xa/LIC e pSV282 em linhagens de *E.coli* DH5-∝, BL21(DE3) e BL21(DE3)pRIL foram eficientemente conduzidas. Em *E.coli* BL21(DE3) as proteínas foram expressas na fração insolúvel e em BL21(DE3)pRIL não houve expressão detectável. Apesar da expressão das proteínas ABC1, ABC1/DOM1 e ABC1/DOM2 fusionadas à cauda do vetor pET32Xa/LIC terem ocorrido nas frações insolúveis, as frações solúveis foram levadas à purificação por cromatografia de afinidade à níquel, confirmando suas expressões na forma totalmente insolúvel. Procedimentos adequados para a solubilização, como a clonagem em outros vetores de expressão e a otimização da expressão e da extração, deverão ser adotados para ser possível a continuação dos estudos.

O gene *PTH3* foi amplificado, e o seqüenciamento evidenciou a existência de introns, o que foi considerado indesejável, tendo em vista os sistemas de expressão procarióticos utilizados neste trabalho. Além disso, não estava previsto a construção de biblioteca de cDNA.

A amplificação do gene *PTH9* visando o seqüenciamento, exigiu o uso de diversos *primers* internos em virtude do tamanho da seqüência. O seqüenciamento de diferentes clones contendo este gene mostrou que as linhagens disponíveis no laboratório apresentam *introns* na seqüência em relação ao banco de dados de referência. Segundo o banco de dados utilizado, na linhagem 4091-5-8 de *M. grisea*, o loco gênico refere-se ao gene *PTH9* (com 2211pb e sem introns), mas em outras linhagens de *M. grisea* (como a linhagem Guy11) o locus gênico refere-se ao gene *NTH1*(GenBank: **AY148092.1**; com 3268pb e com *introns*). Assim, realmente, as linhagens utilizadas no desenvolvimento desta tese contêm o gene *NTH1* e não o gene *PTH9*. Tendo em vista o sistema de expressão procariótico utilizado, investiu-se, portanto, nos genes sem *introns* amplificados com sucesso como apresentado nos tópicos anteriores.

Em função do conjunto de resultados obtidos e das limitações encontradas em diferentes etapas do estudo das proteínas expressas, este trabalho concentrou-se no estudo das proteínas PWL2, PWL2D, XYL5 e XYL5/DOM.

6. RESUMO DOS RESULTADOS

Os resultados apresentados no Manuscrito 1 e nos Resultados Complementares para o gene *PWL2D* e para a proteína PWL2D de *M. grisea* mostram que:

- foi possível clonar o gene *PWL2D* e expressar a proteína por ele codificada em um sistema heterólogo procariótico como *E. coli;*
- tanto o vetor como a linhagem de expressão foram determinantes para a obtenção da proteína solúvel (ainda que parcialmente), sendo que o melhor sistema de expressão foi a linhagem BL21(DE3) e o vetor pET32-Xa/LIC;
- a cauda tioredoxina (ou TRX) fusionada foi menos eficaz em promover a solubilidade da proteína do que a retirada do peptídeo-sinal, ou seja, TRX-PWL2D foi muito menos solúvel do que TRX-*sh*PWL2D;
- a cauda TRX mantém as proteínas estruturadas, isto é, TRX-PWL2D, TRX-*sh*PWL2D e TRX-*or*PWL2D estão estruturadas enquanto PWL2D, *sh*PWL2D e *or*PWL2D mostram o espectro de dicroísmo circular em *random coil*;
- o procedimento de proteólise limitada sobre as proteínas não permitiu a obtenção de um núcleo protéico estável evidenciando o estado desordenado de suas estruturas;
- o espectro de NMR para *or*PWL2D mostrou que esta região está também desestruturada apresentando-se em *random coil*;
- as afirmações mencionadas anteriormente somadas ao fato de que há previsão de que PWL2D seja uma proteína desordenada permite-nos inferir que PWL2D muito provavelmente é uma proteína intrinsecamente desordenada. Tal desordem natural é bem adequada para proteínas efetoras e de avirulência, visto que a

flexibilidade proveniente da desordem estrutural permite uma melhor adaptação aos requisitos conformacionais na interação da proteína com os produtos gênicos da planta hospedeira;

- foi possível propor um modelo de ação da proteína PWL2D em que a susceptibilidade e a resistência de *M. grisea* pudessem ser explicadas para este caso.

Além disso, com base nos resultados apresentados no manuscrito 2 e nos resultados complementares para o gene *XYl5*, para a proteína XYL5 e para o domínio XYL5/DOM vimos que:

- foi possível clonar e expressar uma xilanase ativa do eucarioto *M. grisea* utilizando um sistema heterólogo procariótico como *E. coli;*
- tanto o vetor como a linhagem de expressão foram determinantes para a obtenção da proteína solubilizada (ainda que parcialmente), sendo que o melhor sistema de expressão foi a linhagem BL21(DE3) e o vetor pSV282. Assim obtiveram-se a enzima e seu domínio catalítico como proteínas de fusão purificadas, ou seja, MBP-XYL5 e MBP-XYL5/DOM respectivamente;
- a cauda fusionada MBP não impediu a atividade da enzima, o que pode ser uma vantagem em termos de custo de produção;
- o processo de purificação por cromatografia de afinidade à amilose contribuiu para a perda de atividade tanto de MBP-XYL5 como de MBP-XYL5/DOM;
- os dois métodos utilizados para a avaliação da atividade xilanolítica, a saber, o método do DNS e o método de coloração do meio em placas, foram eficientes.

7. PERSPECTIVAS

7.1. PROTEÍNA PWL2D

- Ensaios *in vivo* em plantas susceptíveis e em plantas resistentes;
- Validação do modelo proposto por meio de ensaios *in vivo*;
- Testes de cristalização com as proteínas PWL2 e PWL2D íntegras, bem como sem o peptídeo sinal, fusionadas à proteína fusionada TRX, visando à comparação das estruturas;
- Mutação dirigida e avaliação das estruturas protéicas consequentes.

7.2. PROTEÍNAS XYL5 E XYL5/DOM

- Testes considerando fatores influentes na atividade de xilanases como a estabilidade temporal da proteína, a concentração de substâncias (componentes dos tampões utilizados na extração e na purificação de proteínas como maltose, imidazol, phenylmethanesulfonylfluoride - PMSF, níquel) sabidamente influentes na atividade xilanolítica, o pH dos tampões, cofatores possíveis, termoestabilidade, etc;
- Testes de atividade com as frações purificadas considerando uma separação não-intencional entre um possível cofator e a enzima remanescente nas frações eluídas;

- Clonagem no vetor pSV268 que fornece a fusão à MBP e a clivagem desta cauda fusionada com trombina; expressão e purificação da proteína;
- Clonagem, expressão e purificação em diferentes vetores e diferentes linhagens de *Pichia pastoris*.

7.3. PROTEÍNA ABC1

- Clonagem e expressão em outros sistemas de expressão ainda não testados para esta proteína (e seus domínios) visando à solubilização.
- Obtenção da estrutura em cristal

8. CONCLUSÕES

Com base nos resultados obtidos neste trabalho, foi possível concluir que:

(1) A proteína PWL2D é super-expressa na forma solúvel se estiver sem o seu peptídeo sinal e for fusionada à cauda de tiorredoxina.

(2) Os resultados estruturais obtidos indicam que PWL2D possivelmente é uma proteína efetora intrinsecamente desordenada.

(3) Os resultados estruturais obtidos permitiram a elaboração de um modelo para explicar a mudança do estado de resistência do hospedeiro de *M. grisea* quando o patógeno fúngico carrega o gene *PWL2*, para o estado de susceptibilidade quando o patógeno fúngico carrega o gene *PWL2D*.

(4) A xilanase XYL5 e seu domínio catalítico apresentam atividade xilanolítica mesmo se fusionadas à MBP e em um sistema de expressão procariótico

(5) Esta é a primeira vez que se descreve a expressão e a caracterização estrutural de PWL2D, e a expressão e a caracterização funcional parcial de XYL5 – duas proteínas evolvidas na patogenicidade de *M. grisea*.

9. REFERÊNCIAS BIBLIOGRÁFICAS

Badhan, AK; Chadha, BS; Kaur, J; Saini, HS; Bhat, MK (2007) - Production of multiple xylanolytic and cellulolytic enzymes by thermophilic fungus *Myceliophthora* sp. *Bioresour Technol* 98:504–510.

Barth, S; Fischer, M; Schmid, RD; Pleiss, J (2004) - Sequence and structure of epoxide hydrolases: a systematic analysis. *Proteins: Structure, Function, and Bioinformatics* 55:846–855.

Bernhardt, TG; Roof, WD; Young, R (2000) - Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. *Proc Natl Acad Sci USA* 97(8):4297-302.

Berrin, JG; Juge, N (2008) - Factors affecting xylanase functionality in the degradation of arabinoxylans. *Biotechnol Lett* 30(7):1139-1150.

Blundell, TL; Johnson, LN (1976)– Crystallization of proteins. – In: Blundell, TL; Johnson, LN - *Protein crystallography.* Academic Press Inc. (London) LTD., p.59-82.

Caracuel-Rios, Z; Talbot, NJ (2007) - Cellular differentiation and host invasion by the rice blast fungus *Magnaporthe grisea*. *Curr Opin Microbiol*, 10(4):339-45.

Cereal Knowledge Bank (2010) – Rice. Disponível em: www.knowledgebank.irri.org. Acesso em Abril de 2010.

Collemare, J; Billard, A; Böhnert, HU; Lebrun, MH (2008) - Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity. *Mycol Res*. 112(Pt 2):207-15.

Collins, T; Gerday, C; Feller, G (2005) - Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev* 29:3–23.

Cortez, EV (1998) - recuperação de enzimas xilanolíticas por precipitação. *Dissertação de Mestrado,* FAENOUI*, Lorena, PR.*

De Wit, PJGM (2007) - How plants recognize pathogens and defend themselves. *Cell. Mol. Life Sci*. 64: 2726.

De Wit PJ; Mehrabi, R; Van den Burg, H; Stergiopoulos, I (2009) - Fungal effector proteins: past, present and future. *Mol Plant Pathol.* 10(6):735-47. Review.

Dean, RA; Talbot, NJ; Ebbole, DJ; Farman, ML; Mitchell, TK; Orbach, MJ; Thon, M; Kulkarni, R; Xu, JR; Pan, H; Read, ND; Lee, YH; Carbone, I; Brown, D; Oh, YY; Donofrio, N; Jeong, JS; Soanes, DM; Djonovic, S; Kolomiets, E; Rehmeyer, C; Li, W; Harding, M; Kim, S; Lebrun, MH; Bohnert, H; Coughlan, S; Butler, J; Calvo, S; Ma, LJ; Nicol, R; Purcell, S; Nusbaum, C; Galagan, JE; Birren, BW (2005) - The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434: 960-966.

Dodds, PN; Rafiqi, M; Gan, PH; Hardham, AR; Jones, DA; Ellis, JG (2009) - Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytol.* 183(4):993-1000.

Donofrio, NM; Oh, Y; Lundy, R; Pan, H; Brown, DE; Jeong, JS; Coughlan, S; Mitchell, TK; Dean, RA (2006) - Global gene expression during nitrogen starvation in the rice blast fungus, *Magnaporthe grisea*. *Fungal Genet. Biol.* 4: 605-617.

Dunker, AK; Brown, CJ; Lawson, JD; Iakoucheva, LM; Obradovic, Z (2002) - Intrinsic disorder and protein function. *Biochemistry* 41(21): 6573-6582.

Dunker, AK; Lawson, JD; Brown, CJ; Williams, RM; Romero, P; Oh, J S; Oldfield, CJ; Campen, AM; Ratliff, CM; Hipps, KW; Ausio, J; Nissen, MS; Reeves, R; Kang, CH; Kissinger, CR; Bailey, RW; Griswold, MD; Chiu, W; Garner, EC; Obradovic, Z (2001) - Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* 19: 26–59.

Dunker, AK; Oldfield, CJ; Meng, J; Romero, P; Yang, JY; Chen, JW; Vacic,V; Obradovic, Z; Uversky, VN (2008) - The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 2008, 9(Suppl 2):S1.

Ellis, JG; Rafiqi, M; Gan, P; Chakrabarti, A; Dodds, PN (2009) - Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Current Opinion in Plant Biology* 12:399–405.

Embrapa (2010) – Pragas e doenças: brusone. Disponível em: www.cnpaf.embrapa.br/arroz/pragasedoencas/brusone.htm. Acesso em Fevereiro de 2010.

Fengxia, L; Mei, L; Zhaoxin, L; Xiaomei, B; Haizhen, Z; Yi, W (2008) - Purification and characterization of xylanase from *Aspergillus ficuum* AF-98. *Bioresour Technol* 99:5938–5941.

Fourrey, CL; Talbot, NJ (2005) – Moving toward a systems biology approach to the study of fungal pathogenesis in the rice blast fungus *Magnaporthe grisea*. *Adv Appl Microbiol.* 57: 177-215.

Goriely, A; Tabor, M (2006) - Estimates of biomechanical forces in *Magnaporthe grisea. Mycol Res.* 110(Pt 7):755-9.

Gray, D; Subramanian, S (2000) - Choice of cellular protein expression system. *Current Protocols In Protein Science.* Proteins-Laboratory manuals: 5.16.1-5.16.34.

Gupta, A; Chattoo, BB (2007) - A novel gene MGA1 is required for appressorium formation in *Magnaporthe grisea*. *Fungal Genet Biol.* 44(11):1157-69.

Howard, RJ; Valent, B (1996) – Breaking and entering: Host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. *Annu. Rev. Microbiol*. (50): 491-512.

Howard, RJ; Valent, B. (1996) - Breaking and entering: host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. *Annu Rev Microbiol.* 50:491-512.

Howie, AJ; Brewer DB; Howell, D; Jones, AP (2008) - Physical basis of colors seen in Congo red-stained amyloid in polarized light. *Laboratory Investigation* 88: 232–242.

Iakoucheva, LM; Brown, CJ; Lawson, JD; Obradovic, Z ; Dunker, AK (2002) - Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J. Mol. Biol.* 323, 573–584.

Idnurm, A; Howlett, B (2001) – Pathogenicity genes of phytopathogenic fungi. *Mol. Plant Pathology* 2 (4): 241-255.

Jeffries, T (1996) - Biochemistry and genetics of microbial xylanases. *Curr Opin Biotechnol* 7:337–342.

Jo, YK; Wang, GL; Boehm, MJ (2007) - Expression analysis of rice defense-related genes in turfgrass in response to *Magnaporthe grisea. Phytopathology:* 97(2):170-8.

Jones, JDG; Dangl, JL (2006) - The plant immune system. *Nature* 444: 323-329.

Kadotani, N; Nakayashiki, H; Tosa, Y; Mayama, S (2003) - RNA silencing in the phytopathogenic fungus *Magnaporthe oryzae*. *Mol Plant Microbe Interact* 16(9):769-76.

Kahmann, R; Basse, C (2001) – Fungal gene expression during pathogenesis-related development and host plant colonization. *Curr. Op. Microbiol.* 4: 374-380.

Kang, S; Sweigard, JA; Valent, B (1995) - The *PWL* host specificity gene family in the blast fungus *Magnaporthe grisea*. *Mol Plant Microbe Interact* 8(6):939-48.

Kasana, RC; Salwan, R; Dhar, H; Dutt, S; Gulati, A (2008) - A rapid and easy method for the detection of microbial cellulases on agar plates using gram's iodine. - *Curr Microbiol.* 57(5):503-7.

Kershaw, MJ; Talbot, NJ (2009) - Genome-wide functional analysis reveals that infection-associated fungal autophagy is necessary for rice blast disease. *Proc Natl Acad Sci USA*. 106(37):15967-72.

Kim, S; Mohamedali, KA; Cheung, LH; Rosenblum, MG (2007) - Overexpression of biologically active VEGF121 fusion proteins in *Escherichia coli*. *J Biotechnol.* 128(3):638-47.

Knogge, W (1996) – Fungal infection of plants. *The Plant Cell* 8: 1711-1722.

Krogh, BL; von Heijne, GE; Sonnhammer, LL (2001) - Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305: 567-580.

Kulkarni, N; Shendye, A; Rao M (1999) - Molecular and biotechnological aspects of xylanases. *FEMS Microbiol Rev* 23:411–456.

Kwon, SW; Cho, YC; Kim, YG; Suh, JP; Jeung, JU; Roh, JH; Lee, SK,; Jeon, JS; Yang, SJ; Lee, YT (2008) - Development of near-isogenic Japonica rice lines with enhanced resistance to *Magnaporthe grisea*. *Mol Cells*. 25(3):407-416.

Laugé, R; De Wit, PJGM (1998) – Fungal avirulence genes: structure and possible functions. *Fungal Gen. Biol.* 24: 285-297.

Li, X; Romero, P; Rani, M; Dunker, AK; Obradovic, Z (1999) - Predicting protein disorder for N-, C-, and internal regions. *Genome Informatics* 10: 30-40.

Li, XH; Yang, HJ; Roy, B; Park, EY; Jiang, LJ; Wang, D; Miao, YG (2010) - Enhanced cellulase production of the *Trichoderma viride* mutated by microwave and ultraviolet. *Microbiol Res.*165(3): 190-198.

Lucas, JA (1998a) – The microbial pathogens. – In: Lucas, JA. *Plant pathology and plant pathogens.* Blackwell Science Ltd., Oxford, 3rd ed., p. 20- 29.

Lucas, JA (1998b) – Host- pathogen specificity. - In: Lucas, JA. *Plant pathology and plant pathogens.* Blackwell Science Ltd., Oxford, 3rd ed., p. 166-190.

Mach, RL; Zeilinger, S (2003) - Regulation of gene expression in industrial fungi: *Trichoderma*. *Appl Microbiol Biotechnol* 60:515–522.

Malakhov, MP; Mattern, MR; Malakhova, OA; Drinker, M; Weeks, SD; Butt, TR (2004) - SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* 5(1-2):75-86.

Mayer, M; Buchner, J (2004) - Refolding of inclusion body proteins. *Methods Mol Med*. 94:239-54.

McPherson, A (1999) - Principles of macromolecular structure and applications of Ray-X crystallography. – In: *Crystallization of biological macromolecules.* Cold Spring Harbor Laboratory Press, New York, p. 31-66.

Miller, GL (1959) - Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry* 31(30):426-428.

Mortimer, DN; Gilbert, J; Shepherd, MJ (1997) - Rapid and highly sensitive analysis of aflatoxins M1, in liquid and powdered milk using affinity column cleanup. *Journal of Chromagraphy* 407: 393-398.

Mosquera, G; Giraldo, MC; Khang, CH; Coughlan, S; Valent, B (2009) - Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. *Plant Cell* 21: 1273-1290.

Nallamsetty, S; Waugh, DS (2007) - A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His6-maltose binding protein fusion tag. *Nat Protoc.* 2(2):383-91.

Nishihara, K; Kanemori, M; Kitagawa, M; Yanagi, H; Yura, T (1998) - Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli. Appl Environ Microbiol.* 64(5):1694-9.

Oliveira, LA; Porto, AL; Tambourgi, EB (2006) - Production of xylanase and protease by *Penicillium janthinellum* CRC 87M-115 from different agricultural wastes. *Bioresour Technol*. 97(6):862-7. 2006

Orbach, MJ; Farrall, L; Sweigard, JA; Chumley, FG; Valent, B (2000) - A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell* 12(11):2019-32.

Pedersen, M; Lauritzen, HK; Frisvad, JC; Meyer, AS (2007) - Identification of thermostable beta-xylosidase activities produced by *Aspergillus brasiliensis* and *Aspergillus niger*. *Biotechnol Lett* 29:743–748.

pET System Manual (2006) - 10th edition. Disponível em www.novagen.com. Acesso em Janeiro 2010.

Polizeli, ML; Rizzatti, AC; Monti, R; Terenzi, HF; Jorge, JA; Amorim, DS (2005) - Xylanases from fungi: properties and industrial applications. *Appl Microbiol Biotechnol.* 67(5):577-91. Review.

Postel, S; Kemmerling, B (2009) - Plant systems for recognition of pathogen-associated molecular patterns. *Seminars in Cell & Developmental Biology* 20: 1025–1031.

QIAGEN Manual. QIAGEN, The Qiaexpressionist: A handbook for high-level expression and purification of 6xHis-tagged proteins - Fifth edition, 2003, pp.102. Hilden, Germany. Available at www.qiagen.com

Ribot, C; Hirsch, J; Balzergue, S; Tharreau, D; Nottéghem, JL; Lebrun, MH; Morel, JB (2008) - Susceptibility of rice to the blast fungus, *Magnaporthe grisea*. *Journal of Plant Physiology* 165(1):114-124.

Romero, P; Obradovic, Z; Li, X; Garner, E; Brown, C; Dunker, AK (2001) – Sequence complexity of disordered protein. *Proteins: Struct. Funct. Gen.* 42: 38-48.

Romero, P; Obradovic, Z; Li, X; Garner, E; Brown, C; Dunker, AK (2004) - Natively disordered proteins: functions and predictions. *Appl. Bioinformatics* 3: 105-113.

Roof, WD; Fang, HQ; Young, KD; Sun, J; Young, R (1997) - Mutational analysis of slyD, an *Escherichia coli* gene encoding a protein of the FKBP immunophilin family. *Mol Microbiol*. 25(6):1031-46.

Saleem, F; Ahmed, S; Jamil, A (2008) - Isolation of a xylan degrading gene from genomic DNA library of a thermophilic fungus *Chaetomium thermophile* ATCC 28076. *Pak J Bot* 40:1225–123.

Santos, A (2010) - Doenças do arroz: brusone. Disponível em www.uesb.br/utilitarios/modelos/monta.asp?site=fitopatologia&tex=iii_09_arroz.html. Acesso em Fevereiro de 2010.

Saunders, DG; Aves, SJ; Talbot, NJ (2010) - Cell cycle-mediated regulation of plant infection by the rice blast fungus. *Plant Cell* 22(2):497-507.

Sesma, A; Osbourn AE (2004) - The rice leaf blast pathogen undergoes developmental processes typical of root-infecting fungi. *Nature* 431: 582-586.

Shengcheng, W; Zhiying, Z; Darvill, AG; Albersheim, P (2001) - Xylanases of *Magnaporthe grisea*: differential requirement for pathogenicity. In: *Host-Parasite Interactions, XXI Fungal Genetics Conference*, Asilomar, California.

Skiner, W; Keon, J; Hargreaces, J (2001) – Gene information for fungal plant pathogens from expressed sequences. *Curr. Op. Microbiol.* 4: 381-386.

Stergiopoulos, I; De Wit PJ (2009) - Fungal effector proteins. *Annu Rev Phytopathol.* 47:233-63. Review.

Studier, FW; Moffatt, BA (1986) - Use of bacteriofage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*. 189: 113-130.

Sweigard, J; Caroll, AM; Kang, S; Chumley, FG; Valent, B (1995) - Identification, cloning and characterization of *PWL*2, a gene for host-species specificity in the rice blast fungus. *Plant Cell* 7: 1221-1233.

Sweigard, JA; Ebbole, DJ (2001) – Functional analysis of pathogenicity genes in a genomic world. *Curr. Op. Microbiol.*4:387-392.

Talbot, NJ (2003) - On The Trail Of A Cereal Killer**:** Exploring the Biology of *Magnaporthe grisea. Annu. Rev. Microbiol.* 57: 177–202.

Talbot, NJ; Ebbole, DJ; Hamer, JE (1993) – Identification and characterization of *MPG1*, a gene involved in a pathogenicity from the rice blast fungus *Magnaporthe grisea*. *Plant Cell* 5: 1575-1590.

Teather, RM, Wood, PJ (1982) - Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol.* 43(4): 777-80.

Thiede, B; Höhenwarter, W; Krah, A; Mattow, J; Schmid, M; Schmidt, F; Jungblut, PR (2005) - Peptide mass fingerprinting. *Methods* 35: 237-247.

Trabulsi, LR; Althertum, F; Gompertz, OF Candeias, J (2008) - Microbiologia, São Paulo: Atheneu.

Tresaugues, L; Collinet,B; Minard, P; Henckes, G; Aufrere, R; Blondeau, K; Liger, D; Zhou, CZ; Janin, J; Van Tilbeurgh, H; Quevillon-Cheruel, S (2004) - Refolding strategies from inclusion bodies in a structural genomics project. *J Struct Funct Genomics.* 5(3):195-204.

Tucker, SL; Talbot, NJ (2004) - Surface attachment and pre-penetration stage development by plant pathogenic fungi. *Annu. Rev. Phytopathol*. 39: 385–418.

Tudzynski, P; Sharon, A (2003) - Fungal pathogenicity genes. In: *Applied Mycology and Biotechnology*, *Vol. 3: Fungal Genomics* (Aurora, DK; Khachatourians, GG; eds), pp. 187–212. Amsterdam: Elsevier.

Uversky, VN (2002) - Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11: 739-756.

Valent, B; Chumley, FG (1991) - Molecular analysis of the rice blast fungus, *Magnaporthe grisea*. *Annu. Rev. Phytopathol,.* 29: 443-467.

Wu, SC; Halley, JE; Luttig, C; Fernekes, LM; Gutiérrez-Sanchez, G; Darvill, AG; Albersheim, P (2006) - Identification of an endo-beta-1,4-D-xylanase from *Magnaporthe grisea* by gene knockout analysis, purification, and heterologous expression. *Appl Environ Microbiol.* 72(2):986-993.

Yara, A; Yaeno, T; Montillet, JL; Hasegawa, M; Seo, S; Kusumi, K; Iba, K (2008) - Enhancement of disease resistance to *Magnaporthe grisea* in rice by accumulation of hydroxy linoleic acid. *Biochem Biophys Res Commun*. 370(2): 344-347.

Yoshida, K; Saitoh, H; Fujisawa, S; Kanzaki, H; Matsumura, H; Yoshida, K; Tosa, Y; Chuma, I,; Takano, Y; Win, J; Kamoun, S; Terauchi, R (2009) - Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae. Plant Cell* 21(5): 1573-91.

Zeigler, R.S (1998) - Recombination in *Magnaporthe grisea*. *Annual Review of Phytopphology* 36:249-276.

Zhang, GM; Huang, J; Huang, GR; Ma, LX; Zhang, XE (2007) - Molecular cloning and expression of a new xylanase gene from *Plectosphaerella cucmerina*. *Appl Microbiol Biotechnol* 74:339–346

Zhou, C; Bai, J; Deng, S; Wang, J; Zhu, J; Wu, M; Wang, W (2008) - Cloning of a xylanase gene from *Aspergillus usamii* and its expression in *Escherichia coli. Bioresour Technol* 99:831–838.

10. ANEXOS

10.1. ARTIGO I

"Molecular and cytogenetic characterization of na AT-rich satellite DNA family in Urvillea chacoensis Hunz. (Paullinieae, Sapindaceae)"

Autores: Juan D. Urdampilleta; Anete Pereira de Souza; **Dilaine R.S. Schneider**; André L.L. Vanzela; Maria F. Ferrucci; Eliana R.F. Martins

Publicado no periódico **Genetica (2009) 136: 171-177**

Genetica (2009) 136:171–177
DOI 10.1007/s10709-008-9332-0

Molecular and cytogenetic characterization of an AT-rich satellite DNA family in *Urvillea chacoensis* Hunz. (Paullinieae, Sapindaceae)

Juan D. Urdampilleta · Anete Pereira de Souza · Dilaine R. S. Schneider · André L. L. Vanzela · María S. Ferrucci · Eliana R. F. Martins

Received: 14 February 2008 / Accepted: 1 November 2008 / Published online: 25 November 2008

Abstract *Urvillea chacoensis* is a climber with $2n = 22$ and some terminal AT-rich heterochromatin blocks that differentiate it from other species of the genus. The AT-rich highly repeated satellite DNA was isolated from *U. chacoensis* by the digestion of total nuclear DNA with *Hin*dIII and *Xba*I and cloned in *Escherichia coli*. Satellite DNA structure and chromosomal distribution were investigated. DNA sequencing revealed that the repeat length of satDNA ranges between 721 and 728 bp, the percentage of AT-base pairs was about 72–73% and the studied clones showed an identity of 92.5–95.9%. Although this monomer has a tetranucleosomal size, direct imperfect repetitions of ~180 bp subdividing it in four nucleosomal subregions were observed. The results obtained with FISH indicate that this monomer usually appears distributed in the terminal regions of most chromosomes and is associated to heterochromatin blocks observed after DAPI staining. These observations are discussed in relation to the satellite DNA evolution and compared with other features observed in several plant groups.

Keywords DAPI · FISH · Heterochromatin · Karyotype · Satellite DNA · *Urvillea*

J. D. Urdampilleta (✉) · E. R. F. Martins
Laboratório de Biossistematica, Departamento de Botânica, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal: 6109, Campinas, SP CEP 13083-970, Brazil
e-mail: juanurdampilleta@hotmail.com

A. P. de Souza · D. R. S. Schneider
Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil

A. L. L. Vanzela
Laboratório de Biodiversidade e Restauração de Ecossistemas, Departamento de Biologia Geral, CCB, Universidade Estadual de Londrina, Londrina, Parana, Brazil

M. S. Ferrucci
Instituto de Botánica del Nordeste, Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste, Corrientes, Argentina

Introduction

Urvillea Kunth (Paullinieae, Sapindaceae) comprises 17 species and is widely distributed in the neotropical regions of America, from southern United States to northern Argentina (Ferrucci 2006). *Urvillea* has two basic chromosome numbers, $x = 12$ (*Stenelytron*), with a single diploid level, and $x = 11$ in section *Urvillea*, with diploid, tetraploid, and octaploid levels (Lombello and Forni-Martins 1998; Ferrucci 2000; Urdampilleta et al. 2006). *Urvillea chacoensis* Hunz. is a climber that occurs in "chaqueña" regions from Argentina, Brazil and Paraguay. As the other species of the *Urvillea* section, its diploid number is $2n = 22$ (Ferrucci 2000; Urdampilleta et al. 2006), but it is cytogenetically differentiated by the presence of large terminal AT-rich heterochromatin blocks and three chromosome pairs bearing NOR (Urdampilleta et al. 2006).

Highly repetitive DNA appears to be one of the main components of plant genomes, and changes in these sequences may be responsible for the variations in genome size and karyotypical features (Flavell 1986; San Miguel and Bennetzen 1998). Different types of repetitive DNA exist within each genome, and the satellite DNA (satDNA), which is formed by tandemly arranged monomers of tens to thousands base pairs (Charlesworth et al. 1994), often constitutes heterochromatin blocks up to 100 Mb.

Additionally, the whole heterochromatin of a genome can be composed of different families of satellite DNA (Kurbis et al. 1998; Sharma and Raina 2005), which can also be located or accumulated in specific chromosomal positions (Flavell 1986). Some new sequences of satDNA have recently been reported and a database was created to understand the evolution and distribution of satDNA in plant genomes (Macas et al. 2002). Since satDNA can be species-specific or typical to some groups of species (Schmidt et al. 1991; Svitashev et al. 1994), it represents a useful tool in the field of taxonomy.

Different techniques can be used to isolate satellite DNA, as centrifugation in density gradients (Deumling 1981), screening of fractions referred to as low Cot DNA (Neumann et al. 2001; Ho and Leung 2002), purification and cloning of band fragments obtained by gel electrophoresis of enzyme-digested genomic DNA (Kato et al. 1984; Ganal and Hemleben 1986; Lakshmikumaran and Ranade 1990; Lorite et al. 2001) and genomic self-priming PCR (GSP-PCR) (Buntjer and Lenstra 1998; Macas et al. 2000). Since we intended to characterize and localize a satellite DNA family of *U. chacoensis* in situ, we isolated and characterized satDNA sequences by purifying and cloning fragments of enzyme-digested genomic DNA. To do so, this marker was FISH mapped and the results were matched with the distribution of 45S rDNA loci and the heterochromatin pattern obtained with DAPI.

Materials and methods

Plant material

Seeds of *U. chacoensis* obtained from the voucher specimen *U. chacoensis*: Bolivia. Dpto. Chuquisaca. Prov. Calvo, 22 km N de Tarairí camino a Boyuibe, 18VII 2003, Ferrucci et al. N° 1763, deposited at the herbarium of the Instituto de Botánica del Nordeste (CTES), Argentina, were collected and cultivated in pots under greenhouse conditions.

Isolation of genomic DNA, cloning and sequencing of repetitive DNA fragments

Leaf genomic DNA was isolated according to the method described by Hoisington et al. (1994) and digested with *Hin*dIII and *Xba*I. The selected bands of about 750 bp were eluted and purified from the agarose gel. The fragments of repetitive DNA were cloned using *Escherichia coli* DH5α as host and pBluescript KS plus (Stratagene) as vector. Colonies containing recombinant plasmids were identified for selection on LB agar medium supplemented with X-gal and IPTG. Recombinant plasmid was isolated using alkaline minipreparation procedure and the insert nucleotide sequences were determined with an ABI377 automated DNA sequencer (Applied Biosystems). Sequences were analysed with Lasergene 7 (DNAStar, Madison, WI, USA) and aligned by using the ClustalW option of the MegAlign program.

Preparation of mitotic chromosomes

Root tips were pretreated with 2 mM 8-hydroxyquinolin for 4–5 h at 15°C, fixed in ethanol–acetic acid (3:1, v:v) and digested at 37°C in a solution composed of 2% cellulase and 20% pectinase. After squashing the meristems in a drop of 45% acetic acid, the preparations were frozen in liquid nitrogen and the coverslips were removed. To identify AT-rich heterochromatin blocks, slides were stained with 2 mg ml^{-1} DAPI for 30 min and mounted with glycerol/McIlvaine buffer pH 7.0, 1:1 (v:v), plus 2.5 mM $MgCl_2$.

Fluorescent in situ hybridization

To compare the distribution of satDNA sequences in relation to genes of rDNA and to DAPI heterochromatin pattern, we used probes marked with nick translation. Recombinant plasmids isolated by alkaline minipreparation containing 18S-5.8S-26S rDNA of wheat (p*Ta*71) (Gerlach and Bedbrook 1979) was labeled with biotin-14-dUTP (Bionick, Invitrogen) and the cloned fragments of satDNA were labeled with DIG (DIG Nick translation mix, Roche). In situ hybridization followed the protocols of Heslop-Harrison et al. (1991) and Schwarzacher and Heslop-Harrison (2000). Slides were incubated in 100 μg ml^{-1} RNAse, post-fixed in 4% (w/v) paraformaldehyde, dehydrated in a 70–100% graded ethanol series and air-dried. Later, 30 μl of the hybridization mixture (4–6 ng μl^{-1} of probe, 50% (v/v) formamide, 10% (w/v) dextran sulfate, 3.3 ng μl^{-1} of calf thymus DNA, 2× SSC and 0.3% (w/v) SDS), previously denatured at 70°C for 10 min, were applied. Samples were denatured/hybridized at 90°C for 10 min, 48°C for 10 min and 38°C for 5 min, using a thermal cycler (Mastercycler, Eppendorf) and the slides were kept overnight in a humid chamber at 37°C. Hybridization signals were detected with avidin-FITC (Sigma) for p*Ta*71 and anti-DIG-Rhodamine (Roche) for satDNA. The slides were counterstained with DAPI and mounted with 25 μl of VectaShield (Vector Laboratories). Photomicrographies were obtained with a BX51 Olympus coupled to an Evolution MT CCD photosystem and Image ProPlus v6 software was used to capture the images.

Results

Isolation of *Uch*725 satellite DNA

The genomic DNA of *U. chacoensis* digested with *Hin*dIII and *Xba*I showed clear electrophoretic bands at ~750 and 1,500 bp (Fig. 1), indicating the presence of tandemly arranged repetitive elements in the genome. Such band pattern suggested the presence of satDNA with a site for *Hin*dIII and another one for *Xba*I in the same repetitive units. The 750 bp repetitive DNA fragments cloned and sequenced represent monomers of satDNA with ~725 bp, since they ranged between 721 and 728 bp in all the studied clones. The p*Uch*1, p*Uch*6 and p*Uch*8 clones contained the monomer element digested with *Hin*dIII, and the p*Uch*11, p*Uch*13 and p*Uch*15 clones contained the monomer element digested with *Xba*I. Figure 2 shows the aligned nucleotide sequences of the studied clones and a sequence analysis revealed some differential characteristics of these sequences. The satDNA family was thus named *Uch*725. This repetition unit possesses a restriction site for *Hin*dIII and another one for *Xba*I, separated by 53–54 bp, which allows to obtain fragments of equal size after the digestion of the genomic DNA. The analyzed sequences were highly AT-rich, with 72.1–73.2% of A or T, and an identity of 92.5–95.9% was observed in the studied clones.

Although no significant inverted repetition was detected within the repetition units, we observed direct imperfect repetitions of ~180 bp. These subrepetitions allow to divide the monomer in four sub-regions (Fig. 2). The alignment of these sub-regions (Fig. 3a) showed an identity of about 70–80% (Fig. 3b).

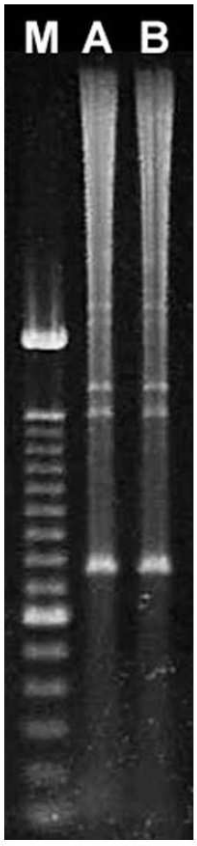


Fig. 1 Restriction enzyme analysis of total genomic DNA of *U. chacoensis*. Genomic nuclear DNA digested with *Hin*dIII (A) and *Xba*I (B) and restriction fragments resolved on 1% agarose gel. M, 100 bp leader

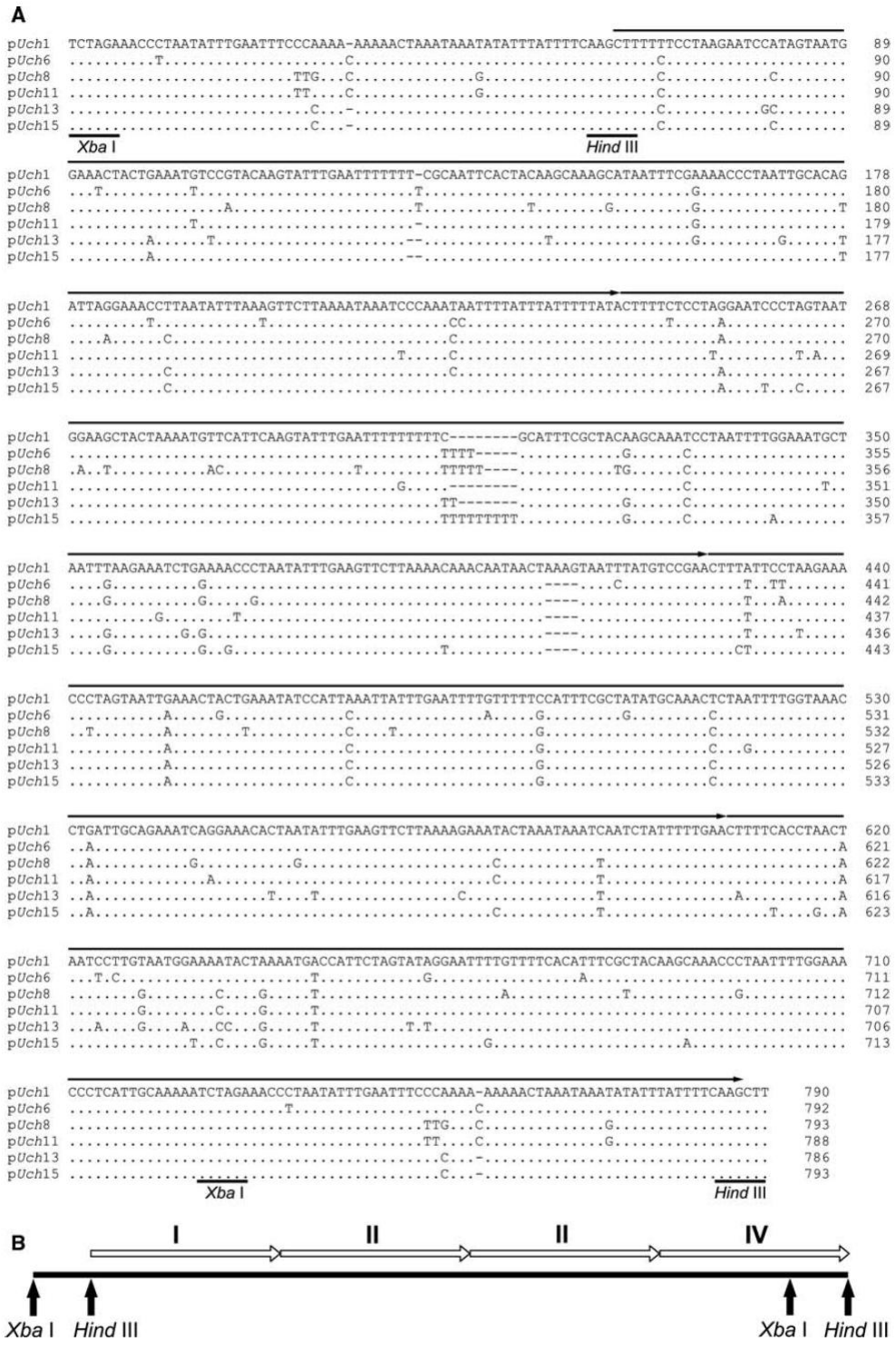
Karyotypical features in *U. chacoensis*

Urvillea chacoensis, with $2n = 22$ chromosomes, which are mostly metacentrics, showed several large terminal heterochromatic blocks after DAPI staining (Fig. 4a). The six longest chromosome pairs showed heterochromatin blocks in both terminal regions, while the smallest chromosomes showed small bands at their tips or no band at all. This banding pattern indicates the presence of a high proportion of AT-rich heterochromatin, usually located in the subtelomeric regions. FISH using the p*Ta*71 probe (18-5.8-26S rDNA) located terminal signals in the short arms of three chromosome pairs (five sites were observed in Fig. 4b, d). $DAPI^+$ bands were not observed in the short arms in NOR carrier chromosomes.

The DAPI banding pattern coincided with the hybridization sites of the p*Uch*6 probe (Fig. 4). The hybridization signals with the p*Uch*6 probe were detected in the two terminal regions of the six largest chromosome pairs. Two chromosome pairs exhibited p*Uch*6 signals in the terminal region of the long arm, and of three small chromosome pairs, only one showed reliable hybridization signals in both terminal regions. Some chromosomes presented hybridization signals with both the p*Uch*6 and p*Ta*71 probes, but p*Ta*71 hybridized at terminal regions of short arms, while p*Uch*6 hybridized at long arms (Fig. 4b–d).

Discussion

The results obtained with DAPI banding and in situ hybridization with the p*Ta*71 probe coincide with the observations by Urdampilleta et al. (2006) for *U. chacoensis*. In addition, in our study, the coincident DAPI banding pattern and FISH with satDNA probes suggest that this satellite DNA represents a structural component of the heterochromatin blocks within the *U. chacoensis* genome. Our results confirmed the presence of a high proportion of subterminal AT-rich heterochromatin with an equilocal distribution.


A
pUch1 TCTAGAAACCCTAATATTTGAATTTCCCAAAA-AAAAACTAAATAAATATATTTATTTTCAAGCTTTTTTCCTAAGAATCCATAGTAATG 89
pUch6 .....T.....C.....C..... 90
pUch8 .....TTG...C.....G.....C.....C..... 90
pUch11 .....TT....C.....G.....C..... 90
pUch13 .....C...-.....C.....GC..... 89
pUch15 .....C...-.....C.....C..... 89
Xba I
Hind III
pUch1 GAAACTACTGAAATGTCCGTACAAGTATTTGAATTTTTTT-CGCAATTCACTACAAGCAAAGCATAATTTCGAAAACCCTAATTGCACAG 178
pUch6 ...T.....T.....T.....G..... 180
pUch8 .....A.....T.....T.....G.....G.....T 180
pUch11 .....T.....-.....G..... 179
pUch13 .....A.....T.....--.....T.....G.....G.....T 177
pUch15 .....A.....--.....T 177
pUch1 ATTAGGAAACCTTAATATTTAAAGTTCTTAAAATAAATCCCAAATAATTTTATTTATTTTTATACTTTTCTCCTAGGAATCCCTAGTAAT 268
pUch6 .....T.....T.....CC.....T.....A..... 270
pUch8 ....A.....C.....C.....A..... 270
pUch11 .....T.....C.....T.....T.A... 269
pUch13 .....C.....C.....A..... 267
pUch15 .....C.....A....T...C..... 267
pUch1 GGAAGCTACTAAAATGTTCATTCAAGTATTTGAATTTTTTTTTC--------GCATTTCGCTACAAGCAAATCCTAATTTTGGAAATGCT 350
pUch6 .....TTTT-----.....G.....C..... 355
pUch8 .A..T.....AC.....T.....TTTTT----.....TG.....C..... 356
pUch11 .....G.....-----.....C.....T.. 351
pUch13 .....TT-------.....G.....C..... 350
pUch15 .....TTTTTTTTT.....G.....C.....A..... 357
pUch1 AATTTAAGAAATCTGAAAACCCTAATATTTGAAGTTCTTAAAACAAACAATAACTAAAGTAATTTATGTCCGAACTTTATTCCTAAGAAA 440
pUch6 ....G.....G.....----....C.....T..TT..... 441
pUch8 ....G.....G.....G.....----.....T...A..... 442
pUch11 .....G.....T.....----.....T..... 437
pUch13 ....G.....G.G.....----.....T.....T..... 436
pUch15 ....G.....G..G.....T.....----.....CT..... 443
pUch1 CCCTAGTAATTGAAACTACTGAAATATCCATTAAATTATTTGAATTTTGTTTTTCCATTTCGCTATATGCAAACTCTAATTTTGGTAAAC 530
pUch6 .....A.....G.....C.....A.....G.....G.....C..... 531
pUch8 ..T.....A.....T.....C....T.....G.....C..... 532
pUch11 .....A.....C.....G.....C...G..... 527
pUch13 .....A.....C.....G.....C..... 526
pUch15 .....A.....C.....G.....C..... 533
pUch1 CTGATTGCAGAAATCAGGAAACACTAATATTTGAAGTTCTTAAAAGAAATACTAAATAAATCAATCTATTTTTGAACTTTTCACCTAACT 620
pUch6 ..A.....A 621
pUch8 ..A.....G.....G.....C.....T.....A 622
pUch11 ..A.....A.....C.....T.....A 617
pUch13 ..A.....T....T.....C.....T.....A.....A 616
pUch15 ..A.....C.....T.....T....G..A 623
pUch1 AATCCTTGTAATGGAAAATACTAAAATGACCATTCTAGTATAGGAATTTTGTTTTCACATTTCGCTACAAGCAAACCCTAATTTTGGAAA 710
pUch6 ...T.C.....T.....G.....A..... 711
pUch8 .....G.....C....G.....T.....A.....T.....G..... 712
pUch11 .....G.....C....G.....T..... 707
pUch13 ...A....G....A...CC...G.....T.....T.T..... 706
pUch15 .....T..C....G.....T.....G.....A..... 713
pUch1 CCCTCATTGCAAAAATCTAGAAACCCTAATATTTGAATTTCCCAAAA-AAAAACTAAATAAATATATTTATTTTCAAGCTT 790
pUch6 .....T.....C..... 792
pUch8 .....TTG...C.....G..... 793
pUch11 .....TT....C.....G..... 788
pUch13 .....C...-..... 786
pUch15 .....C...-..... 793
Xba I
Hind III
B
I
II
II
IV
Xba I
Hind III
Xba I
Hind III

◀ **Fig. 2** Sequence analysis of the studied clones. **a** Alignment of the six clones, p*Uch*1, p*Uch*6, p*Uch*8, p*Uch*11, p*Uch*13 and p*Uch*15. Imperfect direct repeats are indicated by *arrows*. **b** satDNA monomer scheme showing subunit with *arrows*

According to Urdampilleta et al. (2006), some karyotypical features in *U. chacoensis* are important and allow to distinguish this species from other species of the *Urvillea* genus (Urdampilleta et al. 2006). *Urvillea laevis* and *U. filipes*, for example, do not present C-bands, but these authors reported small, scarce DAPI$^+$ sites in *U. laevis*. On the other hand, *U. ulmacea* showed abundant terminal C-bands constituted by $CMA_3{}^+$, and DAPI$^+$ bands in several chromosomes. GC-rich terminal regions ($CMA_3{}^+$) may be associated to NORs in *U. chacoensis*, but not in *U. ulmacea* (Urdampilleta et al. 2006). These results also confirm that the heterochromatin is not homogeneous and can vary both qualitatively and quantitatively among the

Fig. 3 Alignment of all four subrepetitions of p*Uch*6 (**a**) and percent identity of sequences of each sub-region (**b**)

A

```
pUch6 I   CTTTTCTCCTAAGAATCCATAGTAATGGAATCTACTGAAATTTCCGTACAAGTATTTGAATT 62
pUch6 II  .....T...........C..........G.....A....G.T.A.T.............. 62
pUch6 III .....T.TT......A..C.......TA..A..G.......A...A.T...T.......... 62
pUch6 IV  ......A.....C..ATTCCT.........AA....A....G...A.T.T.....GG..... 62

pUch6 I   TT-TTTT---CGCAATTCACTACAAGCAAAGCATAATTTCGGAAACCCTAATTGCACAGATT 120
pUch6 II  ..T....TTTT...T...G....G......C.C......T.....TG.......A.G.A..C 124
pUch6 III ..A....---T...T...G..GT.T.....C.C......T..T..A..........G.A..C 121
pUch6 IV  ..G....---.A......G..........C.C......T.........C......A.A..C 121

pUch6 I   AGGAAATCTTAATATTTAATGTTCTTAAAATAAATCCCAAACCATTTTATTTATTTTTATA  181
pUch6 II  T.....C.C........G.A..........C...CA----.TA.C..A..C...G.CCGA.  181
pUch6 III ......CAC........G.A..........G....A.T...TA.A.CA..C.......GA.  182
pUch6 IV  TA....C..........G.AT...CC....C...AA.T...TA.A.A..........C.AG  182
```

B

	I	II	III	IV
I	***	70.2	75.1	76.2
II		***	79.6	76.2
III			***	74.7
IV				***

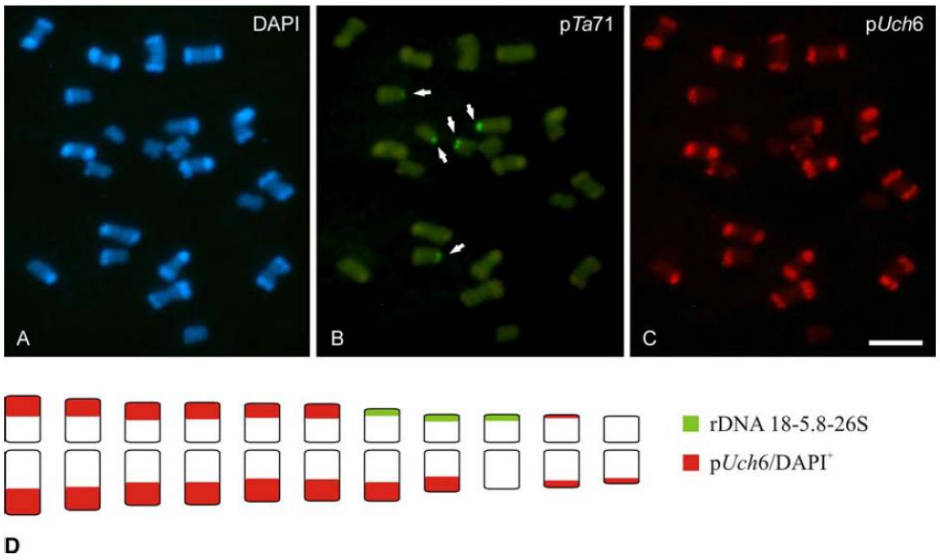


Fig. 4 Chromosome banding and fluorescence in situ hybridization in metaphasic chromosome of *U. chacoensis*. **a** AT-rich heterochromatin terminal blocks observed with DAPI staining; **b** localization of rDNA 18-5.8-26S using p*Ta*71 probe; **c** terminal distribution of *Uch*725 family observed for FISH with p*Uch*6 probe; and **d** idiograms by indicating the localization of rDNA and satDNA. *Bar* = 5 μm

species, within a same group (Guerra 2000), valuing these features as taxonomic markers.

The occurrence of AT-rich terminal repeated sequences is a frequent characteristic in plant genomes (Flavell 1986; Kurbis et al. 1998; San Miguel and Bennetzen 1998). The sequence analysis of plant genomes also emphasizes that ~60% of the satDNA families studied are AT-rich sequences (Macas et al. 2002) and the comparison of banding patterns in different plant groups suggests that the proportion of species with AT-rich heterochromatin is even higher (Guerra 2000). These data thus suggest that the distribution and chemical composition of these sequences are not casual and might be related to still unknown functions of the genome structure.

In plants, the size of the satellite DNA usually varies between 135–195 and 315–375 bp, corresponding to the length of mono and dinucleosomes (Macas et al. 2002; Sharma and Raina 2005). However, some plants have satellite DNA with monomers longer that 600 bp. *Secale cereale*, for example, presents satDNA with units of repetition of 3,900 bp (Langdon et al. 2000). In *U. chacoensis*, monomers of the *Uch*725 family show a size similar to that of the satDNA found in *Aegilops speltoides* (Anamthawat-Jonsson and Heslop-Harrison 1993), *Allium fistulosum* (Seo et al. 2007), *Rumex acetosa* (Shibata et al. 2000), *Sinapis arvensis* (Kapila et al. 1996). Nevertheless, no homology was found in the GenBank and EMBL databases for *Uch*725, indicating that these sequence may be typical of *Urvillea* group.

Heterochromatin was first identified by Heitz (1928) as a cytological event to describe condensed regions in the interphase. Although these structures have long been considered to be DNA without any function, different functions have been attributed to them these last decades (Yunis and Jasmineh 1971; Sýkorová et al. 2001), which guarantees their maintenance. Both in mammalian and plant cells, bulk chromatin presents a nucleosome periodicity of 180 ± 5 bp (Fajkus et al. 1995; Vershinin and Heslop-Harrison 1998) and the satDNA monomers often correspond to mono or dinucleosomes (135 and 195 or 315 and 375 bp) (Macas et al. 2000; Sharma and Raina 2005). Therefore, the satDNA could provide structural genetic codes for the chromatin packing (Trifonov 1989). Sýkorová et al. (2001) found that satDNA sequences favor the transition between telomeric domains and internal chromosomal regions, acting directly in the telomeric stabilization and regulation of genes from subterminal regions. As in *U. chacoensis*, various families of repetitive DNA occur in the subterminal regions of plants and constitute the so called telomere-associated sequences or TASs (Sharma and Raina 2005). In *U. chacoensis*, the presence of four subrepetitions of ~180 bp (tetranucleosome) within the satDNA monomer suggests that this substructure may be directly related to the establishment and packing of chromatin in terminal regions.

The phylogeny within Sapindaceae is not well resolved, since the tribal classification of many genera is confuse (Harrington et al. 2005). Studies on the presence and distribution of satellite DNA can contribute to understand the evolutionary aspects of the genome and thus to establish the taxonomy of some groups (Schmidt et al. 1991; Svitashev et al. 1994). The p*Uch*6 probe, isolated from *U. chacoensis*, has being previously tested in other species of *Urvillea* and genera of Sapindaceae, tribe Paullinieae. For the time being no signals were displayed after of hybridization with p*Uch*6 probe in *Cardiospermum grandiflorum* Sw., *Paullinia elegans* Cambess. and *U. ulmacea* Kunth. However by mean of PCR and by using primers designed from p*Uch*6 were detected some products possibly related to *Uch*725 (Urdampilleta et al. in preparation). As the genomic studies on this family are scarce, the isolation of *Uch725* satellite DNA in *U. chacoensis* offers an important chromosomal marker, whose presence and distribution in related species and genera might contribute to the systematic of Paullinieae.

Acknowledgments The authors are grateful to the Brazilian agencies CNPq, CAPES, and FAPESP for financial support.

References

Anamthawat-Jonsson K, Heslop-Harrison JS (1993) Isolation and characterization of genome-specific DNA sequences in *Triticeae* species. Mol Gen Genet 240:151–158. doi:10.1007/BF00277052

Buntjer JB, Lenstra JA (1998) Self-amplification of satellite DNA in vitro. Genome 41:429–434. doi:10.1139/gen-41-3-429

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220. doi:10.1038/371215a0

Deumling B (1981) Sequence arrangement of a highly methylated satellite DNA of a plant, *Scilla*: a tandemly repeated inverted repeat. Proc Natl Acad Sci USA 78:338–342. doi:10.1073/pnas.78.1.338

Fajkus J, Kovarik A, Kralovics R, Bezdek M (1995) Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. Mol Gen Genet 247:633–638. doi:10.1007/BF00290355

Ferrucci MS (2000) Cytotaxonomy of *Sapindaceae* with special reference to the *Tribe Paullinieae*. Genet Mol Biol 23:941–946. doi:10.1590/S1415-47572000000400039

Ferrucci MS (2006) A new species of *Urvillea* (Sapindaceae) from northwestern Venezuela. Brittonia 58:83–87. doi:10.1663/0007-196X(2006)58[83:ANSOUS]2.0.CO;2

Flavell RB (1986) Repetitive DNA and chromosome evolution in plants. Philos Trans R Soc B-Biol Sci 312:227–242. doi:10.1098/rstb.1986.0004

Ganal M, Hemleben V (1986) Different AT-rich satellite DNAs in *Cucurbita pepo* and *Cucurbita maxima*. Theor Appl Genet 73:129–135. doi:10.1007/BF00273729

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885. doi:10.1093/nar/7.7.1869

Guerra M (2000) Patterns of heterochromatin distribution in plant chromosomes. Genet Mol Biol 23:1029–1041

Harrington MG, Edwards KJ, Johnson SA, Chase MW, Gadek PA (2005) Phylogenetic inference in Sapindaceae *sensu lato* using plastid *matK* and *rbcL* DNA sequences. Syst Bot 30:366–382. doi:10.1600/0363644054223549

Heitz E (1928) Das heterochromatin der moose. I. Jahrb Wiss Bot 69:762–818

Heslop-Harrison JS, Schwarzacher T, Anamthewat Jonsson K, Leitch AR, Shi M, Leitch IJ (1991) In situ hybridization with automated chromosome denaturation. Technique 3:106–109

Ho ISH, Leung FC (2002) Isolation and characterization of repetitive DNA sequences from *Panax ginseng*. Mol Genet Genomics 266:951–961. doi:10.1007/s00438-001-0617-6

Hoisington D, Khairallah M, González De León D (1994) Laboratory protocols—CIMMYT applied molecular genetics laboratory, 2nd edn. CIMMYT, México

Kapila R, Das S, Srivastava PS, Lakshmikumaran M (1996) A novel species-specific tandem repeat DNA family from *Sinapis arvensis*: detection of telomere-like sequences. Genome 39: 758–766. doi:10.1139/g96-095

Kato A, Yakura K, Tanifuji S (1984) Sequence analysis of *Vicia faba* repeated DNA, the FokI repeat element. Nucleic Acids Res 12:6415–6426. doi:10.1093/nar/12.16.6415

Kurbis S, Schimdt T, Heslop Harrison JS (1998) Repetitive DNA elements as major componenet of plant genomes. Ann Bot (Lond) 82:45–55. doi:10.1006/anbo.1998.0779

Lakshmikumaran M, Ranade SA (1990) Isolation and characterization of a highly repetitive DNA of *Brassica campestris*. Plant Mol Biol 14:447–448. doi:10.1007/BF00028781

Langdon T, Seago C, Jones RN, Ougham H, Thomas H, Forster JW, Jenkins G (2000) *De Novo* evolution of satellite DNA on the rye B chromosome. Genetics 154:869–884

Lombello AR, Forni-Martins ER (1998) Chromosomal studies and evolution in Sapindaceae. Caryologia 51:81–93

Lorite P, Garcia MF, Carrillo JA, Palomeque T (2001) A new repetitive DNA sequence family in the olive (*Olea europaea* L.). Hereditas 134:73–78. doi:10.1111/j.1601-5223.2001.00073.x

Macas J, Požárková D, Navrátilová A, Nouzová M, Neumann P (2000) Two new families of tandem repeats isolated from genus *Vicia* using genomic self-priming PCR. Mol Gen Genet 263:741–751. doi:10.1007/s004380000245

Macas J, Mészáros T, Nouzová M (2002) PlantSat: a specialized database for plant satellite repeats. Bioinformatics 18:28–35. doi:10.1093/bioinformatics/18.1.28

Neumann P, Nouzová M, Macas J (2001) Molecular and cytogenetic analysis of repetitive DNA in pea (*Pisum sativum* L.). Genome 44:716–728. doi:10.1139/gen-44-4-716

San Miguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot (Lond) 82:37–44. doi: 10.1006/anbo.1998.0746

Schmidt T, Jung C, Metzlaff M (1991) Distribution and evolution of two satellite DNAs in the genus *Beta*. Theor Appl Genet 82:793–799. doi:10.1007/BF00227327

Schwarzacher T, Heslop-Harrison P (2000) Practical in situ hybridization, 2nd edn. Bios, Oxford

Seo BB, Lee SH, Do GS (2007) Identification and nucleotide sequence of highly repeated DNA sequence using combined RAPD and FISH in *Allium fistulosum*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=3550125 21/2/2007

Sharma S, Raina SN (2005) Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. Cytogenet Genome Res 109:15–26. doi:10.1159/000082377

Shibata F, Hizume M, Kuroki Y (2000) Molecular cytogenetic analysis of supernumerary heterochromatic segments in *Rumex acetosa*. Genome 43:391–397. doi:10.1139/gen-43-2-391

Svitashev S, Bryngelsson T, Vershinin A, Pedersen C, Säll T, von Bothmer R (1994) Phylogenetic analysis of the genus *Hordeum* using repetitive DNA sequences. Theor Appl Genet 89:801–810. doi:10.1007/BF00224500

Sýkorová E, Fajkus J, Ito M, Fukui K (2001) Transition between two forms of heterochromatin at plant subtelomeres. Chrom Res 9:309–323. doi:10.1023/A:1016698713959

Trifonov EN (1989) The multiple codes of nucleotide sequences. Bull Math Biol 51:417–432

Urdampilleta JD, Ferrucci MS, Torezan JMD, Vanzela ALL (2006) Karyotype relationships among four south American species of *Urvillea* (Sapindaceae: Paullinieae). Plant Syst Evol 258:85–95. doi:10.1007/s00606-005-0393-6

Vershinin AV, Heslop-Harrison JS (1998) Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. Plant Mol Biol 36:149–161. doi:10.1023/A:1005912822671

Yunis JJ, Jasmineh WG (1971) Heterochromatin, satellite DNA, and cell function. Science 174:1200–1209. doi:10.1126/science.174.4015.1200

10.2. ARTIGO II

"Functional and small-angle X-ray scattering studies of a new stationary phase survival protein E (SurE) from *Xylella fastidiosa* – evidence of allosteric behaviour"

Autores: Antonio M. Saraiva; Marcelo A. Reis; Suselly Tada; Luciana K. Rosselli-Murai; **Dilaine R.S. Schneider**; Alexandre C. Pelloso; Marcelo A.S. Toledo; Carlos Giles; Ricardo Aparício; Anete Pereira de Souza

Publicado no periódico **FEBS Journal (2009) 276(22): 6751-62**

the FEBS Journal

Functional and small-angle X-ray scattering studies of a new stationary phase survival protein E (SurE) from *Xylella fastidiosa* – evidence of allosteric behaviour

Antonio M. Saraiva[1], Marcelo A. Reis[2,3], Susely F. Tada[1], Luciana K. Rosselli-Murai[1], Dilaine R. S. Schneider[1], Alexandre C. Pelloso[1], Marcelo A. S. Toledo[1], Carlos Giles[3], Ricardo Aparicio[2] and Anete P. de Souza[1]

1 Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas, Brazil
2 Instituto de Química, Universidade Estadual de Campinas, Brazil
3 Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Brazil

Keywords
allosteric behaviour; small-angle X-ray scattering; stationary phase survival protein E (SurE); SurE oligomeric state; *Xylella fastidiosa*

Correspondence
A. P. de Souza, Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas, Campinas São Paulo, Brazil
Fax: +55 19 3521 1089
Tel: +55 19 3521 1132
E-mail: anete@unicamp.br

(Received 5 June 2009, revised 11 August 2009, accepted 18 September 2009)

doi:10.1111/j.1742-4658.2009.07390.x

The genome data of bacterium *Xylella fastidiosa* strain 9a5c has identified several *orfs* related to its phytopathogenic adaptation and survival. Among these genes, the *surE* codifies a survival protein E (XfSurE) whose function is not so well understood, but functional assays in *Escherichia coli* revealed nucleotidase and exopolyphosphate activity. In the present study, we report the XfSurE protein overexpression in *E. coli* and its purification. The overall secondary structure was analyzed by CD. Small-angle X-ray scattering and gel filtration techniques demonstrated that the oligomeric state of the protein in solution is a tetramer. In addition, functional kinetics experiments were carried out with several monophosphate nucleoside substrates and revealed a highly positive cooperativity. An allosteric mechanism involving torsion movements in solution is proposed to explain the cooperative behaviour of XfSurE. This is the first characterization of a SurE enzyme from a phytopathogen organism and, to our knowledge, the first solution structure of a SurE protein to be described.

Structured digital abstract
- MINT-7262492: *XfSurE* (uniprotkb:Q9PF20) and *XfSurE* (uniprotkb:Q9PF20) *bind* (MI:0407) by *x ray scattering* (MI:0826)
- MINT-7262504: *XfSurE* (uniprotkb:Q9PF20) and *XfSurE* (uniprotkb:Q9PF20) *bind* (MI:0407) by *molecular sieving* (MI:0071)

Introduction

Phytopathogens attacking economically important crops are a major concern around the world. It is generally believed that, worldwide, up to 20% of potential crop yields are lost. Bacterial plant diseases are most severe in the tropical regions, where their effects can be disastrous [1]. Throughout the Americas, the bacterium *Xylella fastidiosa* infects a wide range of hosts, including grape, almond, peach and citrus [2,3]. In Brazil, the strain 9a5c is the causal agent of citrus variegated chlorosis [4,5], which is a disease that attacks sweet orange trees and causes financial losses of up to US$ 280–320 million per year (http://www.fundecitrus.com.br).

In the *Xylella* genome, there are 220 genes whose functions have been predicted on the basis of comparative sequence analysis, although they were not classified into a particular metabolic pathway (http://www.

Abbreviations
ASU, asymmetric unit; PDB, Protein Data Bank; pNPP, *p*-nitrophenyl phosphate; SAXS, small-angle X-ray scattering; SurE, stationary phase survival protein E.

xylella.lncc.br) [6]. Among them, *orf XF0703* encodes a stationary phase survival protein E (SurE). The *surE* gene is widely spread among archea, eubacteria and eukaryote species and, apparently, is well conserved [7]. *orf XF0703* has 789 bp and its correlated protein (XfSurE) has 263 amino acids (28.3 kDa) and a theoretical pI of 5.23. The N-terminal primary structure has strong similarity with other previously studied SurEs (Fig. 1).

The *surE* gene was first identified in *Escherichia coli* and belongs to an operon related to the stationary phase survival protein cluster coded by *surE-pcm-nlpD-rpoS* genes [8]. In *Xylella*, this operon has a different genome architecture (*surE-pcm-dedA-nplD*), where the *rpoS* gene, an alternative RNA polymerase sigma factor that is expressed during the stationary phase, is not found [9]. As in *E. coli*, the *Xylella surE* gene has a four nucleotide overlap with *pcm* and the two constitute a bicistronic operon. These genes are highly expressed during the stationary phase in *E. coli*. The PCM protein is involved in rescuing cells from damage caused by isoaspartyl residues that are formed during stress by the conversion of asparagines and aspartate residues [10]. Assays with double mutant *pcm/surE* strains show isoaspartyl residue accumulation but, interestingly, in the *pcm* single mutant strain, the *surE* gene appears to supress this phenotype [11]. These results corroborate the hypothesis that SurE and PCM proteins act in a coordinated manner to maintain cell viability during stress [11,12].

Despite previous studies, the function of the SurE protein is not completely known, nor is its catalytic mechanism. Recently, this protein was reclassified as a nucleotidase (nucleoside 5′-monophosphate phosphohydrolase; EC 3.1.3.5) because of broad specificity to 5′(3′) nucleosides in addition to exopolyphosphatase activity [13]. Thus, SurE protein may be involved in pathway regulation of DNA and RNA synthesis [14] or, in another hypothesis, it comprises a housekeeping enzyme implicated in noncanonical nucleoside catabolism [15,16]. Five crystallographic structures of SurE have so far been characterized, obtained from *Thermotoga maritima* (TmSurE) [17,18], *Pyrobaculum aerophylum* (PaSurE) [7], *Thermus thermophylus* (TtSurE) [19],

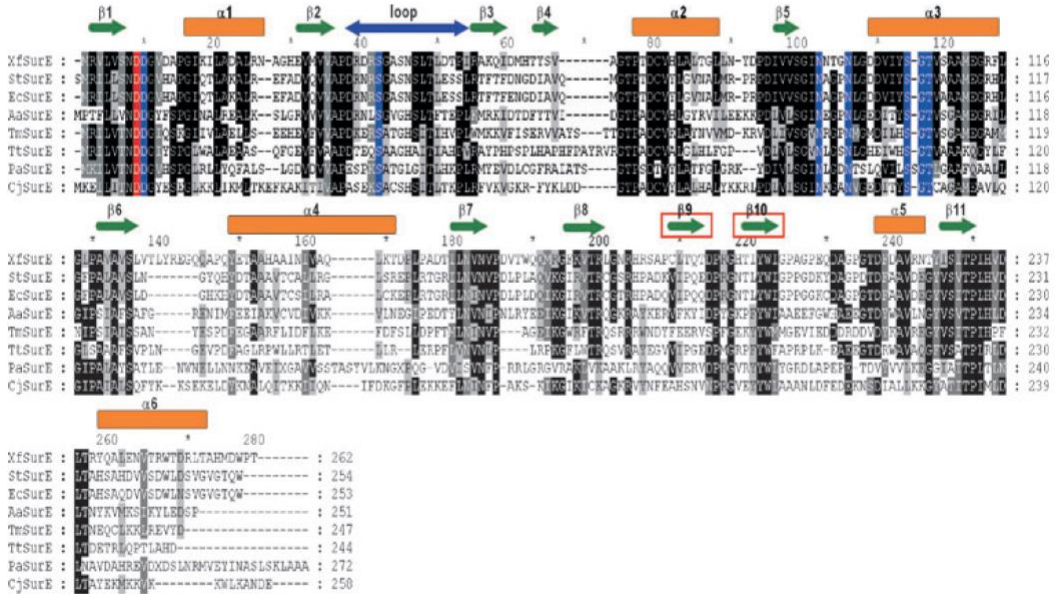


Fig. 1. Multiple sequence alignment of SurEs using CLUSTALW2 [49] and GENEDOC [50], with identities given in parentheses. XfSurE, *X. fastidiosa*; StSurE, *S. typhimurium* (54%); EcSurE, *E. coli* (54%); AaSurE, *A. aeolicus* (42%); TmSurE; *T. maritima* (37%); TtSurE, *T. thermophylus* (36%); PasurE, *P. aerophylum* (30%) and CjSurE, *C. jejuni* (33%). Letters shaded in black indicate residues that are identical in at least four of the SurEs. Conserved substitutions are shaded in grey. Amino acids implicated in metal coordination and catalysis are highlighted in blue and Asp8 is shaded in red. Secondary structure elements (orange sticks for α-helix and green arrows for β-sheet) predicted from XfSurE sequence with PSIPRED [51] are also indicated. The blue double-arrow designates the regions corresponding to the functional loop, and two β-sheets involved in tetramerization are enclosed in red rectangles [19].

Aquifex aeolicus (AaSurE) (not described in the literature) and, recently, *Salmonella typhimurium* (StSurE) [20]. The protein from *Campylobacter jejuni* (CjSurE) is under investigation [16].

In the present study, to provide a better understanding of SurE protein function, and as part of our efforts aiming to characterize possible pathogen-related proteins from *X. fastidiosa* [21–24], we report the overexpression, purification and characterization of a new SurE protein. This is the first characterization of a SurE enzyme from a phytopathogenic organism. In addition, modelling approaches based on small-angle X-ray scattering (SAXS) data were used to obtain information on of the overall shape of this protein, permitting the identification of structural features similar to those of the SurE family of proteins. These studies, associated with kinetics parameters, indicate that XfSurE has different features from the previously described SurE proteins. Therefore, this present study not only contributes to a better understanding of the SurE family of proteins, but also adds new information on the metabolism and phytopathogenicity of *X. fastidiosa*.

Results and Discussion

Cloning, expression and purification

The *X. fastidiosa* genome has provided important information on bacterial metabolism and pathogenesis [6]. In this context, several *orfs* were identified whose functional and cellular roles are so far not well understood. To better comprehend the bacterial metabolism as well to contribute to the current knowledge about the SurE protein family, we have studied the *X. fastidiosa* SurE protein. Despite early functional and biophysical studies of SurE proteins, important questions remain unanswered and the establishment of efficient expression and purification procedures for the production of XfSurE protein was first required in order to obtain sufficient protein for functional and biophysical experiments.

The *orf XF0703* was successfully cloned into the pET29a expression vector. In *E. coli* overexpression assays, the best results were obtaining by harvesting cells 20 h post-induction, using lactose 5.6 mM as the inductor agent. Approximately 20 mg of soluble XfSurE protein was obtained per litre of culture, despite the fact that part of the protein was expressed in insoluble form. The purification of the recombinant XfSurE was performed by immobilized metal ion affinity chromatography using nickel, resulting in a large quantity of purified XfSurE obtained in elution fractions over a imidazol concentration range of 250–500 mM. A single band with approximately 30 kDa (XfSurE plus His-Tag C-terminus) was obtained by SDS-PAGE.

CD

To investigate the secondary structural integrity of recombinant XfSurE, CD spectra were obtained at pH 7.0. Analysis of the data using CDSSTR indicated that XfSurE is mainly composed of α-helices (31%) and β-sheets (20%). These results are in agreement with the predicted secondary contents obtained with PSIPRED (29% α-helices and 20% β-sheets) and also are compatible with the data from the crystal structures of SurEs described in the literature that are composed of Rossmann-like domains with a swapping domain [18], thus indicating that the purified recombinant protein was folded and stable at room temperature.

Size exclusion chromatography

The oligomeric state of SurE proteins in solution is currently not well understood. In *T. maritima*, TmSurE was found as dimers [17] or both dimers and tetramers [18]. PaSuE, CjSurE and StSurE exist as dimers [7,16,20] and TtSurE was described as existing in equilibrium between dimers and tetramers [19]. Size exclusion assays with EcSurE indicated a tetrameric organization [13].

Studies using gel filtration and SAXS analysis point to a tetrameric form of XfSurE in solution. Gel filtration revealed that XfSurE appears during column elution as a single peak between the aldolase (158 kDa) and conalbumin (75 kDa) peaks (Fig. 2). According to a calibration curve, this XfSurE peak corresponds to a calculated molecular mass of 117 ± 1 kDa, therefore being equivalent to the expected mass of a tetramer.

Enzymatic assays

Initial enzymatic assays involving XfSurE were performed using the artificial substrate *p*-nitrophenyl phosphate (pNPP) to determine the best pH and cofactor, as well the kinetic parameters. As shown in Fig. 3A, the enzyme has highest activity at pH 7.0. The optimal pH is the same as those of EcSurE and StSurE [13,20], suggesting that, in mesophylic organisms, the enzyme acts at neutral pH, in contrast to thermophilic bacteria where the activity changes depending on the organism.

Various divalent metals were tested to determine the best cofactor with pNPP (Fig. 3B). The enzyme exhib-

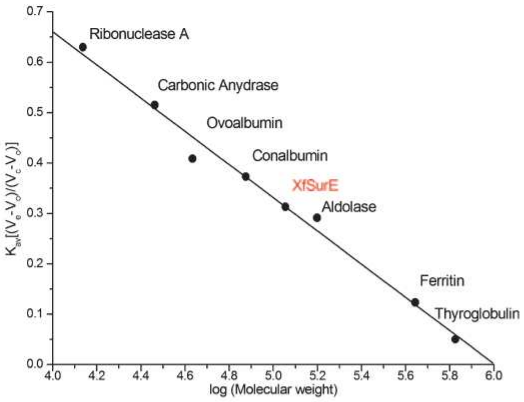


Fig. 2. Calibration curve for XfSurE in buffer B using the molecular mass calibration kits LMW and HMW (GE Healthcare). The estimate of molecular mass for XfSurE is highlighted in red.

ited the highest affinity with Mn^{2+}, followed by $Co^{2+} > Mg^{2+} > Ni^{2+}$. From all characterized SurEs, only EcSurE [13] has a preference for manganese, whereas, in the other SurEs, the preference is for magnesium. Kinetic assays with pNPP (Fig. 3C) showed that, with this substrate, the enzyme has a classical Michaelis–Menten behaviour, with K_m = 3.2 mM. This value is higher than that found for EcSurE (K_m = 2.49) and TtSurE (K_m = 2.9), but below that of StSurE (K_m = 4.8). These results demonstrate that the enzyme has low affinity for this substrate.

Several phosphorylated substrates were screened. The results obtained indicate that XfSurE has activity only against monophosphate nucleosides (Fig. 3A). The enzyme has higher activity with purine nucleosides (3′-AMP > 5′-dGMP > 5′-dAMP > 5′-AMP), but did not exhibit a clear preference between ribo- or deoxyribonucleotide and 3′ or 5′ monophosphate nucleosides. This is consistent with the classification of SurE as a member of 5′(3′) nucleotidase. family [13].

Functional experiments with four natural substrates showed that the best divalent metal was Mn^{2+} for all of them. Moreover, the ions Mg^{2+} and Co^{2+} were demonstrated to be alternative metal options for 3′-AMP, 5′-dAMP and 5′-AMP substrates, and Co^{2+} and Mg^{2+} for 5′-dGMP (Table 1).

Interestingly, kinetic experiments with natural substrates revealed new catalytic properties of XfSurE. It was clearly demonstrated that, in the presence of natural substrates, XfSurE exhibits an allosteric behaviour with a sigmoidal fit for 3′-AMP (Fig. 3C) and other substrates (Table 1). The kinetic values obtained demonstrated that XfSurE possesses the highest affinity for 3′-AMP ($K_{0.5}$ = 0.16), followed by 3′-dAMP, 5′-dGMP and 5′-AMP. These results are similar to the K_m range of EcSurE (0.10–0.37 mM) [13]. Allosteric behaviour was found for EcSurE only in the presence of polyphosphate substrates, where the enzyme had an exopolyphosphate activity with a Hill coefficient of 1.86. In XfSurE, where allosteric behaviour was found for nucleotidase activity, the positive cooperative behaviour was higher, with a Hill coefficient of 2.6 for the 3′-AMP substrate. Other nucleotidases with allosteric properties have been characterized, such as the secreted 5′-nucleotidase from *Trichinella spiralis* (Hill coefficient of 1.56 for AMP) [25] cytosolic nucleotidase from chicken [26] and human cytosolic 5′-nucleotidase II [27–29]. These findings indicate that XfSurE possibly is a regulatory enzyme and that it is modulated most likely by purine nucleoside byproducts originating from an intricate bacterial metabolism.

Similar to EcSurE, the XfSurE enzyme is inactive towards 5′-ATP. To investigate a possible inhibition of XfSurE activity by 5′-ATP, an inhibition curve was obtained in the presence of 3′-AMP (Fig. 3G). The enzyme is inhibited in a dose-dependent manner, similarly to EcSurE [13]. XfSurE lost 50% of this activity in the presence of 0.86 mM 5′-ATP (K_i = 0.43 mM).

Structure analysis

SurE structures

Figure 1 shows a sequence alignment of XfSurE with the seven most identical homologues. The crystallographic structures of five homologous proteins have been previously reported both in apo and holo forms. Apo structures from *P. aerophylum* [PaSurE; Protein Data Bank (PDB) entry 1L5X] [7] and *A. aeolicus* (AaSurE; PDB entry 2PHJ) (not described in the literature) as well as a holo form with magnesium ions of *T. maritima* (TmSurE; PDB entry 1J9J) [17,18] have been determined in a crystal form containing a dimer in the crystal asymmetric unit (ASU). *T. thermophylus* SurE structures were determined in different crystal forms, in particular, the apo structures TtSurE1 (PDB entry 2E69; space group $P3_121$, with a tetramer in the ASU) and TtSurE3 (PDB entry 2E6G; space group $P2_12_12_1$, with three tetramers in the ASU) and the complex TtSurE–MnAMP with Mn^{2+} and AMP (PDB entry 2E6C; space group $P3_121$, with one tetramer in the ASU). Sulfate and phosphate ions from the crystallization buffer were observed in both apo structures [19]. Recently, the crystallographic structure of SurE from *S. typhimurium* in the presence of magnesium ions was determined in two different crystal

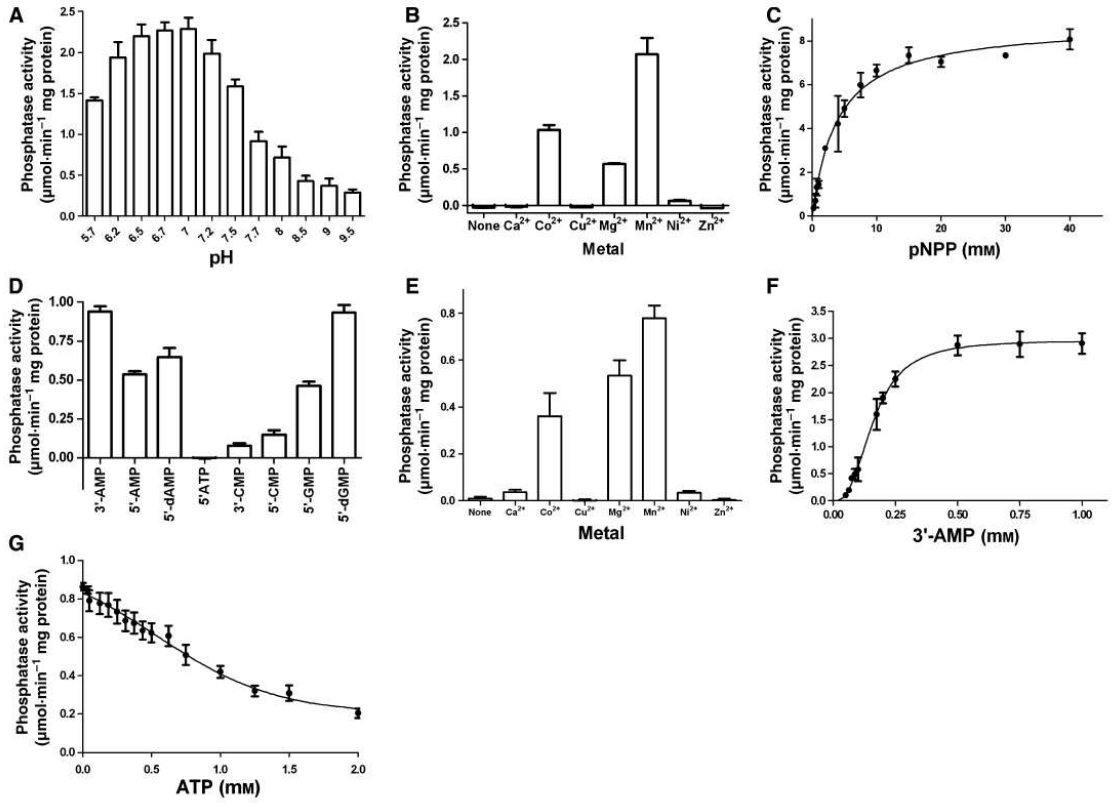


Fig. 3. Functional assays involving XfSurE enzyme. (A–C) Hydrolysis of the artificial substrate pNPP. (A) pH dependence of XfSurE in different buffers: MES-Na (pH 5.5–6.7), Hepes-Na (pH 7.0–8.0) and Tris-HCl (pH 8.2–9.5). (B) Divalent metal preference in buffer Hepes-Na (pH 7.0). Each metal was added at a 0.1 mM concentration, except Mg (added at 1 mM). (C) Kinetics curve of pNPP (0.1–40 mM). The reaction mixture contained 50 mM Hepes-Na buffer, 0.1 mM of Mn^{2+} and 4 mM of pNPP substrate. (D–F) Phosphatase activity with natural substrates. (D) Nucleotidase activity in different phosphorylated nucleosides (at 1 mM). (E) Metal predilection of XfSurE in the presence of 3′-AMP (ion concentrations are the same as those utilized in the pNPP assays). (F) Kinetics curve of 3′-AMP (0.01–1 mM). (G) XfSurE inhibition by ATP (0–2 mM) in the presence of 2.5 mM 3′-AMP.

Table 1. Kinetic parameters and metal predilection for several substrates of XfSurE.

Substrate	$K_{0.5}$ (mM)	V_{max} ([(μM)·(min·mg^{-1})$^{-1}$])	Hill coefficient	Metals
pNPP	3.2 ± 0.4	8.8 ± 0.2	1.18 ± 0.09	$Mn^{2+} > Co^{2+}$
3′-AMP	0.16 ± 0.01	2.96 ± 0.03	2.67 ± 0.09	$Mn^{2+} > Mg^{2+}$
5′-dAMP	0.41 ± 0.03	0.859 ± 0.001	2.9 ± 0.1	$Mn^{2+} > Mg^{2+}$
5′-AMP	0.79 ± 0.05	0.55 ± 0.02	2.6 ± 0.2	$Mn^{2+} > Mg^{2+}$
5′-dGMP	0.50 ± 0.07	2.26 ± 0.08	2.6 ± 0.3	$Mn^{2+} > Co^{2+}$

forms [20], which will be referred to as StSurE1 (PDB entry 2V4O, whose crystals contain a tetramer in the ASU) and StSurE2 (PDB entry 2V4N, where the ASU contains a monomer). In addition, a report on the crystallization of SurE from *C. jejuni* (CjSurE) in two crystal forms containing either a dimer or a tetramer in the ASU is available [16].

It is important to note that the crystal ASU does not necessarily reflect the protein structure in solution. Indeed, different oligomers with the same number of subunits can be generated by the aplication of crystallographic symmetry operations on the molecules contained in the ASU, and several cases have been reported where solution studies were essential to

provide information on the macromolecular assembly in solution [30]. In the case of XfSurE homologues, although various crystal forms were reported, containing a monomer, a dimer, a tetramer or a combination of tetramers in the ASU, additional experimental data other than crystallographic data are necessary to determine which macromolecular contacts mediate multimerization and thus the shape of the probable biologically relevant oligomer. To our knowledge, this is the first solution structure of a SurE protein to be described.

SAXS experiments

SAXS measurements taken at various concentrations showed that the oligomeric state of XfSurE does not depend on the protein concentration. Thus, we report only the data for the sample with a protein concentration of 12.1 mg·mL^{-1}, whose experimental curve is shown in Fig. 4A. The linear behaviour in the Guinier region (Fig. 4A, inset) indicates a monodisperse system, corresponding to a scattering particle with a radius of gyration of 32.7 ± 0.2 Å. The distance distribution function $P(r)$ obtained from the experimental data (Fig. 4B) corresponds to a globular molecule with a maximum intramolecular distance D_{max} = 100 Å. Furthermore, the Kratky plot (Fig. 4B, inset), with a well-defined maximum, indicates that the protein is in a folded native conformation.

In addition to the gel filtration technique, an estimate of the molecular mass of XfSurE in solution was obtained by the method described in the Experimental procedures, using lysozyme (M_{st} = 14.4 kDa) at a concentration of c_{st} = 27.0 mg·mL^{-1} as standard. Values for $I(0)$ obtained by Guinier analysis for the sample and the standard were $I(0) = 1.9564 \times 10^{-3}$ and $I_{st}(0) = 5.8127 \times 10^{-4}$ (arbitrary units on the same scale), respectively, leading to 108 kDa as an estimate for the molecular mass of the protein in solution. This value is in agreement (within the error inherent to the method) with that predicted from the amino acid sequence for four monomers of XfSurE (i.e. 118 kDa). Thus, we assume that XfSurE is a tetramer in solution.

Ab initio shape restoration and structural comparisons

Theoretical scattering curves for different tetramers obtained from the available crystallographic structures of XfSurE homologues were computed by CRYSOL [31] and the values of the parameter χ are provided in Table 2. The best fit was obtained with the homologue StSurE1. The fit to the experimental curve and the distance distribution function are presented in Fig. 4A,B. In both cases, good agreement is observed.

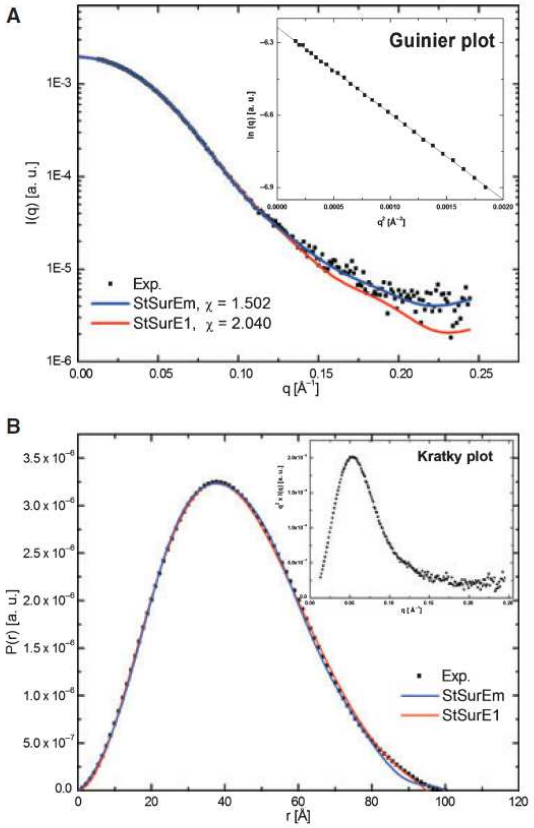


Fig. 4. SAXS experiments. (A) Experimental small-angle X-ray solution scattering curve of XfSurE and the linear behaviour observed in the Guinier plot (inset). The fit to the experimental data calculated with CRYSOL [31] for the models StSurE1 and StSurEm is also shown. (B) Distance distribution function derived from the experimental curve using GNOM [42] and the Kratky plot obtained from the experimental data (inset). $P(r)$ functions for the models StSurE1 and StSurEm calculated using CRYSOL [31] and GNOM [42] are also shown.

In addition to this preliminary analysis, which included calculation of global parameters and analysis of crystallographic structures of homologous proteins, a low resolution envelope for XfSurE was derived from the scattering curve. The best results during the DAMMIF [32] runs were obtained imposing the point-group symmetry P222 (χ = 1.65), followed by P2 (χ = 1.89), P1 (χ = 2.89) and P4 (χ = 10.77). The final average envelope is shown in Fig. 5, superposed onto the model StSurE1.

Taken together, the results of the present study unequivocally indicate that the overall shape of

Table 2. Structural parameters for SurE tetramers. The tetramers of TmSurE, PaSurE and AaSurE were generated by the application of crystallographic symmetry operators. Further details are provided in the text.

Model	R_g (Å)	χ	Area (Å^2)			Dihedrals (°)	d_{AC} Å	d_{BD} Å
			A–B	C–D	AB–CD			
StSurEm	32.40	1.502	2185	2403	2057	23.8	26.7	26.1
StSurE1	32.66	2.040	3781	3731	1257	15.2	26.1	25.8
TmSurE	32.40	2.197	3665	3556	1048	8.9	22.5	22.5
PaSurE	31.68	5.609	2999	2994	2231	−1.9	21.1	21.3
AaSurE	31.10	6.439	3621	3620	1434	−32.6	22.8	22.5
TtSurE–MnAMP	30.71	10.355	3272	2980	1815	−16.3	17.9	22.5
TtSurE1	30.62	10.490	3314	3032	1460	−15.3	18.7	22.4
TtSurE3	29.97	14.473	2913	2730	2147	−24.8	16.9	16.6

XfSurE in solution is very similar to that of StSurE1, the tetramer contained in the ASU of the crystallographic structure of SurE from *S. typhimurium*.

Rigid-body modelling

As shown in Fig. 4A,B, the crystallographic model StSurE1 fits the scattering curve and the distance distribution function in both anisometry and D_{max} quite well. Differences observed at higher angles in the scattering profile and small differences in the $P(r)$ function may be explained by the superposition of the StSurE1 structure onto the XfSurE low resolution envelope. As shown in Fig. 5 (center), there are lateral empty portions that suggest differences in the β-hairpin (Fig. 1) responsible for tetramerization in these proteins. To improve the fit to the experimental solution scattering curve, we searched for a more suitable model, starting from the original StSurE1 structure and adopting an approach where protomers A, B, C and D were treated as rigid bodies subject to two independent movements with respect to Fig. 5 (left): a translation of protomers A and D along the *z*-axis and a rotation of the dimer formed by protomers C and D around the *y*-axis.

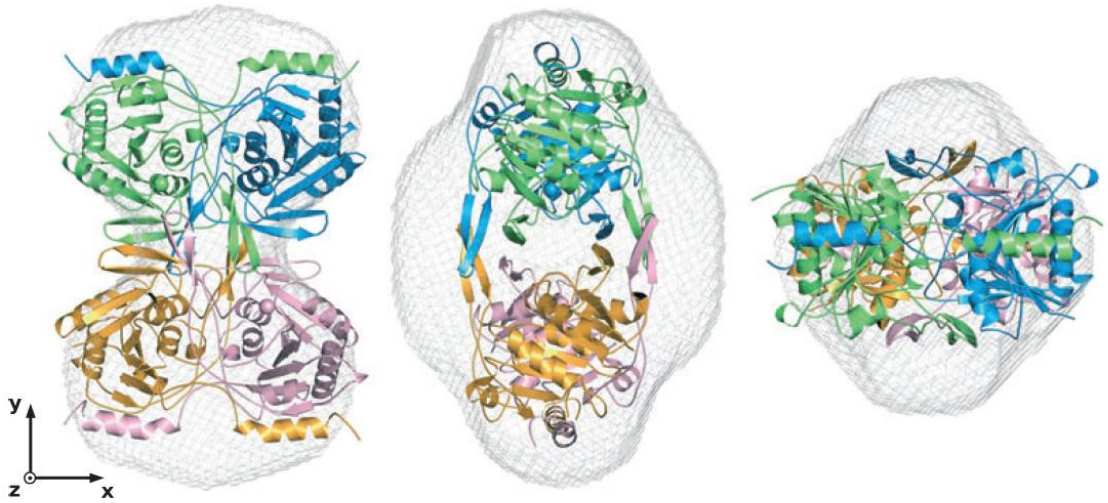


Fig. 5. Cartoon representation of the crystallographic structure of SurE from *S. typhimurium* (StSurE1) superposed onto the XfSurE envelope reconstructed from the experimental solution scattering curve. The center and right views are rotated clockwise by 90° around the *x*- and *y*-axes, respectively. Protomers A, B, C and D are shown in green, blue, orange and pink in this sequence, with the corresponding Asp8 C_α represented as a sphere. With respect to the left panel and the coordinated system of the page, the rigid-body modelling procedure described in the text involved translations (with respect to the original position) of protomers A and D along the *z*-axis and rotations of protomers C and D around the *y*-axis. The drawing was prepared using PYMOL (http://www.pymol.org).

Structural changes introduced during the process were visually inspected by superposing the modified models onto the experimentally-derived envelope.

A characteristic of the StSurE1 tetramer is the distance of the C_α carbons of Asp8 in the monomers A/C (d_{AC}) and B/D (d_{BD}), which are responsible for the symmetry properties observed in some of the reported crystallographic structures [19]. This distance is also correlated with the dihedral angle between the plane formed by Asp8 C_α in protomers ABD and the plane formed by the corresponding atoms in protomers BCD (Fig. S1), as well as with the interface area between the dimers AB and CD. These parameters have proved to be convenient for structural comparison, and the calculated values for various SurE structures are shown in Table 2.

Translational movement in steps of 1 Å and rotations in the range −90° to +90° in steps of 10° were carried out and the χ parameter was computed against the experimental data with CRYSOL [31], resulting in the plot shown in Fig. S2. The distances d_{AC} and d_{BD} remained essentialy constant throughout the process, thus preserving structural symmetry. As shown in Fig. S2, the value of χ for a translation of approximately 7.5 Å is a minimum for several rotation angles, indicating that, in principle, more than one configuration is possible in solution. Of note, there is a central local minimum (translation and rotation of 7.5 Å and +25°, respectively), which will be referred to as StSurEm, for which χ = 1.502, and this is lower than that calculated for the initial model StSurE1 (χ = 2.040). StSurEm has a dihedral angle (23.8°) higher than that of StSurE1 (15.2°) (Table 2) and a higher interaction surface area between dimers AB and CD (2057 Å^2 versus 1257 Å^2 in the case of StSurE1). The fit to the experimental curve and the $P(r)$ function corresponding to the StSurEm model are shown in Fig. 4A,B.

It is important to note that the active site of each protomer as well as other important residues for catalysis (amino acid residues 32–51 in Fig. 1) are located in the core of the tetramer, a region directly affected by changes in the interface area AB/CD. Conceivably, the possible structural configurations mentioned above may be related to a freedom of movement of the apo enzyme in solution, which would itself constitute a mechanism by which XfSurE may regulate the entry of substrate into the active sites of the tetramer. This hypothesis is in agreement with (and could provide an explanation for) the results obtained by the enzymatic assays, where the Hill coefficients unequivocally indicate the highly positive cooperative behaviour of XfSurE. Proteins of dihedral groups provide rich structural variations to build allosteric control [33], as observed in fructose-1,6-bisphosphatase tetramer [34] where, in the context of the two-state Monod–Wyman–Changeux model [35], the T state was associated with a dimer turned 17° from its position in the R state, with both coexisting in equilibrium.

Iwasaki and Miki [19] recently proposed for TtSurE that an observed cooperative behaviour was correlated with the structural asymmetry inherent in the apo enzyme, as a result of differences in the distances d_{AC} and d_{BD} (TtSurE1; Table 2). The correlation with a cooperative behaviour was inferred based on the fact that the structure of a holo enzyme (TtSurE-MnAMP) also exhibted the same asymmetry. Thus, it could not be atributed to the presence of the ligand. However, as shown in Table 2, in another apo structure (TtSurE3) determined by Iwasaki and Miki [19], no asymmetry is observed. Both TtSurE1 and TtSurE3 have an ion near the active site region. The asymmetric apo structure TtSurE1 belongs to a crystal form containing a tetramer in the ASU, whereas the ASU of the apo structure TtSurE3 contains three tetramers, with all of them being symmetric. Thus, from a different point of view, the observed asymmetry cannot be considered as an intrinsic property of the apo enzyme.

In the case of XfSurE, the data suggest that the unequivocal positive cooperative behaviour observed is not related to a permanent structural asymmetry. The results obtained in the present study provide insight into a possible allosteric behaviour of the enzyme that could account for the observed cooperativity. Because XfSurE may exhibit a twisting movement in solution, as discussed above, a simple model for an allosteric mechanism would include two configurations, R and T, similar to the Monod–Wyman–Changeux model [35]. Thus, changes in the dihedral angle and accompanying changes in the contact area between dimers AB/CD would control the access of the substrate to the active sites of the tetramer.

Experimental procedures

Cloning, expression and purification

The *surE* gene from *X. fastidiosa orf XF0703* was amplified by PCR using purified genomic DNA as template. The PCR was carried out using specific primers with restriction enzyme sites *Nde*I and *Xho*I (New England Biolabs, Ipswich, MA, USA) for cloning in pET29a vector (Novagen, Madison, WI, USA): forward primer 5′-ATAAA CATATGCGCGTTCTTGTCAGTAA-3′ and reverse primer 5′-AAACTCGAGTGTTGGCCAGTCCATGTG-3′ (Invitrogen Life Technologies, São Paulo, Brazil). The PCR

amplification product was cloned into pET29a and transformed into *E. coli* DH5-α as the cloning host. Sequencing of cloned vectors revealed *orf XF0703* plus an expected additional eight carboxy terminus amino acid residues, including six histidines (LEHHHHHH).

The plasmid containing the *surE* insert from *X. fastidiosa* was transformed into *E. coli* BL21 (DE3) cells, inoculated overnight at 37 °C and 300 r.p.m. into 3 mL of TB medium containing 40 μg·mL^{-1} of kanamycin antibiotic and transferred to 2 L of TB at the same antibiotic concentration. The cells were grown until D_{560} of 0.6–0.8 was reached, when XfSurE protein overexpression was induced by the addition of 5.6 mM of lactose followed by cultivation for 20 h at 37 °C and 300 r.p.m. The culture was then centrifuged at 3000 ***g*** for 15 min at 4 °C and pelleted cells were resuspended in buffer A (50 mM Tris-HCl, pH 7.5, with 300 mM NaCl) plus 1 mg·mL^{-1} lysozyme and 1 mM phenylmethanelsulfonyl fluoride (Sigma Chemical, St Louis, MO, USA). The cell suspension was let stand for 30 min at 4 °C followed by sonication. Clarification was performed twice by centrifugation at 27 000 ***g*** for 40 min at 4 °C. The XfSurE protein purification was performed in a single chromatography step using a Ni-NTA column (Qiagen, Hilden, Germany) equilibrated with buffer A. The purified XfSurE protein was eluted with five column volumes of buffer A containing 250 mM imidazole and the degree of purity was estimated by SDS-PAGE. Subsequently, the purified XfSurE was dialyzed in buffer B (25 mM Tris-HCl, 50 mM NaCl and 1 mM dithiothreitol). All chemical reagents used were of the highest commercially available grade.

CD

CD spectra of the purified XFSurE protein were measured using a Jasco J-810 Spectropolarimeter dichrograph (Japan Spectroscopic, Tokyo, Japan). The far-UV CD spectra were generated at 24 °C using XfSurE protein at 6.5 μM in 5 mM Tris-H_2SO_4 buffer at pH 7.0. The assays were carried out using a 1 mm pathlength quartz cuvette, and 20 accumulations within the range 260–190 nm at a rate of 50 nm·min^{-1} at 20 °C were recorded for the sample and averaged. Deconvolution and statistical analysis of the CD spectra were performed using the DICHROWEB server [36–38].

Size exclusion chromatography

To assess the oligomeric state of purified XfSurE, gel filtration chromatography was performed using a Superdex 200 HR10/30 (GE Healthcare, Uppsala, Sweden) prepacked column. After equilibration of the column with buffer B (48 mL), the sample was loaded at a flow rate of 0.5 mL·min^{-1}. The Superdex column was calibrated using HMW and LMW calibration kits (GE Healthcare) as standard molecular weight markers.

Enzymatic assays

Assays with artificial substrate pNPP involving pH, cofactors and kinetics were performed at room temperature and the data were obtained spectrophotometrically at 410 nm. The phosphatase activity of SurE protein was screened against different phosphorylated substrates (amino acids, sugars, lipids and nucleosides). Kinetic experiments with four substrates (3′-AMP, 5′-AMP, 5′-dAMP and 5′-dGMP) were carried out aiming to understand the behaviour of XfSurE in the presence of these substrates. The phosphatase activity with natural substrates was measured at 620 nm by the release of free phosphate using the malachite green method [39]. The experiments were performed in 50 mM Hepes-Na buffer (pH 7.0) with 0.1 mM of Mn^{2+}. All the reactions were performed in quadruplicate and repeated three times using 2.5 μg (pNPP assays) and 1.5 μg (natural substrates assays) of fresh purified protein. Kinetic parameters were obtained using GRAPHPAD PRISM (GraphPad Software, San Diego, CA, USA). All substrates were purchased from Sigma Chemical.

SAXS data analysis and modelling

The samples of XfSurE were prepared with concentrations in the range 2.0–12.1 mg·mL^{-1} in buffer B (25 mM Tris-HCl, pH 7.5, 50 mM NaCl and 1 mM dithiothreitol). SAXS data were collected at the D02A-SAXS2 beamline of the Laboratório Nacional de Luz Síncrotron (LNLS, Campinas SP, Brazil) using a 2D position-sensitive MarCCD detector (Marresearch, Hamburg, Germany). The sample-to-detector distance was 1306.85 mm, and the X-ray wavelength was 1.488 Å. A momentum transfer interval $0.01 < q < 0.24$ (Å^{-1}) was covered, where $q = [(4\pi)/(\lambda)]\sin\theta$ and 2θ is the scattering angle. The scattering patterns were measured with a 3 min exposure time at 20 °C. For each concentration, five sucessive frames were recorded. Data reduction using FIT2D [40] included radial integration of the images collected and normalization to the intensity of the transmitted beam to build an average scattering curve. No radiation damage was observed. The average scattering curves underwent conventional Guinier approximation analysis to obtain R_g, the radius of gyration of the molecule and $I(0)$, the scatering intensity at zero angle, assuming Guinier's law $I(q) = I(0)\exp(-q^2[(R_g^2)/3])$ for very small angles ($q < 1.3/R_g$) [41]. More accurate estimates for R_g and $I(0)$ were obtained by the indirect Fourier transform method as implemented in GNOM [42], which also calculates the distance distribution function $P(r)$, allowing assessment of the maximum intramolecular distance (D_{max}) and molecule anisometry. Examination of the Kratky plot [$q^2I(q) \times q$], where the native state of a compact globular protein, has a characteristic profile, was employed as a tool to characterize the conformational state of the protein in solution [43,44]. An estimate for the molecular mass of the protein was obtained using a known

protein collected in the same experimental conditions as a standard for calibration and the equation $[Mc/I(0)]_{protein} = [Mc/I(0)]_{standard}$, where, for each species, M is the molecular mass, c is the concentration and $I(0)$ is the scatering intensity at the origin. This method is capable of providing molecular mass estimates with an error of approximately 10% [45]. The number of subunits (n) and thus the oligomeric state in solution was obtained from the known molecular mass of the protein monomer (P) and the relation $n = M/P$.

Ab initio dummy bead models were calculated from the experimental curves using DAMMIF, a faster version of DAMMIN [32]. Different point-group symmetries (P1, P2, P4 and P222) were imposed. The final low resolution 3D envelope representing the protein was generated by averaging twenty independent DAMMIF runs, with the software DAMAVER and DAMFILT [46] in the automatic mode. With the objective of clearer visualization, the average model was represented by a surface calculated using the software NCSMASK [47].

In parallel to the *ab initio* methods used to construct a low resolution envelope from the experimental curve, rigid-body modelling based on the homologous protein StSurE1 was carried out in an independent fashion. Translation and rotation movements of the protein subunits were performed using PYMOL (http://www.pymol.org). The discrepancy (χ) between the experimental data and the theoretical scattering curve computed for each possible model was assessed with the software CRYSOL [31]. Conventionally, the parameter χ provided by CRYSOL is used as a measurement of the goodness of fit to the experimental data. Structural alignments were carried out with SUPCOMB [48]. Calculation of interaction surface areas between subunits in oligomers was performed using the software AREAIMOL available in the CCP4 package [47]. First, the accessible surface area for each individual subunit (S_i, S_j) and the surface of the complex ($S_{i,j}$) were calculated. The interaction surface area was then obtained as half of the buried surface area, defined as: $S_i + S_j - S_{i,j}$.

Acknowledgements

This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de São, Paulo (FAPESP Process 01/07533-7 and 05/03234-6). A. M. Saraiva, S. F. Tada, L. K. Rosselli-Murai, A. C. Pelloso and M. A. S. Toledo were supported by fellowships from FAPESP. M. A. Reis and D. R. S. Schneider received PhD fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We gratefully acknowledge Laboratório Nacional de Luz Síncrotron (LNLS) for beamline time, Professor Nilson Zanchin for providing CD espectropolarimeter facilities, and Professor Dagmar Stach-Machado and Luis A. Peroni (IB/Unicamp) for technical expertise on the kinetics experiments. We are deeply indebted to Professor Carol Collins (IQ/Unicamp) for carefully reading the manuscript and assisting with language revision.

References

1 Dow JM & Daniels MJ (2000) *Xylella* genomics and bacterial pathogenicity to plants. *Yeast* **17**, 263–271.
2 Lambais MR, Goldman MH, Camargo LE & Goldman GH (2000) A genomic approach to the understanding of *Xylella fastidiosa* pathogenicity. *Curr Opin Microbiol* **3**, 459–462.
3 Purcell AH & Hopkins DL (1996) Fastidious xylem-limited bacterial plant pathogens. *Annu Rev Phytopathol* **34**, 131–151.
4 Pooler MR & Hartung JS (1995) Specific PCR detection and identification of *Xylella fastidiosa* strains causing citrus variegated chlorosis. *Curr Microbiol* **31**, 377–381.
5 Lee RF, Derrick KS, Beretta MJG, Chagas CM & Rosetti V (1991) Citros variegated chlorosis: a new destructive disease of citros in Brazil. *Citros Ind* **000**, 12–15.
6 Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LM, Araya JE, Baia GS, Baptista CS *et al.* (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* consortium of the organization for nucleotide sequencing and analysis. *Nature* **406**, 151–159.
7 Mura C, Katz JE, Clarke SG & Eisenberg D (2003) Structure and function of an archaeal homolog of survival protein E (SurEalpha): an acid phosphatase with purine nucleotide specificity. *J Mol Biol* **326**, 1559–1575.
8 Li C, Ichikawa JK, Ravetto JJ, Kuo HC, Fu JC & Clarke S (1994) A new gene involved in stationary-phase survival located at 59 minutes on the *Escherichia coli* chromosome. *J Bacteriol* **176**, 6015–6022.
9 da Silva Neto JF, Koide T, Gomes SL & Marques MV (2007) The single extracytoplasmic-function sigma factor of *Xylella fastidiosa* is involved in the heat shock response and presents an unusual regulatory mechanism. *J Bacteriol* **189**, 551–560.
10 Li C, Wu PY & Hsieh M (1997) Growth-phase-dependent transcriptional regulation of the *pcm* and *surE* genes required for stationary-phase survival of *Escherichia coli*. *Microbiology* **143**, 3513–3520.
11 Visick JE, Ichikawa JK & Clarke S (1998) Mutations in the *Escherichia coli surE* gene increase isoaspartyl accumulation in a strain lacking the *pcm* repair methyltransferase but suppress stress-survival phenotypes. *FEMS Microbiol Lett* **167**, 19–25.
12 Riehle MM, Bennett AF & Long AD (2001) Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc Natl Acad Sci USA* **98**, 525–530.

13 Proudfoot M, Kuznetsova E, Brown G, Rao NN, Kitagawa M, Mori H, Savchenko A & Yakunin AF (2004) General enzymatic screens identify three new nucleotidases in *Escherichia coli*. Biochemical characterization of SurE, YfbR, and YjjG. *J Biol Chem* **279**, 54687–54694.
14 Reichard P (1988) Interactions between deoxyribonucleotide and DNA synthesis. *Annu Rev Biochem* **57**, 349–374.
15 Galperin MY, Moroz OV, Wilson KS & Murzin AG (2006) House cleaning, a part of good housekeeping. *Mol Microbiol* **59**, 5–19.
16 Gonçalves AMD, Rêgo AT, Thomaz M, Enguita FJ & Carrondo MA (2008) Expression, purification, crystallization and preliminary X-ray characterization of two crystal forms of stationary-phase survival E protein from *Campylobacter jejuni*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 213–216.
17 Lee JY, Kwak JE, Moon J, Eom SH, Liong EC, Pedelacq JD, Berendzen J & Suh SW (2001) Crystal structure and functional analysis of the SurE protein identify a novel phosphatase family. *Nat Struct Biol* **8**, 789–794.
18 Zhang RG, Skarina T, Katz JE, Beasley S, Khachatryan A, Vyas S, Arrowsmith CH, Clarke S, Edwards A & Joachimiak A (2001) Structure of *Thermotoga maritima* stationary phase survival protein SurE: a novel acid phosphatase. *Structure* **9**, 1095–1106.
19 Iwasaki W & Miki K (2007) Crystal structure of the stationary phase survival protein SurE with metal ion and AMP. *J Mol Biol* **371**, 123–136.
20 Pappachan A, Savithri HS & Murthy MRN (2008) Structural and functional studies on a mesophilic stationary phase survival protein (SurE) from *Salmonella typhimurium*. *FEBS J* **275**, 5855–5864.
21 Azzoni AR, Tada SF, Rosselli LK, Paula DP, Catani CF, Sabino AA, Barbosa JA, Guimarães BG, Eberlin MN, Medrano FJ *et al.* (2004) Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*. *Protein Expr Purif* **33**, 297–303.
22 Catani CF, Azzoni AR, Paula DP, Tada SF, Rosselli LK, de Souza AP & Yano T (2004) Cloning, expression, and purification of the virulence-associated protein D from *Xylella fastidiosa*. *Protein Expr Purif* **37**, 320–326.
23 Paula DP, Azzoni AR, Siqueira SF, Catani CF, Rosselli LK & de Souza AP (2003) Expression and purification of a putative H-NS nucleoid-associated protein from the phytopathogen *Xylella fastidiosa*. *Protein Expr Purif* **32**, 61–67.
24 Rosselli LK, Oliveira CL, Azzoni AR, Tada SF, Catani CF, Saraiva AM, Soares JS, Medrano FJ, Torriani IL & Souza AP (2006) A new member of the aldo-keto reductase family from the plant pathogen *Xylella fastidiosa*. *Arch Biochem Biophys* **453**, 143–150.
25 Gounaris K, Selkirk ME & Sadeghi SJ (2004) A nucleotidase with unique atalytic properties is secreted by *Trichinella spiralis*. *Mol Biochem Parasitol* **136**, 257–264.
26 Itoh R, Usami C, Nishino T & Tsushima K (1978) Kinetic properties of cytosol 5′-nucleotidase from chicken liver. *Biochim Biophys Acta* **526**, 154–162.
27 Bretonnet AS, Jordheim LP, Dumontet C & Lancelin JM (2005) Regulation and activity of cytosolic 5′-nucleotidase II. A bifunctional allosteric enzyme of the haloacid dehalogenase superfamily involved in cellular metabolism. *FEBS Lett* **579**, 3363–3368.
28 Spychala J, Madrid-Marina V & Fox IH (1988) High Km soluble 5′-nucleotidase from human placenta. Properties and allosteric regulation by IMP and ATP. *J Biol Chem* **263**, 18759–18765.
29 Walldén K, Stenmark P, Nyman T, Flodin S, Gräslund S, Loppnau P, Bianchi V & Nordlund P (2007) Crystal structure of human cytosolic 5′-nucleotidase II: insights into allosteric regulation and substrate recognition. *J Biol Chem* **282**, 17828–17836.
30 Putnam CD, Hammel M, Hura GL & Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191–285.
31 Svergun D, Barberato C & Koch MHJ (1995) CRYSOL – a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28**, 768–773.
32 Svergun D (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**, 2879–2886.
33 Goodsell DS & Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* **29**, 105–153.
34 Liang JY, Zhang Y, Huang S & Lipscomb WN (1993) Allosteric transtion of fructose-1,6-bisphosphatase. *Proc Natl Acad Sci USA* **90**, 2132–2136.
35 Monod J, Wyman J & Changeux JP (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**, 88–118.
36 Whitmore L & Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668–W673.
37 Whitmore L, Janes RW & Wallace BA (2006) Protein Circular Dichroism Data Bank (PCDDB): data bank and website design. *Chirality* **18**, 426–429.
38 Whitmore L & Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392–400.
39 Geladopoulos TP, Sotiroudis TG & Evangelopoulos AE (1991) A malachite green colorimetric assay for

protein phosphatase activity. *Anal Biochem* **192**, 112–116.
40 Hammersley AP, Svensson SO, Hanfland M, Fitch AN & Häusermann D (1996) Two-dimensional detector software: from real detector to idealised image or two-theta scan. *High Press Res* **14**, 235–248.
41 Glatter O & Kratky O (1982) *Small Angle X-ray Scattering*. Academic Press Inc. (London) Ltd., London.
42 Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**, 495–503.
43 Semisotnov GV, Kihara H, Kotova NV, Kimura K, Amemiya Y, Wakabayashi K, Serdyuk IN, Timchenko AA, Chiba K, Nikaido K *et al.* (1996) Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. *J Mol Biol* **262**, 559–574.
44 Doniach S (2001) Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev* **101**, 1763–1778.
45 Mylonas E & Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J Appl Crystallogr* **40**, s245–s249.
46 Volkov VV & Svergun D (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr* **36**, 860–864.
47 CCP4 (2002) High-throughput structure determination. Proceedings of the 2002 CCP4 (Collaborative Computational Project in Macromolecular Crystallography) Study Weekend. January, 2002. York, United Kingdom. *Acta Crystallogr D Biol Crystallogr* **58**, 1897–1970.
48 Kozin MB & Svergun D (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**, 33–41.
49 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* (2007) ClustalW and ClustalX version 2. *Bioinformatics* **21**, 2947–2948.
50 Nicholas HB, Nicholas KB Jr & Deerfield DW (1997) Genedoc: analysis and visualization of genetic variation. *EMBNEW NEWS* **4**, 14.
51 McGuffin LJ, Bryson K & Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.

Supporting information

The following supplementary material is available:
Fig. S1. Stereoview defining the dihedral angle.
Fig. S2. The parameter χ as a function of the translation and rotation movements.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

10.3. DECLARAÇÃO DE BIOÉTICA

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha tese de Doutorado intitulada **"Expressão, purificação e caracterização parcial de proteínas relacionadas à patogenicidade de *Magnaporthe grisea*"**

( ) não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

(X) tem autorização da(s) seguinte(s) Comissão(ões) de Bioética ou Biossegurança:
Projeto CIBIO IB/UNICAMP, sob Protocolo(s) nº 15/2003.

Campinas, 11 de Junho de 2010.

Dilaine Rose Silva Schneider
Aluna: **Dilaine Rose Silva Schneider**

Orientadora: Dra. **Anete Pereira de Souza**

Para uso da Comissão ou Comitê pertinente:
(X) **Deferido** ( ) **Indeferido**

Nome:

Função: PRESIDENTE CIBIO - CBMEG