

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



**Aline Lima de Oliveira**

**“ESTUDOS ESTRUTURAIS E DINÂMICOS DA PROTEÍNA  
ELICITORA CERATO-PLATANINA”**

Tese apresentada ao Instituto de  
Biologia para obtenção do Título de  
Doutora em Biologia Funcional e  
Molecular, na área de Bioquímica.

Este exemplar corresponde à redação final  
da tese defendida pelo(a) candidato (a)  
*ALINE LIMA DE OLIVEIRA*  
*Celso Benedetti*  
e aprovada pela Comissão Julgadora.

Orientadora: Dra. Thelma de Aguiar Pertinhez

Co-Orientador: Dr. Celso Eduardo Benedetti

Campinas, 2008

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

**Ol41e** Oliveira, Aline Lima de  
Estudos estruturais e dinâmicos da proteína elicitora  
Cerato-Platanina / Aline Lima de Oliveira. – Campinas, SP:  
[s.n.], 2008.

Orientadora: Thelma de Aguiar Pertinhez.  
Co-orientador: Celso Eduardo Benedetti.  
Tese (doutorado) – Universidade Estadual de  
Campinas, Instituto de Biologia.

1. Cerato-platanina. 2. Elicitores. 3. Ressonância  
magnética nuclear. I. Pertinhez, Thelma de Aguiar. II.  
Benedetti, Celso Eduardo. III. Universidade Estadual de  
Campinas. Instituto de Biologia. IV. Título.

**Título em inglês:** Structural and dynamics studies of the protein elicitor Cerato-Platanin.

**Palavras-chave em inglês:** Cerato-platanin; Elicitores; Nuclear magnetic resonance.

**Área de concentração:** Bioquímica.

**Titulação:** Doutora em Biologia Funcional e Molecular.

**Banca examinadora:** Thelma de Aguiar Pertinhez, Gonçalo Amarante Guimarães Pereira,  
Paula Regina Kuser-Falcão, Marcius da Silva Almeida, Richard Charles Garratt.

**Data da defesa:** 15/12/2008.

**Programa de Pós-Graduação:** Biologia Funcional e Molecular.

Campinas, 15 de dezembro de 2008

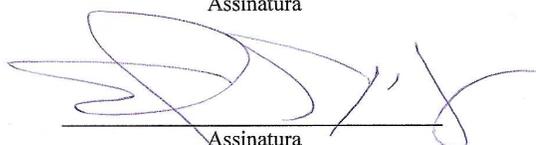
**BANCA EXAMINADORA**

Profª. Dra. Thelma Aguiar Pertinhez (Orientadora)



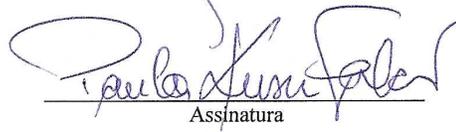
Assinatura

Prof. Dr. Gonçalo Amarante Guimarães Pereira



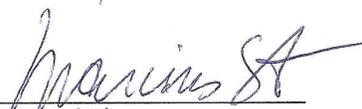
Assinatura

Profª. Dra. Paula Regina Kuser Falcão



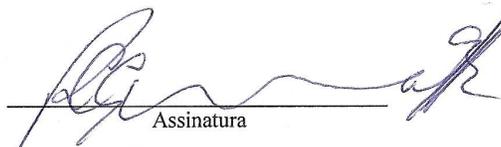
Assinatura

Prof. Dr. Marcius da Silva Almeida



Assinatura

Prof. Dr. Richard Charles Garratt



Assinatura

Prof. Dr. José Andrés Yunes

\_\_\_\_\_

Assinatura

Prof. Dr. Ricardo Aparicio

\_\_\_\_\_

Assinatura

Prof. Dr. Júlio César Borges

\_\_\_\_\_

Assinatura

*Aos meus pais, César e  
Arlinda, este pequeno fruto dos  
meus anos de trabalho.*

## AGRADECIMENTOS

Meus sinceros agradecimentos à Dra. Thelma de Aguiar Pertinhez que sempre acreditou na minha capacidade e me possibilitou, não só os primeiros passos, mas toda a minha caminhada acadêmica. Obrigada por partilhar os seus conhecimentos, por ser sempre disponível, paciente e por me orientar por todos esses anos, mesmo à distância. E principalmente, obrigada por toda a amizade e ensinamentos que foram além da ciência.

Ao Dr. Celso E. Benedetti que, sem medir muito as conseqüências, aceitou ser meu co-orientador. Agradeço-te pela sua disponibilidade, pelas excelentes discussões científicas e por todo o entusiasmo e apoio.

Ao Prof. Dr. Alberto Spisni pela oportunidade em trabalhar nos seus grupos de pesquisa, tanto no brasileiro quanto no italiano, por toda a confiança depositada em mim e por ter-me aberto tantas portas. *Grazie mille!*

À Dra. Luigia Pazzagli e ao Prof. Dr. Gianni Cappugi que nos envolveram nesse projeto e por terem cedido as células de leveduras para a expressão da proteína de estudo, sem o qual esse trabalho não teria iniciado.

Ao Prof. Dr. Daniel O. Cicero que me recebeu em seu laboratório e de forma especial a Dra. Mariana Gallo, que me acompanhou em todos os passos de coleta e análise de dados de RDC e cálculo de estrutura, sempre me apoiando com muita confiança e otimismo. *Mari, grazie mille per tutto l'aiuto, pazienza e generosità!*

Às agências de fomento à pesquisa FAPESP e CAPES pelo apoio financeiro. Ao Laboratório Nacional de Luz Síncrotron, à Universidade Estadual de Campinas, à *Università degli Studi di Parma* e à *Università degli Studi di Roma 'Tor Vergata'*, pela infra-estrutura.

À Profa. Dra. Eneida de Paula, ao Prof. Dr. Cláudio Chrysostomo Werneck e ao Prof. Dr. Mário Tyago Murakami pela disponibilidade e preciosos comentários na minha qualificação.

Aos membros da banca Profa. Dra. Paula Regina Kuser Falcão, Prof. Dr. Marcius da Silva Almeida, Prof. Dr. Richard Charles Garratt, Prof. Dr. José Andrés Yunes, Prof. Dr. Ricardo Aparício e Prof. Dr. Júlio César Borges, por terem aceitado a participar da avaliação desse trabalho.

Ao Rodrigo Bongers, um agradecimento especial pelo excelente apoio técnico na área de informática e por ter me facilitado o uso do *cluster*, sem o qual eu não teria conseguido terminar essa tese. E claro, obrigada pela ótima companhia, por todas as risadas e momentos de descontração.

Ao Tiago e ao Yuri, que mesmo sem entender muito de RMN, me deram um conselho precioso durante uma reunião de grupo. Obrigada meninos! Sem a ajuda de vocês, acho que eu teria demorado mais alguns meses para finalizar o cálculo das estruturas. Além disso, vocês são exemplo claro da importância dessas reuniões onde temos a oportunidade de expor e trocar conhecimento científico e tentar resolver os problemas.

Ao grupo italiano do Prof. Spisni, Profa. Dra Elena Ferrari, Profa. Dra. Emanuela Casali, Profa. Dra. Lorella Franzoni e Prof. Dr. Giorgio Sartor. *Un grande ringraziamento per la vostra simpatia e per avermi accolto tra di voi.*

Ao antigo grupo de RMN, pelos quais eu tenho um carinho grande, Maurício, Leonardo, Sérgio e em especial a Ângela, Patrícia e Rita, obrigada por terem tornado a RMN mais divertida.

Ao grupo GEM, Andrés, André, Adriana, Luli, Malu, Marina, Mariane, Natália, Tiago, Rosi e Yuri, por terem me acolhido de braços abertos, por não terem desistido de mim mesmo eu sendo a estranha do grupo e por me ensinarem quase tudo que eu sei de interação planta-patógeno.

Aos meus queridos amigos Tatiana, Nádia, Gabriela, Yuri, Tiago, Fábio, Marcos, André e Gustavo, pelas preciosas discussões científicas, mas principalmente por tornarem a minha vida muito mais divertida. Vocês sabem que eu tenho muito orgulho de vocês e dos pesquisadores que um dia vocês se tornarão. Vocês me fazem ainda ter esperanças no futuro! Por isso, não me decepcionem!

Aos meus amigos da faculdade de Química, Thaís, André, Alvino e Silvia que me ensinaram a sempre rir dos tropeços da vida.

Aos meus tesouros Lindolfo, Keli, Karen, Kátia, Camila, Ticiane e Ângela, amigos já nem me lembro mais desde quando, pela amizade sincera e incondicional. *Poi, alle mie piccole Ana e Marta, che hanno fatto la mia vita in Italia molto più bella!*

Aos meus avós, cunhados, tios e primos que são meus exemplos de amor, serenidade e dedicação.

À minha família; a minha querida mãe, que sempre esteve presente na minha vida com seu amor incondicional; ao meu pai pelas lições de vida e por sempre se orgulhar daquilo que eu posso fazer de bom; às minhas irmãs, Milena e Laísa, que entre desentendimentos e sorrisos também deram o seu apoio e carinho; e aos meus sobrinhos, Gabriel e Tiago, que mesmo tão pequenininhos já me ensinaram que eu posso acreditar que existe amor à primeira vista.

Muito obrigada a todos vocês que fazem parte da minha vida. Esse trabalho é resultado de muito esforço e eu não tenho dúvidas que ele só foi possível pela ajuda direta e indireta de todos vocês.

## RESUMO

Um dos desafios que envolvem a patologia de plantas é determinar as bases moleculares que relacionam a patogenicidade do fungo na planta hospedeira. Em geral, o mecanismo pelo qual a planta se defende contra uma variedade de patógenos envolve o reconhecimento de moléculas específicas codificadas por esses microorganismos. A ceratoplatanina (CP) é uma proteína moderadamente hidrofóbica que é secretada, porém também está localizada na parede celular do fungo ascomiceto *Ceratocystis fimbriata*. Esse fungo é o agente causador de severas doenças de cancro que atinge um grande número de plantas como *Platanus acerifolia*, *Theobroma cacao*, *Coffea Arabica*, entre outras. A proteína secretada por culturas de *Ceratocystis fimbriata* f. sp. *platani* é capaz de elicitar a síntese de fitoalexinas em folhas de *Platanus acerifolia* e causar morte celular. De acordo com o banco de dados de família de proteínas CP é indicada como a fundadora da “Família Cerato-Platanina”, que inclui outras várias proteínas secretadas por fungos envolvidos em fenômenos fitopatológicos e/ou reações imunológicas. Buscas em banco de dados de proteínas mostram que não existe nenhuma proteína similar a CP com estrutura conhecida. Assim, a elucidação estrutural de moléculas possivelmente envolvidas interação fungo-hospedeiro é de primária importância para a compreensão dos eventos iniciais que levam às respostas de defesa.

A CP foi clonada e expressa em sistema eucariótico e marcada isotópicamente com  $^{13}\text{C}$  e  $^{15}\text{N}$ . Experimentos de ressonância magnética nuclear multinucleares e multidimensionais foram usados para determinar sua estrutura 3D em solução. A estrutura da CP é composta por duas fitas- $\beta$  paralelas e cinco fitas- $\beta$  antiparalelas, reunidas em uma única folha- $\beta$ , e três elementos helicoidais.

A dinâmica da CP foi determinada e os dados experimentais mostraram que a proteína apresenta uma estrutura caracterizada por uma limitada mobilidade interna. Além disso, a proteína apresenta alta estabilidade, mantendo a sua conformação mesmo em temperaturas elevadas ou em uma larga escala de pH e de concentração de uréia.

Com o objetivo de procurar outras proteínas ou domínios conformacionalmente similares ao da CP, sua estrutura foi submetida a análises estruturais utilizando-se a

ferramenta *ProFunc*. Os resultados de alinhamento estrutural mostram que a CP não possui similaridade conformacional a nenhuma proteína conhecida envolvida em avirulência de microorganismos ou a proteínas elicitoras de fungos ou bactérias patogênicas. Porém, surpreendentemente, o enovelamento da CP mostrou-se similar a moléculas pertencentes às famílias das quimiocinas e do complexo de histocompatibilidade principal de classe I (MHC-I). De fato, a estrutura de menor energia da CP se sobrepõe melhor a forma dimérica da interleucina-8 (IL-8) e do peptídeo ativador de neutrófilos-2 (NAP-2), da família de quimiocinas, e aos domínios  $\alpha 1$ ,  $\alpha 2$  da proteína MHC-I. Interessante notar que ambas famílias de proteínas que possuem enovelamento similar ao da CP estão relacionadas com respostas de defesa a patógenos em animais. A semelhança estrutural entre CP e essas moléculas sinalizadoras de defesa sugere que, na célula do hospedeiro, o alvo da CP pode ter similaridade estrutural aos receptores das proteínas IL-8/NAP-2 ou MHC-I. Buscas no genoma de *Arabidopsis* usando a sequência desses receptores forneceram alguns potenciais alvos. A maioria desses alvos são proteínas intracelulares ou de membrana que estão envolvidos na percepção de proteínas de avirulência e de padrões moleculares associados a patógenos, corroborando com a hipótese de que a CP agiria como um elicitor de respostas de defesa e poderia interagir com essas classes de receptores.

Embora a função da CP ainda seja desconhecida, com a elucidação da sua estrutura 3D surgiram novas hipóteses sobre o mecanismo molecular pelo qual a CP interagiria com seus hospedeiros.

## ABSTRACT

One of the most challenging problems facing plant pathology is to understand the molecular basis of the interaction between fungal plant pathogens and their host. Often, the mechanisms by which plants defend themselves against a variety of pathogens involve the recognition of specific molecules encoded by these microorganisms. Cerato-platanin (CP) is a moderately hydrophobic protein that is secreted and localized in the cell wall of *Ceratocystis fimbriata* produced by the ascomycete fungus *Ceratocystis fimbriata*, which is the causative agent of canker stain, a severe disease with incidence in a great number of plants, such as *Platanus acerifolia*, *Theobroma cacao*, *Coffea Arabica*, among others. The protein secreted by culture of *Ceratocystis fimbriata* f. sp. *platani* in the medium is able to elicit phytoalexin synthesis in *Platanus acerifolia* leaves and to enhance cell death. According to protein family database CP has been indicated as the founder of the "Cerato-Platanin Family" that includes other secreted fungal proteins involved in a variety of phytopathological phenomena and/or immunological reactions. The search in the Protein Data Bank revealed this protein family is not structurally characterized. In this way, the resolution of 3D structure of molecules involved in the host-fungus interaction is important to the comprehension of the initial events in the plant defence response.

CP has been cloned and expressed in eukaryotic systems in its  $^{13}\text{C}$ ,  $^{15}\text{N}$  isotopically labelled form. Multinuclear and multidimensional nuclear magnetic resonance has been used to determine its 3D solution structure. The CP structure is composed of two  $\beta$ -strand parallel and five  $\beta$ -strands antiparallel, assembled in a seven-stranded  $\beta$ -sheet, facing three helical elements.

The CP dynamics has been determined and the experimental data reveal that CP shows a structure characterized by limited internal mobility. Moreover, the protein has a good stability, keeping its conformation even at high temperature or in a wide range of both pH and urea concentration.

With the aim to search for other protein or domains conformationally similar to the CP, its 3D structure was submitted to structural analysis using *ProFunc* Server. The results of structure alignment show that CP does not share 3D structural similarity with any known

avirulence proteins or elicitors from bacterial or fungal pathogens. However, surprisingly, the core of CP has a fold similar to molecules belonging to the chemokines and to the major histocompatibility complex class I (MHC-I) families. Indeed, the minimum energy CP structure superposes better to the dimeric form of interleukin-8 (IL-8) and of the neutrophil activating peptide-2 (NAP-2), from the chemokines family, and to the  $\alpha 1$ ,  $\alpha 2$ -domains of MHC-I protein. Interestingly, both families of proteins which have a similar fold to CP are related with defense response to pathogens in animals. The structural resemblance between CP and these defense signaling molecules suggests that, in the host cells, CP targets might have structural similarities to the human IL-8/NAP-2 or MHC-I proteins receptors and a search in the *Arabidopsis* genome with them produced a few hits. Not surprisingly, most of these proteins are surface or intracellular receptors implicated in the recognition of molecular patterns associated to pathogens or avirulence, a finding consistent with the hypothesis that CP, acting as an elicitor, should interact with such class of receptor proteins.

In summary, although the biological function of the CP is still unknown, the elucidation of its 3D structure provides some insights about the molecular mechanism by which CP interact with their hosts.

## LISTA DE ABREVIATURAS

Avr	Avirulência
BMRB	Banco de dados de ressonância magnética de moléculas biológicas ( <i>Biological Magnetic Resonance Data Bank</i> )
CD	Dicroísmo circular
CLEANEX	<i>Clean Chemical Exchange</i>
COSY	<i>Correlated Spectroscopy</i>
CP	Cerato-platanina
CSI	<i>Chemical Shift Index</i>
D	Coefficiente de difusão
DOSY	<i>Diffusion Ordered Spectroscopy</i>
DPBB	<i>Double Psi Beta Barrel</i>
HR	Resposta de hipersensibilidade
HSS	Hibridização subtrativa suprimida
IFN- $\gamma$	Interferon- $\gamma$
IL-8	Interleucina-8
LRR	Domínios ricos em leucina ( <i>leucine rich repeat</i> )
MAMPS	Padrões moleculares associados a microorganismos ( <i>microbial associated molecular patterns</i> )
MAPK	Proteínas quinases associadas à mitose ( <i>mitogen activated protein kinases</i> )
MHC-I	Complexo de histocompatibilidade principal de classe I
MpCP	Cerato-platanina de <i>Moniliophthora perniciosa</i>
MpNEP	Proteínas indutoras de necrose e etileno de <i>Moniliophthora perniciosa</i> ( <i>necrosis and ethylene-inducing protein</i> )
NAP-2	Peptídeo ativador de neutrófilos-2
NBS	Sítio de ligação a nucleotídeo ( <i>nucleotide-binding site</i> )
NOE	Efeito nuclear <i>Overhauser</i> ( <i>nuclear Overhauser effect</i> )
NOESY	<i>Nuclear Overhauser Enhancement Spectroscopy</i>
PDB	Banco de dados de estruturas de proteínas (Protein data bank)
Proteínas R	Proteína de resistência
PRR	<i>Pattern recognition receptors</i>

RDC	Constante dipolar residual ( <i>residual dipolar coupling</i> )
$R_H$	Raio hidrodinâmico
RMN	Ressonância magnética nuclear
$T_1$	Relaxação longitudinal
$T_2$	Relaxação transversa
TCR	Receptores associados a células T ( <i>T cell receptors</i> ),
TFA	Ácido trifluoroacético
TIR	Receptores <i>Toll-interleucine-1</i>
TLR	Receptores <i>Toll-like</i>
TNF- $\alpha$	Fator de necrose tumoral- $\alpha$ ( <i>tumor necrosis factor alpha</i> )
TOCSY	<i>Total Correlation Spectroscopy</i>

# SUMÁRIO

## INTRODUÇÃO

Respostas de defesa das plantas à infecção: o diálogo molecular entre planta-patógeno	1
Proteínas elicitoras de fungos	4
O fungo <i>Ceratocystis fimbriata</i>	7
A proteína Cerato-Platanina	9
A família Cerato-Platanina	12
O clan <i>Double Psi Beta Barrel</i> glucanase	15
A espectroscopia de ressonância magnética nuclear na resolução de estruturas de proteínas	17

## OBJETIVOS E JUSTIFICATIVA 22

### CAPÍTULO 1: Expressão e Purificação da Cerato-Platanina

Introdução	24
Artigo 1	26

### CAPÍTULO 2: Determinação da Estrutura da Cerato-Platanina

Introdução	36
Artigo 2	38
Resultados	
Determinação do raio hidrodinâmico	40
Atribuição das ressonâncias	41
<i>Chemical shift index</i>	47
Constante de acoplamento $J_{HNH\alpha}$	48
Efeito nuclear <i>Overhauser</i>	50
Constante dipolar residual	54
Procedimentos Experimentais	
Determinação do raio hidrodinâmico	57
Preparo da proteína $^{15}\text{N}/^{13}\text{C}$ - CP	57
Ressonância magnética nuclear	58
<i>Chemical shift index</i>	60
Constante de acoplamento $J_{HNH\alpha}$	61

Efeito nuclear <i>Overhauser</i>	61
Constante dipolar residual	62
<b>CAPÍTULO 3: Análises Estrutural e Dinâmica da Cerato-Platanina</b>	
Introdução	64
Artigo 3	65
Resultados Complementares	
Ensaio de interação com potenciais ligantes	98
Acessibilidade ao solvente	100
Procedimentos Experimentais	
Ensaio de interação com potenciais ligantes	103
Acessibilidade ao solvente	103
<b>DISCUSSÃO</b>	104
<b>CONCLUSÕES</b>	113
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	114
<b>ANEXO 1</b>	125
<b>ANEXO 2</b>	130
<b>ANEXO 3</b>	135
<b>ANEXO 4</b>	136
<b>ANEXO 5</b>	142
<b>ANEXO 6</b>	146

## INTRODUÇÃO

### Respostas de defesa das plantas à infecção: o diálogo molecular entre planta-patógeno

Assim como os animais, as plantas são permanentemente expostas a agressões do meio ambiente tanto por agentes bióticos quanto por agentes abióticos. Em particular, o estresse biótico, causado por vírus, bactérias, fungos, insetos, nematóides e outros organismos, pode causar sérias doenças ou danos físicos às plantas. Porém, diferente dos animais que podem se locomover e encontrar lugares mais seguros, as plantas devem permanecer e se adaptar às circunstâncias do local aonde elas brotaram. Além disso, como os vegetais não possuem células circulantes, o reconhecimento do patógeno deve ser uma capacidade de cada célula do organismo. Apesar dessa aparente passividade, as plantas percebem as agressões do meio e a sua alta capacidade de adaptação permite que elas sobrevivam. Parte desse sucesso é devido a rapidez e especificidade dos eventos de reconhecimento e sinalização que são codificados pelas células. Se a resposta é rápida e eficiente, a planta pode conter o processo de infecção, caracterizando uma interação dita 'incompatível'. Ao contrário, quando a resposta não é eficiente e a interação entre o patógeno e a planta leva à doença e ela é dita 'compatível' (Gómez-Gómez, 2004; Jones & Dangl, 2006; Mehta et al., 2008).

Entender como microorganismos e plantas reconhecem um ao outro e estabelecem uma relação é crucial neste campo de investigação. Ao longo dos anos a interação planta-patógeno tem sido estudada extensivamente tanto do ponto de vista do patógeno quanto do hospedeiro. Porém, ainda existem muitas dúvidas com relação aos mecanismos de defesa e as substâncias envolvidas nesse processo.

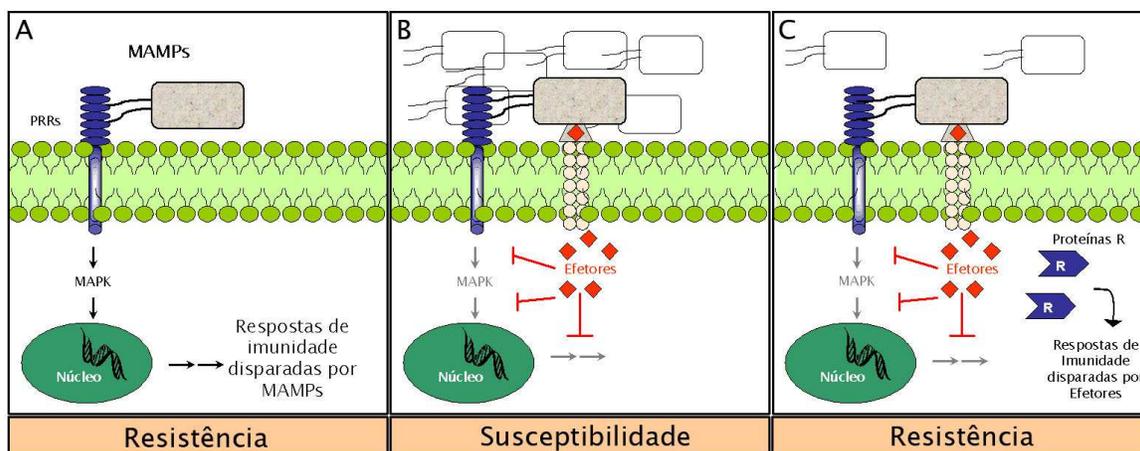
Para ser patogênico o microorganismo deve primeiramente acessar o interior da planta através de feridas da superfície de folhas e raízes, ou de aberturas naturais como os estômatos. Uma vez no interior da planta, o patógeno ainda deve superar outro obstáculo: uma rígida parede celular (Chisholm et al, 2006). A parede celular é uma barreira resistente e estruturalmente heterogênea, constituída principalmente de polissacarídeos, como

celuloses, hemiceluloses, pectinas, e diversas proteínas. Além de prover suporte estrutural à célula, ela também possui uma importante função na defesa da planta contra os patógenos. Por esse motivo, a secreção de enzimas que degradem a parede celular, por exemplo proteases, celulasas e pectina liases, pelos microorganismos é crucial para a colonização da planta hospedeira (Juge, 2006; Mehta et al., 2008).

Uma vez que a parede celular foi atravessada, o microorganismo é exposto à membrana plasmática e entra em contato com receptores de superfície extracelulares que reconhecem padrões moleculares (*pattern recognition receptors* ou PRRs). Esses padrões moleculares associados a microorganismos (*microbial associated molecular patterns* ou MAMPs) são, em geral, compostos altamente conservados em uma grande variedade de micróbios mas que normalmente não estão presentes no hospedeiro. Exemplos clássicos de MAMPs que são comprovadamente reconhecidos por PRRs de plantas são os lipossacarídeos e a flagelina, constituintes da membrana externa e do flagelo de bactérias, respectivamente. A percepção intracelular do patógeno devido ao reconhecimento de MAMPs leva à ativação do sistema de defesa basal e à primeira resposta de resistência da planta. Essa indução está associada a sinalização por MAP quinases (*mitogen activated protein kinases*), tradução de sinal e transcrição de genes de defesa que podem levar a produção de espécies reativas de oxigênio, espessamento da parede celular e acidificação do apoplasto, prevenindo o crescimento microbiano (Figura I.1A). O mecanismo basal e genérico de resposta de defesa tem sido observado em muitas interações incompatíveis planta-microorganismo e acredita-se que isso corrobora com a observação de que as plantas são resistentes à invasão da maioria dos micróbios. Portanto, para colonizar o hospedeiro, o patógeno deve ter mecanismos para interferir na defesa basal da planta e desenvolver a doença (Mehta et al., 2008; Chisholm et al, 2006; Jones & Dangl, 2006).

Existem evidências de que, durante a infecção, os patógenos ativamente suprimem a resposta de defesa disparada através dos MAMPs e aumentam a sua virulência. A supressão da defesa basal ocorre através da secreção de proteínas efetoras no citoplasma da célula da planta que alteram a sinalização ou as manifestações de resposta de resistência (Figura I.1B). As proteínas efetoras, tanto de bactérias quanto de fungos, podem apresentar atividade enzimática e ser responsáveis por modificar as proteínas do hospedeiro, evadindo

a resistência. Apesar da atividade enzimática já ter sido demonstrada para algumas proteínas, a função da maioria dos efetores permanece desconhecida e existem poucas evidências de sua atividade na virulência ou na supressão da defesa da planta. Sabe-se que, durante a infecção, microorganismos fitopatogênicos podem secretar aproximadamente de 20 a 30 efetores (Chisholm et al, 2006). No caso de bactérias, cinco sistemas de secreção já foram descritos, mas o principal sistema usado para secretar proteínas diretamente no citoplasma da célula hospedeira é o sistema secretório tipo III. No caso de fungos, o mecanismo de secreção de efetores não é muito bem conhecido, mas especula-se que o haustório, uma estrutura especializada em alimentação, seja responsável pela liberação de proteínas no apoplasto e estas entram na célula da planta mediadas por um peptídeo sinal (Yip & Strynadka, 2006; Ellis et al., 2006).



**Figura I.1.** Modelos de respostas de defesa das plantas à infecção. (A) Reconhecimento de MAMPs (como a flagelina) por receptores de superfície extracelulares PRRs que disparam uma resposta de imunidade basal. (B) Patógeno secreta proteínas efetoras com objetivo de suprimir a resposta da planta. (C) Proteínas de resistência da planta (proteínas R) reconhecem a atividade dos efetores e restauram a resistência através de respostas de hipersensibilidade (Chisholm et al, 2006).

Não surpreendentemente, uma vez que o patógeno adquire a capacidade de suprimir o sistema de defesa basal do hospedeiro, a planta desenvolve um mecanismo mais especializado para se defender do microorganismo. A imunidade disparada por efetores envolve a ação de proteínas de resistências (R) em reconhecer direta ou indiretamente

efetores codificados por genes de avirulência (Avr) dentro da célula da planta (Figura I.1C). Inicialmente foi proposto que a resistência da planta a um patógeno seria condicionada a presença específica de genes determinantes de avirulência no patógeno e de genes de resistência no hospedeiro. Essa correspondência em par, R-Avr, foi descrita como resistência gene-a-gene (Flor, 1971). Porém, a identificação de interações diretas entre os produtos de genes R-Avr é muito limitada, sugerindo que possivelmente não exista uma correspondência gene-a-gene, mas é provável que proteínas R possam reconhecer indiretamente a atividade de múltiplos efetores Avr. A “hipótese de guarda” foi proposta por Dangl e Jones (2001) e postula que proteínas R do hospedeiro monitoram ou “guardam” a integridade de alvos particulares com o objetivo de detectar a manipulação desses alvos por efetores de avirulência. Como resultado, a ativação de proteínas R media a resistência e suprime o crescimento do microorganismo através de uma resposta de hipersensibilidade (HR) e subsequentemente morte celular programada. A função da resposta de hipersensibilidade é conter o patógeno e pode ser exemplificada por várias caracterizações bioquímicas, incluindo mudança no fluxo de íons, hiperperoxidação de lipídeos, fosforilação de proteínas, geração de espécies reativas de oxigênio e de substâncias com propriedades antimicrobianas, como os compostos fenólicos e as fitoalexinas (Jones & Dangl, 2006; Mehta et al., 2008; Gómez-Gómez et al., 2004).

### Proteínas elicitoras de fungos

Ao contrário de bactérias, pouco se sabe sobre a identidade e função de elicitores de fungos e oomicetos patogênicos. Consideráveis avanços foram alcançados nos últimos 10 anos com relação à identificação de moléculas determinantes da interação planta-fungo. Atualmente, mais de 25 genomas de fungos já foram elucidados, incluindo genomas de fungos patogênicos de planta e de humano, como por exemplo dos fungos *Aspergillus fumigatus* e *Magnaporthe grisea*, respectivamente. O desafio da biologia moderna é analisar a expressão, função e regulação desse grupo de proteínas que são codificadas por esses genomas (Mehta et al., 2008; Chisholm et al, 2006; O’Connell and Panstruga, 2006).

Proteínas secretadas por fungos patogênicos têm sido isoladas e caracterizadas bioquimicamente. Muitas dessas proteínas se localizam primariamente no apoplasto da

planta e são enzimas inibidoras ou proteases que provavelmente funcionam contra a defesa disparada pela plantas (O'Connell and Panstruga, 2006). Por outro lado, fungos patogênicos e oomicetos secretam várias proteínas não catalíticas, em geral de baixo peso molecular, que são caracterizadas como MAMPs. O sistema fungo-planta que tem sido melhor estudado é o do tomate (*Solanum lycopersicum*) e seu patógeno *Cladosporium fulvum*. Os genes *Cf* de tomate conferem resistência ao *C. fulvum* através da codificação de proteínas de membrana com domínios extracelulares ricos em leucina (*leucine rich repeat* ou LRR), sugerindo interações extracelulares com os correspondentes determinantes Avr. Este é o caso da proteína AVR2 que se liga e inibe a ação da protease extracelular Rcr3 e propõe-se que o complexo enzima-AVR2 é o responsável por ativar a resistência *Cf-2*-dependente (Rooney et al., 2005). Um outro exemplo é a proteína AVR9, a primeira proteína Avr a ser identificada em fungos patogênicos. O modelo proposto para a indução de resistência pelo AVR9 assume a formação de um complexo trimérico envolvendo um receptor primário não específico, o ligante AVR9 e um tradutor de sinal específico para iniciar o processo de sinalização de resistência da planta (Luderer et al., 2001).

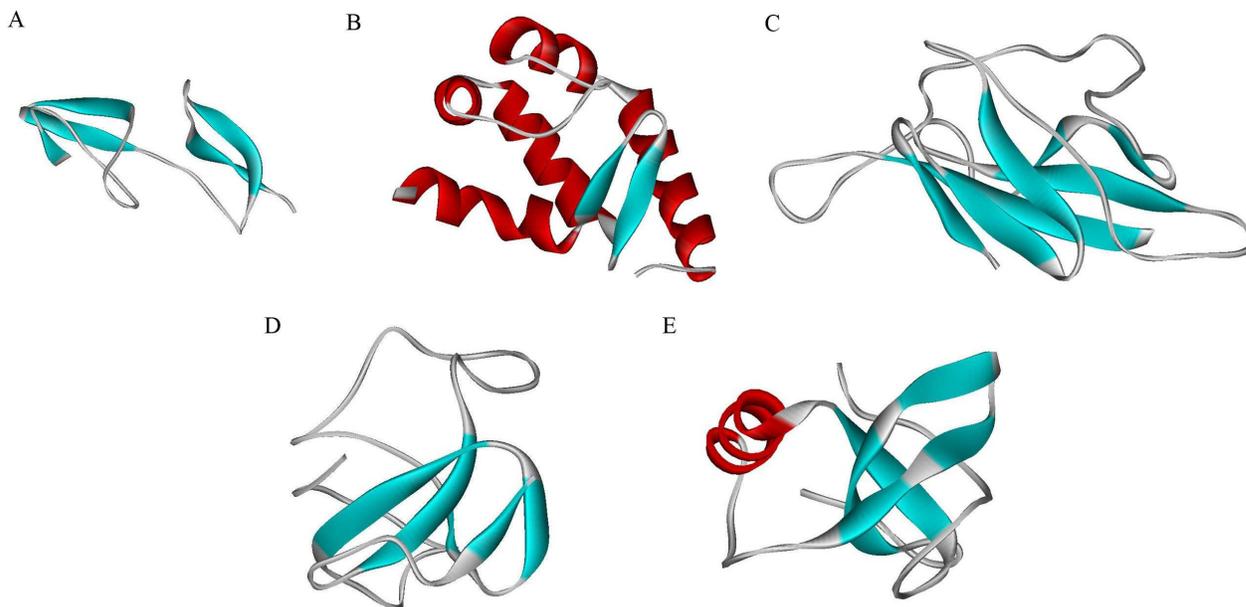
Apesar desses recentes avanços, o número de estruturas tridimensionais de elicitores de fitopatógenos é limitado. Atualmente, entre as estruturas relacionadas a mecanismos de infecção de fungos estão: a NIP1, as elicinas  $\beta$ -criptogeína and  $\beta$ -cinamonina, as AvrL567-A e AvrL567-D e as hidrofobinas EAS e HFBII. Esses elicitores não parecem ter propriedades estruturais em comum e devem possuir diferentes funções na planta hospedeira.

A proteína NIP1 é uma proteína elicitora composta por 60 resíduos de aminoácido, entre os quais são 10 resíduos de cisteína que formam 5 ligações dissulfeto, secretada pelo patógeno de cevada *Rhynchosporium secalis*. A associação da NIP1 com sítios de ligação da proteína de membrana Rrs1 de cevada é o evento inicial na tradução do sinal que leva a indução de necrose e resistência da planta (van't Slot et al., 2007). A sua estrutura é caracterizada por dois domínios de folhas- $\beta$ , compostos, respectivamente, de duas e três fitas- $\beta$  antiparalelas (PDB: 1KG1, Figura 1.2A). As 5 ligações dissulfeto estabilizam esses domínios e definem a posição entre eles (van't Slot et al., 2003).

Elicitinas são proteínas secretadas por espécies de oomicetos *Phytophthora* e *Pythium* que são parasitas de diversas plantações com importância econômica, como batata, feijão e tabaco. Essas proteínas extracelulares são capazes de se ligar e transportar esterol e ácido-graxos através da membrana plasmática da planta, elicitando respostas de defesa. A  $\beta$ -criptogeína (PDB: 1BEO) e  $\beta$ -cinamonina (PDB: 1LJP) possuem um enovelamento composto por cinco e seis  $\alpha$ -hélices, respectivamente, e um motivo *beak-like*, formado por duas pequenas fitas- $\beta$  antiparalelas e um loop  $\Omega$  (Figura I.2B). O motivo *beak-like* das elicinas provavelmente está envolvido no reconhecimento receptor/ligante (Boissy et al., 1996; Rodrigues et al., 2002; Fefeu et al., 1997).

Proteínas AvrL-567 são proteínas de avirulência do fungo *Melampsora lini*, causador de bolor em culturas de linho. A estrutura de ambas as proteínas AvrL567-A (PDB: 2OP6) e AvrL567-D (PDB: 2QVT) são bastante similares e revelam um enovelamento em  $\beta$ -*sandwich* (Figura I.2C). Essas duas proteínas possuem 92% de identidade na sequência de aminoácidos, mas diferem na especificidade do reconhecimento das proteínas de resistência (proteínas L5 e L6). Os dados estruturais revelaram que contatos de diferentes cadeias laterais dos resíduos levam a efeitos na estabilização ou desestabilização da interação e, conseqüentemente, na afinidade, da proteína com o alvo. Um modelo do complexo AvrL567-A e o domínio LRR da proteína L5 também foi proposto (Wang et al., 2007).

As hidrofobinas são famílias de proteínas secretadas por fungos filamentosos, possuem baixo peso molecular e baixa homologia de sequência, mas são caracterizadas por um padrão muito similar de hidrofobicidade e de ligações dissulfeto. Elas se agregam em interfaces hidrofóbicas/hidrofílicas, formando estruturas anfipáticas, que possuem funções distintas no desenvolvimento do fungo e na interação planta-patógeno. As únicas estruturas tridimensionais resolvidas dessas famílias são as proteínas EAS (PDB: 2FMC) e HFBII (PDB: 2B97) (Figura I.2D), de *Trichoderma reesei* e *Neurospora crassa*, respectivamente. Ambas estruturas apresentam um domínio  $\beta$ -*barrel*, porém elas diferem entre si significativamente: EAS possui um *loop* longo e desordenado que não está presente na HFBII, além de duas fitas- $\beta$  na estrutura da proteína EAS que são substituídas por uma  $\alpha$ -hélice na HFBII (Kwan et al., 2006; Hakanpää et al., 2004).



**Figura I.2.** Estruturas 3D de proteínas elicitoras de fitopatógenos: (A) NIP1 (PDB: 1KG1), (B) elicitorina  $\beta$ -cinamonina (PDB: 1LJP), (C) AvrL567-A (PDB: 2OP6) e as hidrofobinas (D) EAS (PDB: 2FMC) e (E) HFBII (PDB: 2B97). Em vermelho estão representadas as estrutura em  $\alpha$ -hélice e em ciano as estrutura em folha- $\beta$ .

### O fungo *Ceratocystis fimbriata*

*Ceratocystis fimbriata* é um ascomiceto que pode causar doença em plantas caracterizada pelo murchamento de folhas, apodrecimento de raízes e cancro (Figura I.3). Este agente fitopatógeno, também conhecido por *Ceratostomela fimbriata*, *Ophiostoma fimbriatum* ou *Endoconidiophora fimbriata*, foi descrito pela primeira vez em 1890 e é a espécie mais difundida do gênero *Ceratocystis*. Ele ataca várias espécies de plantas de grande interesse econômico por todo o mundo, apresentando focos de infecção e perdas nas colheitas principalmente na América Latina. Um total de 31 espécies, representando 14 famílias, tem sido confirmado como hospedeiras. Entre elas estão: manga (*Mangifera indica*), cacau (*Theobroma cacao*), café (*Coffea arabica*), eucalipto (*Eucalyptus* ssp.), seringueira (*Hevea brasiliensis*), entre outros (Harrington, 2000; Baker et al., 2003). Apesar do amplo número de espécies hospedeiras e da sua grande difusão geográfica, tem sido observado que diversas linhagens do *C. fimbriata* têm se especializado em diferentes hospedeiros. Por exemplo, a manga, que é reportada como uma planta hospedeira no Brasil,

crece livre deste fungo na América Central, onde o *C. fimbriata* comumente ataca plantações de café e cacau. Análises filogenéticas de sequências de DNA revelaram que existem outros clados dentro do grupo *C. fimbriata*. (Baker et al., 2003; Johnson et al., 2005; Engelbrecht et al., 2004; Harrington, 2000).



**Figura I.3.** Plátano sadio (esquerda) e com sintomas da doença (direita) após a infecção do fungo *C. fimbriata* (painel central) (<http://www.public.iastate.edu/~tcharrin/>)

Outro hospedeiro do fungo *C. fimbriata* que tem recebido considerável atenção é o plátano (*Platanus acerifolia*), uma planta ornamental encontrada em meios urbanos. A doença tem se mostrado potencialmente a mais perigosa para o futuro desta árvore uma vez que um grande número de plantas, principalmente na Europa e nos Estados Unidos, têm sido atacadas pela forma *Ceratocystis fimbriata* f. sp. *platani*, podendo ser destruída no período de 4 a 7 anos (Alami et al., 1998; Pazzagli et al., 1999). Em particular, a forma *platani* é um grupo morfologicamente específico do fungo *C. fimbriata* que ataca o gênero *Platanus* (Engelbrecht et al., 2004).

No Brasil, a doença causada pelo *C. fimbriata* é conhecida como murcha-de-ceratozystis. Ela foi verificada pela primeira vez em 1978, em Rondônia, e mais recentemente, em 1997-1998, em enxertos e cacauzeiros do sul da Bahia. Desde a sua constatação, a doença tem ocorrido de forma endêmica, causando perdas significativas em plantações e dificultando a atividade agrícola (Silva et al., 2004; Santos & Ferreira, 2003).

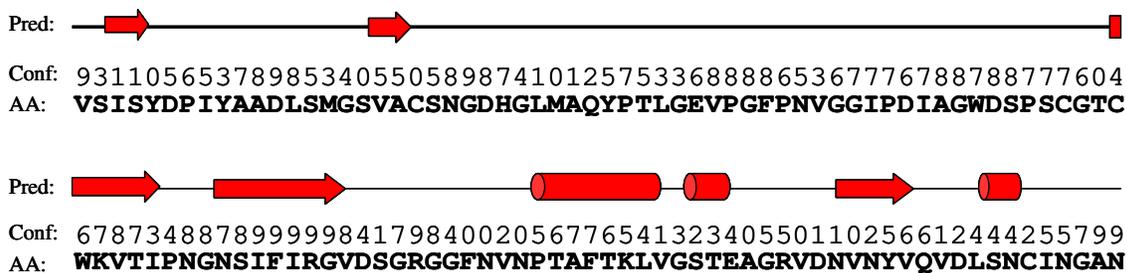
O fungo *C. fimbriata* penetra na planta necrosando os raios medulares, comprometendo o xilema, causando murcha e secamento das folhas, que permanecem aderidas à planta por muito tempo, mesmo após a sua morte. Os sintomas no tronco e nas raízes se expressam em forma de lesões de coloração marrom acastanhada (Silva et al., 2004; Ferreira et al., 2006) (Figura I.3). O patógeno espalha-se por meio de insetos, pelas raízes de plantas adjacentes e através de feridas no tronco, galhos e raízes. Porém, o fungo é transmitido primariamente por injúrias mecânicas causadas por ferramentas contaminadas que são usadas durante a poda. Uma vez na planta, o fungo move-se através do xilema e pode sobreviver por vários meses ou até mesmo anos (Baker et al., 2003; Silva et al., 2004; Pazzagli et al., 1999).

Até o momento, as únicas formas de controle da doença são: a vistoria da lavoura; a não realização de tratos culturais nem de colheitas em plantas com sintomas suspeitos; a eliminação imediata de toda a plantação doente ou morta e, se possível, a queima no local; a desinfecção das ferramentas usadas na execução das práticas de controle; e o replante somente após 60 dias (Ferreira et al., 2005; Ferreira et al., 2006).

### A proteína Cerato-Platanina

Diversas proteínas extracelulares e de baixo peso molecular produzidas por fungos fitopatógenos têm apresentado atividade fitotóxica. A cerato-platanina (CP) é uma proteína liberada abundantemente pelo ascomiceto *Ceratocystis fimbriata* f. sp. *platani*. Em 1999, Pazzagli e co-autores reportaram que 0,8 nmol de CP é capaz de induzir necrose em folhas de tabaco e que a incubação desta proteína pura em folhas de tabaco e de plátano provoca a síntese de substâncias fluorescentes, sugerindo a liberação de fitoalexinas (Pazzagli et al., 1999). Esse comportamento não define exatamente a função da CP na biologia do *C. platani*, porém sugere um envolvimento potencial na interação fungo-hospedeiro.

A proteína CP foi isolada a partir de culturas de fungos *Ceratocystis fimbriata* f. sp. *platani* e sua estrutura primária foi determinada (Figura I.4). Ela contém 120 resíduos de aminoácidos, peso molecular de 12,4 kDa, contendo 40% de resíduos hidrofóbicos e duas ligações dissulfeto (Cys<sup>20</sup>-Cys<sup>57</sup> e Cys<sup>60</sup>-Cys<sup>115</sup>) (Pazzagli et al., 1999). A sequência primária da CP e a predição da sua estrutura secundária estão apresentadas na Figura 4.



**Figura I.4.** Sequência primária (AA) e predição da estrutura secundária (Pred) da proteína cerato-platanina. Os cilindros representam estrutura em  $\alpha$ -hélice, as flechas representam estrutura em folha- $\beta$  e a linha representa regiões sem estrutura definida. A confiabilidade da predição (Conf) está representada através da escala de 0 a 9, sendo 0 indica baixa confiabilidade e 9 indica de alta confiabilidade (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>).

O gene *cp* consiste em um único exon de 360 pares de base que codifica uma proteína de 134 resíduos de aminoácidos. Um peptídeo sinal de 14 resíduos é removido durante a secreção da proteína madura. Além do fato da CP ser secretada no meio, ela também está presente nas paredes dos ascósporos, hifas e conídias do *C. platani*, indicando que ela tem uma importância na formação da parede celular do fungo (Pazzagli, et al., 1999; Boddi et al., 2004). Assim como outras proteínas da superfície de fungos, foi observado que a CP é capaz de agregar-se *in vitro*, após a sua incubação em 30% de trifluoretanol a 37°C por 30 dias, condições essas necessárias para a desestabilização inicial da proteína (Carresi et al., 2006). Através de imagens de microscopia de força atômica, foi observado que a CP tende a formar oligômeros com forma anelada, eventualmente resultando em agregados bem organizados, conhecidos como fibrilas. Um modelo simples, compatível com a forma e tamanho do agregado observado pelas imagens da microscopia,

foi proposto baseado na hipótese de que a agregação é energeticamente favorável quando a superfície exposta é reduzida (Sbrana et al., 2007).

Em 2004, Scala e co-autores confirmaram, através de análises espectroscópicas, a habilidade da CP em elicitar a síntese de fitoalexinas em plantas hospedeiras e não-hospedeiras. Foi demonstrado que uma pequena quantidade de CP (< 2,0 nmol) induz a síntese de umbeliferona e gliceolina, que são referências de fitoalexinas em plátano e soja, respectivamente (Scala et al., 2004). Recentemente, também foi mostrado que a concentração da proteína CP diminui de 150  $\mu$ M a 0 M após 48h quando aplicada sobre a superfície de folhas de plátano. Nesse mesmo intervalo de tempo os níveis de fitoalexina são aumentados. Além disso, foi observado que quando a superfície de folhas de plátano é tratada com CP ocorre uma inibição do crescimento do fungo *C. platani* (Fontana et al., 2008). A partir desses resultados, especula-se que a CP é absorvida pela superfície da folha e que a planta, reconhecendo uma substância que não é da sua própria natureza, dispara uma resposta de defesa.

A hipótese de que a CP é um elicitor efetivo de respostas de defesa foi confirmada por ensaios de hibridização subtrativa suprimida (HSS). Os resultados de HSS revelaram que genes de folhas de plátano relacionados com defesa são transcritos após o tratamento com CP. Um total de 78 sequências de genes diferencialmente expressos foi identificado, sendo que 19,2% dessas sequências estão associadas a genes que codificam proteínas envolvidas em defesa e/ou estresse (Fontana et al., 2008). Por exemplo, foi identificado o gene Rar1 que codifica proteínas que regulam o gene-R, responsável por mediar respostas de resistências em várias espécies de plantas (Freialdenhoven et al., 1994); além dos genes que codificam a enzima  $\beta$ -1,3 glucanase, que é uma endo-glicohidrolase, uma enzima envolvida na resistência da planta contra a invasão de fungos (Neuhaus et al., 1992); ou ainda a ribulose-fosfato 3-epimerase e translocase, que são duas enzimas chaves para a via das pentoses-fosfato, a principal fonte de compostos fenólicos em plantas e portanto também associadas a mecanismos de resistência (Buchanan et al., 2000), entre outros (Fontana et al., 2008).

## A família Cerato-Platanina

No banco de dados de família de proteínas *Pfam* (Finn et al., 2008) a CP é descrita como o primeiro membro da “Família Cerato-Platanina” (PF07249). Até o momento a Família Cerato-Platanina é constituída por 44 proteínas de aproximadamente 150 resíduos de aminoácido com alta similaridade a CP, secretadas por 30 espécies de fungos. Em alguns casos essas proteínas estão relacionadas a fenômenos fitopatológicos e reações imunológicas, como por exemplo a proteína *snodprot1* de *Phaeosphaeria nodorum* que é produzida durante a infecção de folhas de trigo, a SP1 de *Leptosphaeria maculans*, causador da doença canela preta em canola (Wilson et al., 2002), a MSP1 de *Magnaporthe grisea*, causador da doença de brusone em arroz (Jeong et al., 2007), a Asp f 15 de *Aspergillus fumigatus* causador de reações alérgicas em humanos (Rementeria et al., 2005), a ACA1 de *Antrodia camphorata*, que possui atividades imunomodulatória (Sheu et al., 2005), e o antígeno CS-Ag de *Coccidioides immitis*, causador de doenças respiratórias (Cole et al., 1989). Em outros casos, essas proteínas estão associadas a fungos de vida livre, como por exemplo as proteínas Sm1 e Epl1 que são secretadas pelos fungos *Thricoderma virens* (Djonović et al., 2006) e *Thricoderma atroviridis* (Seidl et al., 2006), respectivamente, e que não possuem atividade tóxica a plantas, microorganismos ou animais.

Pouco se sabe sobre a função das proteínas que compõem a família CP. Porém, recentes estudos têm demonstrado o envolvimento dessas moléculas durante o processo de colonização e/ou infecção pelo fungo. Por exemplo, Jeong e co-autores demonstram que a proteína MSP1 é requerida para a virulência do fungo *Magnaporthe grisea*. Foi reportado que mutantes do fungo *M. grisea* com gene *mSP1* suprimido possuem um fenótipo indistinguível ao do fungo nativo porém, esse mutante possui virulência bastante reduzida (Jeong et al., 2007).

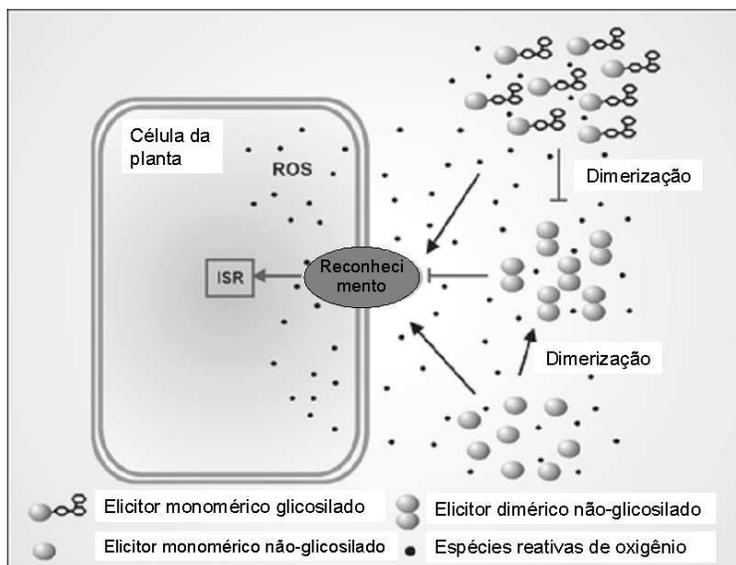
Resultados interessantes também foram mostrados para a Sm1, uma proteína secretada pelo fungo *Thricoderma virens*. O fungo *T. virens* é utilizado com sucesso como agente de controle biológico em plantas, uma vez que ele é capaz de induzir resistência a doenças causadas por outros fungos, como *Rhizoctonia solani*, *Pythium ultimum* e *Botrytis cinera*. Foi demonstrado que a Sm1 é uma proteína que não tem atividade tóxica ou

enzimática, mas que ela é requerida para a habilidade do fungo *T. virens* em induzir resistência sistêmica em algodão e milho e que ela é capaz de elicitar repostas de defesa em plantas. A atividade protetora de Sm1 está relacionada à indução e ao acúmulo de espécies reativas de oxigênio e compostos fenólicos e ao aumento da transcrição de genes de defesa e de genes envolvidos na biosíntese de fitoalexinas. De fato, foi reportado que a super expressão de Sm1 aumenta os níveis de proteção sistêmica da planta (Djonović et al., 2007; Djonović et al., 2006; Vargas et al., 2008).

Os mecanismos moleculares de reconhecimento da Sm1 e tradução do sinal ainda não foram descritos, porém alguns discernimentos sobre o controle da atividade elicitora dessa proteína foram feitos. Vargas e co-autores demonstraram que somente a forma monomérica das proteínas Sm1 e Epl1 (uma proteína homóloga a Sm1) é capaz de ativar repostas de defesa em planta. Ao contrário, a dimerização dessas proteínas faz com que elas percam a sua propriedade elicitora, não protegendo a planta contra a ação do patógeno. Dessa forma, especula-se que a planta possua mecanismos específicos para o reconhecimento de elicitores. Esses receptores seriam capazes ou de discriminar entre as formas monoméricas e diméricas da proteína ou de reconhecer motivos particulares da forma monomérica da proteína que não estariam mais expostos quando as subunidades se agregassem, com consequente alteração da tradução do sinal e da ativação de mecanismos de resistência (Vargas et al., 2008).

Caracterizações bioquímicas da Sm1 e Epl1 demonstraram que dois fatores principais podem controlar o estado oligomérico dessas proteínas e portanto as suas atividades elicitora. O primeiro deles é a glicosilação. Foi mostrado que a forma monomérica da Sm1 é produzida como uma glicoproteína. A presença da molécula de açúcar proporcionaria um impedimento estérico à associação das subunidades protéicas e, consequentemente, impediria a sua dimerização (Figura I.5). O outro fator que controla o estado oligomérico dessas proteínas é a oxidação de resíduos de triptofano. Sabe-se que a oxidação de resíduos de Trp pode estar envolvida na agregação e inativação de algumas proteínas. Nesse caso, foi demonstrado que a forma dimérica da Epl1 apresenta um dos seus resíduos de Trp no estado oxidado. Assim, propõe-se que o acúmulo de espécies reativas de oxigênio em células de planta poderia estimular a oxidação dessas proteínas e,

consequentemente, a sua agregação, alterando o reconhecimento e ativação de respostas de respostas (Figura I.5) (Vargas et al., 2008).



**Figura I.5.** Processo bioquímico proposto para o controle de indução de resistência por elicitores proteicos. Elicitores produzidos na forma monomérica podem ser reconhecidos pela célula e respostas de defesas, como a produção de espécies reativas de oxigênio, são ativadas. Porém, a forma monomérica dos elicitores não-glicosilados é susceptível a dimerização oxidativa e, nesse caso, eles não serão reconhecidos pela célula, se tornando inativos. Quando o elicitor monomérico é produzido na forma glicosilada, a presença do açúcar impede a dimerização da proteína e mantém o elicitor ativo, permitindo o disparo de respostas de defesa da planta (Vargas et al., 2008).

Recentemente, também foi encontrado no genoma do fungo *Moniliophthora perniciosa*, causador da vassoura-de-bruxa em plantações de cacau, pelo menos cinco fases abertas de leitura (*open reading frame* ou ORFs) homólogas ao gene que codifica a proteína CP. Um dos genes *cp* de *M. perniciosa* foi expresso em *Escherichia coli* e a habilidade da proteína recombinante obtida (MpCP1) de induzir necrose em folhas de cacau e tabaco foi demonstrada. Porém, foi reportado que o perfil da lesão causada pela MpCP1 apresenta características diferentes da necrose observada quando a planta é infectada pelo fungo *M. perniciosa*. Entretanto, foi mostrado que uma mistura de MpCP1 com proteínas indutoras de necrose e etileno-2 (*necrosis and ethylene-inducing protein-2* ou MpNEP2), uma outra

família de elicitores expressas por *M. pernicioso*, além de possuir um efeito sinérgico na progressão da doença, apresenta, nesse caso, os sintomas da necrose similares àqueles encontrados em plantas naturalmente infectadas pelo patógeno. Isso indicaria que a presença dessas duas famílias de proteínas pode ser relevante para o progresso da infecção (Zaparoli et al., 2008).

### O clan *Double Psi Beta Barrel* glucanase

No banco de dados de família de proteínas *Pfam* (Finn et al., 2008) a CP também é reportada como membro da superfamília das *Double Psi Beta Barrel* glucanases (DPBB, CL0199). O clan das DPBB glucanases é constituído por cinco famílias de proteínas: o domínio 3D (PF06725), a família 45 das glicosil hidrolases (PF02015), a família *Rare lipoprotein A (RlpA)-like double psi beta barrel* (PF03330), a família barwin (PF00967) e a família cerato-platanina (PF07249). De maneira geral, os membros que constituem o clan DPBB apresentam domínios enzimáticos, porém as proteínas que constituem essas famílias se mostram de forma bastante diversa em relação à similaridade estrutural e funcional (Castillo et al., 1999).

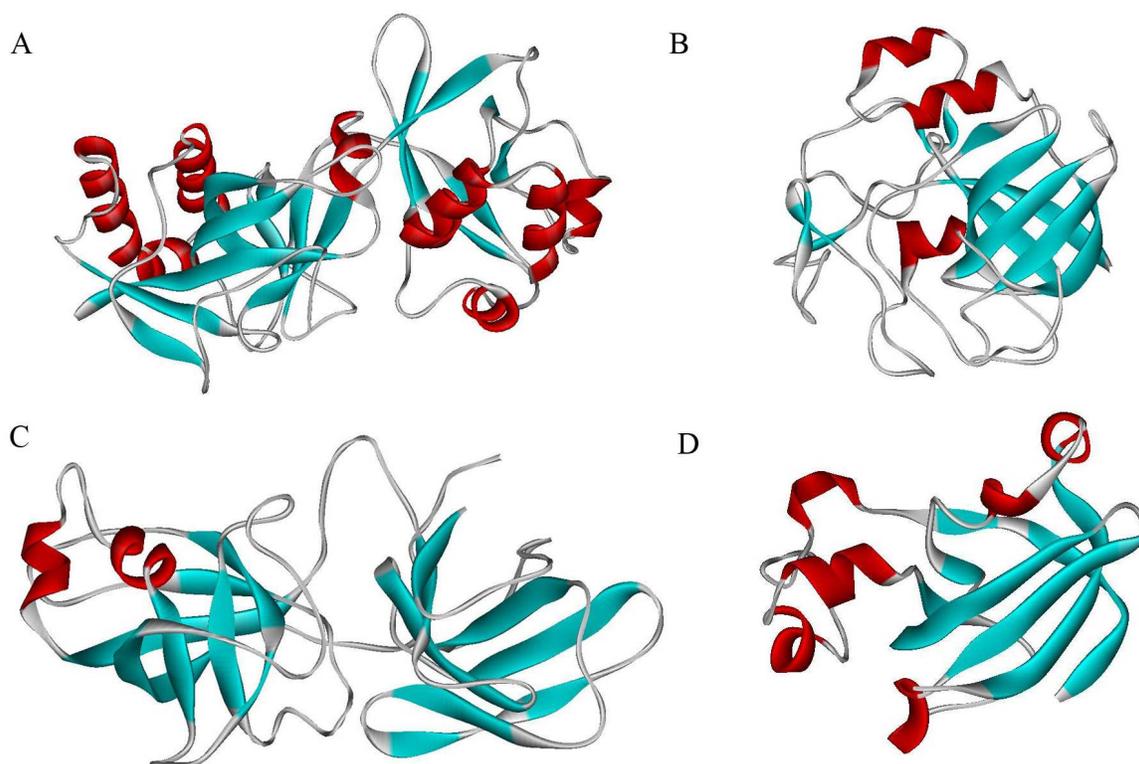
Os membros da família domínio 3D são proteínas que estão envolvidas na manutenção e crescimento da parede celular bacteriana através da clivagem de ligações  $\beta$ -(1,4)-glicosídicas de peptídeos glicanos. Essas proteínas apresentam três aspartatos conservados na sua estrutura primária, que podem estar envolvidos na ligação a cátions, e por isso o nome 3D. Além disso, o ácido aspártico central se encontra em um motivo DTG que é característico de sítios ativos de peptidases. A única proteína com estrutura resolvida dessa família é a proteína MltA de *Escherichia coli* (PDB: 2AE0, Figura I.6A). Ela é composta por dois domínios separados, um maior que está envolvido no enovelamento *double-psi  $\beta$ -barrel* e um domínio menor que também tem uma topologia  *$\beta$ -barrel*. Uma grande cavidade separa os dois domínios onde as cadeias de glicanos se ligam (van Straaten et al., 2005).

As glicosil-hidrolases são um grupo de enzimas que hidrolisam ligações glicosídicas e portanto estão envolvidas na degradação de celuloses e xilanos. Em particular, a família

45 reúne enzimas que possuem apenas atividade endoglucanase. A região mais conservada dessa família está localizada na região N-terminal, que contém um aspartato que está envolvido no mecanismo catalítico. Outra característica dessas proteínas é a presença de várias cisteínas envolvidas em ligações dissulfeto (Henrissat & Bairoch, 1993). Até o momento, 6 estruturas dessa família já foram resolvidas, entre elas, a estrutura da endoglucanase-5 de *Humicola insolens* (PDB: 1HD5, Figura I.6B) que é constituída de seis fitas- $\beta$ , formado um motivo  $\beta$ -barrel, que são conectadas por três estruturas helicoidais.

A família *RlpA-like* DPBB é pouco estudada e assim a função das proteínas que constituem esse grupo ainda não é bem entendida. Porém, os membros que compõem a família *RlpA-like* DPBB são bastante diversos e podem conter várias funções. A única proteína da família com estrutura resolvida é a do pólen alergênico de grama PHL P1 (PDB: 1N10, Figura I.6C), que é constituída de dois domínios  $\beta$ -barrel (Castillo et al., 1999).

A proteína fundadora da família barwin, e que dá nome a família, foi isolada de extratos aquosos de sementes de cevada. Análises comparativas de sequência mostram que essa proteína é altamente similar à proteína alergênica heveína de seringueira, as regiões C-terminais de proteínas codificadas pelos genes *win1* e *win2*, relacionados ao murchamento de batatas, e a proteínas relacionadas a patogenicidade de tabaco. A alta similaridade entre essas proteínas e a habilidade da proteína barwin em se ligar a tetrâmeros de N-acetilglucosamina sugerem que esse domínio esteja envolvido em mecanismos de defesa da planta (Svensson et al., 1992). A estrutura em solução da proteína barwin foi resolvida (PDB: 1BW3, Figura I.6D) e é composta de quatro fitas- $\beta$  antiparalelas, duas fitas- $\beta$  paralelas e quatro pequenas  $\alpha$ -hélices (Ludvigsen & Poulsen, 1992).



**Figura I.6.** Estruturas 3D dos membros do clan *Double Psi Beta Barrel* glucanases: (A) MltA (PDB: 2AE0), (B) endoglucanase-5 (PDB: 1HD5), (C) PHL P1 (PDB: 1N10) e (D) barwin (PDB: 1BW3).

### A espectroscopia de ressonância magnética nuclear na resolução de estruturas de proteínas

A disponibilidade de estruturas tridimensionais de macromoléculas tem aumentado imensamente o conhecimento de como os processos biológicos ocorrem, principalmente nas relações entre estrutura-função desses compostos. Nas últimas décadas, a espectroscopia de ressonância magnética nuclear (RMN) emergiu e se estabeleceu como uma técnica poderosa para a determinação de estrutura 3D de proteínas e ácidos nucleicos em solução, contribuindo para estudos de mecanismos moleculares.

No início dos anos 80, estudos feitos por R. R. Ernst e K. Wüthrich demonstraram que era possível obter, com resolução atômica, a estrutura de pequenas proteínas utilizando a técnica de RMN. A incorporação da marcação uniforme de isótopos de  $^{13}\text{C}$ ,  $^{15}\text{N}$  e  $^2\text{H}$  em sistemas biológicos de interesse, acoplado ao desenvolvimento de gradientes e das

sofisticadas sequências de pulso e ao aumento significativo da força dos campos magnéticos, proporcionaram o aumento da resolução de espectros, tornando-se possível, nos dias de hoje, a resolução de estruturas de proteínas com massa molecular de até ~40 kDa. Paralelamente ao desenvolvimento experimental, aperfeiçoamentos metodológicos, como a automação computacional, também se tornaram disponíveis, permitindo novos melhoramentos na qualidade das estruturas resolvidas por RMN (Kay, 1997, Spronk et al., 2004; Claridge, 1999).

A base de todos os experimentos de RMN é uma propriedade mecânica quântica do núcleo: o *spin*. Associado ao *spin* está o momento magnético, que nesse caso pode ser interpretado como um dipolo magnético. Esses dipolos atômicos podem se orientar paralela- ou antiparalelamente a um campo magnético externo. As duas orientações possíveis correspondem a energias diferentes e os *spins* podem “pular” de uma orientação à outra, absorvendo ou emitindo a diferença de energia na forma de radiação eletromagnética. O desbalanço do momento magnético nuclear orientado paralela- e antiparalelamente dá origem a uma pequena polarização do *spin* nuclear na amostra, resultando em uma rede macroscópica de magnetização (Wider, 2000; Kay, 1997; Claridge, 1999).

Dessa forma, os experimentos de RMN são realizados em espectrômetros compostos de um magneto supercondutor, que produz um campo magnético estático extremamente forte e homogêneo, e um console que pode gerar ondas eletromagnéticas. Com a irradiação eletromagnética, a magnetização pode ser manipulada pelos espectroscopistas de RMN. A frequência da irradiação deve ser igual a diferença de energia entre os dois estados possíveis do *spin* e, tipicamente abrange a radio-frequência entre 50-900 MHz. Após a perturbação do equilíbrio dos *spins*, através de vários pulsos de radiação de radio-frequência o sinal de RMN pode ser observado (Wider, 2000; Kay, 1997; Claridge, 1999).

Uma propriedade fundamental explorada na técnica de RMN é o fato de que o momento magnético de núcleos individuais pode interagir com pequenos campos magnéticos criados por *spins* vizinhos. Essa interação *spin-spin* é usada para correlacionar diferentes núcleos em uma molécula. Os núcleos podem interagir diretamente, através do espaço, ou indiretamente, através da ligação química. As análises de um espectro de RMN

são baseadas nessas correlações entre ligação química e espaço. Porém, em particular, as informações obtidas através do espaço provêm a base para as informações geométricas requeridas para determinar a estrutura de moléculas em solução e são medidas através do Efeito Nuclear *Overhauser* (*nuclear Overhauser effect* ou NOE). O NOE reflete a transferência de magnetização entre *spins* que estão interagindo via seus dipolos associados. Assim, no caso de proteínas, os núcleos que estão envolvidos na correlação de NOE podem pertencer a resíduos de aminoácidos que estão separados na estrutura primária, mas que estão próximos no espaço a uma distância menor que 5 Å. Para moléculas com peso molecular maior que 5 kDa a intensidade do NOE é proporcional a  $r^{-6}$ , onde  $r$  é a distância entre os dois *spins* que estão interagindo (Wider, 2000; Kay, 1997; Claridge, 1999). Recentemente, um avanço importante para o refinamento de estrutura de biomoléculas tem sido o uso das constantes dipolares residuais (RDCs) como restrições de orientação internuclear relativa ao campo magnético. Ao contrário dos dados de NOEs, as constantes dipolares contêm informações valiosas que não dependem da distância e assim melhoram significativamente a qualidade das estruturas obtidas (de Alba & Tjandra, 2002).

Além da espectroscopia de RMN, a principal técnica experimental utilizada para derivar modelos estruturais é a difração de raios-X de cristais de proteínas. A técnica de difração de raios-X foi introduzida cinco décadas atrás com a determinação da estrutura 3D do DNA e da mioglobina. Ela se tornou uma técnica bem estabelecida, produzindo cada vez mais estruturas acuradas, apesar de ter como fator limitante a necessidade de obtenção de cristais da proteína de interesse. A espectroscopia de RMN em alta resolução é uma técnica muito mais jovem e, como descrito anteriormente, foi usada pela primeira vez para a determinação de estrutura de proteínas e ácidos nucléicos apenas 20 anos atrás. Desde então, cerca de 7500 estruturas 3D determinadas através da técnica de RMN já foram depositadas no banco de dados de estruturas de proteínas (*Protein Data Bank* ou PDB), o que corresponde a 15% do total de estruturas do banco. Mais importante que isso porém, é o fato de que o número de enovelamentos únicos determinados por RMN é maior, aproximadamente 25%, e portanto as estruturas resolvidas através dessa técnica são uma importante fonte de informação no enovelamento de proteínas (Spronk et al., 2004). Além disso, a maioria das estruturas resolvidas por RMN (~75%) não possui estrutura

cristalográfica correspondente, tornando-se assim uma técnica complementar e única para determinação de estruturas de proteínas de difícil cristalização. Outra vantagem de resolver estruturas por RMN é que esses estudos são feitos em solução aquosa, e portanto, em condições bastante similares às condições fisiológicas na qual a proteína age. Essa característica também permite que sejam feitas comparações entre diferentes condições do meio que, por exemplo, poderiam modular a relação estrutura-função, assim como avaliar a flexibilidade física e a dinâmica de interação com outras moléculas. O estudo de proteínas por RMN fornece um conjunto de conformações de moléculas de baixa-energia que satisfazem critérios geométricos específicos determinados por protocolos experimentais. Enquanto cada conformação pode ser considerada como uma “fotografia” estática da molécula, o conjunto de estruturas representa a dinâmica da proteína (Montelione et al., 2000; Cozzini et al., 2008).

É importante ressaltar que os modelos estruturais não são gerados meramente para serem publicados e depositados em banco de dados públicos, mas supõe-se que eles possam explicar funções biológicas e servir como guias para estudos posteriores. A resolução de estruturas também é relevante para aumentar a descrição dos enovelamentos possíveis das proteínas existentes na natureza, além de facilitar a predição de estrutura, através dos métodos de bioinformática, de uma grande fração de sequências cujas estruturas não serão determinadas experimentalmente. Para cada nova estrutura descoberta em média 100 sequências de proteínas sem nenhuma caracterização estrutural podem ser modeladas (Baker & Sali, 2001). Porém, é sempre importante perceber que a representação 3D de uma molécula não é a sua estrutura real, mas um modelo de estrutura construído para representar dados experimentais. O modelo da estrutura pode ou não ser uma representação útil da verdade, dependendo da qualidade dos dados e das habilidades do experimentador em traduzir os dados experimentais em coordenadas estruturais (Spronk et al., 2004).

Por fim, além das informações estruturais, a espectroscopia de RMN também provê informações sobre a dinâmica de macromoléculas. Apesar da estrutura estática de proteínas ser bastante informativa, ela é frequentemente insuficiente para explicar completamente o sistema. A mobilidade tem uma contribuição crítica para a energia de interação das macromoléculas e o custo entrópico dessa interação ajuda a garantir a seletividade das

ligações. Portanto, identificar as regiões flexíveis das macromoléculas pode ajudar a esclarecer os mecanismos de interação e função da proteína. Vale a pena enfatizar que a RMN é a única técnica disponível que pode monitorar e discriminar esses processos moleculares. A ampla abrangência de escalas de tempo acessíveis a RMN permite a observação de modos de movimento tão diversos quanto os movimentos dos vetores de ligação ou mudanças conformacionais. As propriedades de relaxação, como  $T_1$  (relaxação longitudinal),  $T_2$  (relaxação transversa) e heteroNOEs, do núcleo  $^{15}\text{N}$  são os parâmetros de medida mais comumente utilizados para estudos de dinâmica de proteínas na escala de tempo de pico a nanosegundos. A velocidade de transferência de magnetização entre prótons com diferentes deslocamentos químicos e de troca próton/deutério, ao contrário, reportam movimentos muito lentos (ms a dias) e permite o discernimento de mudanças conformacionais. Essa combinação única de estrutura e dinâmica leva a RMN ser uma ferramenta poderosa para caracterizar propriedades termodinâmicas e cinéticas de interação de moléculas (Ishima & Torchia, 2000; Ferentz & Wagner, 2000; Kay, 1997; Cozzini et al., 2008).

## OBJETIVOS E JUSTIFICATIVA

O controle de doenças de plantas é largamente realizado por meio de pesticidas e o uso racional desses produtos pode ter, em curto prazo, um efeito positivo para o produtor. Porém, a longo prazo, além do surgimento de fitopatógenos resistentes a essas substâncias químicas, os resultados à sociedade e ao meio ambiente podem se tornar negativos devido à poluição causada pelos resíduos. A indução da defesa natural de plantas contra fitopatógenos é uma ferramenta atrativa para o controle alternativo de pragas. Porém, o desenvolvimento desses mecanismos de defesa requer um conhecimento detalhado da biologia molecular e da bioquímica de interação microorganismo-hospedeiro.

Atualmente, a pesquisa tem-se direcionado a identificar moléculas e seus respectivos receptores capazes de induzir respostas de defesa. Apesar de diferentes compostos já terem sido relatados como indutores do aumento da resistência das plantas aos patógenos, estudos da função dessas substâncias na interação parasita-hospedeiro são limitados, principalmente no caso de fungos. Assim, a elucidação estrutural de moléculas possivelmente envolvidas nesta interação é de primária importância para a compreensão dos eventos iniciais que levam às respostas de defesa do hospedeiro. Um completo entendimento das bases moleculares que envolvem a resistência de plantas poderá permitir a aplicação dessas descobertas na construção de plantas que contenham formas resistentes a um maior número de patógenos.

A incubação da proteína CP pura em folhas de tabaco e de *Platanus acerifolia* provoca a ativação do mecanismo de defesa dessas plantas. Esse comportamento sugere um envolvimento potencial na interação fungo-hospedeiro. Neste contexto, este trabalho teve como objetivo determinar a estrutura tridimensional da CP e estudar suas propriedades dinâmicas através das técnicas de Dicroísmo Circular e Ressonância Magnética Nuclear.

Além disso, bancos de dados revelam que não há proteínas homólogas a CP com estrutura tridimensional resolvida. Assim, a elucidação da estrutura desta proteína poderá ser de grande interesse para o aumento de informações disponíveis em banco de dados específicos, como o *Protein Data Bank*, facilitando o estudo de proteínas homólogas à CP. O depósito de novas estruturas 3D em bancos de dados é fundamental para o aumento do

impacto dessas bibliotecas, que são utilizadas como ferramentas essenciais para a troca de conhecimento entre membros da comunidade científica.

Por fim, é de grande importância ressaltar que o desenvolvimento desse projeto permitiu também a formação e capacitação da aluna em uma área de pesquisa com potencial ainda pouco explorado no Brasil, e com carência de pessoal qualificado. Devido à recente implantação e utilização das técnicas de determinação de estrutura de proteínas por RMN no país, o desenvolvimento da pesquisa descrita nessa tese visava contribuir significativamente para o crescimento brasileiro nesta área de pesquisa.

## CAPÍTULO 1: Expressão e Purificação da Cerato-Platanina

### Introdução

Este primeiro capítulo apresenta os resultados de clonagem em *Pichia pastoris*, expressão e purificação da proteína de estudo. O trabalho está apresentado na forma do artigo científico que foi publicado na revista *Protein Expression and Purification*.

As células de *P. pastoris* foram clonadas pelos nossos colaboradores Dra. Barbara Pantera, Dra. Luigia Pazzagli e Prof. Gianni Cappugi do *Dipartimento di Scienze Biochimiche, Università di Firenze*, Itália, e disponibilizadas ao nosso grupo para subsequente expressão em larga escala, enriquecimento isotópico e análises estruturais. A proteína CP recombinante foi obtida através da expressão do gene *cp* em cepas GS115 de leveduras de *P. pastoris*, utilizando-se o vetor de expressão extracelular pPIC9, contendo uma sequência sinal que endereça a secreção da proteína. A principal vantagem em se utilizar esse sistema é que a levedura secreta um alto nível de proteínas heterólogas. Esse fato, combinado com a pouca quantidade de proteína no meio mínimo de crescimento, fez com que a CP recombinante fosse a principal proteína do meio de crescimento das células, servindo com um primeiro passo de purificação. Outra característica importante é o fato da proteína ser expressa em sua forma solúvel e enovelada.

Devido à característica hidrofóbica da proteína CP, a sua purificação foi realizada utilizando-se como estratégia a cromatografia de fase reversa. Esse método separa proteínas e peptídeos com diferentes hidrofobicidades devido a sua interação reversível com a superfície hidrofóbica da fase estacionária da coluna, neste caso com os grupos octanodecil-alifático ligados à superfície da sílica. Com a alteração das condições da fase móvel, ocasionada pelo aumento da concentração de acetonitrila, as moléculas ligadas são eluídas em tempos diferentes. Esse único passo de purificação foi suficiente para obter a proteína pura.

Apesar da proteína recombinante obtida apresentar três resíduos de aminoácidos adicionais na região N-terminal quando comparada a sequência primária da proteína nativa,

os espectros de CD mostraram que a conformação estrutural da CP recombinante obtida é semelhante a da CP nativa.

**ARTIGO 1:**

Cerato-platanin, a phytotoxic protein from *Ceratocystis fimbriata*:  
Expression in *Pichia pastoris*, purification and characterization

Lara Carresi<sup>a</sup>, Barbara Pantera<sup>a</sup>, Camilla Zoppi<sup>a</sup>, Gianni Cappugi<sup>a</sup>, Aline L. Oliveira<sup>b</sup>,  
Thelma A. Pertinhez<sup>c</sup>, Alberto Spisni<sup>c</sup>, Aniello Scala<sup>d</sup>, Luigia Pazzagli<sup>a</sup>

<sup>a</sup>*Dipartimento di Scienze Biochimiche, Università di Firenze, Florence, Italy*

<sup>b</sup>*Centro de Biologia Molecular Estrutural, LMLS, Campinas, Brazil*

<sup>c</sup>*Dipartimento di Medicina Sperimentale, Università di Parma, Parma, Italy*

<sup>d</sup>*Dipartimento di Biotecnologie Agrarie, Università di Firenze, Florence, Italy*

Protein Expression and Purification (2006) 49: 159–167

## Cerato-platanin, a phytotoxic protein from *Ceratocystis fimbriata*: Expression in *Pichia pastoris*, purification and characterization

Lara Carresi<sup>a</sup>, Barbara Pantera<sup>a</sup>, Camilla Zoppi<sup>a</sup>, Gianni Cappugi<sup>a</sup>, Aline L. Oliveira<sup>b</sup>,  
Thelma A. Pertinhez<sup>c</sup>, Alberto Spisni<sup>c</sup>, Aniello Scala<sup>d</sup>, Luigia Pazzagli<sup>a,\*</sup>

<sup>a</sup> Dipartimento di Scienze Biochimiche, Università di Firenze, Florence, Italy

<sup>b</sup> Centro de Biologia Molecular Estrutural, LMLS, Campinas, Brazil

<sup>c</sup> Dipartimento di Medicina Sperimentale, Università di Parma, Parma, Italy

<sup>d</sup> Dipartimento di Biotecnologie Agrarie, Università di Firenze, Florence, Italy

Received 26 January 2006, and in revised form 30 June 2006

Available online 16 July 2006

### Abstract

Cerato-platanin (CP) is a phytotoxic protein secreted by the Ascomycete *Ceratocystis fimbriata* f.sp. *platani*. This Ascomycete causes canker stain which is a severe disease with a high incidence in the European *Platanus acerifolia*. CP probably plays a role in the disease, eliciting defence-related responses in the host plants. CP is a 120 amino acid protein, containing 40% hydrophobic residues and two S-S bridges. In the EMBL data bank CP is the first member of a new fungal protein family known as the Cerato-Platanin Family. The N-terminal region of CP shows a high similarity with that of cerato-ulmin, a phytotoxic protein produced by the Ophiostoma species and that belongs to the hydrophobin family. Hydrophobins are hydrophobic proteins secreted by many saprophytic or pathogenic fungi and have a remarkable ability to self-assemble into a rodlet structure takes part in physiological and/or pathological processes. The methyltrophic yeast *Pichia pastoris* was used to obtain a high-level expression of recombinant CP (rCP) and the pPIC9 vector was chosen to bring about extra-cellular secretion of the protein. The preliminary structural and functional characterization presented here reveals no significant differences between the native and the recombinant protein. We also show that CP self-assembles in solution. The availability of rCP will allow its three-dimensional structure to be determined, facilitating an understanding of the role of CP in the pathogenesis of canker stain. It is also an excellent model for investigating the mechanism of action of the other proteins related to CP.

© 2006 Elsevier Inc. All rights reserved.

**Keywords:** Hydrophobins; Fungal protein; Canker stain disease; Protein aggregation

Phytopathogenic fungi produce several extra-cellular small proteins having phytotoxic activity. These proteins may induce cell necrosis either by suppressing the plant defensive systems, or by modifying the structure of the plant cellular membrane. In many cases these proteins act as elicitors that stimulate plant defensive systems and inducing the synthesis of molecules such as phytoalexins, that act directly against the pathogen [1]. One of these proteins is cerato-platanin, which is released abundantly in culture by *Ceratocystis fimbriata* f.sp. *platani*, an Ascomycete that causes canker stain in the plane-tree, a severe disease

with a high incidence in the European *Platanus acerifolia*. Symptoms of canker stain include foliar wilting and spreading lesions of phloem, cambium and sapwood that may lead to the death of the tree [2]. CP<sup>1</sup> is involved in the development of the disease since it elicits defence-related responses such as phytoalexin synthesis and cell death in both host and non-host plants [3,4].

<sup>1</sup> Abbreviations used: CP, cerato-platanin; rCP, recombinant CP; YNB, yeast nitrogen base; HRP, horseradish peroxidase; RP, reverse-phase; AOX1, alcohol oxidase; MD, minimal dextrose; MM, minimal methanol; Mut<sup>S</sup>, methanol utilizing slow; Mut<sup>+</sup>, methanol utilizing plus; OD, optical density; PMSF, phenylmethanesulfonyl fluoride; MGY, minimal glycerol yeast; PVDF, polyvinylidene difluoride, TFA, trifluoroacetic acid.

\* Corresponding author. Fax: +39 055 4222725.

E-mail address: [luigia.pazzagli@unifi.it](mailto:luigia.pazzagli@unifi.it) (L. Pazzagli).

CP has been purified from the culture filtrate of *C. fimbriata*; it is a 120 amino acid protein (Mw. 12.4 kDa), with 40% hydrophobic residues, two S–S bridges between Cys 20–57 and Cys 60–115, and a 14-residue signal peptide that is removed during secretion of the mature protein [3,5]. In the EMBL data bank, CP is the first member of the Cerato-Platanin Family, a new protein family that consists of seven secreted fungal proteins (EBI-InterPro IPR010829) characterized by a high amino acid sequence similarity, which is unusual among the other, better known, fungal protein families. On the other hand, the proteins in this family are not invariably characterized by a clear functional similarity although, in some cases, they are involved in phyto-pathological phenomena and immunological reactions. For example, the product of the *snodprot1* gene from *Phaeosphaeria nodorum* is expressed when this fungus infects wheat leaves [6]; the allergen Asp f13 causes an allergic reaction in humans [7]; the CS antigen is produced by *Coccidioides immitis*, which is the causative agent of a human respiratory disease [8] and the spl protein by *Leptosphaeria maculans*, which is the blackleg pathogen of *Brassica napus* [9].

CP also shares some structural and functional properties in common with the hydrophobins; for example it has a high similarity with the phytotoxic protein cerato-ulmin [10,11]. Hydrophobins are small to medium size (75–120 amino acids) hydrophobic proteins with a low sequence similarity and they are produced by many saprophytic or pathogenic fungi where they play a role in a broad range of growth and development processes [12]. Hydrophobins are usually secreted into the medium and are also localized on the cell walls of hyphae and conidia where they mediate the interaction between the fungus and the environment. On aerial hyphae they form a water-repellent coating characterized by a typical rodlet pattern that reduce the surface tension, allowing the fungus to leave the liquid medium and to grow in the air [13]. Rodlets are also important in the first stage of fungal pathogenesis, when the fungus must adhere to the hydrophobic surface of the plant before it can penetrate and infect the host [19]. Finally, hydrophobins have the remarkable ability *in vitro*, to self-assemble into a polymeric amphipathic monolayer at the hydrophobic/hydrophilic interface that is similar to the rodlet pattern on the surface of aerial hyphae [14,15]. For some highly hydrophobic members of this protein family, self-assembly is characterized by an increase in the  $\beta$ -sheet content and by the formation of amyloid-like structures [16–18].

To explore the role played by CP in canker stain, for which significant amounts of protein are needed, we aimed to produce CP by recombinant methods. *Escherichia coli*, the most common heterologous protein expression system, does not correctly process many polypeptides, so that proteins are formed which are insoluble and incorrectly folded. In the case of the recombinant CP we were able to obtain the correct product only after a refolding step and at low yields [5]. The methylotrophic yeast *Pichia pastoris* is

known to be an efficient expression system for proteins containing Cys residues, it can able to produce soluble proteins with a very high yield and having the right post-translational maturation [20]. In the present work we report the expression of rCP in *P. pastoris* and some details of the structure and biological activity of rCP are compared with those of the native CP.

## Materials and methods

### Materials

The *Pichia* expression kit, restriction enzymes, T4-DNA ligase and *Taq* DNA polymerase were purchased from InVitrogen (San Diego, CA, USA). The host *P. pastoris* strain GS115 (*his4*) from the *Pichia* expression kit was taken for expression of CP and grown, transformed and analyzed according to the manufacturer's instructions [21]. The shuttle vector pPIC9 carrying the secretion signal of the  $\alpha$ -mating factor sequence from *S. cerevisiae* was used.

*Escherichia coli* DH5 $\alpha$  (InVitrogen) was used for routine plasmid amplification, the pGEX-2T was from Pharmacia (Uppsala, Sweden) and the pGEM-T Easy vector was from Promega (Madison, WI, USA).

Peptone, tryptone, yeast extract, agar-Y, yeast nitrogen base (YNB) with ammonium sulphate without amino acid and dextrose, D-biotin and L-histidine were from BIO 101 system inc. (Carlsband, CA, USA); D-glucose, methanol and glycerol were purchased from Sigma-Fluka (St. Louis, MO, USA). The MiniElute Gel Extraction kit, QUIAQuick Gel Extraction kit, HiSpeed Plasmid Midi kit and the PCR Purification kit were purchased from Qiagen (Valencia, CA, USA); the GenElute Plasmid Miniprep kit was from Sigma. The restriction enzymes were from In vitrogen.

PCR amplification was performed with a Gene Amp PCR System 2400 from Perkin Elmer (Wellesley, MA, USA). A 1 Kb plus DNA Ladder (InVitrogen) was used for 1% agarose gel electrophoresis. Protein markers and the Coomassie blue stain reagent (Bio-Rad, Hercules, CA, USA) were used for SDS-PAGE gel electrophoresis; horseradish peroxidase (HRP)-conjugated secondary antibodies and PVDF membranes were from Amersham Biosciences (Piscataway, NJ, USA); the enhanced chemiluminescence (ECL) system for visualizing Western blot was purchased from Pierce (Rockford, IL, USA).

Protein concentration was determined using the Bicinchoninic acid method from Pierce. The semi-preparative Reverse-Phase (RP) HPLC column (C4, 5  $\mu$ m, 250  $\times$  10 mm) was from Vydac (Grace, Columbia, MD, USA).

The Bio-Gel P-10 column and the Sephadex G75 column employed in the purification of the native protein were from Bio-Rad and Pharmacia, respectively.

The YM-3 membrane and the ultra-filtration unit were from Millipore (Bedford, MA, USA). All reagents were of analytical grade unless otherwise stated.

### Construction of the expression plasmid

Standard molecular biology methods were used for the construction of the expression plasmid [21,22]. The pGEM-T Easy plasmid from *E. coli* DH5 $\alpha$  cells containing the *cp* gene [5] was used to isolate the cDNA fragment encoding the mature CP. This construct was digested with *Bam*HI and *Eco*RI restriction enzymes and the fragment was amplified by PCR using the primers:

PIC9cpFor (5'-CTCGAGAAAAGAGTCTCGATCTC-3') and PIC9cpRev (5'-GAATTCCTAATTAGCGCCGTTAATG-3')

that were designed on the basis of the *cp* and pPIC9 sequences [21]. The sense primer includes a *Xho*I restriction site (underlined) and a KEX2 site (in italic). The anti-sense primer contains a stop codon and an *Eco*RI restriction site (underlined). The PCR product was purified from a 1% agarose gel and subcloned into pGEM-T easy vector.

The pGEM-*cp*, purified by precipitation with ethanol, was electroporated into *E. coli* DH5 $\alpha$  using a Gene Pulser II (Bio-Rad). *E. coli* cells were grown in SOC medium and then selected on LB dishes. Positive colonies were analyzed by Colony PCR using the above primers and gene insertion was verified by agarose gel electrophoresis and by sequence analysis carried out at Bio Molecular Research c/o CRIBI, University of Padua, Italy. Subcloning in pGEM yielded a plasmid useful for the next step of restriction cleavage.

The pGEM-*cp* was extracted from an overnight liquid culture of one positive clone; the construct was digested with *Eco*RI and *Xho*I and cloned into the same enzyme-digested vector pPIC9 using T4 DNA ligase. The resulting plasmid pPIC9-*cp* contained the *cp* gene under the control of AOX1 (Alcohol Oxidase) promoter and was in-frame to the  $\alpha$ -factor secretion signal. Finally, pPIC9-*cp* was electroporated into *E. coli* DH5 $\alpha$  for the amplification of the recombinant plasmid and positive colonies were selected. The pPIC9-*cp* plasmid was purified and subjected to DNA sequence analysis.

### Transformation of *P. pastoris* cells and expression of CP

Approximately 60  $\mu$ g of pPIC9-*cp* was linearized with the *Bgl*II restriction enzyme; an aliquot of the reaction mixture was used to confirm the complete digestion by 1% agarose gel.

The digestion products were purified from salts and enzyme by extraction with phenol/chloroform/isoamyl alcohol (25/24/1) and precipitated with ethanol. Then, the fragments were re-suspended in water and used to transform *P. pastoris* GS115 His<sup>-</sup> cells, made competent in accordance with manufacturer's instructions [21]. Transformants were screened for methanol utilization by patching about 90 colonies on Minimal Dextrose (MD) dishes and Minimal Methanol (MM) dishes incubated for 3 days at 30°C, and scored as Mut<sup>S</sup> (methanol utilizing slow) or

Mut<sup>+</sup> (methanol utilizing plus). Integration of *cp* into the *Pichia* genome was confirmed by PCR. Colony PCR was performed using the primers:

5' AOX1 (5'-GACTGGTTCCAATTGACAAGC-3') and 3' AOX1 (5'-GCAAATGGCATTCTGACATCC-3').

Mut<sup>+</sup> integrants produce two bands, while Mut<sup>S</sup> integrants show only one band corresponding to CP gene cloned into pPIC9 (360 bp + 492 bp).

For the expression of CP four clones of the His<sup>+</sup> Mut<sup>S</sup> transformed strains were selected and stored at -80°C in YNB containing 20% glycerol. Each clone was patched on an MD dish and incubated at 30°C for 3 days. A single colony of each clone was used to inoculate 100 mL of MGY in a one-litre flask and grown at 30°C in a shaking incubator at 200 rpm until the cultures reached an optical density (OD) of 7.0 at 600 nm. Cells were then harvested by centrifuging, re-suspended in 10 mL of MM medium and transferred to 100 mL flasks. The cultures were maintained at 28°C in a shaking incubator and supplemented daily with 100% methanol to a final concentration of 0.5%. Aliquots (500  $\mu$ L) of the expression cultures were centrifuged at 8000g for 5 min at different times. The cell pellet was re-suspended in 200  $\mu$ L of PBS containing 1 mM EDTA and 1 mM PMSF (phenylmethanesulfonyl fluoride) and was then homogenized by adding 50  $\mu$ L of glass beads in a Bio101 Thermo Savant Fast Prep (Q Biogen Inc.). rCP was detected in both the supernatant and in the cell pellet by SDS-PAGE and Western blot using anti-CP antibodies as previously described [4]; protein concentration was determined by the bicinchoninic acid method (Pierce).

### Large scale production and purification of rCP and native CP

Large-scale production of rCP was carried out using the clone that had given the best yield in the small-scale experiments. Production started by patching the colony (stored at -80°C in YNB containing 20% glycerol) on MD and incubating at 30°C for 3 days. A single colony was then inoculated in 1.0 L of MGY (Minimal Glycerol Yeast) until the OD at 600 nm was 7.0. After centrifugation at 3000g for 15 min the pellet was re-suspended in 100 mL of MM, placed in a 1.0 L baffled flask and returned to the incubator for further growth. Five hundred microliters of 100% methanol was added every 24 h to maintain induction. Aliquots (500  $\mu$ L) of the expression cultures were centrifuged at 8000g for 5 min at regular intervals (2, 4, 6, 8, 10, 12 days) and the supernatants were applied to a semi-preparative reverse phase HPLC column (Vydac, C4, 5  $\mu$ m, 250  $\times$  10 mm) to analyze the rCP concentration. Elution was performed with a water/acetonitrile gradient in 0.1% (v:v) trifluoroacetic acid.

After 10 days of culture, the entire medium was harvested by centrifugation (8000g for 15 min) and the supernatant was concentrated by ultra-filtration on a YM3

membrane. Aliquots were then applied to a semi-preparative reverse phase HPLC column as above to obtain the purified protein. The purity of the CP was analyzed by SDS-PAGE and Western blot. Composition and protein concentration of the purified CP were determined by amino acid analysis: 1–2 nmol of protein was hydrolyzed with 6.0 M HCl and analyzed on an amino acid analyzer equipped with a Na<sup>+</sup>-high-performance column (4 × 120 mm, Beckman, Fullerton, CA, USA) and a post-column derivatization with *o*-phthaldialdehyde and 2-mercaptoethanol [23]. Cys residues were determined as carboxymethylcysteine [3].

Native CP was obtained from the culture filtrate of *C. fimbriata* using a modification of a purification procedure previously reported [3]. Briefly, the virulent strain Cf AF 100 of *C. fimbriata* f.sp. *platani*, [2], was placed in 1.0 L of sterilized potato dextrose broth medium (Sigma) and incubated for 21 days at 25 °C. The culture medium was filtered on a 0.45 μm membrane (Millipore), lyophilized, redissolved in 10 mL of 1% acetic acid, and then loaded on a Bio-Gel P-10 (Bio-Rad) column (2.5 × 70 cm), equilibrated, and eluted with 1% acetic acid. The CP-containing fractions were pooled, concentrated to 10 mL by ultra-filtration on a YM3 membrane, applied to a Sephadex G-75 column (2.4 × 90 cm) (Pharmacia), equilibrated and eluted with 0.05 M Na-acetate buffer, pH 5.0, containing 1.0 M NaCl. The fractions containing the protein were pooled, lyophilized, and applied to a reverse phase HPLC column (Vydac, C4, 5 μm, 250 × 10 mm). The elution was carried out using a water/acetonitrile gradient in 0.1% (v:v) trifluoroacetic acid. During the purification steps the occurrence of CP was determined by SDS-PAGE and the yield was determined with the bicinchoninic acid protein assay (Pierce).

### Structural characterization of recombinant CP

#### Mass spectrometry

Twenty picomole either of CP or of rCP were dissolved in 50% acetonitrile containing 0.05% TFA, diluted 1:1 in a sinapinic acid matrix and deposited on the target to be analyzed on a MALDI-TOF mass spectrometer (Omniflex, Bruker-Franzen Analytik, Bremen, Germany) [24].

#### N-terminal amino acid sequence

rCP was reduced and carboxymethylated (CM-rCP) as previously described [3]. Intact CM-rCP, immobilized on a polyvinylidene difluoride (PVDF) membrane (Millipore), was subjected to automated Edman degradation on a Procise model 491 gas-phase protein sequencer (Applied Biosystems, Foster City, CA, USA) [25].

#### Circular dichroism

CD spectra of CP and rCP were obtained by dissolving the proteins in 50 mM Na-phosphate buffer, pH 7.0, to a final concentration of 6.25 μM. CD spectra were recorded in the far-UV region from 190 to 260 nm using a JASCO J-810 spectropolarimeter (Easton, MD, USA). The spectra

were obtained using a 1 mm path length quartz cell with 0.5 nm step resolution, 50 nm/min speed and 8 s response time, at 25 °C. After baseline correction, the observed ellipticity,  $\theta$  (mdeg) was converted to the mean residue molar ellipticity  $[\theta]$  (deg cm<sup>2</sup> dmol<sup>-1</sup>).

#### <sup>1</sup>H-NMR

Monodimensional NMR was obtained by dissolving rCP to the final concentration of 0.7 mM in 600 μL of phosphate buffer 10 mM at pH 5.8, and 5% of D<sub>2</sub>O has been recorded in a Varian INOVA AS500 spectrometer operating at 499.730 MHz for <sup>1</sup>H frequency, at 20 °C.

#### Biological activity

The ability of CP to induce the synthesis of phytoalexins in the leaves of the host *Platanus acerifolia* was tested as previously described [3,4]. One nanomole of CP was dissolved in 10 μL of water and applied to the lower surface of a plane-tree leaf. After 24 h, samples were recovered and fluorescence was recorded with a Perkin Elmer spectrofluorimeter 650-10S (Perkin Elmer, Wellesley, MA, USA),  $\lambda_{\text{ex}} = 365$  nm,  $\lambda_{\text{em}} = 460$  nm and slit = 5. The results were expressed as nmol/mL of umbelliferone equivalents. Data obtained were analyzed by ANOVA, fixed model. Homogeneous groups were identified by means of Tukey's HDS test.

#### In vitro aggregation of CP

Self-aggregation of native and recombinant CP was induced by incubating the protein (0.54 mM) in a 30% TFE:H<sub>2</sub>O (v:v) solution [26]. Incubation was performed at 37 °C for 30 days. Aliquots were used to test for thioflavine T binding. [27]. Five microliter of the sample was withdrawn at 0, 10 and 30 days of incubation and mixed with 495 μL of 6.4 μM ThT (thioflavine T from Sigma) in 50 mM glycine-NaOH buffer, pH 8.5 (final concentration of CP 0.54 μM). A Perkin-Elmer LS-55 spectrofluorimeter with an excitation wavelength of 435 nm and emission wavelengths of 400–600 nm was used to determine ThT fluorescence spectra.

To detect the specificity of the ThT response, different aliquots of the aggregated sample were centrifuged at 15,000g for 15 min; the pellets were re-suspended in the aggregation solution and assayed for protein concentration with bicinchoninic acid. Different amounts of aggregated protein were then analyzed for ThT fluorescence.

Two microgram of the aggregated sample were then analyzed by SDS-PAGE to detect the integrity of the protein after 30 days of incubation. To dissolve the aggregates 2 μg was treated with 20 μL of anhydrous TFA (Trifluoroacetic acid) [18]; after 15 min the samples were dried under a gentle N<sub>2</sub> pressure, and then applied on a 15% gel. Native and recombinant protein (2 μg) in their soluble form were also treated by TFA and applied to SDS-PAGE.

## Results

### Construction of the expression vector

The cDNA fragment encoding the mature CP was isolated from the pGEM-*cp* plasmid obtained previously[5]. Previous experiments with the EA repeats gave a protein with four N-terminal additional residues (EAEA) indicating that the dipeptidylaminopeptidase (STE13 gene product) was not able to eliminate this sequence in our experimental conditions. To obtain rCP without no additional amino acid residues at the N-terminus, the *cp* gene was cloned with *XhoI/EcoRI* cloning sites in the pPIC9 plasmid without inserting the sequence Glu-Ala-Glu-Ala after the KEX2 site, as suggested by Brake [28], (Fig. 1).

The pGEM-*cp* construct was digested with the *Bam*HI and *Eco*RI restriction enzymes and the fragment was PCR-amplified using primers PIC9cpFor (containing the *Xho*I restriction site and the KEX2 site) and PIC9cpRev (containing a stop codon and an *Eco*RI restriction site). After this fragment had been cloned in the pGEM-T-easy vector, it was *Xho*I/*Eco*RI digested and cloned in the pPIC9 plasmid digested by the same enzyme. The pPIC9-*cp* construct was linearized and used to transform *P. pastoris* GS115 His<sup>-</sup> cells.

### Expression and purification of the recombinant CP

CP is a rather hydrophobic protein and its rapid secretion in aqueous solution could lead to precipitation; the transformants were therefore screened for Mut<sup>S</sup> and Mut<sup>+</sup> phenotype and only the first ones were chosen for expression. In fact, the His<sup>+</sup> Mut<sup>S</sup> strains grow slowly on methanol and the expressed protein is secreted slowly into the culture medium; on the contrary, the protein secretion in the Mut<sup>+</sup> clones reaches a maximum only after 3–4 days. The clone that produced the largest amount of protein was tested by SDS-PAGE and Western blot (data not shown) for scale-up expression. Aliquots of the recombinant protein were purified from the culture-broth supernatant by analytical C-18 reverse phase HPLC. The time course analysis of the purified protein was monitored by bicinchoninic acid protein assay after 2, 4, 6, 8, 10 and 12 days of induction (Fig. 2). Maximum CP expression occurred 10 days after induction; no increase in the pro-

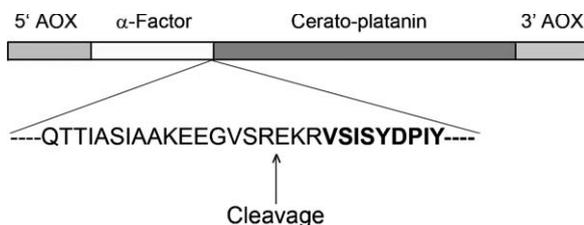


Fig. 1. Schematic representation of the pPIC9-*cp* construct. Bold letters indicate the cerato-platanin sequence. The arrow shows the cleavage site of KEX2 in our expression conditions.

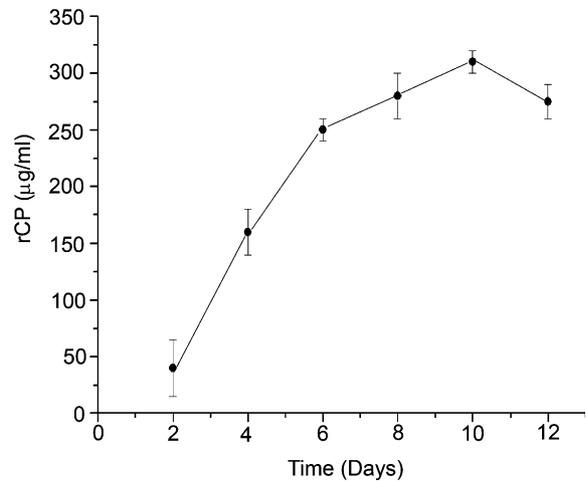


Fig. 2. Time-course analysis of the rCP production in the induction medium (100 mL). Five hundred microliter of culture medium was collected at different days of incubation and applied to a semi-preparative reverse phase column. Protein concentration was determined on the purified samples by the bicinchoninic acid assay. Dots on the graph represent the medium value obtained from three different purifications.

duction of recombinant protein was obtained with a longer incubation period.

A single-step purification by semi-preparative RP-HPLC (Fig. 3) yielded 32 mg of pure protein starting from 1 L of MGY culture medium. In contrast, purification of the native protein, starting from 1 L of culture filtrate of *C. fimbriata*, produced only 2.3 mg (Table 1). The purity of the recombinant and native protein was determined by amino acid analysis (data not shown), by SDS-PAGE and by Western blot analysis (Fig. 4).

### Characterization of purified rCP

The molecular weight of rCP, as determined by MALDI-TOF mass spectrometry, is 12,809.8 Da. This value is 412.6 Da more than that of CP at 12,397.2 Da (Fig. 5). To investigate this discrepancy we subjected the carboxymethylated protein to Edman degradation; no multiple residues were found during sequencing, confirming the purity of the recombinant protein, but unfortunately the N-terminal sequence showed that rCP contains the additional sequence EKR. This is probably because, in our experimental conditions, the protease recognizes the glutamate residue located upstream of the KEX2 site, preventing the production of rCP free of additional N-terminus residues (Fig. 1). Alternatively, the *kex2* cleavage might be inefficient in the context of this fusion sequence, and that another protease might be responsible for the observed cleavage pattern.

CD analysis nevertheless showed that the spectra of CP and rCP are similar, indicating that the recombinant protein was correctly folded (Fig. 6). The <sup>1</sup>H NMR spectrum of rCP (data not shown) showed intense signals with a good dispersion, confirming that the protein has a well defined fold and indicating that, in phosphate buffer, it do not

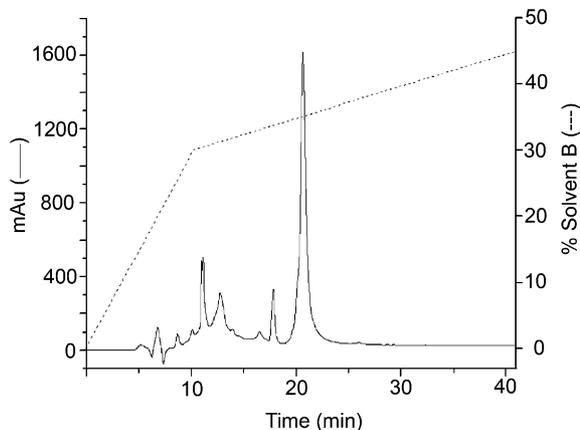


Fig. 3. RP-HPLC purification of the culture medium from 10 days induced *Pichia* growth. The column was a Vydac, C4, 5  $\mu$ m, 250  $\times$  10 mm; Solvent A, 10 mM TFA in water; Solvent B, 10 mM TFA in acetonitrile. Flow rate was 2 ml/min. The elution was performed by an acetonitrile gradient as indicated by the dotted line (---).

aggregate at the high concentration used in the NMR experiments. Next, production of phytoalexins from the leaves of *P. acerifolia* was compared, in order to determine the biological activity of CP [3,4]; no statistically significant differences were found between the action of CP and rCP (Table 2).

#### Self-aggregation of CP and rCP

The first step in finding the mechanism by which CP interacts with the host cells was to determine whether, like other hydrophobins, CP self-assembled into polymeric and amphipathic monolayers which would be similar to amyloid fibrils and characterized by a  $\beta$ -sheet content higher than the one found in the monomeric form [16,17,27]. An *in vitro* study was carried out using the experimental procedures employed for other proteins that undergo aggregation [26]. Since aggregation often requires destabilizing conditions, a 30% TFE:H<sub>2</sub>O (v:v) mixture was used to obtain a partly unfolded CP that was then incubated at 37 °C.

To monitor protein aggregation we used the reagent thioflavine T (ThT) that binds to the ordered and aggregated protein structures resulting in an increase in its fluorescence spectrum at 485 nm [27]. Fig. 7A shows the fluorescence increase for CP and rCP; the fluorescence is first detectable at

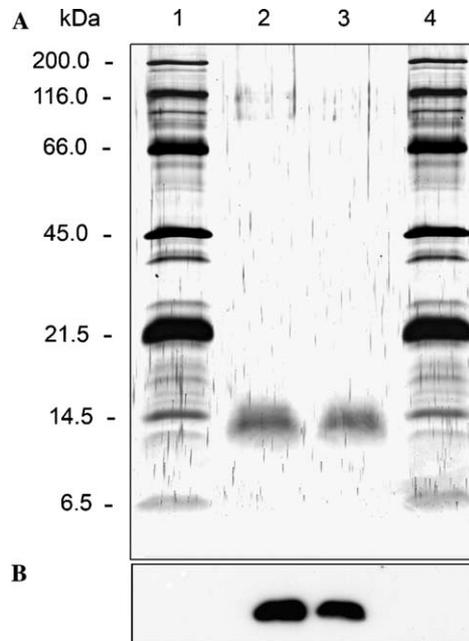


Fig. 4. SDS-PAGE and Western blot of native and recombinant CP. (A) SDS-PAGE and (B) Western blot analysis. with anti-CP polyclonal antiserum (A) lane 1, protein standard; lane 2, CP 1  $\mu$ g; lane 3, rCP 1  $\mu$ g; lane 4, protein standard; (B) lane 2, CP 1  $\mu$ g; lane 3, rCP 1  $\mu$ g. SDS-PAGE was silver stained, western blot analysis was performed with anti-CP polyclonal antiserum.

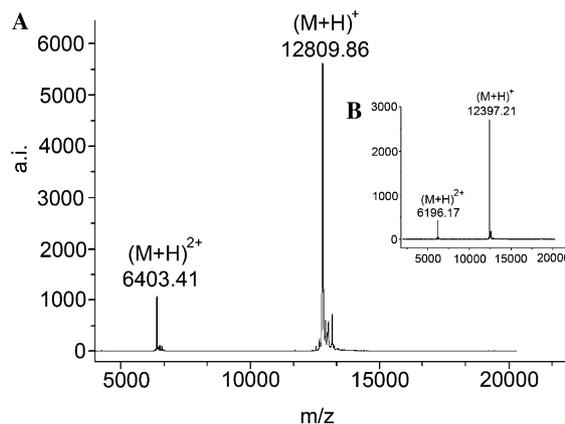


Fig. 5. MALDI-TOF mass spectrum of purified rCP (A) and CP (B). Analysis was performed using a LINEAR method. The mass spectrometer was calibrated with a 6–18 kDa protein standard.

Table 1  
Purification of native and recombinant CP

CP from <i>C. fimbriata</i> <sup>a</sup>		rCP expressed in <i>E. coli</i> <sup>b</sup>		rCP secreted by <i>P. pastoris</i> <sup>a</sup>	
Cultural filtrate (1.0 L)	153	Cell culture (1.0 L)	268	Induction medium (100 mL)*	96
Bio-gel P10	35	Sepharose- GSH	16.5	RP-HPLC	32
Sephadex G75	92	RP-HPLC	2.1		
RP-HPLC	23	Refolding step	1.2		

Values are expressed as mg of total protein determined by bicinchoninic assay.

<sup>a</sup> for details see Materials and Methods.

<sup>b</sup> for details see ref. [5].

\* from 1.0 L of MGY cell culture.

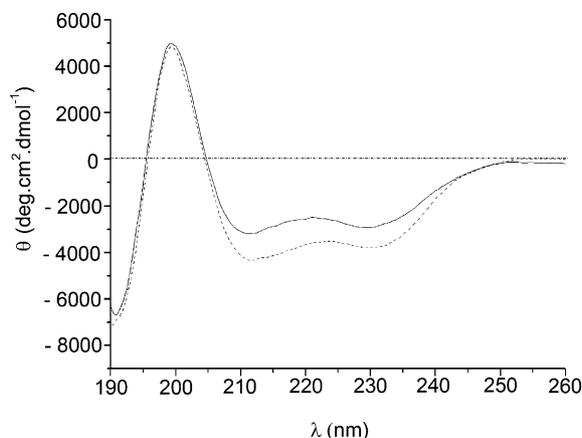


Fig. 6. CD spectra of native and recombinant CP. Far-UV CD spectra of CP (bold line) and rCP (dotted line). Sample concentration was  $6.25 \mu\text{M}$  in  $50 \text{ mM Na-phosphate buffer, pH } 7$ , at  $25^\circ\text{C}$ .

Table 2  
Biological activity of native and recombinant CP

Protein	Activity	
Native CP	$3.90 \pm 0.97$	$n = 9$
Recombinant CP	$4.13 \pm 2.27$	$n = 10$

Results are reported as nmol/ml of phytoalexins. Data are the mean  $\pm$  D of values obtained by three independent experiments. The values are not significantly different at  $P < 0.01$  according to Tukey's test.

day 10, and shows a 35-fold increase after 30 days of incubation for both CP and rCP. Moreover, as shown in Fig. 7B, the fluorescence intensity is directly proportional to the concentration of the aggregated CPs, implying that the reaction of ThT is specific for the aggregated form of the protein [16]. To investigate the molecular state of CP and rCP during aggregation, after 30 days of incubation the samples were dissolved in TFA and analyzed by SDS-PAGE. As it is shown in Fig. 7C a shortening of the polypeptide chain occurs in both native and recombinant CP.

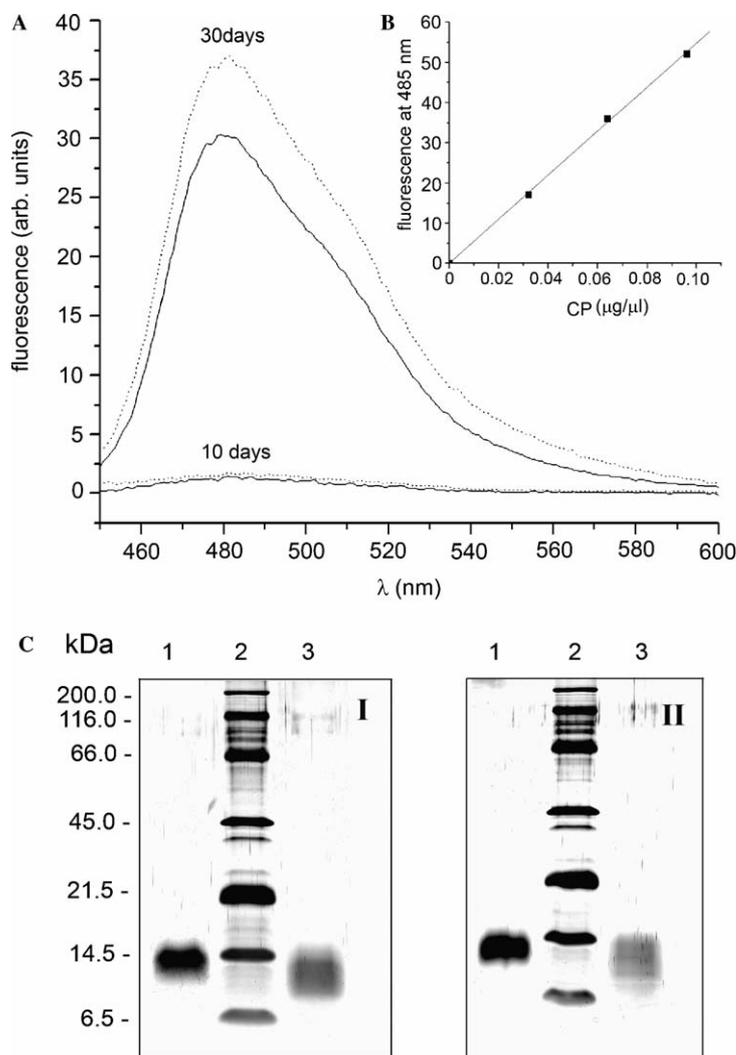


Fig. 7. ThT fluorescence and CP aggregation. (A) Fluorescence spectra of ThT in the presence of aliquots of CP (solid line) and rCP (dotted line) incubated in 30% TFE:H<sub>2</sub>O (v:v) mixture at  $37^\circ\text{C}$ . The samples contain  $6.4 \mu\text{M}$  ThT in a  $50 \text{ mM glycine-NaOH buffer pH } 8.5$  in the presence of  $5.4 \mu\text{M}$  CP or rCP. (B) Fluorescence intensity vs. the concentrations of aggregated CP.  $\lambda_{\text{ex}} = 435 \text{ nm}$ ;  $\lambda_{\text{em}} = 485 \text{ nm}$ . (C) SDS-PAGE of the 30 days aggregated protein. I- lane 1: native CP ( $2 \mu\text{g}$ ), lane 2: protein standard, lane 3: aggregated CP ( $2 \mu\text{g}$ ). II- lane 1: recombinant CP ( $2 \mu\text{g}$ ), lane 2: protein standard, lane 3: aggregated rCP ( $2 \mu\text{g}$ ).

## Discussion

Cerato-platanin is the first member of the Cerato-Platanin Family, a new protein family that has been added to the list of the many fungal protein families already known [5,12–15,29,30]. CP is secreted by the phytopathogenic fungus *C. fimbriata* f.sp. *platani*, which is the causal agent of the canker stain disease that affects *Platanus acerifolia* [2,3]. Like other fungal proteins such as hydrophobins, it is also present in the cell wall [31]. Many studies have recently investigated the role of CP in the interaction between the fungus and the plane-tree, and have sought to explain how the other proteins in this new protein family interact with the membrane of the host cell and exert their biological effect [6–9].

A large amount of protein is needed to carry out such a study, but the purification yield of the native protein, though increased by recent improvements in the purification procedure, is insufficient for the experiments involved (Table 1). Use of a recombinant protein can overcome these problems. The expression of CP in *E. coli* produced an incorrectly folded protein probably because the two disulfide bridges formed too quickly and/or incorrectly, interfering with the hydrophobic collapse during folding [5]. However, the biologically active form of the protein could be achieved by a refolding step that led to a low purification yield (Table 1).

Since recombinant expression in eukaryotic organism often allows a correct posttranslational processing of protein with disulfide bridges and/or other posttranslational modification [20,28], we tested the methylotrophic yeast *P. pastoris* using the pPIC9 vector in order to obtain a high level of expression. By exploiting the extracellular secretion of CP, a single-step purification gave a protein yield of about 40% which is more than 10 times the yield of native CP and rCP expressed in *E. coli*.

The amino acid sequence of rCP from *P. pastoris* shows three additional residues (EKR) at N-terminus. These are probably due to the presence of a sequence in the *Xho*I restriction site that is recognized by a protease, causing the cleavage of the signal peptide. However, the presence of the EKR sequence does not affect the function and structure of rCP. In fact, the biological activity of rCP, according to its ability to induce phytoalexin synthesis in plane-tree leaves, is comparable to that of CP. The CD spectra of the native and recombinant proteins, they do not differ significantly, suggesting the presence of the same secondary structure elements. Finally, the NMR spectrum of rCP shows that the protein has a well defined fold. These facts, together with the knowledge that *P. pastoris* grows very readily in a minimal medium in which it is possible to introduce a source of <sup>15</sup>N and/or <sup>13</sup>C, suggest that it will be possible to produce a double-labeled protein suitable for multi-dimensional NMR experiments.

Starting from the similarity of CP to the hydrophobins, we observed that native CP and recombinant CP formed aggregates that were ThT positive. The self-assembling of hydrophobins occurs at the hydrophobic/hydrophilic interface and it is obtained, *in vitro*, by vortexing the protein solution for a

short time [14,18,32]; on the contrary aggregation of cerato-platanin requires long time of incubation and the use of solvents that are employed in the study of human amyloid-like proteins [26,33]. Some of these proteins undergo to a fragmentation of the polypeptide chain and the aggregation process involves short peptides such as the 1–40 and 1–43 peptides of the A $\beta$  amyloid precursor protein and the peptides derived from the fragmentation of the PrP prion protein [34–38]. The SDS-PAGE analysis of CP aggregates shows that, during aggregation, a shortening of the polypeptide chains occurs suggesting that CP could follow the behaviour of amyloid-like proteins which self assemble after fragmentation. However, these aggregates need a deeper investigation to clarify the mechanism of self assembling and the mechanism by which CP, localized in the cell wall of *C. fimbriata* [31], interacts with the cell wall of *P. acerifolia*. An understanding of protein aggregation should also clarify the action mechanism of the other members of the Cerato-Platanin Family.

## Acknowledgments

This work was partly supported by MIUR (Italy –PRIN 2003, n. 2003070128\_005) and FAPESP (Brazil- AS (#99/11030-9 and #99/07574-3) and TAP #00/02026-7). We greatly acknowledge Dr. Giuseppe Orsomando (Istituto di Biotecnologie Biochimiche; Università Politecnica delle Marche, Ancona, Italy) for his assistance in the Edman degradation experiments.

## References

- [1] M.D. Templeton, A. Rikkerink, R.E. Beever, Small cysteine-rich proteins and recognition in fungal-plant interactions, *Mol. Plant Microbe Interact.* 7 (1994) 320–325.
- [2] A. Panconesi, Canker stain of plane trees: a serious danger to urban plantings in Europe, *J. Plant Pathol.* 81 (1999) 3–15.
- [3] L. Pazzagli, G. Cappugi, G. Manao, G. Camici, A. Santini, A. Scala, Purification, characterization, and amino acid sequence of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata* f. sp. *platani*, *J. Biol. Chem.* 274 (1999) 24959–24964.
- [4] A. Scala, L. Pazzagli, C. Comparini, A. Santini, S. Tegli, G. Cappugi, Cerato-platanin an early-produced protein by *Ceratocystis fimbriata* f.sp. *platani* elicits phytoalexin synthesis in host and non-host plants, *J. Plant Pathol.* 86 (2004) 23–29.
- [5] L. Pazzagli, B. Pantera, L. Carresi, T.A. Pertinhez, A. Spisni, S. Tegli, C. Zoppi, A. Scala, G. Cappugi, Cerato-platanin, the first member of a new fungal protein family: cloning, expression and characterization, *Cell Biochem. Biophys.* 44 (2006) 512–521.
- [6] A.L. Sharen, G. Krupinski, Cultural and inoculation studies of *Septoria nodorum*, cause of Glume Blotch of wheat, *Phytopathology* 60 (1970) 1480–1485.
- [7] V.P. Kurup, B. Banerjee, S. Hemmann, P.A. Greenberger, K. Blaser, R. Crateri, Selected recombinant *Aspergillus fumigatus* allergens bind specifically to IgE in ABPA, *Clin. Experim. Allergy* 30 (2000) 988–993.
- [8] S. Pan, G.T. Cole, Molecular and biochemical characterization of a *Coccidioides immitis*-specific antigen, *Infect. Immun.* 63 (1995) 3994–4002.
- [9] L.M. Wilson, A. Idnurm, B.J. Howlett, Characterization of a gene (sp1) encoding a secreted protein from *Leptosphaeria maculans*, the blackleg pathogen of *Brassica napus*, *Mol. Plant Pathol.* 3 (2002) 487–493.

- [10] M.A. Stringer, W.E. Timberlake, Cerato-ulmin, a toxin involved in Dutch elm disease, is a fungal hydrophobin, *Plant Cell* 2 (1993) 145–146.
- [11] G. Del Sorbo, A. Scala, F. Scala, S. Tegli, Cerato-ulmin, a toxin produced by the pathogens of the dutch elm disease, in: Rajeev K. Upadhyay (Ed.), *Advances in Microbial Toxin Research and its Biotechnological Exploitation*, Kluwer Academic/ Plenum Publishers, New York, 2002, pp. 93–103.
- [12] H.A.B. Wosten, M.L. de Vocht, Hydrophobins, The fungal coat unrevealed, *BBA* 1469 (2000) 79–86.
- [13] H.A.B. Wosten, Hydrophobins: multipurpose proteins, *Annu. Rev. Microbiol.* 55 (2001) 625–646.
- [14] H.A.B. Wosten, O.M.H. de Vries, J.G.H. Wessels, Interfacial self-assembly of a fungal hydrophobin into an hydrophobic rodlet layer, *Plant cell* 5 (1993) 1567–1574.
- [15] M.B. Linder, G.R. Szilvay, T. Nakari-Setälä, M.E. Penttilä, Hydrophobins: the protein-amphiphiles of filamentous fungi, *FEMS Microbiol. Rev.* 29 (2005) 877–896.
- [16] P. Butko, J.P. Buford, J.S. Goodwin, P.A. Stroud, C.L. McCormick, G.C. Cannon, Spectroscopic evidence for amyloid-like interfacial self-assembly of hydrophobin Sc3, *Biochem. Biophys. Res. Commun.* 280 (2001) 212–215.
- [17] J.P. Mackay, J.M. Matthews, R.D. Winefield, L.G. Mackay, R.G. Haverkamp, M.D. Templeton, The hydrophobin EAS is largely unstructured in solution and functions by forming amyloid-like structures, *Structure (Camb.)* 9 (2001) 83–91.
- [18] P.A. Stroud, J.S. Goodwin, P. Butko, G.C. Cannon, C.L. McCormick, Experimental evidence for multiple assembled states of Sc3 from *Schizophyllum commune*, *Biomacromolecules* 4 (2003) 956–967.
- [19] R.J. Whiteford, P.D. Spanu, Hydrophobins and the interaction between fungi and plants, *Mol. Plant Pathol.* 3 (2002) 391–400.
- [20] C.E. White, N.M. Kempf, E.A. Komives, Expression of highly disulfide-bonded proteins in *Pichia pastoris*, *Structure* 2 (1994) 1003–1005.
- [21] *Pichia* Expression Kit Version M: A Manual of Methods for Expression of Recombinant Proteins in *Pichia pastoris*, InVitrogen, 1999.
- [22] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, second ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA, 1989.
- [23] J.R. Benson, P.E. Hare, *O*-phthalaldehyde: fluorogenic detection of primary amines in the picomole range. Comparison with fluorescamine and ninhydrin, *Proc. Natl. Acad. Sci. USA* 72 (1975) 619–622.
- [24] P. Roepstorff, M.R. Larsen, H. Rahbeck-Nielsen, E. Nordhoff, Sample preparation for matrix assisted laser desorption/ionization mass spectrometry for peptides, protein and nucleic acid, in: Second ed., *Cell Biology: A Laboratory Handbook*, vol. 4, Academic Press, 1998.
- [25] J.B.C. Findlay, D.J. Pappin, D.N. Keen, Automated solid phase microsequencing, *Protein Sequencing: A Practical Approach*, in: J.B.C. Findlay, M.D. Geisow (Eds.), IRL Press at the Oxford University Press, New York, 1989, pp. 69–84.
- [26] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, C.M. Dobson, Designing conditions for in vitro formation of amyloid protofilaments and fibrils, *Proc. Natl. Acad. Sci. USA* 96 (1999) 3590–3594.
- [27] H. LeVine, Thioflavine T interaction with amyloid  $\beta$ -sheet structures, *Amyloid: Int. J. Exp. Clin. Invest.* III 2 (1995) 1–6.
- [28] A.J. Brake, J.P. Merryweather, D.G. Coit, U.A. Heberlein, G.R. Masiarz, G.T. Mullenbach, M.S. Urdea, P. Valenzuela, P.J. Barr,  $\alpha$ -Factor-Directed Synthesis and Secretion of Mature Foreign Proteins in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. USA* 81 (1984) 4642–4646.
- [29] L.M. Yu, Elicitins from *Phytophthora* and basic resistance in tobacco, *Proc. Natl. Acad. Sci. USA* 92 (1995) 4088–4094.
- [30] R.D. Kulkarni, H.S. Kelkar, R.A. Dean, An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins, *Trends Biochem. Sci.* 28 (2003) 118–121.
- [31] S. Boddi, C. Comparini, R. Calamassi, L. Pazzagli, G. Cappugi, A. Scala, Cerato-platanin protein is located in the cell walls of ascospores, conidia and hyphae of *Ceratocystis fimbriata f. sp. platani*, *FEMS Microbiol. Lett.* 233 (2004) 341–346.
- [32] A.H. Kwan, R.D. Winefield, M. Sunde, J.M. Matthews, R.G. Haverkamp, M.D. Templeton, J.P. Mackay, Structural basis for rodlet assembly in fungal hydrophobins, *Proc. Natl. Acad. Sci. USA* 103 (2006) 3621–3626.
- [33] G. Marcon, G. Plakoutsi, C. Canale, A. Relini, N. Taddei, C.M. Dobson, G. Ramponi, F. Chiti, Amyloid formation from HypF-N under conditions in which the protein is initially in its native state, *J. Mol. Biol.* 347 (2005) 323–335.
- [34] D.J. Selkoe, Alzheimer's disease—genotypes, phenotype and treatments, *Science* 275 (1997) 630–631.
- [35] G. Multhaup, T. Ruppert, A. Schlicksupp, L. Hesse, E. Bill, R. Pipkorn, C.L. Masters, C. Beyreuther, Copper binding amyloid precursor protein undergoes a site-specific fragmentation in the reduction of hydrogen peroxide, *Biochemistry* 37 (1998) 7224–7230.
- [36] P. Westermark, Aspects on human amyloid forms and their fibril polypeptides, *FEBS J.* 272 (2005) 5942–5949.
- [37] M. Ettaiche, R. Pichot, J.P. Vincent, J. Chabry, In vivo cytotoxicity of the prion protein fragment 106–126, *J. Biol. Chem.* 275 (2000) 36487–36490.
- [38] K. Tsirolunikov, J.M. Chobert, T. Haertlé, Copper-dependent degradation of recombinant ovine prion protein, *FEBS J.* 273 (2006) 1959–1965.

## CAPÍTULO 2: Determinação da Estrutura da Cerato-Platanina

### Introdução

Neste segundo capítulo serão apresentados os procedimentos que envolveram a determinação e caracterização da estrutura da proteína CP por RMN. Entre as etapas para a resolução da estrutura podem ser destacadas: aquisição de dados experimentais, assinalamento sequencial das ressonâncias dos átomos na estrutura molecular, assinalamento das interações observadas entre pares de átomos e a tradução dessas interações em informações conformacionais. A atribuição das ressonâncias é a etapa inicial do trabalho e erros cometidos nessa fase serão propagados para as etapas seguintes de cálculo e validação das estruturas. A ausência de proteínas homólogas a CP com estrutura resolvida tornou o assinalamento das interações internucleares crucial e difícil, principalmente porque não foi possível a comparação dos resultados obtidos com modelos previamente conhecidos.

Após a análise dos espectros, 100% das ressonâncias da cadeia principal ( $^1\text{H}$ ,  $^{15}\text{N}$  e  $^{13}\text{C}$ ) e 92% das ressonâncias de  $^{13}\text{C}$  e 95% de  $^1\text{H}$  das cadeias laterais da CP foram atribuídas. Os valores de deslocamentos químicos referentes a cada átomo estão apresentados no Anexo 1 e foram depositadas no banco de dados BMRB (*Biological Magnetic Resonance Data Bank*; <http://www.bmrb.wisc.edu>), com o número de acesso 6990. A notificação desses resultados foi feita através de uma carta ao editor da revista *Journal of Biomolecular NMR*, apresentado a seguir. O material suplementar enviado a revista está apresentado como Anexo 2.

É importante também destacar que, além das restrições clássicas de distância e de ângulos de torção, as constantes dipolares residuais (RDCs) foram utilizadas para a determinação da estrutura da CP. Diferente das restrições tradicionais, os RDCs descrevem a orientação internuclear relativo ao campo magnético e, portanto, são restrições independentes da distância. Apesar desse método ter sido desenvolvido apenas nos últimos anos, ele tem se tornado uma valiosa fonte de informações e tem ganhado um amplo destaque na determinação de estruturas por RMN, melhorando significativamente a

acurácia e a precisão das estruturas obtidas. Os dados de RDC foram coletados e analisados com a ajuda da Dra. Mariana Gallo e do Prof. Dr. Daniel O. Cicero, do *Dipartimento di Scienze e Tecnologie Chimiche*, na *Università degli Studi di Roma "Tor Vergata"*, Itália.

**ARTIGO 2:**

*Letter to the Editor:  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonance assignments of cerato-platanin, a phytotoxic protein from *Ceratocystis fimbriata*.*

Aline L. Oliveira<sup>a,b</sup>, Luigia Pazzagli<sup>c</sup>, Barbara Pantera<sup>c</sup>, Gianni Cappugi<sup>c</sup>, Celso E. Benedetti<sup>b</sup>, Alberto Spisni<sup>a</sup> & Thelma A. Pertinhez<sup>a</sup>

<sup>a</sup>*Department of Experimental Medicine, University of Parma, Parma, Italy.*

<sup>b</sup>*Center for Molecular Structural Biology, LNS, Campinas, Brazil*

<sup>c</sup>*Department of Biochemical Sciences, University of Florence, Florence, Italy*

Journal of Biomolecular NMR (2006) 36:50

## Letter to the Editor

**<sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C resonance assignments of cerato-platanin, a phytotoxic protein from *Ceratocystis fimbriata***  
DOI 10.1007/s10858-006-9023-z

Cerato-platanin (CP) is a recently described phytotoxic protein released by ascomycete fungus *C. fimbriata* which is the causative agent of canker stain in a number of plants (Pazzagli et al., 1999). CP elicits defence-related responses, suggesting its involvement in the recognition process and the elicitation of plant defence mechanisms. Since CP functions remain speculative and it shows no structural homology with other proteins, the determination of its 3D structure is of obvious interest. In fact, we expect it will provide new clues on its biological functions and may also disclose some new fold. Based on these considerations we undertook the determination of the 3D solution structure of the recombinant CP <sup>15</sup>N and <sup>13</sup>C double-labelled. 100% of the backbone atoms and, for the side chains, 92% of <sup>13</sup>C and 95% of <sup>1</sup>H chemical shifts have been assigned. CSI analysis indicate the presence of seven  $\beta$ -strands and two  $\alpha$ -helices. Resonance assignments have been deposited in the BioMagResBank under the accession number 6990.

Reference: Pazzagli et al. (1999) *J. Biol. Chem.*, **274**, 24959–24964.

A.L. Oliveira<sup>a,b</sup>, L. Pazzagli<sup>c</sup>, B. Pantera<sup>c</sup>, G. Cappugi<sup>c</sup>, C.E. Benedetti<sup>b</sup>, A. Spisni<sup>a</sup> & T.A. Pertinhez<sup>a,\*</sup>  
<sup>a</sup>*Dept. Experimental Medicine, University of Parma, Parma, 43100, Italy;* <sup>b</sup>*Center for Molecular Structural Biology, LNLS, Campinas, Brazil;* <sup>c</sup>*Dept. Biochemical Sciences, University of Florence, Florence, Italy*

\*To whom correspondence should be addressed. E-mail: [thelma@unipr.it](mailto:thelma@unipr.it)

**Supplementary material** is available at <http://dx.doi.org/10.1007/s10858-006-9023-z>

## Resultados

### Determinação do raio hidrodinâmico

Na caracterização de estados nativos de proteínas em solução, medidas das dimensões moleculares do sistema são consideradas importantes, principalmente por que esses dados podem informar sobre estados conformacionais da proteína, como por exemplo a presença de dímeros ou oligômeros maiores (Wilkins et al., 1999).

Utilizando-se a equação de Stokes-Einstein (Equação 2.1) é possível relacionar o coeficiente de difusão ( $D$ ) de uma partícula esférica, em solução e a diluição infinita, com o seu raio hidrodinâmico ( $R_H$ ):

$$D = \frac{kT}{6\pi\eta R_H} \quad (\text{Eq. 2.1})$$

onde  $\eta$  é a viscosidade do solvente,  $T$  é a temperatura da solução e  $k$  é a constante de Boltzmann (Jones et al., 1997). Porém, para se eliminar o efeito da viscosidade do solvente, as medidas de difusão podem ser efetuadas com a adição de um padrão interno e, portanto, com valor de  $R_H$  conhecido. Assim, o  $R_H$  da proteína de estudo pode ser determinado de maneira simples a partir da seguinte relação (Equação 2.2) (Wilkins et al., 1999):

$$R_H^{\text{proteína}} = \frac{D_{\text{referência}}}{D_{\text{proteína}}} \times R_H^{\text{referência}} \quad (\text{Eq.2.2})$$

Dessa forma, o  $R_H$  da CP foi calculado através de medidas do coeficiente de difusão. Através dos espectros de DOSY, foi calculado que a difusão da CP é de  $0,53 \times 10^{-10} \text{ m}^2/\text{s}$ , enquanto que o coeficiente de difusão do dioxano, utilizado como molécula de referência, é de  $4,88 \times 10^{-10} \text{ m}^2/\text{s}$ . A partir desses resultados e sabendo-se que o  $R_H$  do dioxano é de  $2,12 \text{ \AA}$  (Wilkins et al., 1999), pode-se calcular, através da Equação 2.2, que o  $R_H$  da proteína CP é  $19,5 \text{ \AA}$ .

Estudos empíricos de proteínas com estrutura conhecida essencialmente esférica mostram que há uma forte relação entre a dimensão molecular e o tamanho da cadeia polipeptídica. Essa relação é dada pela Equação 2.3.

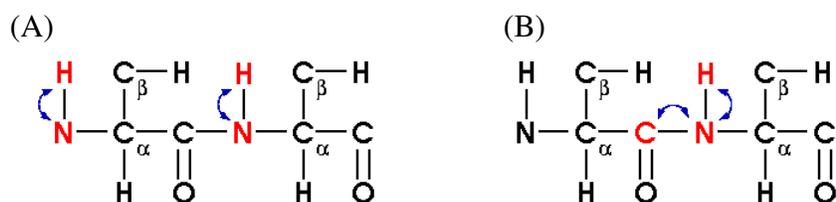
$$R_H = (4,75 \pm 1,11) N^{0,29 \pm 0,02} \quad (\text{Eq. 2.3})$$

onde  $N$  é o número de resíduos da cadeia polipeptídica e  $R_H$  é o raio hidrodinâmico em Å (Wilkins et al., 1999). A partir dessa relação, calculou-se que o  $R_H$  teórico da proteína CP recombinante (123 resíduos de aminoácido) é 19,2 Å. Comparando-se  $R_H$  calculado com os dados obtidos experimentalmente, observa-se que esses valores são próximos, confirmando que a proteína CP, nessas condições experimentais, é monomérica.

### Atribuição das ressonâncias

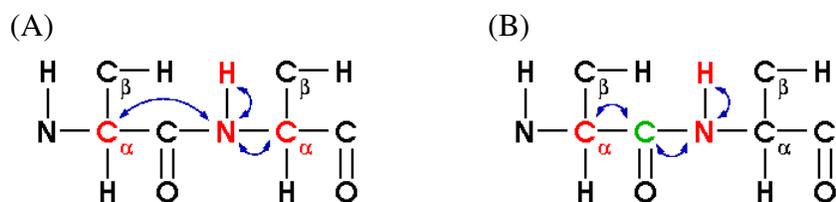
A estratégia para a atribuição sequencial das ressonâncias explora o acoplamento escalar  $J$  entre os átomos de  $^1\text{H}$  e os núcleos de  $^{15}\text{N}$  e  $^{13}\text{C}$  das cadeias principal e lateral da proteína ao qual eles estão diretamente ligados. Os espectros 2D e 3D heteronucleares obtidos para a proteína CP apresentaram uma boa resolução, com pouca sobreposição de sinais, proporcionando informações suficientes para uma completa e inequívoca atribuição dos sistemas de *spins*. Além disso, a sensibilidade das sequências de pulso e a concentração da proteína (0,7 mM) foram suficientes para a aquisição dos espectros com uma boa relação sinal/ruído.

Inicialmente foi efetuada a análise do espectro 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC, de onde se obteve informações da correlação entre os  $^1\text{H}$  e os  $^{15}\text{N}$ , tanto da cadeia principal da proteína CP (Figura 2.1A) quanto dos resíduos que possuem  $^{15}\text{N}$  na sua cadeia lateral (Trp, Arg, Asn, Gln e Lys). Uma vez que o número de sinais no  $^1\text{H}$ - $^{15}\text{N}$  HSQC corresponde a soma de todos os resíduos de amino ácidos menos as prolínas e o N-terminal, utilizou-se este espectro como ponto de partida para a análise dos experimentos 3D e a identificação de todos os sinais. O primeiro experimento 3D a ser analisado foi o HNCO, que correlaciona o deslocamento químico do  $\text{H}^{\text{N}}$  e  $^{15}\text{N}$  de um aminoácido ( $i$ ) com a carbonila ( $\text{C}'$ ) do resíduo precedente ( $i-1$ ), conforme ilustrado na Figura 2.1B.



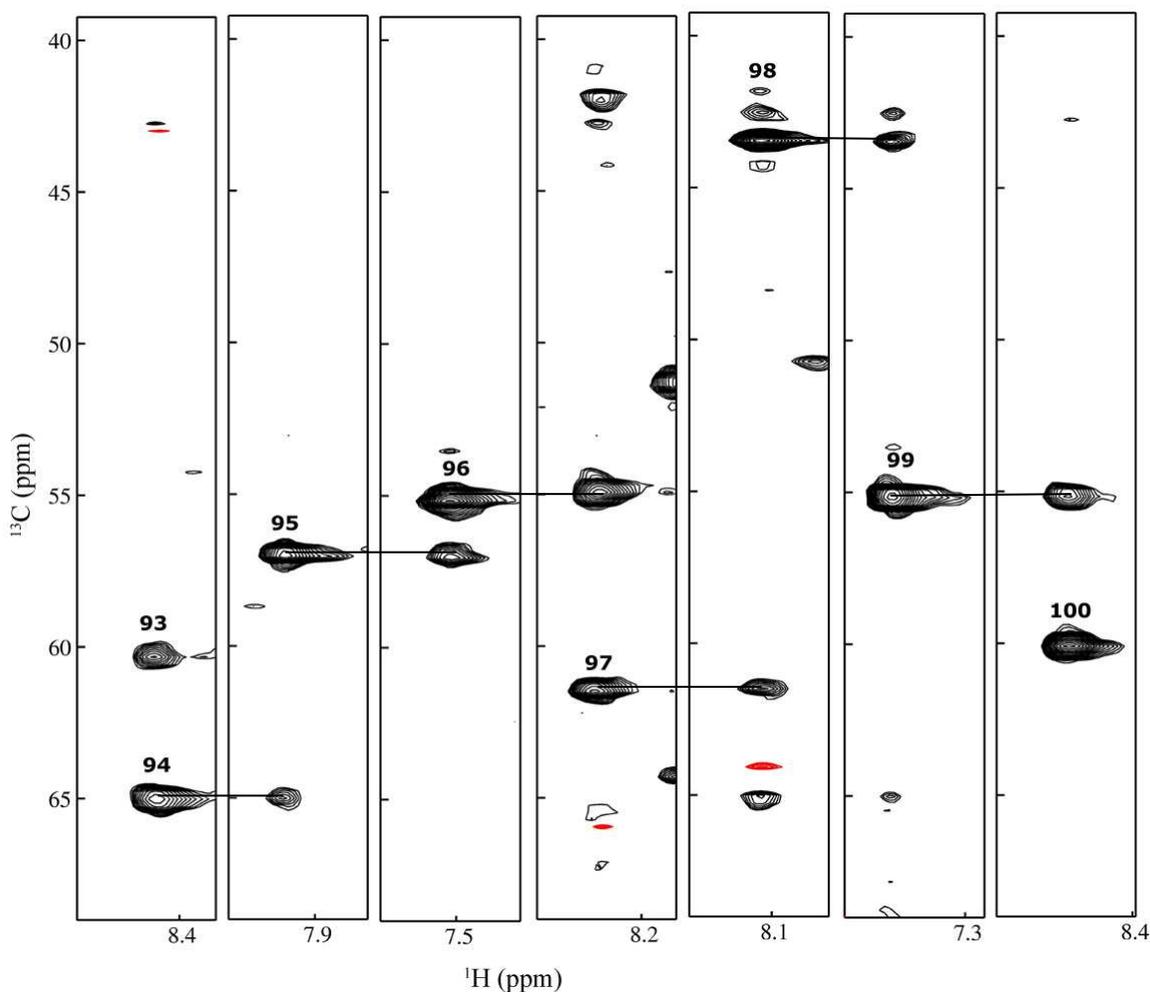
**Figura 2.1.** Ilustração da transferência de magnetização que ocorre nos experimentos (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC e (B) HNCO. Os átomos em vermelho são os núcleos que serão observados e as setas em azul indicam a direção da transferência de magnetização (<http://www.nanuc.ca/research/pulse.php>).

O experimento HNCA correlaciona o deslocamento do  $\text{H}^{\text{N}}$  com o deslocamento do  $^{15}\text{N}$  e do  $\text{C}^{\alpha}$  intra-residual (*i*) e do aminoácido anterior (*i-1*) (Figura 2.2A). Para distinguir qual pico deveria ser atribuído ao resíduo *i* e qual deveria ser atribuído ao resíduo *i-1*, comparou-se os planos desse espectro com os respectivos planos do espectro HN(CO)CA, que estabelece correlações entre as ressonâncias dos núcleos  $^{15}\text{N}$ - $\text{H}^{\text{N}}$  e  $-\text{C}^{\alpha}$  do resíduo precedente (*i-1*), como mostra a Figura 2.2B.



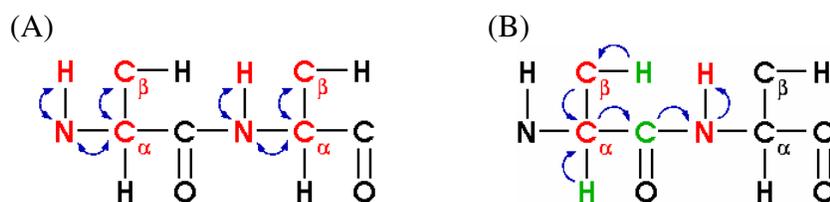
**Figura 2.2.** Ilustração da transferência de magnetização que ocorre nos experimentos (A) HNCA e (B) HN(CO)CA. Os átomos em vermelho são os núcleos que serão observados, os átomos em verde são os núcleos que estão envolvidos na magnetização, mas que não são observados no espectro e as setas em azul indicam a direção da transferência de magnetização (<http://www.nanuc.ca/research/pulse.php>).

A Figura 2.3 apresenta seções dos planos  $^{13}\text{C}/^1\text{H}$  do espectro 3D HNCA, com as respectivas frequências dos núcleos  $\text{H}^{\text{N}}$  e  $^{15}\text{N}$  (terceira dimensão) de seis resíduos consecutivos da CP (Thr $^{94}$ -Thr $^{100}$ ). Esse resultado mostra a aplicação do espectro HNCA. Para cada grupo amida é possível identificar um sinal mais intenso para o  $\text{C}^{\alpha}$  intra-residual e outro, com uma menor intensidade, para o aminoácido anterior.



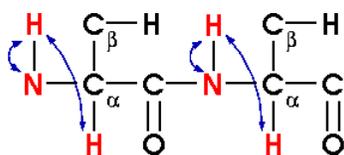
**Figura 2.3.** Seções do plano  $^{13}\text{C}/^1\text{H}$  do espectro HNCA de 0,7 mM da proteína  $^{13}\text{C}/^{15}\text{N}$ -CP, 10 mM tampão fosfato, 5%  $\text{D}_2\text{O}$ , pH 5,8 e  $20^\circ\text{C}$ . Os números indicam a qual resíduo da proteína o sinal se refere. Cada plano do espectro apresentado na figura, corresponde a frequência de  $^{15}\text{N}$  dos aminoácidos indicados (Anexo 1).

A atribuição do  $\text{C}^\beta$  é particularmente importante para a atribuição sequencial da proteína. Conhecendo-se os valores de deslocamento químico do  $\text{C}^\alpha$  e  $\text{C}^\beta$  pode-se determinar o tipo de aminoácido que está relacionado a cada  $^{15}\text{N}$  da cadeia principal, facilitando assim a atribuição sequencial da proteína. Com esse objetivo, foram adquiridos e analisados os espectros HNCACB e CBCA(CO)NH que correlacionam os spins dos  $\text{C}^\alpha$  e  $\text{C}^\beta$  com o par  $\text{H}^{\text{N}}\text{-}^{15}\text{N}$  intra-residual (*i*) e do resíduo antecedente (*i-1*), respectivamente (Figura 2.4).



**Figura 2.4.** Ilustração da transferência de magnetização que ocorre nos experimentos (A) HNCACB e (B) CBCA(CO)NH. Os átomos em vermelho são os núcleos que serão observados, os átomos em verde são os núcleos que estão envolvidos na magnetização, mas que não são observados no espectro e as setas em azul indicam a direção da transferência de magnetização (<http://www.nanuc.ca/research/pulse.php>).

Por fim, através do experimento HNHA foi possível obter a correlação entre  $H^N$ - $H^\alpha$  e  $^{15}N$  (Figura 2.5) e, conseqüentemente os valores de deslocamento químico do  $H^\alpha$ .

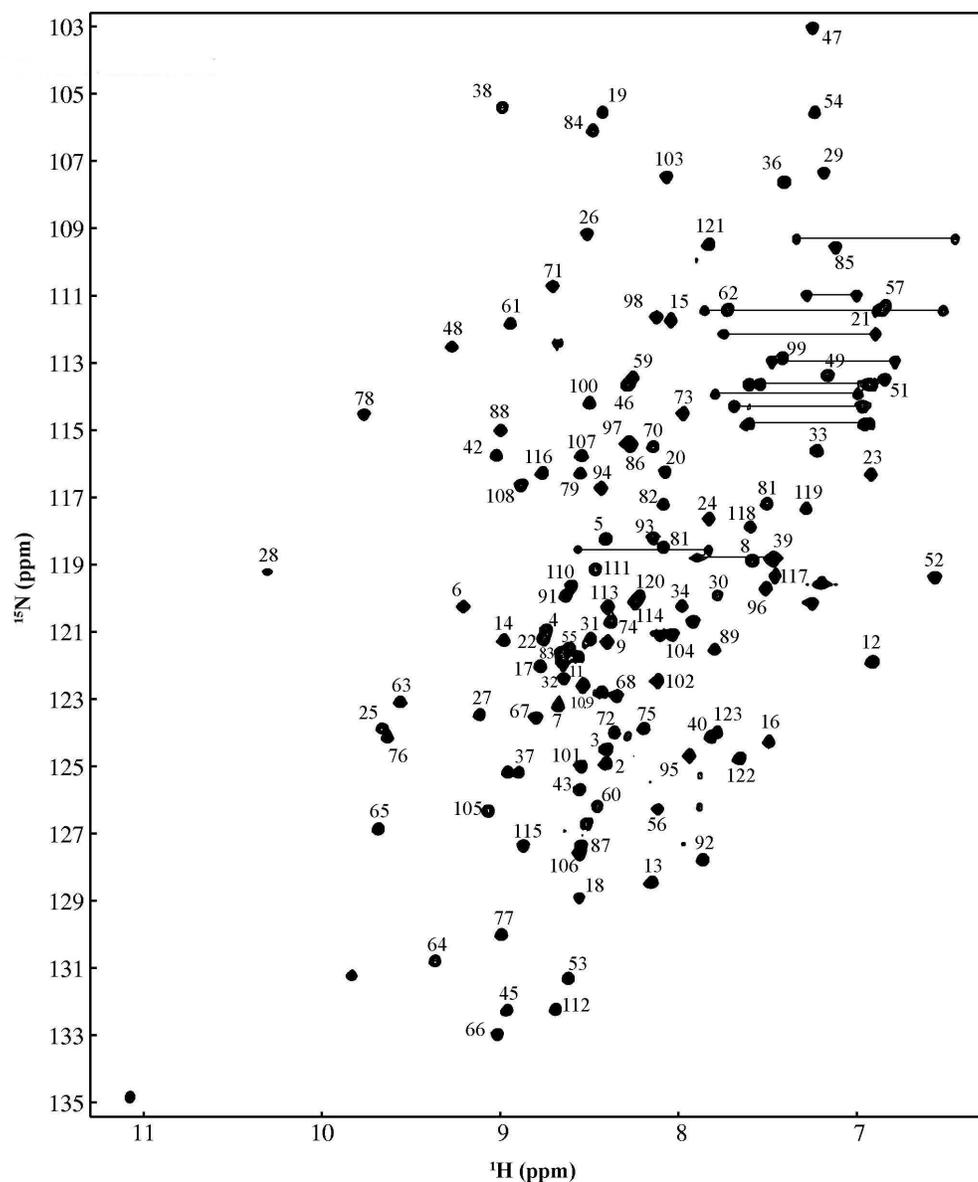


**Figura 2.5.** Ilustração da transferência de magnetização que ocorre no experimento HNHA. Os átomos em vermelho são os núcleos que serão observados e as setas em azul indicam a direção da transferência de magnetização (<http://www.nanuc.ca/research/pulse.php>).

Em posse desses dados, a conectividade sequencial das ressonâncias da cadeia principal da proteína pode ser feita correlacionando todas as informações obtidas ( $\delta$  de  $H^N$ ,  $H^\alpha$ ,  $^{15}N$ ,  $C^\alpha$ ,  $C^\beta$  de resíduos  $i$  e  $i-1$ ). Paralelamente, a identificação de cada aminoácido pode ser alcançada através da comparação da conectividade sequencial com os valores de deslocamento químicos que são característicos para cada tipo de aminoácido. Uma vez que as ressonâncias são atribuídas ao seu correspondente tipo de aminoácido, as porções sequenciais de deslocamento químico podem ser mapeadas na estrutura primária da proteína.

Cem por cento das ressonâncias dos núcleos  $^1H$ ,  $^{15}N$  e  $^{13}C$  da cadeia principal da proteína CP foram atribuídas. Os valores de deslocamento químico ( $\delta$ , ppm) de cada um

desses átomos estão reunidos no Anexo 1. A Figura 2.6 mostra o espectro de  $^1\text{H}$ - $^{15}\text{N}$  HSQC com a atribuição dos resíduos da proteína CP.

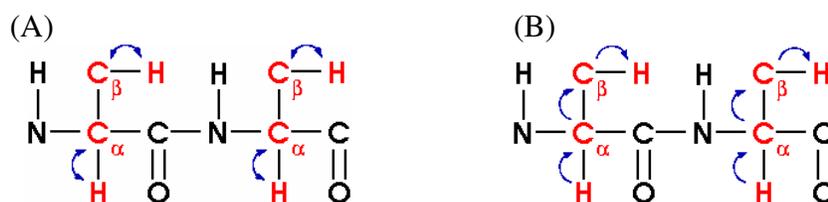


**Figura 2.6.** Espectro de  $^1\text{H}$ - $^{15}\text{N}$ -HSQC de 0,7 mM da proteína  $^{13}\text{C}/^{15}\text{N}$ -CP, 10 mM tampão fosfato, 5%  $\text{D}_2\text{O}$ , pH 5,8 e  $20^\circ\text{C}$ . As linhas conectam os  $^1\text{H}$  dos grupos  $\text{NH}_2$  da cadeia lateral de Asn e Gln. Os picos não identificados referem-se as cadeias laterais de Trp e Arg.

O espectro  $^1\text{H}$ - $^{15}\text{N}$  HSQC da CP apresenta sinais intensos com uma boa dispersão, confirmando que a proteína tem uma estrutura bem definida e indicando que ela não agrega em altas concentrações, como nas condições requeridas para os experimentos de RMN. Neste espectro podem-se identificar as ressonâncias atribuídas aos  $\text{H}^{\text{N}}$  e ao  $^{15}\text{N}$  de cada um dos resíduos da proteína. Também foram identificados treze pares de ressonâncias de  $\text{NH}_2$  (totalizando 26 sinais, que correlacionam cada  $^1\text{H}$  com seu respectivo  $^{15}\text{N}$ ) da cadeia lateral de onze resíduos de Asn e dois resíduos de Gln presentes na CP. Além disso, dois sinais foram atribuídos ao  $\text{H}^{\epsilon}$  do anel indólico dos dois triptofanos e outros três picos foram atribuídos ao  $\text{H}^{\epsilon}$  da cadeia lateral de argininas.

Em posse da atribuição das ressonâncias da cadeia principal da proteína CP, iniciou-se a coleta e análise dos espectros de HCCH-TOCSY, HCCH-COSY e  $^1\text{H}$ - $^{13}\text{C}$  HSQC para a atribuição dos sistemas de *spins* da cadeia lateral dos resíduos alifáticos da proteína CP. Para isso, a amostra foi liofilizada e ressuspensa em 100% de  $\text{D}_2\text{O}$  para ser possível observar aqueles sinais que se encontram próximo ao sinal da água (4-5 ppm).

A partir do experimento 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC obteve-se informações da correlação de todos os  $^1\text{H}$  ligados a  $^{13}\text{C}$  (Figura 2.7A) envolvidos na cadeia lateral da CP. Os experimentos 3D HCCH-TOCSY e HCCH-COSY correlacionam os núcleos de  $^{13}\text{C}$  e  $^1\text{H}$  através do acoplamento escalar  $J$ , ou seja, através da ligação química (Figura 2.7B). Semelhante aos espectros 2D TOCSY e COSY de  $^1\text{H}$ , no espectro HCCH-TOCSY para cada  $^{13}\text{C}$  é possível observar todos  $^1\text{H}$  envolvidos na cadeia de um específico resíduo enquanto que no experimento HCCH-COSY essa correlação ocorre apenas entre os átomos envolvidos a três ligações químicas. Como os valores dos deslocamentos químicos de  $\text{H}^{\alpha}$ ,  $\text{H}^{\beta}$  e  $\text{C}^{\alpha}$ ,  $\text{C}^{\beta}$  de cada resíduo da CP são conhecidos, foi possível determinar o sistema de *spins* do restante da cadeia lateral através da análise dos planos de  $^1\text{H}$ - $^1\text{H}$  ( $^{13}\text{C}$  na terceira dimensão) do espectro HCCH-TOCSY. Para distinguir os  $^1\text{H}$  e  $^{13}\text{C}$  sequenciais de um mesmo resíduo utilizou-se o espectro HCCH-COSY. As ressonâncias das cadeias laterais dos resíduos aromáticos foram atribuídas através dos espectros NOESY  $^1\text{H}$ - $^{13}\text{C}$  CHSQC que correlacionam os átomos de  $^1\text{H}$  do anel aromático com o  $\text{C}^{\alpha}$  e  $\text{C}^{\beta}$ , previamente conhecidos, através do espaço.



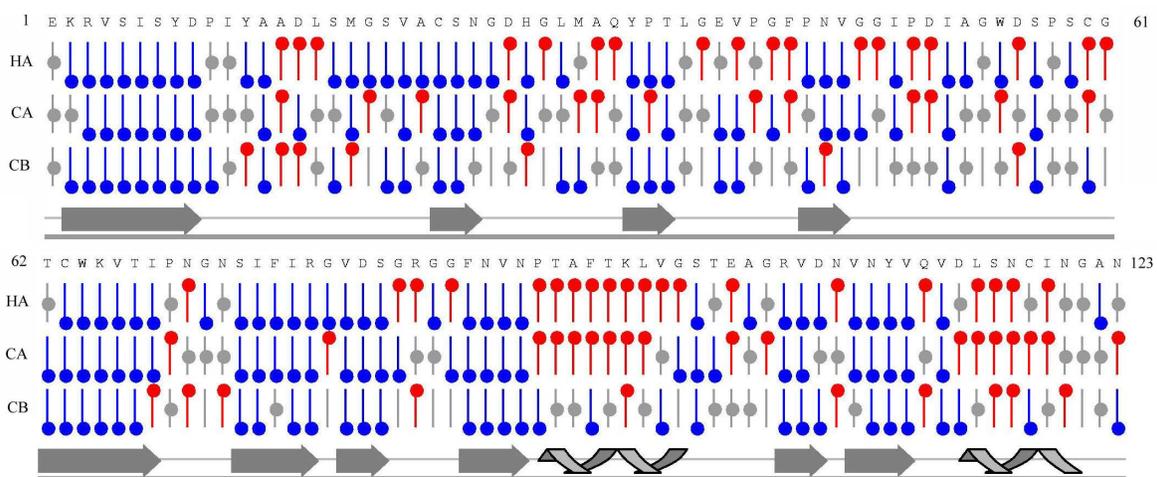
**Figura 2.7.** Ilustração da transferência de magnetização que ocorre nos experimentos (A)  $^1\text{H}$ - $^{13}\text{C}$  CHSQC e (B) HCCH-TOCSY e HCCH-COSY. Os átomos em vermelho são os núcleos que serão observados e as setas em azul indicam a direção da transferência de magnetização (<http://www.nanuc.ca/research/pulse.php>).

Dessa forma, 92% das ressonâncias de  $^{13}\text{C}$  e 95% dos  $^1\text{H}$  da cadeia lateral da CP foram atribuídas. Devido a sobreposição ou ausência de sinal, algumas ressonâncias de cadeia lateral não puderam ser atribuídas. Os valores de deslocamentos químicos referentes a cada átomo da proteína estão apresentados no Anexo 1.

### Chemical shift index

O deslocamento químico obtido através dos espectros de RMN tem sido utilizado como um parâmetro para indicar elementos de estrutura secundária presente numa proteína, uma vez que esses dados são influenciados pela geometria local da cadeia polipeptídica (Wishart & Nip, 1998). Em posse da atribuição das ressonâncias de  $\text{H}^\alpha$ ,  $\text{C}^\alpha$  e  $\text{C}^\beta$ , o CSI da proteína CP foi calculado (Figura 2.8).

O CSI da CP indica que a proteína apresenta 8 fitas- $\beta$  que compreendem as regiões entre os resíduos Lys<sup>2</sup>-Asp<sup>9</sup>, Cys<sup>23</sup>-Ans<sup>25</sup>, Tyr<sup>34</sup>-Thr<sup>36</sup>, Pro<sup>44</sup>-Val<sup>46</sup>, Thr<sup>62</sup>-Ile<sup>68</sup>, Ser<sup>73</sup>-Ser<sup>81</sup>, Phe<sup>86</sup>-Asn<sup>89</sup> e Arg<sup>104</sup>-Val<sup>111</sup> e duas  $\alpha$ -hélices entre os resíduos Pro<sup>90</sup>-Gly<sup>98</sup> e Leu<sup>115</sup>-Gly<sup>121</sup>.



**Figura 2.8.** *Chemical Shift Index* para os átomos de  $H^\alpha$ ,  $C^\alpha$  e  $C^\beta$  da proteína  $^{13}C/^{15}N$ -CP. As barras azuis indicam um CSI negativo, as barras vermelhas indicam um CSI positivo e as barras cinzas indicam um CSI igual a zero. As setas indicam a presença de estrutura em fita- $\beta$  e as espirais a presença de estrutura em  $\alpha$ -hélice.

### Constante de acoplamento $J_{HNH\alpha}$

A constante homonuclear  $J$  surge do acoplamento entre dois spins nucleares devido a influência dos spins dos elétrons envolvidos na ligação. A constante de acoplamento homonuclear a três ligações  $J_{HNH\alpha}$  contém informações valiosas com relação ao ângulo dihedro  $phi$  ( $\phi$ ) da cadeia principal de proteínas e conseqüentemente da sua estrutura secundária (Vuister & Bax, 1993).

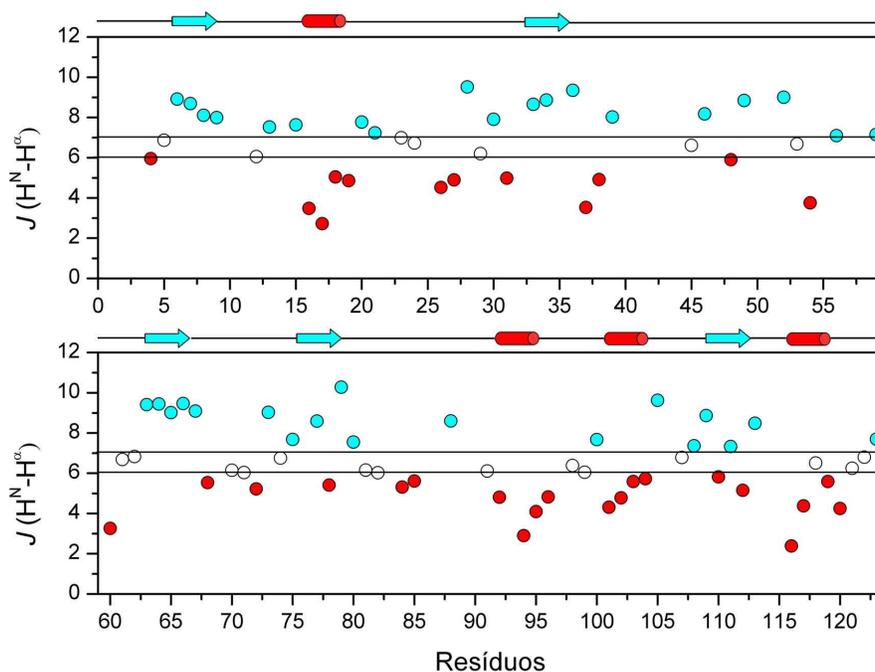
Assim, com o propósito de medir a constante homonuclear  $J_{HNH\alpha}$  foram coletados espectros de HNHA da proteína  $^{13}C/^{15}N$ -CP e os valores medidos estão reunidos da Figura 2.9.

A magnitude da constante  $J_{HNH\alpha}$  pode ser relacionada com o ângulo  $\phi$  aplicando-se a conhecida equação de Karplus (Equação 2.4):

$$J = A \cos^2(\phi - 60) + B \cos(\phi - 60) + C \quad (\text{Eq. 2.4})$$

onde  $A = 6,51$ ;  $B = -1,76$  e  $C = 1,60$ . De acordo com essa equação, valores de constante  $J$  menores que 6 Hz corresponderão a ângulos  $\phi$  de aproximadamente  $-60^\circ \pm 40^\circ$ , que são característicos de estruturas secundárias do tipo  $\alpha$ -hélice, enquanto que valores de constante  $J$  maiores que 7 Hz corresponderão a ângulos entre  $-120^\circ \pm 50^\circ$ , que são típicos de estruturas em folhas- $\beta$  (Vuister & Bax, 1993).

Dos 123 resíduos da proteína CP, foram calculadas 90 constantes  $J_{HNH\alpha}$ . Para o restante dos resíduos, os picos de correlação apresentavam pouca resolução, ausência ou sobreposição de sinal, não sendo possível calcular as suas respectivas constantes. As regiões entre os resíduos Ile<sup>6</sup>-Asp<sup>9</sup>, Cys<sup>63</sup>-Thr<sup>67</sup>, Phe<sup>75</sup>-Asp<sup>80</sup> e Val<sup>108</sup>-Val<sup>113</sup> possuem valores de  $J_{HNH\alpha}$  maiores que 7 Hz, indicando que essas regiões estão envolvidas em estrutura do tipo  $\beta$ . Os resíduos Leu<sup>16</sup>-Gly<sup>19</sup>, Ala<sup>92</sup>-Leu<sup>96</sup>, Glu<sup>101</sup>-Arg<sup>104</sup> e Ser<sup>116</sup>-Asn<sup>120</sup> apresentam constantes com valores menores que 6 Hz, típicos de estrutura helicoidal (Figura 2.9).



**Figura 2.9.** Gráfico da constante de acoplamento  $J_{HNH\alpha}$  obtidos para a proteína CP. Os círculos ciano indicam valores de  $J_{HNH\alpha}$  maiores que 7 Hz, característicos de estrutura em folha- $\beta$ , os círculos vermelhos indicam valores de  $J_{HNH\alpha}$  menores que 6 Hz, característicos de estrutura em  $\alpha$ -hélice, e os círculos brancos indicam valores de  $J_{HNH\alpha}$  entre 6-7 Hz, correspondentes a estrutura randômica.

De forma geral, esses resultados são coerentes aos resultados de CSI. Uma exceção é a região entre os resíduos Leu<sup>16</sup>-Gly<sup>19</sup> e Glu<sup>101</sup>-Arg<sup>104</sup> que, ao contrário dos resultados de  $J_{HNH\alpha}$ , o CSI não indica a presença de uma estrutura definida envolvendo esses resíduos. Outra discordância entre os resultados se refere a porção entre os resíduos Cys<sup>23</sup>-Asn<sup>25</sup> e Pro<sup>44</sup>-Val<sup>46</sup> cujo CSI indica a presença de estrutura tipo  $\beta$  porém esses resíduos não possuem valores de  $J_{HNH\alpha}$  maiores que 7 Hz.

### Efeito nuclear Overhauser

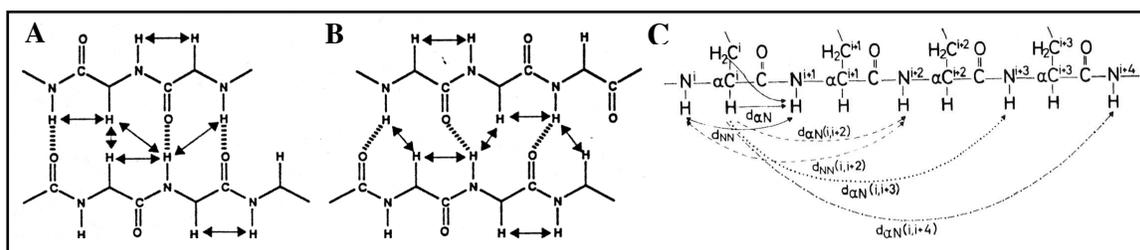
Na determinação de estruturas 3D de proteínas em solução por RMN o único dado que é capaz de fornecer informações sobre os limites máximos de distância entre dois átomos é derivado do NOE. Por definição, NOE é a variação da população de equilíbrio, e consequentemente da intensidade de ressonância, de um spin nuclear  $I$  devido perturbação provocada pela diferença de população de um vizinho próximo  $S$ . Geralmente essa perturbação ocorre por saturação ou por inversão de população. O sistema perturbado  $I$  altera a sua população de spin com objetivo de retornar a situação de equilíbrio. O mecanismo que contribui para essas mudanças de população de spin nucleares é o acoplamento dipolar, ou seja, é a interação direta e magnética através do espaço. E o que permite obter as informações de distância é o fato de que a força do acoplamento dipolar entre dois *spins* é inversamente proporcional a sua distância ( $1/r^6$ ) (Claridge, 1999).

Para extrair as restrições de distância do espectro de NOESY, os picos de correlação cruzada devem ser atribuídos, ou seja, o par de átomos de hidrogênios que estão interagindo deve ser identificado. Para isso, a atribuição dos espectros de <sup>15</sup>N-NOESY-HSQC e <sup>13</sup>C-NOESY-HSQC obtidos para a proteína <sup>15</sup>N/<sup>13</sup>C-CP baseou-se nos valores de deslocamento químico determinado previamente (Anexo 1). O espectro bidimensional <sup>1</sup>H NOESY e seções de planos dos espectros <sup>15</sup>N-NOESY-HSQC e <sup>13</sup>C-NOESY-HSQC estão apresentados no Anexo 3.

Para ajudar na atribuição dos NOEs e a determinar o enovelamento da proteína, os resultados de CSI e de constantes  $J_{HNH\alpha}$  foram usados como referência. Além disso, usando-se o programa TALOS (Cornileuscu et al., 1999) foi efetuada uma predição dos

ângulos dihedros  $\phi$  e  $\psi$ , que estão intimamente relacionados às características de estrutura secundária. O programa TALOS (*Torsion Angle Likelihood Obtained from Shift and sequence similarity*) é um sistema de banco de dados para a predição empírica dos ângulos dihedros  $\phi$  e  $\psi$  com base nos valores de deslocamento químico. As análises são feitas através de uma combinação de cinco tipos de deslocamento químico ( $H_\alpha$ ,  $C_\alpha$ ,  $C_\beta$ , CO, N) para uma dada sequência de proteína. Os valores dos ângulos dihedros  $\phi$  e  $\psi$  preditos estão reunidos no Anexo 4.

Desta forma, inicialmente controlou-se a presença de NOEs característicos de elementos de estrutura secundária helicoidal e de folha- $\beta$ , como ilustra a Figura 2.10.



**Figura 2.10.** Esquema dos padrões de NOEs (representados por flechas) observados para estruturas em fitas- $\beta$  antiparalela (A), paralela (B) e  $\alpha$ -hélice (C) (Wüthrich, 1986).

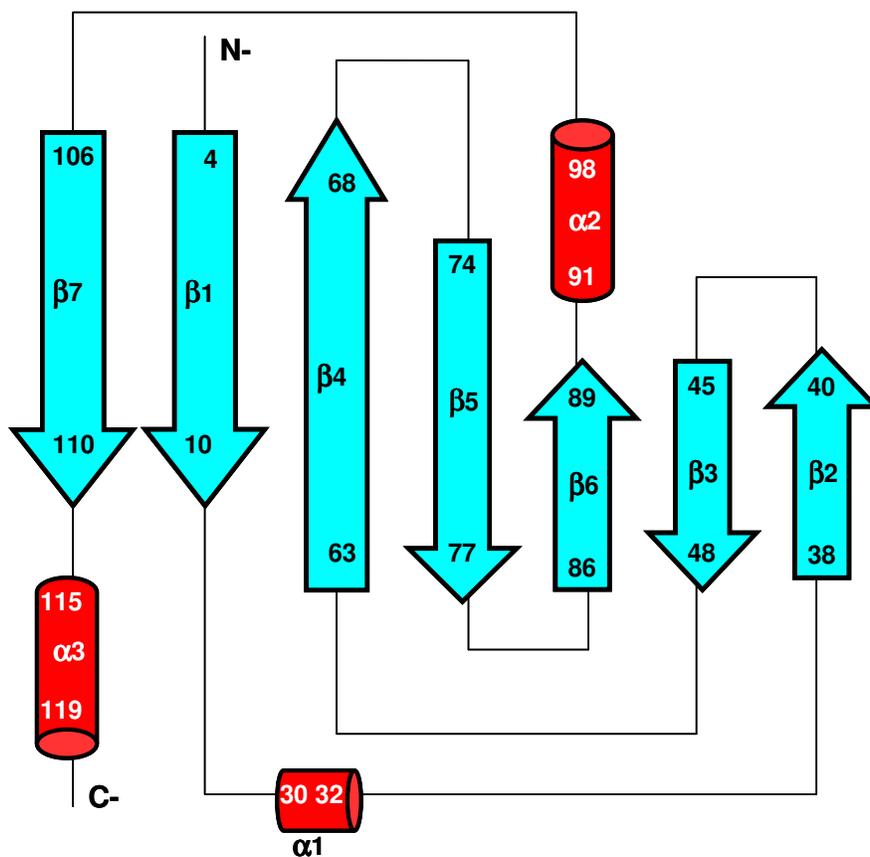
Devido à precisão limitada da posição dos picos e dos valores de deslocamento químico, muitos NOEs não puderam ser atribuídos como um par único de *spins*, apresentando uma atribuição ambígua, composta por múltiplos pares de *spins* (sobreposição de sinal). O número de NOEs, principalmente entre os átomos da cadeia principal da proteína CP, com atribuição única é pequeno. Além disso, foram encontradas dificuldades devido a ausência de sinais esperados, provavelmente pela rápida relaxação ou mudança conformacional da proteína, a artefatos dos espectros e a ruído.

Esses inevitáveis efeitos tornaram difícil a análise dos dados de RMN. Porém, as dificuldades na atribuição dos NOEs foram superadas através das análises da coerência entre os sinais de cadeia principal e lateral da proteína, assim como pela atribuição feita de forma interativa, onde as estruturas preliminares foram utilizadas como referência para a redução de ambiguidades no assinalamento dos NOEs. Assim, a partir dos dados iniciais de

restrições conformacionais, iniciou-se o cálculo das estruturas que por sua vez, foram analisadas para identificar erros nas restrições e assinalamentos. Tais erros são tipicamente identificados através das análises de violações das restrições com relação ao conjunto de estruturas, seguido da verificação das restrições contra os dados experimentais. As violações de distância podem ocorrer devido a uma atribuição errada, sobreposição de picos que não foram consideradas inicialmente ou a erros no volume do sinal, e conseqüentemente na distância entre os átomos, devido ao efeito conhecido como difusão de *spin*. Essas análises foram utilizadas para decidir quando a restrição deveria ser modificada, por exemplo quando falsos assinalamentos foram feitos, ou quando a distância das restrições deveriam ser ajustadas. É nesse passo que a determinação de estrutura por RMN é extremamente vulnerável a erros humanos, como por exemplo no caso de remoções injustificadas de restrições que apresentam distância violada (Spronk et al., 2004; Ferentz & Wagner, 2000; Kay, 1997).

Os passos de assinalamento e análise de erros foram feitos de maneira interativa. Após os dados serem otimizados, eles foram utilizados para calcular o conjunto de estruturas finais que é subsequentemente sujeito a validação para se obter uma indicação da qualidade e estatística das estruturas. A validação pode também apontar para possíveis problemas, como conformações erradas, e foi usada para re-examinar cuidadosamente os dados ou o procedimento de cálculo da estrutura (Spronk et al., 2004).

Foram atribuídos um total de 1250 NOEs, sendo 451 intra-residuais, 415 de média distância e 384 de longa distância. A partir das análises dos NOEs de longa distância, que fornecem maiores informações sobre o enovelamento da proteína, pôde-se determinar a topologia da proteína CP, apresentada na Figura 2.11. A estrutura da CP é composta por duas fitas- $\beta$  paralelas, Asp<sup>106</sup>-Tyr<sup>110</sup> ( $\beta$ 7) e Val<sup>4</sup>-Pro<sup>10</sup> ( $\beta$ 1), e cinco fitas- $\beta$  anti-paralelas, Cys<sup>63</sup>-Ile<sup>68</sup> ( $\beta$ 4), Ile<sup>74</sup>-Arg<sup>77</sup> ( $\beta$ 5), Phe<sup>86</sup>-Asn<sup>89</sup> ( $\beta$ 6), Asn<sup>45</sup>-Gly<sup>48</sup> ( $\beta$ 3) e Gly<sup>38</sup>-Val<sup>40</sup> ( $\beta$ 2), formando uma única folha- $\beta$ , além de três elementos helicoidais, sendo uma hélice  $3_{10}$ , Leu<sup>30</sup>-Ala<sup>32</sup> ( $\alpha$ 1), e duas  $\alpha$ -hélices, Thr<sup>91</sup>-Gly<sup>98</sup> ( $\alpha$ 2) e Leu<sup>115</sup>-Ile<sup>119</sup> ( $\alpha$ 3).



**Figura 2.11.** Topologia da CP baseada na atribuição dos NOEs de longa distância.

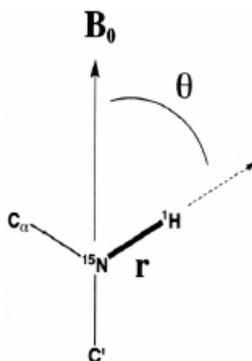
As indicações de estrutura secundária obtidas através do CSI e da constante  $J_{HNH\alpha}$  apresentam concordância com a topologia obtida, foram observadas apenas pequenas variações com relação a extensão dos resíduos envolvidos nas estruturas. Apenas a porção que envolve os resíduos Leu<sup>16</sup>-Asn<sup>25</sup> apresenta diferentes previsões dependendo da análise que se é efetuada. O CSI indica a presença de uma fita- $\beta$  entre os resíduos Cys<sup>23</sup>-Asn<sup>25</sup> enquanto que a constante  $J_{HNH\alpha}$  dos resíduos Leu<sup>16</sup>-Gly<sup>19</sup> é característica de estrutura helicoidal. Interessante notar que também foram encontradas dificuldades na análise dos NOEs dessa região, principalmente devido à ausência de informação. Essa característica evidencia a presença de estruturas desordenadas e que é refletida na contradição entre as análises de CSI e  $J_{HNH\alpha}$  desse segmento.

### Constante dipolar residual

Os RDCs surgem das mesmas interações básicas que dão origem ao NOE, a interação dipolo-dipolo através do espaço ou de acoplamento dipolar. Para um par de núcleos com spin 1/2 acoplados fracamente esta interação pode ser representada pela Equação 2.5:

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h S}{(2\pi r)^3} \left\langle \frac{3 \cos^2 \theta - 1}{2} \right\rangle \quad (\text{Eq. 2.5})$$

onde  $r$  é a distância entre os núcleos  $i$  e  $j$ ,  $\gamma_{ij}$  são as razões giromagnéticas para os núcleos  $i$  e  $j$ ,  $\mu_0$  é a permeabilidade no espaço,  $h$  é a constante de Plank,  $S$  é o parâmetro de ordem (medido em graus de mobilidade interna) e  $\theta$  é o ângulo entre o vetor internuclear e o campo magnético externo, como mostra a Figura 2.12. Para um par de núcleos diretamente ligados,  $r$  é fixo e  $\theta$  torna-se a fonte primária de informação estrutural. Assim, a obtenção da restrição de ângulo  $\theta$  para cada vetor de ligação ao longo da cadeia principal da proteína torna-se uma ferramenta poderosa para determinar a geometria molecular (de Alba & Tjandra, 2002; Prestegard et al., 2005).



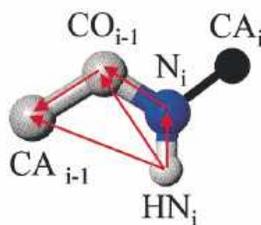
**Figura 2.12.** Representação esquemática da orientação da ligação N-H com relação ao campo magnético externo  $B_0$ .

As chaves  $\langle \rangle$  na Equação 2.5 denotam uma média do tombamento molecular que ocorre em solução. Em meio isotrópico, devido ao rápido tombamento da molécula, a média de interação do vetor internuclear tende a zero e nenhuma interação dipolo-dipolo é

observada. Porém, essa média pode ser diferente de zero pela inclusão de um alinhamento parcial à molécula estudada. O alinhamento geralmente é feito com o uso de um cristal líquido aquoso, como bicelas ou bacteriófagos filamentosos, que interagem fracamente com a molécula de interesse. Infelizmente, o alinhamento parcial introduz parâmetros adicionais à Equação 2.5, que precisam ser caracterizados. Para extrair esses parâmetros de alinhamentos, a dependência angular do RDC, a quantidade  $\langle \cos^2(\theta) - 1 \rangle$ , pode ser decomposta em um termo não dependente do tempo, que descreve a orientação relativa do vetor internuclear a uma referência arbitrária, e em um termo dependente do tempo. Esse segundo termo mede a população e a orientação média das moléculas alinhadas ao campo magnético e é descrito por uma quantidade tensorial, a matriz de Saupe de ordem  $3 \times 3$  (Saupe, 1968). Essa matriz pode ser reduzida a dois componentes independentes que contém toda a informação sobre a orientação média do tensor com respeito ao eixo  $z$  e sobre a população de moléculas que estão, em média, alinhadas ao longo do eixo  $z$ . Os dois tensores são usualmente expressos como a componente axial  $D_a$  e o componente rômico  $D_r$ . A rombicidade mede a assimetria do tensor e o componente axial mede o grau de alinhamento molecular residual (de Alba & Tjandra, 2002; Prestegard et al., 2005).

Na prática é fácil medir as RDCs. Quando todos os parâmetro da equação estão no sistema de unidade internacional (SI),  $D_{ij}$  é dado em Hz e corresponde a contribuição dipolar do desdobramento de dubletos que pode ser observado para cada membro de um par de spin fracamente acoplados. Quando os spins são diretamente ligados o acoplamento escalar entre a ligação ( $J_{ij}$ ) também existe, e o desdobramento observado é a soma do acoplamento escalar e dipolar. Assim, os RDCs são medidos como a diferença entre o desdobramento observado em espectros da molécula de estudo em meio parcialmente alinhado ( $J_{ij} + D_{ij}$ ) e isotrópico ( $J_{ij}$ ) (de Alba & Tjandra, 2002; Prestegard et al., 2005).

Para a proteína CP foram determinadas RDCs das ligações da cadeia principal:  $^1\text{H}_\text{N}$ - $^{15}\text{N}$ ,  $^1\text{H}_\text{N}$ - $^{13}\text{C}_\text{O}$ ,  $^{15}\text{N}$ - $^{13}\text{C}_\text{O}$ ,  $^{13}\text{C}_\alpha$ - $^{13}\text{C}_\text{O}$ , representadas na Figura 2.13, e  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$ .



**Figura 2.13.** Representação dos vetores de ligação, representados por flechas, de cada RDC medido (de Alba & Tjandra, 2002).

Com a adição de 16 mg/mL do bacteriófago filamentosso Pf1 foi observado o alinhamento da proteína CP e, conseqüentemente, foi possível determinar 106 RDCs de  $^1\text{H}_\text{N}$ - $^{15}\text{N}$ , 90 RDCs de  $^1\text{H}_\text{N}$ - $^{13}\text{CO}$ , 92 RDCs de  $^{15}\text{N}$ - $^{13}\text{CO}$ , 112 RDCs de  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  e 90 RDCs de  $^{13}\text{C}_\alpha$ - $^{13}\text{CO}$ . Para os demais resíduos, não foi possível calcular as suas respectivas constantes devido a pouca resolução, ausência ou sobreposição dos sinais.

Uma vez que o valor da constante dipolar depende da razão giromagnética e da distância entre os átomos envolvidos na ligação (Equação 2.5), as magnitudes intrínsecas das constantes medidas são diferentes. Assim, para o uso dessas medidas como restrições no cálculo das estruturas, os valores das RDCs foram normalizados com relação a ligação  $^1\text{H}_\text{N}$ - $^{15}\text{N}$  pelos fatores mostrados na Tabela 2.1. Os valores de RDCs normalizados estão reunidos no Anexo 5.

**Tabela 2.1.** Fator de normalização para as RDCs das ligações  $^1\text{H}_\text{N}$ - $^{13}\text{CO}$ ,  $^{15}\text{N}$ - $^{13}\text{CO}$ ,  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  e  $^{13}\text{C}_\alpha$ - $^{13}\text{CO}$ .

Ligação	Fator de normalização
$^1\text{H}_\text{N}$ - $^{13}\text{CO}$	0,319
$^{15}\text{N}$ - $^{13}\text{CO}$	0,121
$^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$	2,020
$^{13}\text{C}_\alpha$ - $^{13}\text{CO}$	0,198

## Procedimentos Experimentais

### Determinação do raio hidrodinâmico

O valor do raio hidrodinâmico da CP foi calculado através de medidas do coeficiente de difusão. Para isso, foi preparada 600 µL de solução 0,2 mM CP em D<sub>2</sub>O e 0,02% de dioxano, que foi utilizado como padrão interno para eliminar o efeito da viscosidade do solvente. O experimento foi coletado a 20°C utilizando-se o espectrômetro de RMN Varian Inova de 500 MHz e a sequência de pulso Dbppste (*Diffusion Ordered Spectroscopy Bipolar Pulse Pair Stimulated Echo*) (Wu et al., 1995). 24 espectros foram adquiridos com a força do gradiente variando entre 5 e 100% do seu valor máximo (30000 Gauss) de modo que o decaimento do sinal da proteína fosse entre 80 e 90%. Os espectros de RMN foram adquiridos com 2048 pontos complexos e uma janela espectral de 8000 Hz. O tempo total de coleta do experimento, utilizando-se 512 transientes, foi de 6 horas. O decaimento dos sinais de ressonância da proteína e do dioxano foi calculado pelo ajuste de equações gaussianas.

### Preparo da proteína <sup>15</sup>N/<sup>13</sup>C- CP

A proteína recombinante CP foi obtida através da clonagem e expressão do gene cp em cepas de levedura *Pichia pastoris* (linhagem GS115). Foi utilizado o vetor de expressão extracelular pPIC9, contendo uma sequência sinal que endereça a secreção da proteína (Carresi *et al.*, 2006).

A marcação isotópica com <sup>13</sup>C e <sup>15</sup>N da proteína foi feita através do crescimento da levedura em meio MGY (3,4 g/L YNB, 1% <sup>15</sup>NH<sub>4</sub>SO<sub>4</sub>, 10% <sup>13</sup>C-glicerol, 0,02% biotina) a 30°C e agitação constante (220 rpm). No momento em que a OD<sub>600</sub> atingiu valores entre 4-7 (aproximadamente após 20 h de crescimento), as células foram coletadas por centrifugação a 4000 rpm por 20 min a temperatura de 4 °C. O sobrenadante foi cuidadosamente descartado e as células foram re-suspendidas em meio MM (3,4g/L YNB, 1% <sup>15</sup>NH<sub>4</sub>SO<sub>4</sub>, 0,5% <sup>13</sup>C-metanol, 0,02% biotina) e mantidas sob agitação constante (220 rpm) a 28°C por 12 dias para induzir a expressão da proteína heterológa. A cada 24 h foram

adicionados 0,5% de  $^{13}\text{C}$ -metanol com o objetivo de repor o álcool evaporado e assim manter a indução da expressão da proteína CP.

Ao final de 12 dias de indução, o sobrenadante foi coletado por centrifugação (13000 rpm, 30 min, a  $4^{\circ}\text{C}$ ) e concentrado até atingir um volume de aproximadamente 1 mL. Para isso, foi utilizado concentrador *Centriprep*<sup>®</sup> (Amicon, Millipore), com membranas com corte de 3 kDa (YM-3). O concentrador foi primeiramente centrifugado por 95 min, a  $25^{\circ}\text{C}$  e 3000 g com 15 mL de água deionizada MiliQ com o objetivo de eliminar todo o glicerol presente na membrana, que poderia atrapalhar a aquisição dos dados de RMN. Após eliminar o glicerol, alíquotas de 15 mL do sobrenadante foram centrifugadas ( $25^{\circ}\text{C}$ , 3.000 g) usando concentrador *Centriprep*<sup>®</sup>, eliminando-se o filtrado, até que se atingisse o volume desejado.

A proteína  $^{13}\text{C}/^{15}\text{N}$ -CP foi purificada utilizando-se a estratégia de cromatografia líquida de fase reversa. Para isso, foi adicionado ao sobrenadante concentrado 2% de acetonitrila e 0,065% de ácido trifluoroacético (TFA). A solução resultante foi clarificada por centrifugação a 13000 rpm por 15 min e filtração em membrana Millipore com poro de  $0,45\ \mu\text{m}$ . O sobrenadante foi aplicado a uma coluna C18 – 238TP (Vydac), com dimensão de 10 x 250 mm e partículas com diâmetro de  $5\ \mu\text{m}$ . A coluna foi equilibrada com 200 mL de solução de 2% de acetonitrila e 0,065% de TFA em água deionizada MiliQ (eluente A). A coluna foi eluída com fluxo de 5 mL/min com um gradiente de acetonitrila e 0,05% de TFA (eluente B). O perfil de eluição foi monitorado a 280 e 216 nm.

Para a coleta dos dados as frações puras de  $^{15}\text{N}/^{13}\text{C}$ - CP, analisadas através de espectrometria de massas, foram liofilizadas e ressuspendidas em 600  $\mu\text{L}$  de 10 mM tampão fosfato, pH 5,8 a uma concentração final de 0,7 mM e 5% ou 100% de  $\text{D}_2\text{O}$ .

### Ressonância magnética nuclear

Os experimentos de RMN foram adquiridos nos espectrômetros de Ressonância Magnética Nuclear Varian Inova de 500 MHz e 600 MHz do Centro de Biologia Molecular e Estrutural (CeBiMe/LNLS). Foram realizados experimentos 1D homonuclear, 2D e 3D heteronucleares de tripla ressonância utilizando-se as sequências de pulso do *software*

*Varian Inc. ProteinPack*. Todos os dados foram coletados a 20°C e processados em estações *Silicon Graphics Octane 2* utilizando-se os *softwares* NMRpipe/NMRview (Delaglio et al., 1995; Johnson & Blevins, 1994).

Os espectros monodimensionais foram coletados com janela espectral de 8000 Hz e 48 transientes com 16 K de pontos para cada indução de decaimento livre. A supressão da água foi feita através da pré-saturação.

Os experimentos bidimensionais  $^1\text{H}$ - $^{15}\text{N}$  HSQC (Kay et al., 1992) foram coletados com janela de 8000 Hz e 512 pontos complexos para  $^1\text{H}$  e 1725,25 Hz, com 128 incrementos para  $^{15}\text{N}$ .

A atribuição das ressonâncias da cadeia principal foi feita através da coleta e análise dos espectros 3D heteronucleares HNCA (Ikura et al., 1990), HN(CO)CA (Ikura et al., 1990), HNCO (Grzesiek & Bax, 1992a), CBCA(CO)NH (Grzesiek & Bax, 1992b), HNCACB (Wittekind & Mueller, 1993) e HNHA (Vuister & Bax, 1993). Os parâmetros e o tempo total de coleta de cada experimentos de tripla ressonância estão reunidos na Tabela 2.2.

**Tabela 2.2.** Parâmetros e tempo total de aquisição dos espectros de RMN.

	HNCA	HN(CO)CA	HNCO	CBCA(CO)NH	HNCACB
Janela espectral (Hz)					
$^1\text{H}$	8000,00	8000,00	8000,00	8000,00	8000,00
$^{15}\text{N}$	1275,25	1275,25	1275,25	1275,25	1275,25
$^{13}\text{C}$	3769,92	3769,92	3017,10	7540,78	7540,78
Número de pontos (complexos)					
$t_1$ ( $^1\text{H}$ )	43	43	48	44	43
$t_2$ ( $^{15}\text{N}$ )	100	100	48	48	64
$t_3$ ( $^{13}\text{C}$ )	512	512	512	512	512
<b>Transientes</b>	4	4	4	4	4
<b>Tempo de coleta</b>	~27 h	~27 h	~14 h	~26 h	~34 h

Para a atribuição das ressonâncias das cadeias laterais, a amostra foi liofilizada e ressuspendida em 600  $\mu\text{L}$  de  $\text{D}_2\text{O}$ . Foram coletados os experimentos heteronucleares 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC (Kay et al., 1993), 3D HCCH-TOCSY (Kay et al., 1993) (*Total Correlation Spectroscopy*) e HCCH-COSY (Patt, 1992) (*Correlated Spectroscopy*) para região de

carbonos alifáticos (região entre 5 e 80 ppm) e 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC (Kay et al., 1993) e 3D NOESY  $^{13}\text{C}$ -HSQC (*Nuclear Overhauser Enhancement Spectroscopy*) (Pascal et al., 1994) para região de carbonos aromáticos (região entre 90 e 140 ppm). Os parâmetros utilizados nesses experimentos e o tempo total de coleta estão reunidos na Tabela 2.3. Os experimentos  $^1\text{H}$ - $^{13}\text{C}$  HSQC foram coletados utilizando-se o modo *phase cycle* e *time constant*. O espectro de NOESY  $^{13}\text{C}$ -HSQC foi coletado com *mixing time* de 100 ms.

**Tabela 2.3.** Parâmetros e tempo total de aquisição dos espectros de RMN.

	$^{13}\text{C}$ -HSQC alifático	$^{13}\text{C}$ -HSQC aromático	HCCH- TOCSY	HCCH- COSY	NOESY $^{13}\text{C}$ -HSQC aromático
Janela espectral (Hz)					
$^1\text{H}$	8000,00	8000,00	8000,00	8000,00	8000,00
$^{13}\text{C}$	9324,4	6204,02	9324,01	9324,01	6203,96
$^1\text{H}$	**	**	8000,00	8000,00	8000,00
Número de pontos (complexos)					
$t_1$ ( $^1\text{H}$ )	1024	512	80	80	80
$t_2$ ( $^{15}\text{N}$ )	128	128	68	68	56
$t_3$ ( $^1\text{H}$ )	**	**	512	512	512
<b>Transientes</b>	32	32	16	16	4
<b>Tempo de coleta</b>	~3 h	~3 h	~110 h	~110 h	~29 h

### Chemical shift index

Em posse das atribuições das ressonâncias foi realizada a análise da estrutura secundária da proteína pelo método do CSI (Wishart & Nip, 1998). Este método é baseado na diferença dos valores de deslocamento químico do  $^1\text{H}$  ( $\text{H}^\alpha$ ) e  $^{13}\text{C}$  ( $\text{C}^\alpha$  e  $\text{C}^\beta$ ) de cada aminoácido da proteína com respeito aos valores tabelados para os mesmos em estrutura randômica ( $\Delta\delta = \delta_{\text{proteína}} - \delta_{\text{randômica}}$ ). Um grupo de quatro ou mais valores de CSI maiores que 0,1 ppm para  $\text{H}^\alpha$  e 0,5 ppm para  $^{13}\text{C}$  indica a presença de estrutura em  $\alpha$ -hélice e valores maiores que -0,1 ppm para  $\text{H}^\alpha$  e -0,5 ppm para  $^{13}\text{C}$  indica estruturas em folha- $\beta$ . O deslocamento químico da cadeia polipeptídica em conformação randômica usado como referência foi publicado por Wishart & Skykes (1994). Esses cálculos foram realizados automaticamente com o *software* NMRview (Johnson & Blevins, 1994).

### Constante de acoplamento $J_{HNH\alpha}$

As constantes de acoplamento  $J_{HNH\alpha}$  foram determinadas através do espectro de HNHA (Vuister & Bax, 1993) coletado com uma janela espectral de 8000,00 Hz para  $^1\text{H}$  em ambas as dimensões e de 1725,25 Hz para  $^{15}\text{N}$  e uma matrix com número de pontos de 64 (complexos) x 48 (complexos) x 512 (complexos) em  $t_1$  ( $^1\text{H}$ ),  $t_2$  ( $^{15}\text{N}$ ) e  $t_3$  ( $^1\text{H}$ ), respectivamente. O tempo total de coleta do experimento, utilizando-se 16 transientes, foi de 67 horas. A constante de acoplamento  $J_{HNH\alpha}$  está relacionada com a intensidade do pico de correlação ( $I_c$ ) e da diagonal ( $I_d$ ) de acordo com a Equação 2.6:

$$\frac{I_c}{I_d} = -\tan^2(2\pi J_{HH} \zeta) \quad (\text{Eq. 2.6})$$

onde  $2\zeta$  corresponde à duração do tempo de *dephasing delay*, que neste caso foi de 25 ms.

Como descrito por Vuister & Bax (1993), valores de  $J_{HNH\alpha}$  apresentam um erro sistemático de 11% para  $T_{1\text{sel}}$  de 100 ms devido a efeitos de relaxação. Assim as constantes obtidas experimentalmente apresentam um valor menor do que o valor da constante real. Para corrigir essa diferença e obter o valor real de  $J_{HNH\alpha}$ , as constantes obtidas experimentalmente foram multiplicadas por 1,1.

### Efeito nuclear Overhauser

Para a aquisição dos NOE, experimentos 3D  $^{15}\text{N}$ -NOESY-HSQC e  $^{13}\text{C}$ -NOESY-HSQC (Pascal et al., 1994) para região de carbonos alifáticos (região entre 5 e 80 ppm) e aromáticos (região entre 90 e 140 ppm) foram coletados utilizando-se *mixing time* de 100 ms, empregando uma amostra 0,7 mM CP em 10 mM tampão fosfato em  $\text{D}_2\text{O}$ , pH 5,8. Os espectros foram obtidos em espectrômetros de Varian Inova de 600 MHz da *Università degli Studi di Parma* (Itália) e/ou Bruker Avance 700 MHz da *Università degli Studi di Roma 'Tor Vergata'* (Itália). Os parâmetros utilizados nesses experimentos e o tempo total de coleta estão reunidos na Tabela 2.4.

**Tabela 2.4.** Parâmetros e tempo total de aquisição dos espectros de RMN.

	NOESY <sup>15</sup> N-HSQC	<sup>13</sup> C-HSQC Alifático	NOESY <sup>13</sup> C-HSQC Alifático	<sup>13</sup> C-HSQC Aromático	NOESY <sup>13</sup> C-HSQC Aromático
<b>Janela espectral (Hz)</b>					
<sup>1</sup> H	8000,0	8000,0	8000,0	8000,0	8000,0
<sup>15</sup> N/ <sup>13</sup> C	2035,0	9324,4	11901,0	6204,0	6204,0
<sup>1</sup> H	8000,0	**	8000,0	**	8000,0
<b>Número de pontos (complexos)</b>					
t <sub>1</sub> ( <sup>1</sup> H)	1024	1024	2048	512	80
t <sub>2</sub> ( <sup>15</sup> N/ <sup>13</sup> C)	96	128	180	128	56
t <sub>3</sub> ( <sup>1</sup> H)	180	**	272	**	512
<b>Transientes</b>	8	32	4	32	4
<b>Tempo de coleta</b>	~46 h	~3 h	~61 h	~3 h	~29 h

Distâncias interprotônicas aproximadas foram derivadas das intensidades dos sinais dos NOE. Para isso, os picos de correlação foram divididos em três grupos, relacionando-se as seguintes distâncias interatômicas: 1,8-2,8 Å para NOEs de intensidade forte, 1,8-4,0 Å para NOEs de intensidade média e 1,8-5,0 Å para NOEs de intensidade fraca.

### Constante dipolar residual

Foram determinadas as constantes dipolares residuais das seguintes ligações da cadeia principal da proteína: <sup>1</sup>H<sub>N</sub>-<sup>15</sup>N, <sup>1</sup>H<sub>N</sub>-<sup>13</sup>CO, <sup>15</sup>N-<sup>13</sup>CO, <sup>13</sup>C<sub>α</sub>-<sup>1</sup>H<sub>α</sub> e <sup>13</sup>C<sub>α</sub>-<sup>13</sup>CO. Para isso foram coletados espectros bidimensionais <sup>1</sup>H-<sup>15</sup>N HSQC modificados (Bax et al., 2001; Ottiger & Bax, 1999; Delaglio et al., 1991; de Alba & Tjandra, 2002), HNCOC com intensidade modulada (McFeeters et al., 2005) e o espectro 3D HACACO (Cicero et al., 2006) da proteína <sup>15</sup>N/<sup>13</sup>C- CP em meio isotrópico e parcialmente alinhado. Para a coleta dos experimentos em meio anisotrópico foi preparada 450 µL de solução da proteína <sup>15</sup>N/<sup>13</sup>C- CP a uma concentração final de 0,4 mM nas mesmas condições de tampão e pH da amostra em meio isotrópico, porém contendo 16 mg/mL do bacteriófago filamentosso Pf1 (Asla Labs) para induzir o alinhamento molecular (Hans et al., 1998). Os dados foram adquiridos em espectrômetros Bruker Avance 400 MHz e 700 MHz (*Università degli Studi*

di Roma 'Tor Vergata', Itália). Os parâmetros utilizados nesses experimentos e o tempo total de coleta estão reunidos na Tabela 2.5.

**Tabela 2.5.** Parâmetros de aquisição dos espectros de RMN para determinação das constantes dipolares residuais das respectivas ligações.

	$^1\text{H}_\text{N}-^{15}\text{N}$	$^1\text{H}_\text{N}-^{13}\text{CO}$	$^{15}\text{N}-^{13}\text{CO}$	$^{13}\text{C}_\alpha-^1\text{H}_\alpha$	$^{13}\text{C}_\alpha-^{13}\text{CO}$
<b>Janela espectral (Hz)</b>					
$^1\text{H}$	11261,0	11261,0	11261,0	9124,0	6410,0
$^{15}\text{N}/^{13}\text{C}$	2837,9	2837,9	2837,90	5281,0	1056,0
$^{13}\text{CO}$	**	**	**	2112,0	**
<b>Número de pontos (complexos)</b>					
$t_1 (^1\text{H})$	1024	4096	4096	1024	2048
$t_2 (^{15}\text{N}/^{13}\text{C})$	300	896	896	296	116
$t_3 (^{13}\text{CO})$	**	**	**	44	**
<b>Transientes</b>	24	16	16	16	16

## CAPÍTULO 3: Análises Estrutural e de Dinâmica da Cerato-Platanina

### Introdução

Neste terceiro capítulo serão apresentadas a estrutura e as propriedades dinâmicas da CP. Surpreendentemente, o enovelamento da CP mostrou-se similar à forma dimérica das quimiocinas interleucina-8 (IL-8) e peptídeo ativador de neutrófilos-2 (NAP-2) e ao complexo de histocompatibilidade principal de classe I (MHC-I), que possuem funções importantes na sinalização de resposta do sistema imune inato de animais contra patógenos. Além disso, os dados de dinâmica mostram que a CP possui uma estrutura compacta e com pouca flexibilidade.

Estudos de estabilidade da CP foram seguidos através da técnica espectroscópica de dicroísmo circular. A principal vantagem dessa técnica está na obtenção rápida e eficiente de informações estruturais, permitindo estudos de mudanças conformacionais de peptídeos e proteínas pela variação das condições do sistema.

Os resultados estão apresentados na forma de artigo científico que foi submetido a revista *ChemBioChem* e são os dados de maior relevância obtidos durante o desenvolvimento da tese.

Na tentativa de se encontrar possíveis funções da CP, ensaios de atividade de glucanase, celulase e poligalacturonase foram realizados para determinar se a proteína seria capaz de quebrar as ligações 1,3 e 1,4 de  $\beta$ -glucano,  $\beta$ -1,4-glicosídicas de celulose e  $\alpha$  1-4 de ácido galacturônico, respectivamente. Os ensaios de atividade enzimática foram realizados pelos nossos colaboradores Dra. Luigia Pazzagli e Prof. Dr. Gianni Cappugi do *Dipartimento di Scienze Biochimiche, Università di Firenze, Itália*. Também foram testadas as hipóteses da CP interagir com ligantes como cobre ou N-acetilglucosamina.

Por fim, os resíduos acessíveis ao solvente foram mapeados através do experimento de CLEANEX. Os estudos de interação com ligante e de acessibilidade ao solvente estão apresentados como resultados complementares em sequência ao artigo.

**ARTIGO 3:**

The solution structure of Cerato-Platanin reveals a new fold among  
the protein fungal elicitor

**Aline Lima de Oliveira<sup>1</sup>, Luigia Pazzagli<sup>2</sup>, Gianni Cappugi<sup>2</sup>, Aniello Scala<sup>3</sup>, Mariana Gallo<sup>4</sup>, Daniel Oscar Cicero<sup>4</sup>, Barbara Pantera<sup>2</sup>, Alberto Spisni<sup>5</sup>, Celso Eduardo Benedetti<sup>1</sup> and Thelma Aguiar Pertinhez<sup>5</sup>**

<sup>1</sup>*Center for Molecular and Structural Biology, Brazilian Synchrotron Light Laboratory, Campinas, Brazil*

<sup>2</sup>*Department of Biochemical Sciences, University of Florence, Florence, Italy*

<sup>3</sup>*Department of Agricultural Biotechnologies, University of Florence, Florence, Italy*

<sup>4</sup>*Department of Chemical Science and Technology, University of Rome "Tor Vergata", Rome, Italy*

<sup>5</sup>*Department of Experimental Medicine, University of Parma, Parma, Italy*

ChemBioChem (2008) *Submitted*

# The solution structure of cerato-platanin reveals a new fold among the protein fungal elicitors

Aline L. de Oliveira,<sup>[a]</sup> Luigia Pazzagli,<sup>[b]</sup> Gianni Cappugi,<sup>[b]</sup> Aniello Scala,<sup>[c]</sup> Mariana Gallo,<sup>[d]</sup> Daniel O. Cicero,<sup>[d]</sup> Barbara Pantera,<sup>[b]</sup> Alberto Spisni,<sup>\*,[e]</sup> Celso E. Benedetti,<sup>[a]</sup> and Thelma A. Pertinhez<sup>[e]</sup>

<sup>[a]</sup>Center for Molecular and Structural Biology, Brazilian Synchrotron Light Laboratory, Campinas, Brazil

<sup>[b]</sup>Department of Biochemical Sciences, University of Florence, Florence, Italy

<sup>[c]</sup>Department of Agricultural Biotechnologies, University of Florence, Florence, Italy

<sup>[d]</sup>Department of Chemical Science and Technology, University of Rome "Tor Vergata", Rome, Italy

<sup>[e]</sup>Department of Experimental Medicine, University of Parma, Parma, Italy

\*Correspondence to: Alberto Spisni, Università degli Studi di Parma, Via Volturno, 39, 43100, Parma, Italy, Tel. +39 0521033807; Fax +39 0521033802. E-Mail: [alberto.spisni@unipr.it](mailto:alberto.spisni@unipr.it)

**Supported by:** Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Conselho Nacional de Pesquisa e Tecnologia (CNPq), Brazil and MIUR, Italy

*Short title:* The solution structure of cerato-platanin

*Keywords:* Cerato-platanin, fungal elicitor, IL-8, NMR, structural biology.

*Footnote/abbreviation:* CP: recombinant CP protein corresponding to the wild-type from *Ceratocystis platani*, but including also a N-terminus EKR-tag derived from the cloning vector

## Abstract

Cerato-platanin (CP) is a protein elicitor secreted by the fungal pathogen *Ceratocystis platani*, the causal agent of the plane tree canker. CP elicits defence reactions in host and non-host plants; but its molecular mode of action and host targets are still unknown. CP does not share 3D structural similarity with any known avirulence proteins or elicitors from bacterial or fungal pathogens. Its NMR solution structure shows a rigid core formed by two parallel and five anti-parallel  $\beta$ -strands, assembled in a seven-stranded  $\beta$ -sheet that faces two short  $\alpha$ -helices. An additional short  $3_{10}$  helix in N-terminus is exposed to the solvent. Two disulfide bonds contribute to stabilize the overall structure that, as a result, exhibits thermal and pH stability.

Though the Pfam database indicates that CP ought to belong to the double Psi beta barrel glucanase clan, the protein has no glycosidase activity. Surprisingly, however, the CP fold resembles the one of proteins such as the human interleukins IL-8, NAP-2 and MHC-I, which play central roles in the innate immune response against animal pathogens. Overall, these results indicate a route that should lead to the comprehension of the mode of propagation of the pathogen in the plant.

## Introduction

Plants are often infected by pathogens with different lifestyles, thus they evolved diverse forms of host-microbe interactions.<sup>[1]</sup> The most studied plant defence mechanism involves the action of plant resistance proteins (R) that recognize pathogen factors encoded by the avirulence genes (Avr). In this type of interaction, the so-called gene-for-gene mechanism, the direct association of Avr and R proteins triggers an R-mediated secondary defence leading to a hypersensitive response: a localized plant cell death that effectively restricts pathogen growth.<sup>[2-4]</sup> Avr products have also been shown to interact with or modify non-R proteins, thus giving support to the “guard hypothesis”, which states that Avr proteins target host factors that are “guarded” by one or more R proteins.<sup>[5]</sup>

In addition to Avr recognition, plants have developed ways to sense other signalling molecules, called elicitors, which may be common to entire groups or classes of microorganisms. This mechanism of non-host resistance represents the most common form of plant resistance and it is the first line of defence against the majority of potential pathogens.<sup>[6,7]</sup> As a result, plants recognize microbial/pathogen-associated molecular patterns (MAMPs/PAMPs), comprising lipopolysaccharides, peptidoglycans, mannans, chitin, ergosterol and proteins relevant for the life cycle of the pathogen, by pattern recognition receptors (PRRs) that activate a primary defence response. These responses consist in cell wall reinforcements, accumulation of defence-related proteins, synthesis of phytoalexins and secondary metabolites, to mention some.<sup>[6,8,9]</sup> Besides PAMPs, plant pathogens may secrete an array of cell wall degrading enzymes that play diverse functions in pathogenicity but also elicit localized or systemic defence reactions in the host.<sup>[9,10]</sup> Finally, phytopathogenic fungi and oomycetes, such as *Cladosporium fulvum*, *Fusarium oxysporum*, *Melampsora lini*, *Magnaporthe grisea*, *Phytophthora sojae* and *Peronospora parasitica*, secrete several non-catalytic proteins, generally characterized by low molecular weight, which are also classified as PAMPs.<sup>[3,7,11,12]</sup> These non-catalytic proteins belong to five protein families including elicitors (PF00964), the class I (PF01185) and class II hydrophobins (PF06766),

the NIP-1 protein family (PF08995) and cerato-platanin (PF07249). These proteins affect plants acting as toxins or as suppressors of host defensive systems. For instance, they can modify the membrane permeability, elicit resistance-related events and induce necrosis and/or phytoalexins synthesis<sup>[3,12]</sup>

The cerato-platanin family (CPF) consists of several secreted proteins found not only in phytopathogenic fungi but also in a number of animal pathogens and non-pathogenic fungal species (Figure 1). These proteins have been shown to act primarily as virulence factors, such as the snodprot1 homolog MSP1 from the rice blast fungus *Magnaporthe grisea*, or as elicitors of defence responses in plants,<sup>[13-15]</sup> such as the proteins Sm1 and Epl1, from soilborne *Trichoderma virens* and *Hypocrea atroviride*, respectively.<sup>[16-18]</sup> Other members of the CPF, however, appear to act in animal-fungus interactions, showing allergenic properties and inducing strong immunological reactions. This is particularly the case of the antigen CS-Ag from *Coccidioides immitis* and Asp13 protein from *Aspergillum fumigatus*, both causative agents of human lung diseases.<sup>[19,20]</sup>

The founder member of the CPF is cerato-platanin, CP (Figure 1), a moderately hydrophobic protein that is secreted and localized in the cell wall of *Ceratocystis platani*, the causal agent of the canker stain disease of *Platanus acerifolia*.<sup>[15,21]</sup> CP is thought to play a role in the host-fungus interaction since it induces cell necrosis, phytoalexin synthesis and over-expression of defence related genes in host and non-host plants.<sup>[14,22,23]</sup> Although the biological function of the CP homologues MSP1, Sm1 and Epl1 have recently begun to be unveiled,<sup>[16-18]</sup> very little is still known about the molecular mechanisms by which CP and other uncharacterized members of the CPF interact with their hosts. Additionally, no structure-function relationship data is yet available for any of the CPF members. Here we report the NMR-derived solution 3D structure of *C. platani* CP, and we show that, surprisingly, the fold is remarkably similar to the one of human cytokines and MHC class I proteins involved in innate immunity.

## Results

Protein sequence alignments show that CP from *C. platani* shares significant similarities to fungi proteins only (Figure 1). All the identified proteins are currently classified as members of the CPF. In fact, several amino acid residues within the cerato-platanin domain are conserved among the protein sequences (Figure 1). Most significantly are the four cysteines, which form two disulfide bridges in CP.<sup>[21]</sup> They are invariant among all proteins and are located in regions of high conservation (Figure 1).

### The solution structure of CP

The far UV-CD spectrum indicates that the recombinant protein is folded. The broad negative band with a minimum at 212 nm and the intense positive band at 200 nm are suggestive of a protein with predominant  $\beta$ -sheet structure containing some helical elements (Figure 2A). In addition, the negative band centred at 230 nm together with the intense negative bands in the near-UV spectrum spanning the region 260-310 nm (Figure 2B) are consistent with the presence of disulfide bonds as well as of aromatic residues organized in a stable tertiary structure.

The good dispersion of the backbone amides in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum confirmed that the protein is folded and the complete resonance assignments have been previously deposited in the BMRB, entry 6990.<sup>[24]</sup>

DOSY experiments indicated that CP has a hydrodynamic radius of 19.5 Å, consistent with the existence of the protein in solution in the monomeric form.

The 3D structure of CP has been determined on the basis of 1250 non redundant NOE-derived distance restraints, a set of 197 RDCs, involving HN-N and  $\text{H}_\alpha$ - $\text{C}_\alpha$  inter-nuclear vectors, and 127 dihedral angles derived from chemical shift analyses using the TALOS program (Table I).<sup>[25]</sup> The structural statistics of the 20 lower energy models show good stereochemical features. The quality of the structural definition is highlighted by an overall RMSD of  $1.7 \pm 0.2$  Å.

Significantly, if we consider only the structured regions, the RMSD falls to  $0.6 \pm 0.1$  Å for backbone and  $1.02 \pm 0.09$  Å for the heavy atoms (Table I). The poor definition of the loops originates from a reduced numbers of NOEs available (Figure 3A).

The CP structure is composed of two parallel, Asp<sup>106</sup>-Tyr<sup>110</sup> ( $\beta$ 7) and Val<sup>4</sup>-Pro<sup>10</sup> ( $\beta$ 1), and five anti-parallel  $\beta$ -strands, Cys<sup>63</sup>-Ile<sup>68</sup> ( $\beta$ 4), Ile<sup>74</sup>-Arg<sup>77</sup> ( $\beta$ 5), Phe<sup>86</sup>-Asn<sup>89</sup> ( $\beta$ 6), Asn<sup>45</sup>-Gly<sup>48</sup> ( $\beta$ 3) and Gly<sup>38</sup>-Val<sup>40</sup> ( $\beta$ 2), assembled in a seven-stranded  $\beta$ -sheet which faces two anti-parallel  $\alpha$ -helical elements, residues Thr<sup>91</sup>-Gly<sup>98</sup> ( $\alpha$ 2) and Leu<sup>115</sup>-Ile<sup>119</sup> ( $\alpha$ 3) (Figure 3B). In addition, there is a short  $3_{10}$ -helix element, Leu<sup>30</sup>-Ala<sup>32</sup> ( $\alpha$ 1) located in the protein N-terminal region that, in spite of being rather hydrophobic, is exposed to the solvent. The protein presents also loops of variable sizes. The two longer loops, L1 (Ile<sup>11</sup>-Gly<sup>29</sup>) and L4 (Ile<sup>49</sup>-Thr<sup>62</sup>) connecting  $\beta$ 1- $\alpha$ 1 and  $\beta$ 3- $\beta$ 4, respectively, are linked by a disulfide bond (Cys<sup>23</sup>-Cys<sup>60</sup>). The second disulfide bond, Cys<sup>63</sup>-Cys<sup>118</sup>, connects  $\alpha$ 3 with  $\beta$ 4. The N- and C-terminal regions, residues Glu<sup>1</sup>-Arg<sup>3</sup> and Asn<sup>120</sup>-Asn<sup>123</sup>, turn out to be unstructured (Figure 3B).

Overall, the protein is rather compact as suggested by the H/D exchange experiments (Figure 4). In particular, the amide protons of Ser<sup>7</sup>-Ala<sup>13</sup>, Leu<sup>37</sup>-Val<sup>40</sup>, Val<sup>46</sup>-Ile<sup>49</sup>, Trp<sup>64</sup>-Ile<sup>68</sup>, Ile<sup>74</sup>-Asp<sup>80</sup>, Asn<sup>87</sup>-Asn<sup>89</sup> and Asn<sup>109</sup>-Val<sup>113</sup>, most of them belonging to elements of secondary structure, do not exchange even after one month. Exceptions are the short  $\alpha$ 1  $3_{10}$ -helix, located in the N-terminal region and part of  $\alpha$ 3, located at the C-terminal end of the protein.

Inspection of the surface hydrophobicity shows that, despite the high content in hydrophobic residues, there is not a net segregation of hydrophobic or charged residues on its surface (Figure 3C), as observed in the hydrophobins.<sup>[26]</sup> Instead there is a distribution of hydrophobic, hydrophilic and neutral patches.

## Backbone relaxation data and reduced spectral density analysis

Analysis of the backbone amides  $^{15}\text{N}$ -relaxation reveals that CP is rather rigid (Figure 5A). However, the N- and C-terminal residues, Gly<sup>54</sup>, Phe<sup>86</sup> and the stretch Ser<sup>99</sup>-Arg<sup>104</sup> exhibit either  $T_1$  or  $T_2$  and/or  $\{^1\text{H}-^{15}\text{N}\}$  NOE values significantly different from the average, suggesting they experience fast internal motion. Most of the residues belonging to the loops show relaxation and  $\{^1\text{H}-^{15}\text{N}\}$  NOE values comparable to those of the structured regions (Figure 5A). Nonetheless, in spite of their expected reduced mobility, their structural definition is, in general, limited (Figure 3A). This fact can be due to the solvent exposition of those regions that justifies the reduced number of NOEs that has been measured.

Reduced spectral density mapping of the  $^{15}\text{N}$ -relaxation measurements confirmed the previous interpretation. In fact, recognizing that  $J(0)$  reports slow motions (milli to microsecond) whereas  $J(0.87\omega_{\text{H}})$  is sensitive to fast internal motions (pico to nanosecond),<sup>[27]</sup> the large values of  $J(0)$  and small values of  $J(0.87\omega_{\text{H}})$  (Figure 5B) confirm that CP structure is characterized by a limited internal mobility. Exceptions are the residues and regions indicated before (Figure 5B), thus confirming a certain degree of structural instability in those protein stretches.

## CP stability

Protein stability has been analyzed as a function of pH, temperature and urea concentration. The CD spectra of CP at different pHs show that the protein maintains its conformation in a wide pH range (Figure 6A). On the other hand, heat denaturation shows that the protein is stable up to 60°C. In fact, CP unfolding is complete only at about 82°C and the melting temperature turns out to be 76°C (Figure 6B). The unfolding, however, is not reversible as highlighted by the different shape of the cooling curve and by the fact that only 61% of the secondary structure is recovered (Figure 6B).

Figure 6C shows that even at 8.0 M urea it is not possible to reach a plateau. Nonetheless, the  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum of CP in 8.0 M urea shows characteristics of a protein with a disordered and largely unfolded structure (Figure 6D, middle panel). Noteworthy, differently from thermal unfolding, when urea is removed by dialysis, the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum (Figure 6D, right panel) exhibits the same fingerprint of the native protein (Figure 6D, left panel), indicating that the chemical unfolding is reversible.

### **CP glycosidase activity**

According to the current classification of protein families reported in the Pfam database,<sup>[28]</sup> CPF belongs to the double Psi beta barrel (DPBB) glucanase clan. Nonetheless, CP does not show glycosidase activity even when it is assayed at high protein concentration (data not shown). Moreover, CP is not able to bind the tetramer of N-acetylglucosamine (data not shown), a typical feature of the barwin protein,<sup>[29]</sup> the closest CP sequence similarity within the DPBB family.

### **CP shows structural similarity but no sequence homology with cytokines and class I MHC proteins**

Structural analysis using the ProFunc server<sup>[30]</sup> revealed a good fold alignment with human proteins belonging to both the chemokine and class I major histocompatibility complex (MHC-I) families. The structured regions of CP can be superposed to the dimeric form of interleukin-8 (IL-8), to the neutrophil activating peptide-2 (NAP-2) and to the  $\alpha 1, \alpha 2$ -domain of MHC-I protein<sup>[31-33]</sup> (Figure 7). However, despite those fold similarities, CP does not share any sequence homology as well as any functional sequence motif with those proteins.<sup>[34,35]</sup>

## Discussion

Despite recent advances in the structural characterization of plant pathogens Avr and elicitor proteins, little is still known about the structural relationship between these effectors and their corresponding targets or cognate PRRs. In fact, to date, only a few pairs of effector/targets and PAMP/PRRs have been identified and structurally characterized.<sup>[3,12]</sup>

Although CP has been shown to elicit plant defence reactions in plants,<sup>[15,21,23]</sup> its biological function remains unknown. Nonetheless, recent data have shown that the CP homologues MSP1 (Figure 1), from the fungal pathogen *M. grisea*, is required for virulence,<sup>[16]</sup> while the Sm1 protein, from the bio-control fungus *T. virens*, is capable of inducing defence responses and systemic resistance in host and non-host plants.<sup>[17,36]</sup> Sm1, which lacks phytotoxic activity, triggers the production of reactive oxygen species and induces the expression of several defence-related genes.<sup>[36]</sup> Most significantly, oligomerization of Sm1 and its homolog Epl1 (Figure 1) determines the activity of these proteins as elicitors of systemic resistance.<sup>[18]</sup> Only the monomeric forms of Epl1 and Sm1 have the ability to induce plant defences, while dimers fail to protect plants against a pathogen.<sup>[18]</sup> Interestingly, CP and several other members of the CPF lack the sequence motifs which are involved in Sm1 and Epl1 dimerization and, in fact, CP exists as a monomer in solution, even at high concentrations. We can predict, therefore, that CP will maintain a high elicitor activity in the host.

Here we report, for the first time, the structure of the CP protein from *C. platani*, the founder of the CPF. According to Pfam database<sup>[28]</sup> the CPF belongs to the DPBB glucanase clan. However, CP has no glycosidase activity and it does not show any structural similarity with members of the DPBB clan, including barwin (1BW4), a plant protein that shares 50% sequence similarity with CP. More than that, no structural similarity has been observed between CP and any of the available 3D structures of Avr proteins and elicitors of bacterial and fungal pathogens. Instead, the CP fold is surprisingly similar to those of the human cytokines IL-8 (1ILQ), and NAP-2 (1TVX), as well as to the MHC-I proteins (1M6O).

IL-8 and NAP-2 are well characterized pro-inflammatory chemokines that promote leukocyte movement to sites of microbial infection, whereas class I MHC molecules mediate antiviral immune responses through binding and subsequent display of short viral peptides for T cell recognition.<sup>[37,38]</sup> Thus, a common feature between the cytokines and MHC-I is their central role in innate immunity against a variety of microorganisms and animal pathogens.

The structural resemblance between CP and these defence signalling molecules suggests that, in the host cells, CP targets might have structural similarities to the human IL-8/NAP-2 or MHC-I G proteins receptors. A search in the *Arabidopsis* genome with the human IL-8 receptor CXCR1<sup>[34]</sup> produced a few hits, including the hypothetical protein CAB40985 and the genomic sequence AT1G42595 which is flanked by the *mlo9* gene (*NP\_174980*), a homolog of the barley mildew resistance locus O (MLO). Noteworthy, homologues of barley MLO are the only family of seven-transmembrane proteins in plants to have a topology that is reminiscent of mammalian G-protein coupled receptors.<sup>[39]</sup>

When the search in the *Arabidopsis* genome was carried out using the human MHC-I receptors P43626, P43629, LIR-1 (*NP\_006660*) and LIR-2 (*NP\_001075106*), we found significant similarities with leucine-rich repeat receptor-like kinases (RLK), such as SRF7 (Q9LUL4),<sup>[40]</sup> and with putative disease-resistance proteins, including RPP5-like (*BAD43012*) and ATPERK1 (*NP\_189098*), a receptor kinase induced by wounding and fungal infection.<sup>[41]</sup> The C-terminal region of LIR-1 exhibits sequence similarities to an RLK protein that has been identified as a component of a plasma membrane micro-domain required for the interaction between the endophytic fungus *Piriformospora indica* and *Arabidopsis*.<sup>[42]</sup> Not surprisingly, most of these proteins are surface or intracellular receptors implicated in PAMPs or Avr recognition, a finding consistent with the hypothesis that CP, acting as an elicitor, should interact with such class of receptor proteins.

In CP, the regions of structural instability located in few exposed regions of the protein, and accessible from only one side of the molecule, involve the N- and C terminus and L7 that connects

$\alpha$ 2- $\beta$ 7: indeed, they could be the sites of CP with its receptors. In fact, regions with those conformational and dynamic features are often involved in protein-protein interaction processes.

Extending the search to other fungi, we found that the opportunist fungal pathogen *A. fumigatus*, which can cause deadly invasive infections in humans, secretes several allergens,<sup>[20]</sup> including a protein that is 83% similar to CP (Fig. 1). Although the role of *A. fumigatus* CP in the fungal pathogenicity has not yet been demonstrated, it is worth noting that *A. fumigatus* induces the synthesis of IL-8 and tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ), which seem crucial for the pro-inflammatory response against the fungus.<sup>[43]</sup> Such inflammatory response has been linked to the pathogenesis of invasive aspergillosis,<sup>[44]</sup> and a high level of IL-8 has been correlated to increased severity of the respiratory diseases<sup>[45]</sup> as well as to destructive responses in the lung, including epithelial barrier dysfunction.<sup>[46,47]</sup> Therefore, it is tempting to speculate that *A. fumigatus* CP, because of its fold similarity with IL-8, could contribute to the on set of an acute inflammatory response that would ultimately favour pathogen invasion.

In line with this notion is the fact that the medicinal fungus *Antrodia camphorata*, used in traditional Chinese medicine for treating cancer,<sup>[48]</sup> does produce an immunomodulatory protein, named Aca1, that is a homolog of *C. platani* and *A. fumigatus* CPs (Figure 1). Significantly, purified Aca1 (US patent n° 2005/0164931A1) induces TNF- $\alpha$  and  $\gamma$ -interferon (IFN- $\gamma$ ) production, both of which are known to stimulate IL-8 secretion and modulate the recruitment of lymphocytes in mycobacterial diseases,<sup>[49]</sup> as observed in *A. fumigatus* infection. Thus, the finding that CP has an IL-8-like fold seems not fortuitous and it may in fact represent another example of molecular mimicry involving IL-8 receptor.<sup>[50]</sup>

In conclusion, we feel that, our results while they contribute to the understanding of the role of this class of fungal elicitors and to increase the appreciation of their biological function in a number of fungal species that have diverse lifestyles, they indicate a route that should lead to the comprehension of the mode of propagation of the pathogen in the plant. Thus, they open the way to the identification of new targets to fight, in the safest way for humans, this plant disease.

## Experimental Section

### Expression and purification of cerato-platanin

The recombinant CP protein was expressed in *Pichia pastoris* and purified by reverse-phase liquid chromatography, as previously described.<sup>[51]</sup> The resulting 123-residue protein includes a non-native N-terminal Glu<sup>1</sup>-Lys<sup>2</sup>-Arg<sup>3</sup> tag and it is functionally indistinguishable from native CP.<sup>[51]</sup>

For isotopic labelling, CP was expressed in minimum medium supplemented with <sup>15</sup>N-ammonium sulphate and, in the case of double labelled proteins with <sup>13</sup>C-glycerol, and <sup>13</sup>C-methanol, as described previously.<sup>[24]</sup> The purified protein was estimated to be ~95% pure by SDS-PAGE and mass spectrometry.

### Circular dichroism spectroscopy

Circular dichroism (CD) measurements were performed on a JASCO J-810 spectropolarimeter equipped with a Peltier temperature control system. The far-UV spectra were recorded at 20°C in a 1 mm optical path length cell in the wavelength range 190-260 nm with a bandwidth of 1 nm. The near-UV spectra were recorded at 20°C in a 10 mm optical path length cell in the wavelength range 250-350 nm. For each measurement, the average of four spectra was taken to improve the signal to noise ratio and each spectrum was corrected by subtracting its blank. Typically, 15 µM of purified CP solution in 10 mM sodium phosphate buffer, pH 5.8, was used for the CD measurements. The stability of the protein at different pH conditions was also assayed. The pH values were adjusted by adding few drops of concentrated HCl or NaOH.

Thermal denaturation curves were recorded over a 15-90°C temperature interval, following the intensity of the CD signal at 212 nm. The temperature range was monitored at a temperature scan rate of 1°C.min<sup>-1</sup>, using 1 mm optical path length and 1 nm bandwidth. Reversibility of

denaturation was tested by an inverse temperature scanning immediately after the end point temperature (90°C) was reached and using the same temperature scan rate.

Urea titration experiments were performed using, for each urea concentration, a separately prepared protein solution left overnight at 25°C for equilibration. Urea induced denaturation was monitored by CD spectroscopy following the signal intensity at 230 nm and using a 1 mm optical path length.

## **NMR measurements**

NMR samples contained 0.7 mM of double-labelled  $^{13}\text{C}/^{15}\text{N}$  CP in 10 mM phosphate buffer, pH 5.8, 0.05%  $\text{NaN}_3$  and 5% or 100%  $\text{D}_2\text{O}$  (Cambridge Isotope Laboratories Inc). Two and three-dimensional NMR experiments were collected at 20°C on spectrometers Varian Inova AS 500 MHz or 600 MHz at the CeBiME-LNLS, Brazil, on a Varian Inova AS 600 MHz at the CIM, University of Parma, Italy or on a Bruker Avance 700 MHz at the University of Rome “Tor Vergata”, Italy. All NMR spectrometers were equipped with a  $z$  pulsed-field gradient unit and a triple resonance probe. NOE-derived distance restraints were obtained from 3D  $^1\text{H}$ - $^{15}\text{N}$ -NOESY<sup>[52]</sup> and 3D  $^1\text{H}$ - $^{13}\text{C}$ -NOESY, separately optimized for aliphatic and aromatic side chains<sup>[53]</sup> collected with 100 ms mixing times.

Approximate interproton distances were derived and the corresponding restraints subdivided into three groups: 1.8-2.8 Å for strong, 1.8-4.0 Å for medium and 1.8-5.0 Å for weak NOEs. IPAP-HSQC<sup>[54]</sup> and 3D HACACO experiments<sup>[55]</sup> were used to measure the HN-N and H $\alpha$ -C $\alpha$  RDCs, respectively. To induce molecular alignments for RDC measurements, the protein was mixed with filamentous phage Pf1 (Asla Labs) to a final concentration of 16 mg.mL<sup>-1</sup>. All spectra were processed and analyzed using the NMRView/NMRPipe softwares.<sup>[56,57]</sup>

DOSY experiments were carried out using the BPPSTE (bipolar pulse pairs stimulated echo) method.<sup>[58]</sup> The duration of the total diffusion-phase encoding gradient pulse was 2 ms, the diffusion delay was 0.05 s and the minimum gradient strength was set to 0.3 Gauss/cm. Diffusion coefficients

were measured using 0.2 mM CP in 10 mM phosphate buffer, pH 5.8 and 0.02% dioxane was used as a radius standard.<sup>[59]</sup>

## Structure calculation

Structure calculation and refinement were performed with XPLOR-NIH<sup>[60]</sup> using a simulated annealing protocol. The structures were also refined against a database of Ramachandran plot dihedral angles. Backbone dihedral restraints ( $\phi$  and  $\psi$  angles) were derived from chemical shift analyses using the TALOS program.<sup>[25]</sup> A set of 93 HN-N and 104 H $_{\alpha}$ -C $_{\alpha}$  residual dipolar couplings were included in the refinement. Fitting of the observed RDCs to a preliminary structure, using the singular value decomposition (SVD) method<sup>[61]</sup> with the PALES program,<sup>[62]</sup> provided an initial guess of the magnitude and orientation of the molecular alignment tensor. Subsequent optimization during the simulated annealing process led to the following final values for the N-H normalized magnitude of the RDC tensor and rhombicity, respectively:  $Da^{NH} = -18.20$  Hz and  $R = 0.66$ .

The ensemble of the 20 lower energy structures (from a total of 500 calculated structures), with no residues in the disallowed region of Ramachandran plot, was chosen to represent the protein solution 3D structure and, together with the constraints list, were deposited in the Protein Data Bank (2K8W). The programs AQUA and PROCHECK<sup>[63]</sup> were used to analyze the structures. Table I reports the structural statistics. To carry out a complete cross-validation of RDCs, a series of simulated annealing calculations were performed, each one of them lacking 10% of the RDCs (randomly chosen from the complete dataset.<sup>[64]</sup> By means of the SVD method the missing RDCs in each run were back-calculated to determine how well each RDC could be predicted.

## Hydrogen exchange measurements

Proton-deuterium (H/D) exchange measurements were performed with a lyophilized <sup>15</sup>N-labeled sample dissolved in 100% D<sub>2</sub>O. The loss of intensity of labile protons was monitored by recording a series of <sup>1</sup>H-<sup>15</sup>N HSQC spectra<sup>[65]</sup> at 600 MHz. The elapsed time between dissolving the

lyophilized protein and collecting the first  $^1\text{H}$ - $^{15}\text{N}$ -HSQC experiment was 37 min and a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded every 23 min for 30 days.

## Backbone $^{15}\text{N}$ -relaxation measurements and analysis

Longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation times as well as  $\{^1\text{H}$ - $^{15}\text{N}\}$  NOEs of backbone amides were determined using experiments acquired at 500 MHz with a  $^{15}\text{N}$ -labelled CP sample.<sup>[66]</sup> The  $\{^1\text{H}$ - $^{15}\text{N}\}$  NOE values were taken as the ratio of the peak volumes of the experiments recorded with and without  $^1\text{H}$  saturation using a saturation time of 3 s and a recycle delay of 2 s. Relaxation delays of 10, 210, 510, 710, 1010, 1310, 1510 and 1710 ms were employed for  $T_1$  measurements, and 10, 50, 90, 130, 170, 210, 250 and 290 ms for  $T_2$  measurements.  $T_1$  and  $T_2$  values were extracted by fitting the peak intensities to a single exponential decay function in NMRView.<sup>[56]</sup> The uncertainties of peak intensities were evaluated using the standard deviation of the spectral noise measured in a region free of cross peaks. The reduced spectral density mapping analysis was performed according to the equations described by Farrow et al. (1995):<sup>[27]</sup>

$$\sigma_{\text{NH}} = R_1(\text{NOE} - 1)(\gamma_{\text{N}}/\gamma_{\text{H}})$$

$$J(0.87\omega_{\text{H}}) = 4\sigma_{\text{NH}}/(5d^2)$$

$$J(\omega_{\text{N}}) = (4R_1 - 5\sigma_{\text{NH}})/(3d^2 + 4c^2)$$

$$J(0) = (6R_2 - 3R_1 - 2.72\sigma_{\text{NH}})/(3d^2 + 4c^2)$$

where  $d = (\mu_0 h \gamma_{\text{N}} \gamma_{\text{H}} / 8\pi^2)(r_{\text{NH}}^{-3})$  and  $c = \omega_{\text{N}} \Delta\sigma / \sqrt{3}$ .  $\mu_0$  is the permeability of free space,  $h$  is Planck's constant,  $\gamma_{\text{N}}$  and  $\gamma_{\text{H}}$  are the gyromagnetic ratios of the  $^1\text{H}$  and  $^{15}\text{N}$  nuclei, respectively;  $r_{\text{NH}} = 1.02 \text{ \AA}$  is the average amide bond length,  $\omega_{\text{N}}$  is the  $^{15}\text{N}$  Larmor frequency and  $\Delta\sigma = -160 \text{ ppm}$  is the chemical shift anisotropy for  $^{15}\text{N}$ .

## Glycosidase assay

Endo-1,3- $\beta$ -glucanase, polygalacturonase and cellulase activity were determined using as substrates laminarin, polygalacturonic acid and carboxymethyl-cellulose, respectively, according to

the methods reported in literature.<sup>[67-69]</sup> The released reducing sugars were detected using the PAHBAH procedure<sup>[70]</sup> with glucose as a standard. Enzymes (laminarinase from *Trichoderma* sp., pectinase from *Aspergillus niger* and cellulase from *Trichoderma reesei*) and substrates blanks were also included. Chitinase activity was detected at 575 nm using chitin azure as a substrate (chitinase from *Trichoderma viride* was used as positive control).

The binding of the  $\beta$ -(1-4) tetramer of N-acetylglucosamine (0.5 mM) has been studied by CD spectroscopy and chemical shift perturbation in  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum obtained after addition of 0.54 mM  $^{15}\text{N}$ -labelled CP.

## Acknowledgements

This work was partially supported by FAPESP grants (NMR project 99/11030-9, TAP 00/02026-7, Smolbnet 00/10266-8), Brazil, and MIUR (PRIN 2005 n° 20050717.5-002), Italy. ALO received fellowships from FAPESP (03/12418-8) and CAPES (BEX3666/05-0), whereas CEB received a fellowship from CNPq.

We acknowledge that part of the CD and NMR experiments have been carried out using equipment of the CIM of the University of Parma, Italy.

## References

- [1] P. J. de Wit, *Cell. Mol. Life Sci.* **2007**, *64*, 2726-2732.
- [2] H. H. Flor, *Annual Review of Phytopathology* **1971**;9, 275-296.
- [3] S. T. Chisholm, G. Coaker, B. Day, B. J. Staskawicz, *Cell* **2006**, *124*, 803-814.
- [4] P. N. Dodds, G. J. Lawrence, A. M. Catanzariti, T. Teh, C. I. Wang, M. A. Ayliffe, B. Kobe, J. G. Ellis, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8888-8893.
- [5] R. Marathe, S. P. Dinesh-Kumar, *Mol. Cell* **2003**, *11*, 284-286.
- [6] K. S. Mysore, C. M. Ryu, *Trends Plant Sci.* **2004**, *9*, 97-104.
- [7] C. Zipfel, G. Felix, *Curr. Opin. Plant Biol.* **2005**, *8*, 353-360.
- [8] J. G. D. Jones, J. L. Dangl, *Nature* **2006**, *444*, 323-329.
- [9] T. Nürnberger, V. Lipka, *Mol. Plant Pathol.* **2005**, *6*, 335-345.
- [10] N. Juge, *TRENDS Plant Sci.* **2006**, *11*, 359-367.
- [11] P. Skamnioti, C. J. Ridout, *Mol. Plant Pathol.* **2005**, *6*, 551-559.
- [12] C. Zipfel, *Curr. Opin. Immunol.* **2008**, *20*, 10-16.
- [13] L. M. Wilson, A. Idnurm, B. J. Howlett, *Mol. Plant Pathol.* **2002**, *3*, 487-493.
- [14] A. Scala, L. Pazzagli, C. Comparini, A. Santini, S. Tegli, G. Cappugi, *J. Plant Pathol.* **2004** *86*, 23-29.
- [15] L. Pazzagli, B. Pantera, L. Carresi, C. Zoppi, T. A. Pertinhez, A. Spisni, S. Tegli, A. Scala, G. Cappugi, *Cell Biochem. Biophys.* **2006**, *44*, 512-521.
- [16] J. S. Jeong, T. K. Mitchell, R. A. Dean, *FEMS Microbiol. Lett.* **2007**, *273*, 157-165.
- [17] S. Djonović, W. A. Vargas, M. V. Kolomiets, M. Horndeski, A. Wiest, C. M. Kenerley, *Plant Physiol.* **2007**, *145*, 875-889.
- [18] W. A. Vargas, S. Djonović, S. A. Sukno, C. M. Kenerley, *J. Biol. Chem.* **2008**, *283*, 19804-19815.
- [19] S. Pan S, G. T. Cole, *Infect. Immun.* **1995**, *63*, 3994-4002.

- [20] A. Rementeria, N. López-Molina, A. Ludwig, A. B. Vivanco, J. Bikandi, J. Pontón, J. Garaizar, *Rev. Iberoam. Micol.* **2005**, *22*, 1-23.
- [21] L. Pazzagli, G. Cappugi, G. Manao, G. Camici, A. Santini, A. Scala, *J. Biol. Chem.* **1999**, *274*, 24959-24964.
- [22] S. Boddi, C. Comparini, R. Calamassi, L. Pazzagli, G. Cappugi, A. Scala, *FEMS Microbiol.Lett.* **2004**, *233*, 341-346.
- [23] F. Fontana, A. Santini, M. Salvini, L. Pazzagli, G. Cappugi, A. Scala, M. Durante, R. Bernardi, *J. Plant Pathol.* **2008**, *90*, 295-306.
- [24] A. L. Oliveira, L. Pazzagli, B. Pantera, G. Cappugi, C. E. Benedetti, A. Spisni, T. A. Pertinhez, *J. Biomol. NMR* **2006**, *36*, 50.
- [25] G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* **1999**, *13*, 289-302.
- [26] A. H. Kwan, R. D. Winefield, M. Sunde, J. M. Matthews, R. G. Haverkamp, M. D. Templeton, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3621-3626.
- [27] N. A. Farrow, O. Zhang, A. Szabo, D. A. Torchia, L. E. Kay, *J. Biomol. NMR* **1995**, *6*, 153-162.
- [28] R. D. Finn, J. Tate, J. Mistry, P. C. Coghill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer, A. Bateman, *Nucleic Acids Res.* **2008**, *36*, D281-D288.
- [29] S. Ludvigsen, F. M. Poulsen, *Biochemistry* **1992**, *31*, 8783-8789.
- [30] R. A. Laskowski, J. D. Watson, J. M. Thornton, *Nucleic Acids Res.* **2005**, *33*, W89-W93.
- [31] C. Eigenbrot, H. B. Lowman, L. Chee, D. R. Artis, *Proteins* **1997**, *27*, 556-566.
- [32] M. G. Malkowski, J. B. Lazar, P. H. Johnson, B. F. P. Edwards, *J. Mol. Biol.* **1997**, *266*, 367-380.
- [33] H. D. Hickman-Miller, W. Bardet, A. Gilb, A. D. Luis, K. W. Jackson, D. I. Watkins, W. H. Hildebrand, *J. Immunol.* **2005**, *175*, 367-375.
- [34] N. J. Skelton, C. Quan, D. Reily, H. Lowman, *Structure* **1999**, *7*, 157-168.

- [35] W. A. Macdonald, A. W. Purcell, N. A. Mifsud, L. K. Ely, D. S. Williams, L. Chang, J. J. Gorman, C. S. Clements, L. Kjer-Nielsen, D. M. Koelle, S. R. Burrows, B. D. Tait, R. Holdsworth, A. G. Brooks, G. O. Lovrecz, L. Lu, J. Rossjohn, J. McCluskey, *J. Exp. Med.* **2003**, *198*, 679-691.
- [36] S. Djonović, M. J. Pozo, L. J. Dangott, C. R. Howell, C. M. Kenerley, *Mol. Plant. Microbe Interact.* **2006**, *19*, 838-853.
- [37] J. Koch, R. Tampé, *Cell. Mol. Life Sci.* **2006**, *63*, 653-662.
- [38] Y. Kobayashi, *Front Biosci.* **2008**, *13*, 2400-2407.
- [39] A. Devoto, H. A. Hartmann, P. Piffanelli, C. Elliott, C. Simmons, G. Taramino, C. S. Goh, F. E. Cohen, B. C. Emerson, P. Schulze-Lefert, R. Panstruga, *J. Mol. Evol.* **2003**, *56*, 77-88.
- [40] B. Eyüboğlu, K. Pfister, G. Haberer, D. Chevalier, A. Fuchs, K. F. Mayer, K. Schneitz, *BMC Plant Biol.* **2007**, *7*, 16.
- [41] N. F. Silva, D. R. Goring, *Plant Mol. Biol.* **2002**, *50*, 667-685.
- [42] B. Shahollari, J. Vadassery, A. Varma, R. Oelmüller, *Plant J.* **2007**, *50*, 1-13.
- [43] L. Y. Chiang, D. C. Sheppard, F. N. Gravelat, T. F. Patterson, S. G. Filler, *Infect. Immun.* **2008**, *76*, 3429-3438.
- [44] I. Shalit, D. Halperin, D. Haite, A. Levitov, J. Romano, N. Osherov, I. Fabian, *J. Antimicrob. Chemother.* **2006**, *57*, 230-235.
- [45] S. Chollet-Martin, P. Montravers, C. Gibert, C. Elbim, J. M. Desmonts, J. Y. Fagon, M. A. Gougerot-Pocidalo, *Infect. Immun.* **1993**, *61*, 4553-4559.
- [46] J. Wessely-Szponder, *Acta Vet. Hung.* **2008**, *56*, 187-196.
- [47] H. Li, P. Liu, J. Cepeda, D. Fang, R. B. Easley, B. A. Simon, L. Q. Zhang, S. Q. Ye, *J. Inflamm. (Lond).* **2008**, *5*, 15.
- [48] Y. C. Hseu, F. Y. Wu, J. J. Wu, J. Y. Chen, W. H. Chang, F. J. Lu, Y. C. Lai, H. L. Yang, *Int. Immunopharmacol.* **2005**, *5*, 1914-1925.
- [49] Z. Hasan, A. Mahmood, S. Zafar, A. A. Khan, R. Hussain, *Int. J. Lepr. Other. Mycobact. Dis.* **2004**, *72*, 35-44.

- [50] S. K. Ahuja, J. L. Gao, P. M. Murphy, *Immunol. Today* **1994**, *15*, 281-287.
- [51] L. Carresi, B. Pantera, C. Zoppi, G. Cappugi, A. L. Oliveira, T. A. Pertinhez, A. Spisni, A. Scala, L. Pazzagli, *Protein Expr. Purif.* **2006**, *49*, 159-167.
- [52] O. Zhang, L.E. Kay, J. P. Olivier, J. D. Forman-Kay, *J. Biomol. NMR* **1994**, *4*, 845-858.
- [53] D. R. Muhandiram, N. A. Farrow, G. Xu, S. H. Smallcombe, L. E. Kay, *J. Mag. Reson.* **1993**, *102*, 314-321.
- [54] M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* **1998**, *131*, 373-378.
- [55] D. O. Cicero, G. M. Contessa, M. Paci, R. Bazzo, *J. Magn. Reson.* **2006**, *180*, 222-228.
- [56] B. A. Johnson, R. A. Blevins, *J. Biomol. NMR* **1994**, *4*, 603-614.
- [57] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277-293.
- [58] A. R. Waldeck, P. W. Kuchel, A. J. Lennon, B. E. Chapman, *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *30*, 39-68.
- [59] J. A. Jones, D. K. Wilkins, L. J. Smith, C. M. Dobson, *J. Biomol. NMR* **1997**, *10*, 199-203.
- [60] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65-73.
- [61] J. A. Losonczi, M. Andrec, M. W. F. Fischer, J. H. Prestegard, *J. Magn. Reson.* **1999**, *138*, 334-342.
- [62] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 3791-3792.
- [63] R. A. Laskowski, J. A. C. Rullman, M. W. MacArthur, R. Kaptein, J. M. Thornton, *J. Biomol. NMR* **1996**, *8*, 477-486.
- [64] G. M. Clore, D. S. Garrett, *J. Am. Chem. Soc.* **1999**, *121*, 9008-9012.
- [65] L. E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* **1992**, *114*, 10663-10665.
- [66] R. Ishima, D. A. Torchia, *Nat. Struct. Biol.* **2000**, *7*, 740-743.
- [67] Y. Gueguen, W. G. B. Voorhorst, J. van der Oost, W. M. de Vos, *J. Biol. Chem.* **1997**, *272*, 31258-31264.

- [68] F. Cervone, M. G. Hahn, G. De Lorenzo, A. Darvill, P. Albersheim, *Plant Physiol.* **1989**, *90*, 542-548.
- [69] T. M. Wood, K. M. Bhat, *Methods Enzymol.* **1988**, *160*, 87-112.
- [70] W. S. York, A. G. Darvill, M. McNeil, T. I. Stevenson, P. Albersheim, *Methods Enzymol.* **1985**, *118*, 3-40.

**Table I.** Experimental restraints and structural statistics of the 20 models of CP structure.

Number of experimental restrains	1609
Distance restraints from NOEs	1250
Short Range	451
Medium range	415
Long range	384
Hydrogen bond distance restraints <sup>[a]</sup>	35
Dihedral angle restraints	127
Phi constrains	66
Psi constrains	61
H-N Residual dipolar coupling constants	93
H $\alpha$ -C $\alpha$ Residual dipolar coupling constants	104
Average number of restraints per residue	13.08
XPLOR-NIH energie (kcal/mol) <sup>[b]</sup>	
<i>E</i> <sub>total</sub>	-417 $\pm$ 55
<i>E</i> <sub>bond</sub>	37 $\pm$ 4
<i>E</i> <sub>angle</sub>	487 $\pm$ 23
<i>E</i> <sub>improper</sub>	90 $\pm$ 6
<i>E</i> <sub>vdw</sub>	111 $\pm$ 8
<i>E</i> <sub>dih</sub>	2.6 $\pm$ 0.9
<i>E</i> <sub>rama</sub>	-1310 $\pm$ 35
<i>E</i> <sub>noe</sub>	64 $\pm$ 7
<i>E</i> <sub>sani</sub>	101 $\pm$ 7
RMS deviation from experimental restrains	
Average H-N RDC constant violation (Hz)	0.97 $\pm$ 0.06
Average H $\alpha$ C $\alpha$ RDC constant violation (Hz) <sup>[c]</sup>	0.71 $\pm$ 0.03
<i>R</i> <sub>free</sub>	0.33 $\pm$ 0.06
Ramachandran Analysis	
Residues in most favored regions (%)	71 $\pm$ 3
Residues in additionally allowed regions (%)	26 $\pm$ 3
Residues in generously allowed regions (%)	3 $\pm$ 2
Residues in disallowed regions (%)	0 $\pm$ 0
RMSD (Å)	
All	Backbone - Heavy-atoms
All	1.7 $\pm$ 0.2      2.6 $\pm$ 0.7
Structured regions <sup>[d]</sup>	0.6 $\pm$ 0.1      1.02 $\pm$ 0.09

<sup>[a]</sup>HN-O and N-O distances were constrained to 2.1  $\pm$  0.5 Å and 3.0  $\pm$  0.5 Å, respectively.

<sup>[b]</sup>Final values for the force constants used for the experimental constraints are:  $\kappa_{noe}$ =30.0;  $\kappa_{dih}$ =10.0;  $\kappa_{RDC}$ = 0.5.

<sup>[c]</sup>Normalized to the H-N values.

<sup>[d]</sup>Values refer to residues: 4-10, 30-32, 38-40, 45-48, 63-68, 74-77, 86-89, 91-98, 106-110 and 115-119.

## Figure Legends

**Figure 1.** Sequence alignment of CPF members from different fungal species, generated by ClustalW (<http://www.ebi.ac.uk/clustalW>). Invariable residues are shaded and marked with asterisks. The additional sequence “EKR” at the N-terminal of the recombinant CP from *C. platani* (P81702) is underlined and the cysteines involved in disulfide bridges are linked. The N-terminal secretion signal sequences of all proteins were removed to optimize the alignment of CP with other CP variants from *Ceratocystis fimbriata*, CF-COF (ABM63509), CF-MANG (ABM63508), CF-THEO (ABM63511), CF-IPO (ABM63510), CF-CRO (ABM63507), and with related proteins from *Ceratocystis populicola* (ABM63506), *Sclerotinia sclerotiorum* (XP\_001588549), *Botryotinia fuckeliana* (XP\_001559499), *Hypocrea atroviridis* (ABE73692), *Trichoderma virens* (AAZ80388), *Trichoderma antroviride* (CAL80754), *Gibberella zeae-1* (XP\_391381), *Neurospora crassa* (XP\_958708), *Podospora anserine* (XP\_001911154), *Magnaporthe grisea* (XP\_359969), *Gibberella zeae-2* (AAV83791), *Cochliobolus lunatus* (AAQ87930), *Phaeosphaeria nodorum* (XP\_001803929), *Leptosphaeria maculans* (AAM33130), *Botryotinia fuckeliana* (XP\_001552257), *Coprinopsis cinerea* (XP\_001839821), *Antrodia camphorata* (AAT11911), *Aspergillus clavatus* (XP\_001275544), *Aspergillus fumigatus* (XP\_755595), *Neosartorya fischeri* (XP\_001260717), *Aspergillus terreus* (XP\_001208607), *Aspergillus oryzae* (XP\_001821825), *Aspergillus niger* (XP\_001399295), *Aspergillus nidulans* (XP\_663877), *Coccidioides immitis* (XP\_001247410) and *Coccidioides posadani* (Q00398). The secondary structure elements and loops determined in the 3D structure of CP are shown.

**Figure 2.** Far-UV (A) and near-UV (B) circular dichroism spectra of recombinant CP at 15  $\mu$ M in 10 mM sodium phosphate buffer, pH 5.8 at 20°C.

**Figure 3.** The three-dimensional structure of CP. (A) Superimposition of the backbone atoms of 20 lower energy models. (B) Ribbon representation of the minimum energy structure. The disulfide bonds are marked in red. (C) Surface of CP with hydrophilic and hydrophobic areas shown in blue and red, respectively.

**Figure 4.** Solvent accessibility of backbone amide protons compared to the protein secondary structure. Plots of the persistence time of the amides signals in the H/D exchange experiment vs. the protein residues.

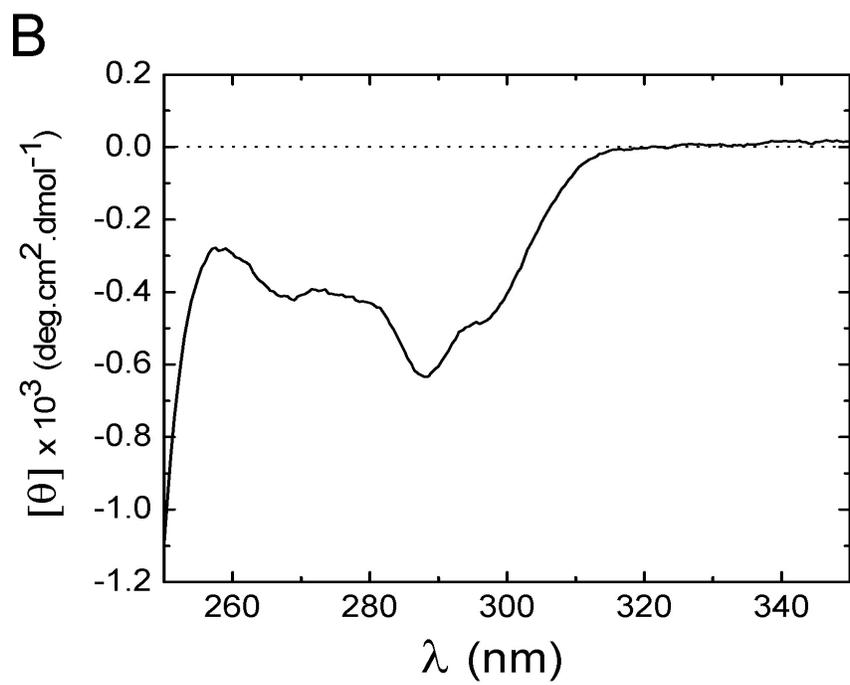
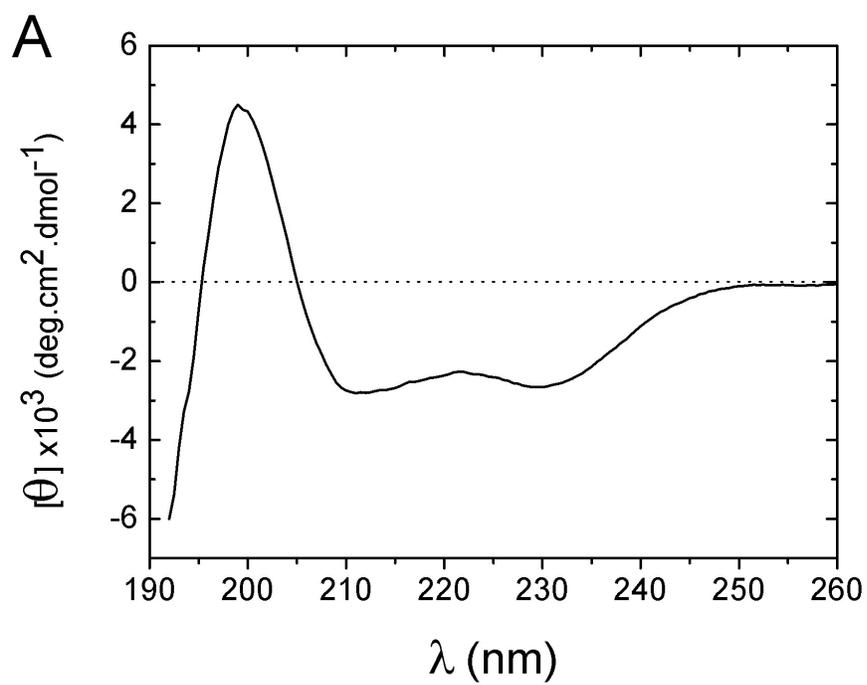
**Figure 5.** Backbone relaxation analysis of CP. (A) Plots of  $^1\text{H}$ - $^{15}\text{N}$  NOE (upper panel),  $T_1$  (middle panel) and  $T_2$  (lower panel) vs. the protein residues. (B) Spectral densities of backbone motions vs. the protein residues.  $J(0)$  (upper panel),  $J(\omega_N)$  (middle panel) and  $J(0.87\omega_H)$  (lower panel).

**Figure 6.** Characterization of CP stability. (A) Far-UV CD spectra of CP at different pH conditions. (B) Temperature-induced denaturation (filled squares) and renaturation (opened circles) curves of CP. The curves were recorded over a 15–90°C temperature interval, following the CD signal at 212 nm. (C) Urea denaturation curve obtained by monitoring the rate  $[\theta]_0/[\theta]$  at 230 nm. (D)  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of CP in the absence (left panel), presence (middle panel) and after dialysis (right panel) of 8.0 M urea.

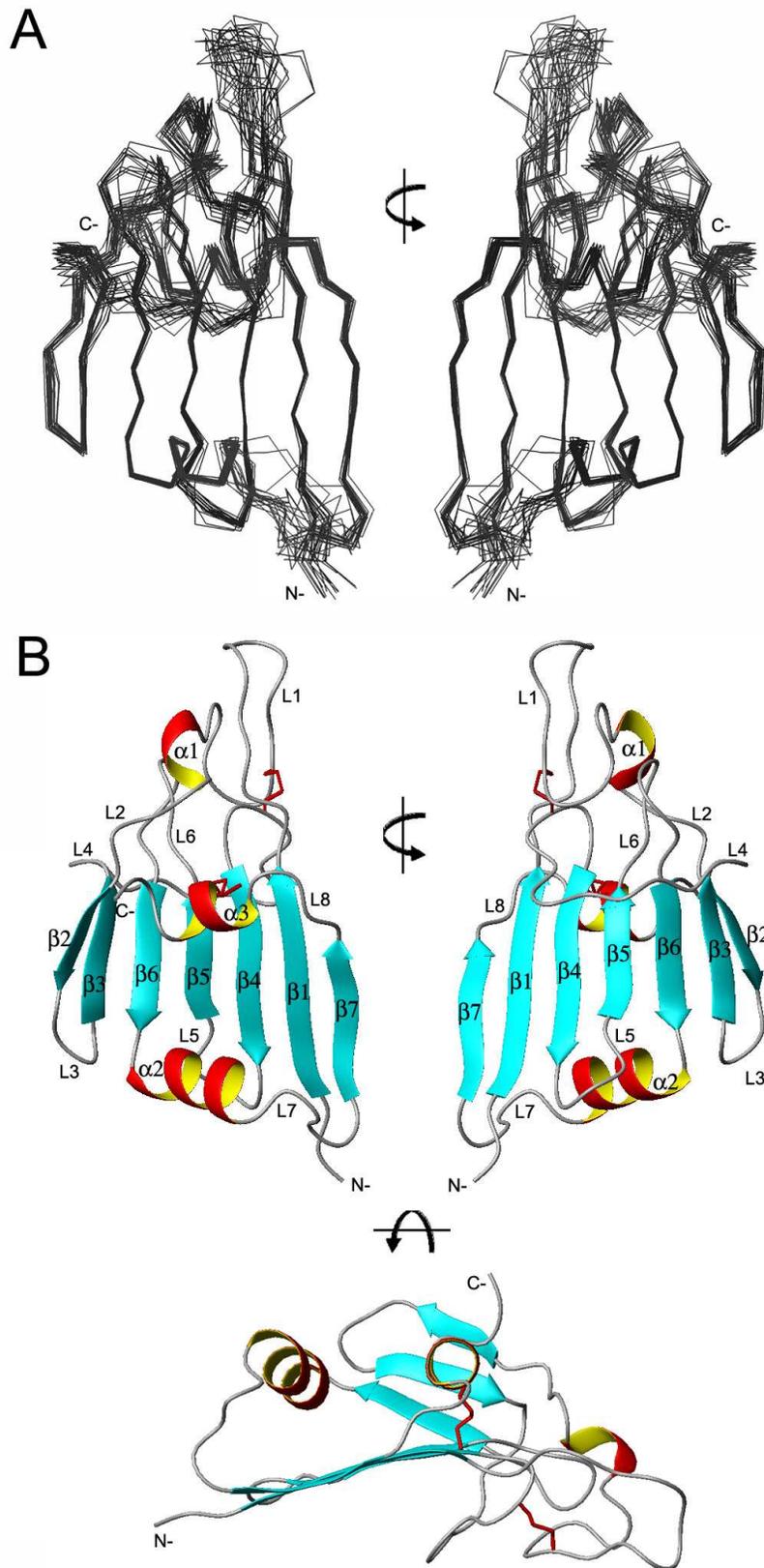
**Figure 7.** Superposition of  $\text{C}\alpha$  atoms of CP minimum energy structure (2K8W, red) with (A) dimeric IL-8 (1ILQ, yellow, RMSD 3.3 Å), (B) NAP-2 (1TVX, cyan, RMSD 3.3 Å) and (C) MHC-I (1M6O, green, RMSD 3.4 Å) human proteins.



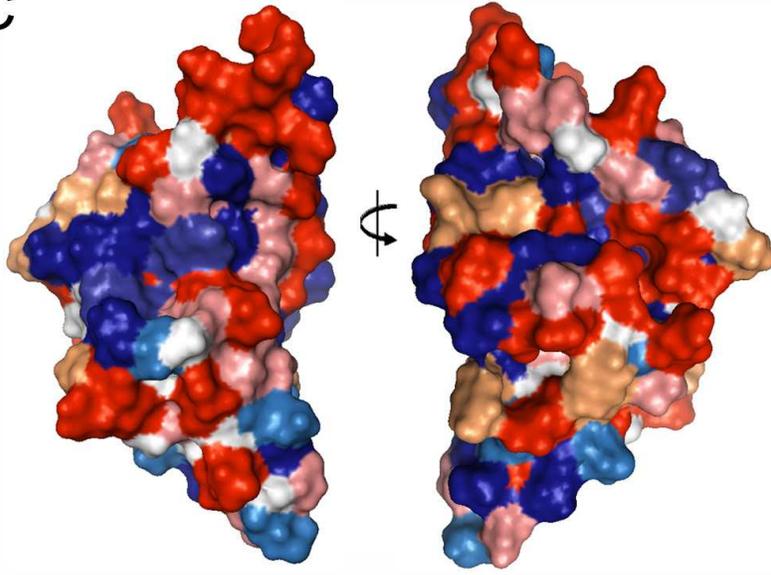
- Figure 2 -



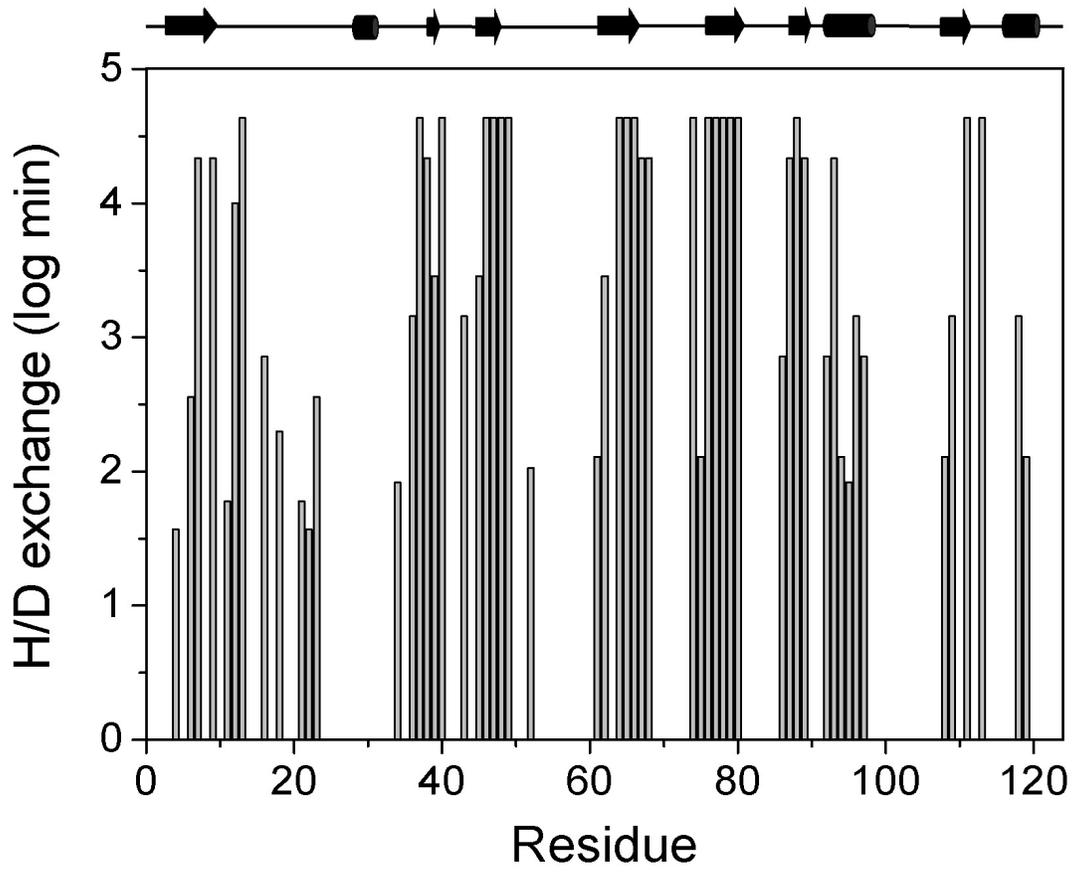
- Figure 3 -



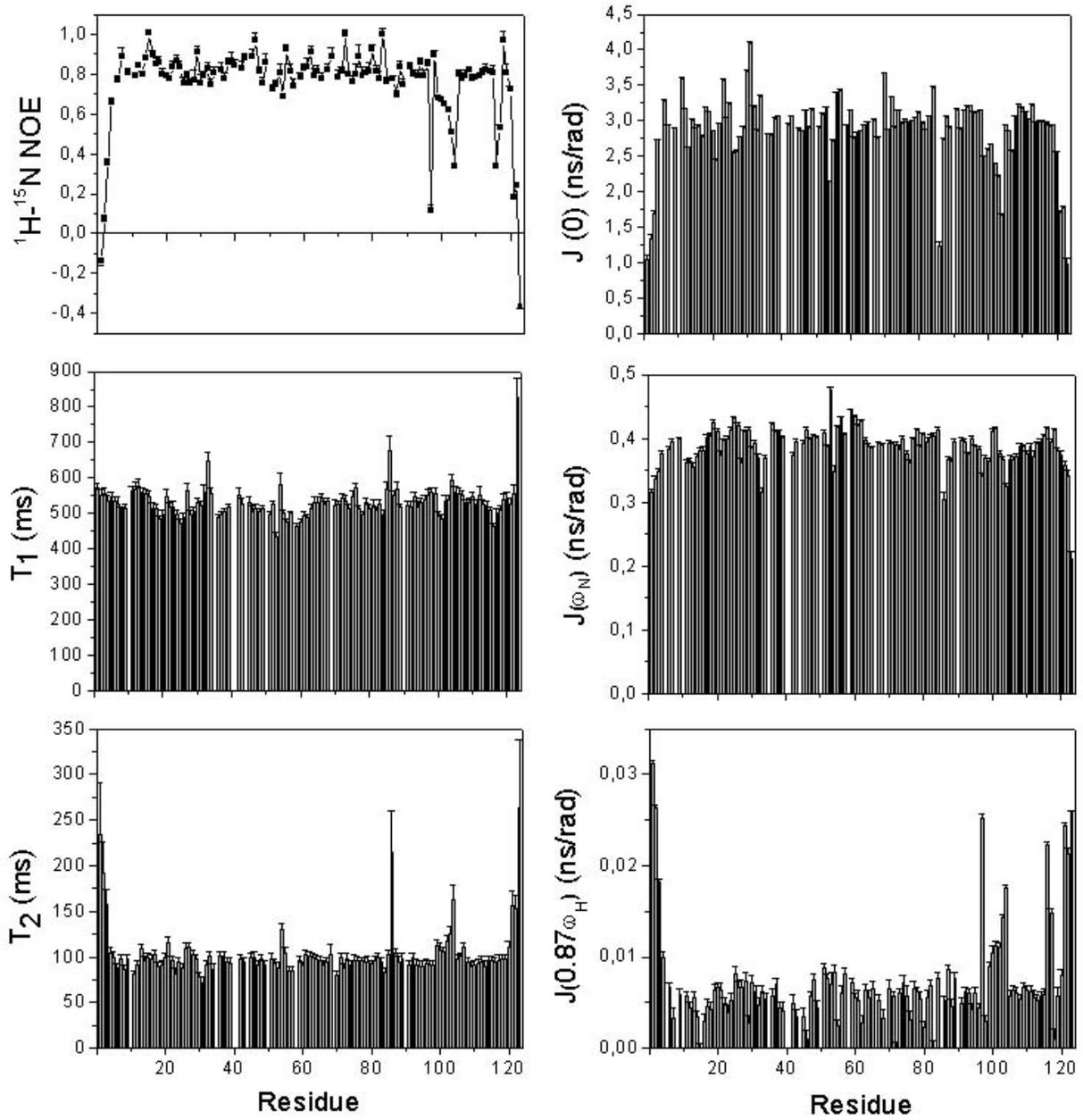
C



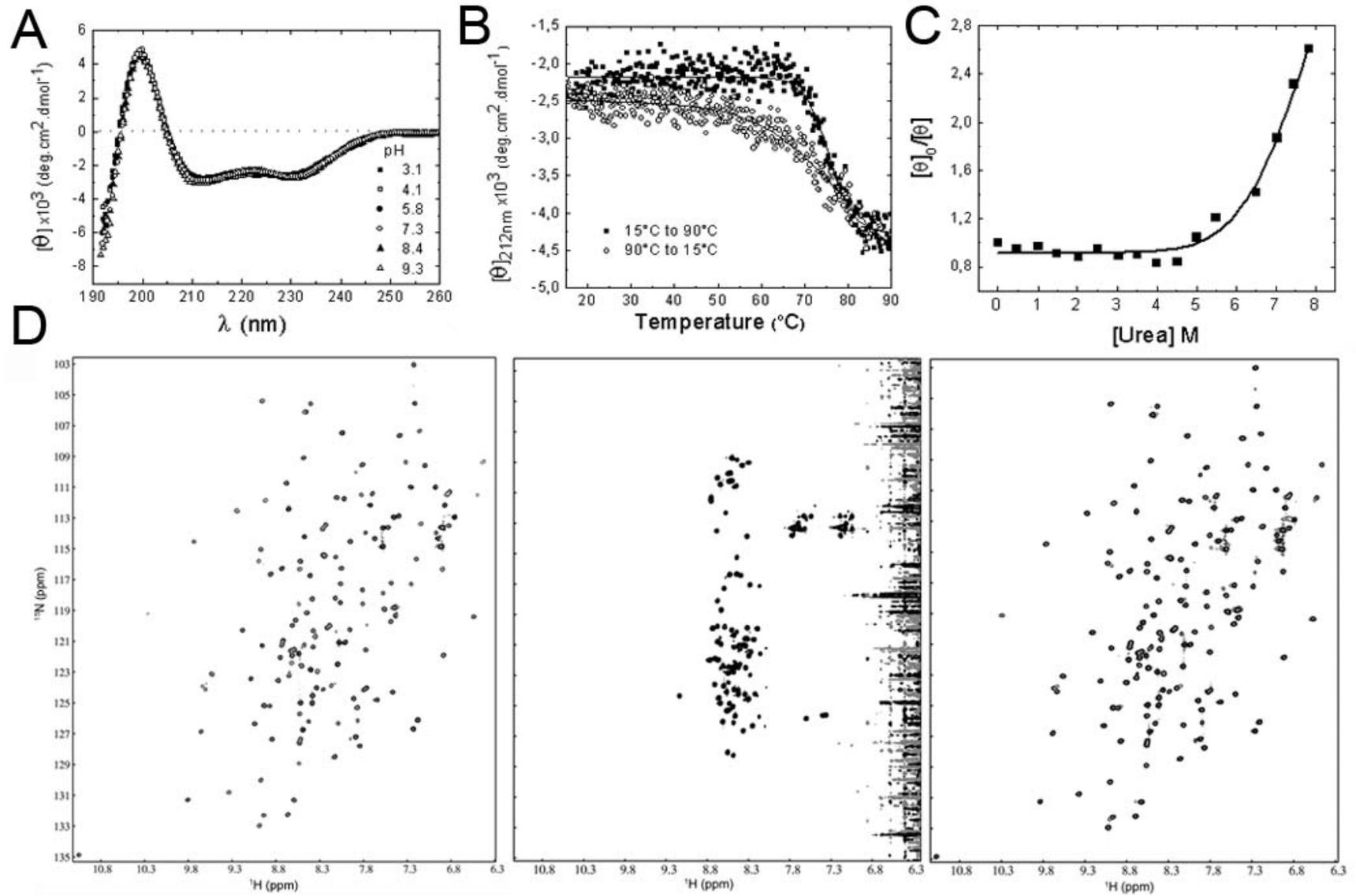
- Figure 4 -



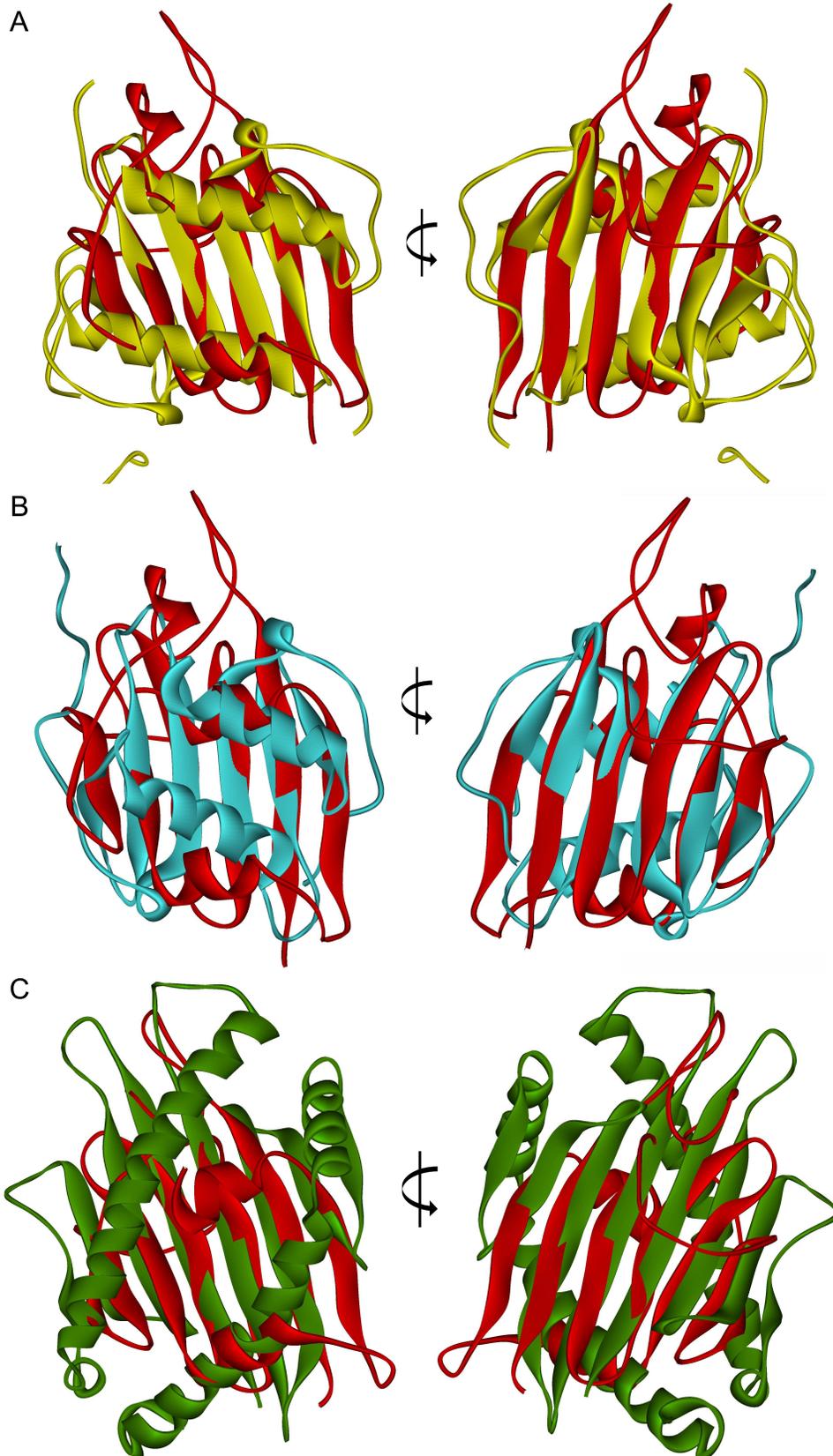
- Figure 5 -



- Figure 6 -



- Figure 7 -



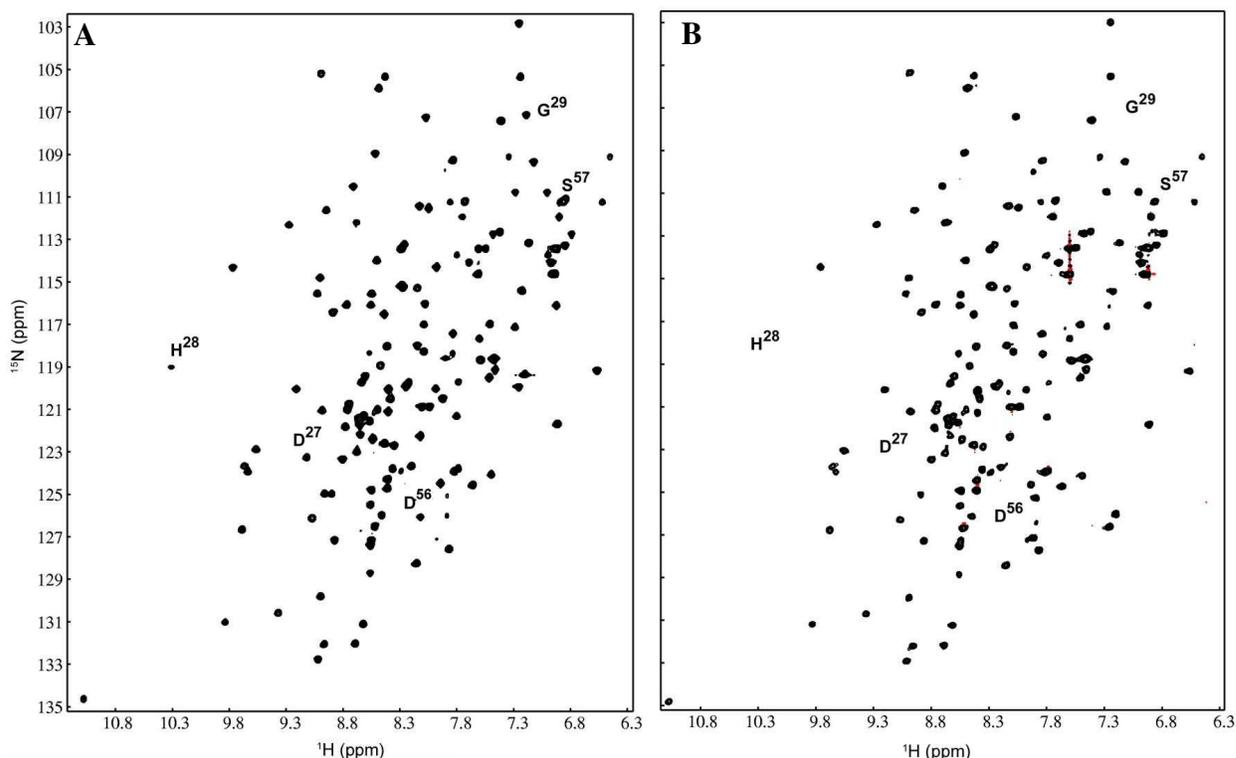
## Resultados Complementares

### Ensaio de interação com potenciais ligantes

Outro resultado interessante obtido através da ferramenta *ProFunc* foi a localização de um possível sítio de ligação ao íon cobre, caracterizado pela seguinte assinatura:

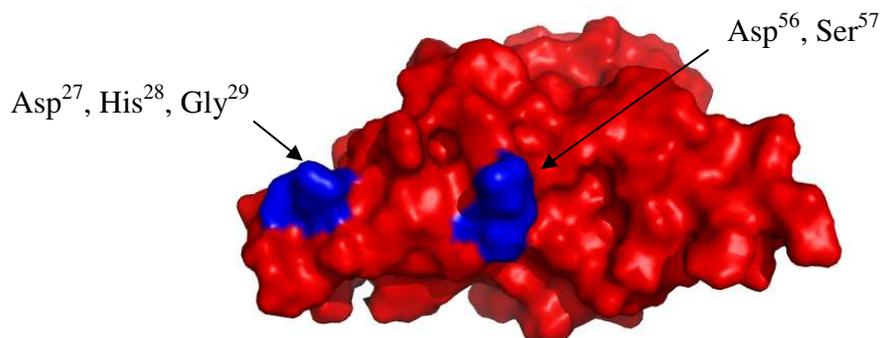
[GA]-x(0,2)-[YSA]-x(0,1)-[VFY]-{SEDT}-C-x(1,2)-[PG]-x(0,1)-H-x(2,4)-[MQ]

É interessante notar que esse sítio de ligação envolveria os resíduos Gly<sup>19</sup>-Gln<sup>33</sup>, que, em sua maioria, não apresentam uma estrutura definida e que, portanto, têm uma maior plasticidade para a interação com cobre. Para verificar a possível interação da proteína com Cu<sup>2+</sup>, foram coletados espectros de <sup>1</sup>H-<sup>15</sup>N HSQC da CP nas proporções de 0,5:1, 1:1 e 2:1 de [CuSO<sub>4</sub>]:[CP] (Figura 3.1). Os resíduos envolvidos na interação com o cobre terão seu sinal perturbado, alterando o valor de seu deslocamento químico.



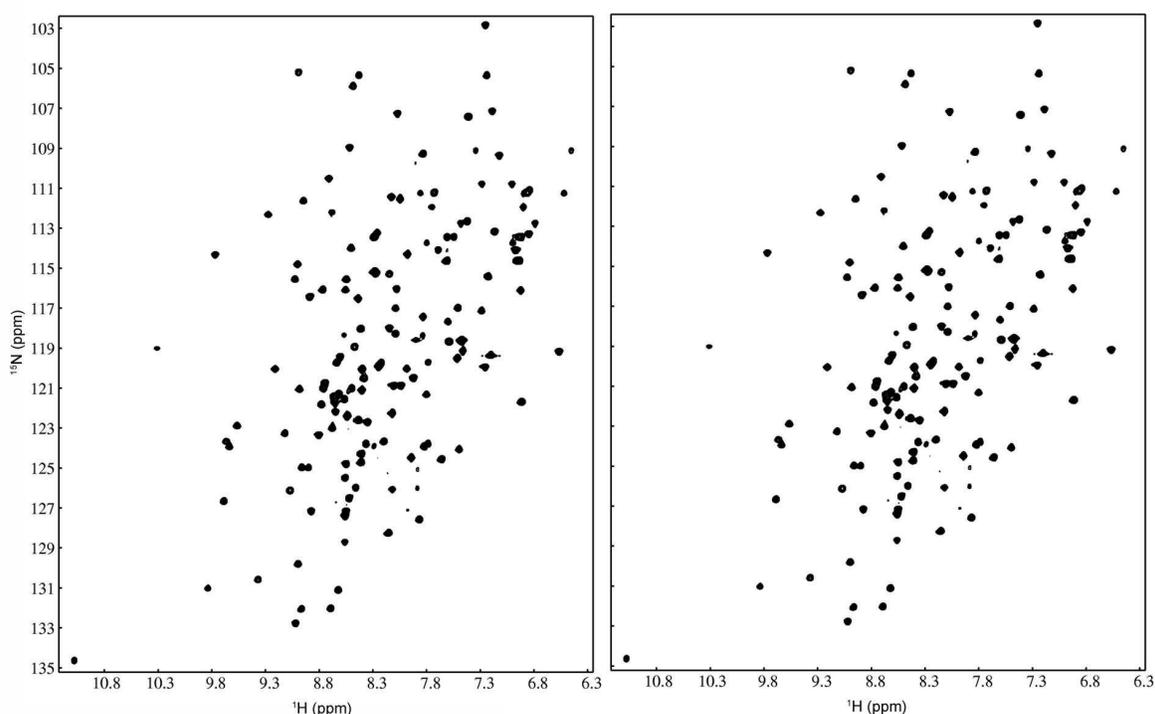
**Figura 3.1.** Espectros de <sup>1</sup>H-<sup>15</sup>N HSQC de 0,2 mM de proteína <sup>15</sup>N-CP em 10 mM tampão fosfato, pH 5,8 a 20°C (A) na ausência e (B) na presença de CuSO<sub>4</sub> na proporção 1:1.

A análise dos espectros mostrou que os sinais referentes aos resíduos Asp<sup>27</sup>, His<sup>28</sup>, Gly<sup>29</sup>, Asp<sup>56</sup> e Ser<sup>57</sup> desaparecem do espectro de <sup>1</sup>H-<sup>15</sup>N HSQC (Figura 3.1), sugerindo uma possível ligação desses resíduos com Cu<sup>2+</sup>. Porém, ao contrário dos resíduos Asp<sup>27</sup>, His<sup>28</sup> e Gly<sup>29</sup>, os resíduos Asp<sup>56</sup> e Ser<sup>57</sup> não estão incluídos na assinatura que caracteriza a ligação com cobre e estão a uma distância de ~20 Å (Figura 3.2). Além disso, para que aja ligação da proteína com o cobre, os resíduos de cisteína devem estar na sua forma reduzida e, portanto, não envolvidos em ligações dissulfeto (Ma et al., 2005), o que não é o caso da CP. Levando-se tudo isso em consideração, acredita-se que, provavelmente, a ligação da proteína CP a íons cobre seja inespecífica. Porém outros experimentos devem ser feitos para se comprovar essa hipótese.



**Figura 3.2.** Representação em superfície da estrutura da proteína CP. Em azul estão destacados os resíduos que tiveram os seus sinais perturbados pela presença de íons Cu<sup>2+</sup>.

Entre as estruturas do clan das DPBB glucanases, a CP apresenta uma maior similaridade às proteínas da família barwin. É descrito que essas proteínas são capazes de se ligar tetrâmeros de N-acetilglucosamina (Ludvigsen & Poulsen, 1992). Assim, também foi testada a capacidade da CP em se ligar a esse sacarídeo. Para isso foram coletados espectros <sup>1</sup>H-<sup>15</sup>N HSQC da proteína CP na proporção de 1:1 de [N-acetilglucosamina]:[CP]. Porém, o espectro obtido não apresenta nenhuma perturbação no deslocamento químico dos sinais da proteína (Figura 3.3), indicando que a CP não interage com esse ligante.



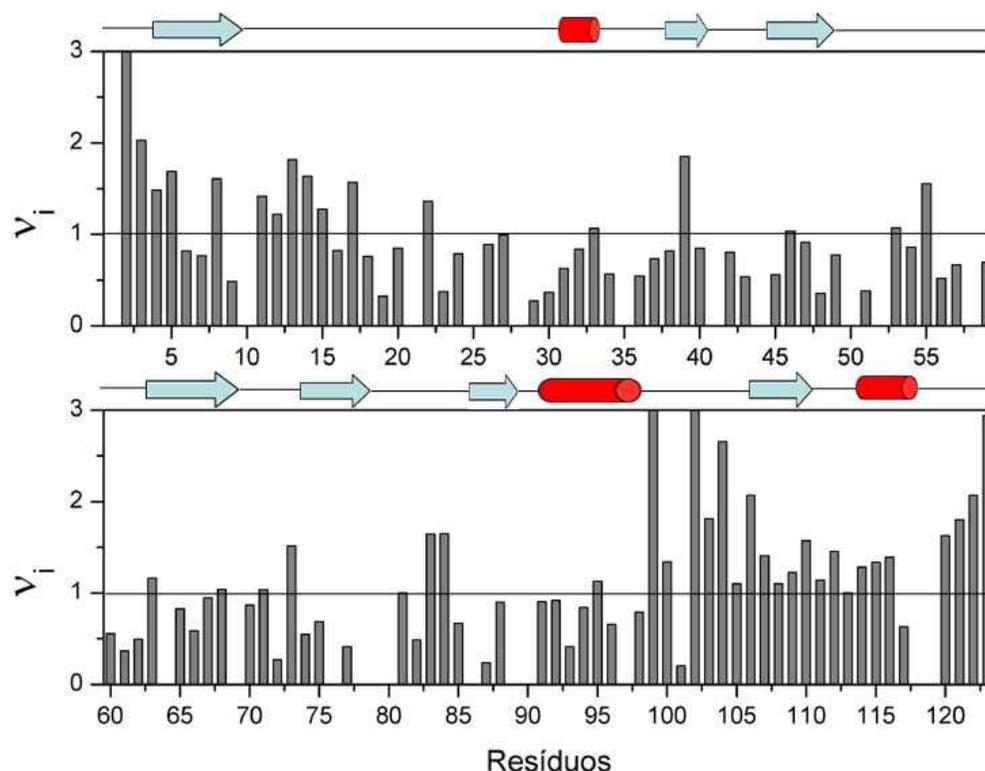
**Figura 3.3.** Espectros  $^1\text{H}$ - $^{15}\text{N}$  HSQC de 0,2 mM da proteína  $^{13}\text{C}/^{15}\text{N}$ -CP em 10 mM tampão fosfato, pH 5,8 a 20°C (A) na ausência e (B) na presença N-acetilglucosamina na proporção 1:1.

### Acessibilidade ao solvente

Conhecer as diferentes exposições e acessibilidade dos resíduos ao solvente podem fornecer informações valiosas sobre uma proteína principalmente quando existem poucas informações estruturais, como é o caso da CP. Assim, para identificação dos resíduos que estão localizados na superfície da molécula, foi realizado o experimento CLEANEX (*Clean Chemical Exchange*) para região da cadeia principal ( $^1\text{H}$ - $^{15}\text{N}$ ) da proteína. Esse experimento detecta a transferência de magnetização entre grupo amídico da cadeia principal da proteína e a água através do NOE. Dessa forma, a cadeia principal dos resíduos que estão próximos a moléculas de água, e expostos ao solvente, apresentam um pico de correlação no espectro (Hwang et al., 1998).

Para uma análise qualitativa do resultado, o espectro de CLEANEX obtido foi comparado com o espectro de  $^1\text{H}$ - $^{15}\text{N}$  HSQC da proteína CP (Anexo 6). Para a análise quantitativa, foram medidos os volumes dos picos de correlação do espectro de CLEANEX

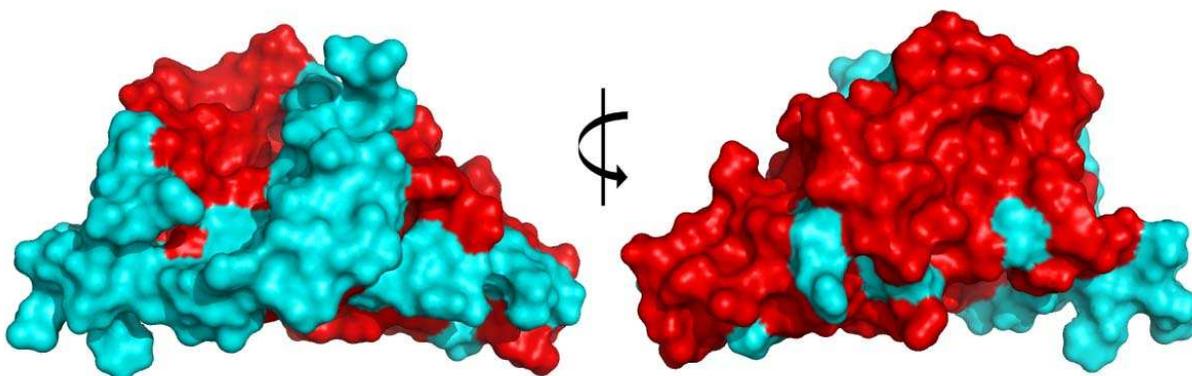
para cada resíduo da proteína e foram calculados os seus respectivos volumes auto-escalado  $v_i$  (Figura 3.4). A cadeia principal dos resíduos que apresentam um volume  $v_i$  maiores que 1 se encontram em regiões de maior acessibilidade ao solvente.



**Figura 3.4.** Gráfico do volume auto-escalado  $v_i$  obtidos para 0,7 mM da proteína  $^{13}\text{C}/^{15}\text{N}$ -CP, 10 mM tampão fosfato, pH 5,8 a 20°C comparado com a estrutura secundária da proteína. As flechas em ciano indicam estrutura em fita- $\beta$  e os cilindros vermelhos indicam estrutura em  $\alpha$ -hélice.

De maneira geral, os dados de acessibilidade ao solvente corroboram os resultados de troca H/D, apresentados no Artigo 3 (Figura 4). Já era esperada uma exposição ao solvente da cadeia principal das regiões N-terminal, compreendendo os resíduos Glu<sup>1</sup>-Ser<sup>5</sup>, e C-terminal, entre os resíduos Ser<sup>99</sup>-Asn<sup>123</sup>. Por outro lado, a cadeia principal dos resíduos entre Met<sup>18</sup>-Gly<sup>98</sup> apresentam, de modo geral, sinais menos intensos e valores de  $v_i$  menores que 1, indicando que estes resíduos estão localizados em porções mais internas da proteína CP.

É interessante notar que, além dos resíduos N- e C-terminais, os resíduos Tyr<sup>8</sup>, Ile<sup>11</sup>-Ser<sup>17</sup>, Ala<sup>22</sup>, Glu<sup>39</sup>, Trp<sup>55</sup>, Cys<sup>63</sup>, Ser<sup>73</sup>, Arg<sup>83</sup>, Gly<sup>84</sup> e Lys<sup>95</sup> apresentam sinais no CLEANEX de maior intensidade e valores de  $v_i$  maiores que 1. Como a cadeia principal desses resíduos representam as regiões com maior acesso a moléculas de água e, portanto, de maior acessibilidade a outras espécies químicas, poderia-se hipotisar que eles estejam envolvidos em interações específicas da CP com seu alvo. A localização preferencial desses resíduos nas porções  $\beta 1$ ,  $\beta 7$ ,  $\alpha 3$  e L7 e, como pode ser observado na figura 3.5, em uma face da proteína corrobora essa hipótese.



**Figura 3.5.** Representação da superfície da estrutura da proteína CP. Em ciano estão destacados os resíduos que apresentaram maior acessibilidade ao solvente.

## Procedimentos Experimentais

### Ensaio de interação com potenciais ligantes

Os experimentos de RMN de interação com ligante foram adquiridos no espectrômetro de Ressonância Magnética Nuclear Varian Inova 600 MHz da *Università degli Studi di Parma* (Itália). Foram realizados experimentos bidimensionais  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (Kay et al., 1992) com janela de 8000 Hz e 512 pontos complexos para  $^1\text{H}$  e 1725,25 Hz, com 128 incrementos para  $^{15}\text{N}$ . Todos os dados foram coletados a 20°C e processados em estações *Silicon Graphics Octane 2* utilizando-se os *softwares* NMRpipe/NMRview (Delaglio et al., 1995; Johnson & Blevins, 1994). Para a coleta dos dados foi preparada 600  $\mu\text{L}$  de solução de  $^{15}\text{N}$ -CP a uma concentração final de 0,2 mM em 10 mM tampão fosfato, pH 5,8 e 5% de  $\text{D}_2\text{O}$ . Para a análise da interação com ligante, foram adicionadas 0,1, 0,2 e 0,4 mM de  $\text{CuSO}_4$  ou 0,2 mM de N-acetilglucosamina.

### Acessibilidade ao solvente

Os experimento CLEANEX (Hwang et al., 1998) para região da cadeia principal ( $^1\text{H}$ - $^{15}\text{N}$ ) foi coletado no espectrômetro de Ressonância Magnética Nuclear Varian Inova 600 MHz da *Università degli Studi di Parma* (Itália) com uma janela espectral de 8000 Hz e 2048 pontos complexos para  $^1\text{H}$  e de 2430 Hz e 256 incrementos para  $^{15}\text{N}$ . O tempo total de coleta do experimento, utilizando-se 176 transientes, foi de ~4 horas. Os dados foram coletados a 20°C e processados em estações *Silicon Graphics Octane 2* utilizando-se os *softwares* NMRpipe/NMRview (Delaglio et al., 1995; Johnson & Blevins, 1994). Para a coleta do experimento foi preparada 600  $\mu\text{L}$  de solução de  $^{15}\text{N}$ -CP a uma concentração final de 0,7 mM em 10 mM tampão fosfato, pH 5,8 e 5% de  $\text{D}_2\text{O}$ .

Para a análise dos dados, foi medido o volume dos picos de correlação para cada resíduo da proteína e calculado o seu respectivo volume auto-escalado, de acordo com a seguinte equação (Equação 3.1):

$$v_i = \frac{V_i}{\left(\frac{1}{n}\right) \sum_i V_i} \quad (\text{Eq. 3.1})$$

## DISCUSSÃO

O estudo dos mecanismos de defesa de planta é um assunto emergente e tem ganhado interesse da comunidade científica nos últimos 20 anos. Enquanto muitos pesquisadores vêem o sistema imune de mamíferos como o auge da evolução, a seleção natural tem afinado as estratégias de defesa de plantas nesses 1,6 bilhões de anos sem recorrer a anticorpos, células T e provavelmente produzindo um sistema não menos sutil ou eficiente (SurrIDGE, 2001).

Apesar dos avanços nos estudos relacionados à biologia molecular de plantas, poucas moléculas envolvidas na patogenicidade de fungos são caracterizadas. Uma visão geral das estruturas descritas até o momento, evidencia a falta de similaridade entre elas e encoraja estudos com o objetivo de encontrar novos enovelamentos que possam ajudar a explicar os mecanismos moleculares das respostas de defesa em plantas. Em particular, a estrutura 3D da CP poderá aumentar o conhecimento de MAMPs não catalíticos, que tem uma importante função em resposta de plantas a hospedeiros e não-hospedeiros.

Para realizar os estudos estruturais e dinâmicos da CP, um passo essencial foi a produção heteróloga dessa proteína em células de *P. pastoris* na sua forma pura, enovelada e com estrutura semelhante a da CP nativa. Sem essa etapa inicial, não seria possível realizar a marcação da proteína CP com isótopos  $^{15}\text{N}$  e  $^{13}\text{C}$  para subsequente análise estrutural através da técnica de RMN.

A fase sucessiva de análise e atribuição das ressonâncias de  $^1\text{H}$ ,  $^{15}\text{N}$  e  $^{13}\text{C}$  da cadeia principal e lateral da proteína CP também foi facilitada pela boa resolução dos espectros de  $^1\text{H}$ - $^{15}\text{N}$  HSQC,  $^1\text{H}$ - $^{13}\text{C}$  HSQC, HNCA, HNCO, HN(CO)CA, CBCA(CO)NH, HNCACB, HCCH-TOCSY, HCCH-COSY, permitindo o assinalamento inequívoco e quase que total dos deslocamentos químicos referentes a cada átomo da proteína.

Superadas as etapas iniciais, a estrutura da CP foi calculada com base nas restrições de distância, extraídas dos NOEs, das orientações internucleares relativas ao campo magnético derivados dos RDCs da ligação  $^1\text{H}_\text{N}$ - $^{15}\text{N}$  e  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  e de ângulos diedros  $\phi$  e  $\psi$  derivados da análise do deslocamento químico usando-se o programa TALOS (Cornileuscu et al., 1999). Apesar de ter sido feita a coleta dos dados de RDCs referentes às ligações

$^1\text{H}_\text{N}$ - $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$ ,  $^1\text{H}_\text{N}$ - $^{13}\text{CO}$ ,  $^{15}\text{N}$ - $^{13}\text{CO}$  e  $^{13}\text{C}_\alpha$ - $^{13}\text{CO}$ , apenas as constantes referentes às ligações  $^1\text{H}_\text{N}$ - $^{15}\text{N}$  e  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  foram utilizadas como restrições para o cálculo das estruturas. As restrições de RDCs das ligações  $^1\text{H}_\text{N}$ - $^{13}\text{CO}$ ,  $^{15}\text{N}$ - $^{13}\text{CO}$ ,  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  e  $^{13}\text{C}_\alpha$ - $^{13}\text{CO}$  não foram adicionadas devido a dificuldades no cálculo da magnitude dos componentes axial  $D_a$  e rômbo  $D_r$  do tensor de alinhamento. Como descrito no Capítulo 2, a orientação do tensor de alinhamento dos RDCs experimentais com relação a correspondente estrutura 3D é essencial para a utilização dessas constantes como restrição. Outros fatores que aumentaram a dificuldade do cálculo do tensor de alinhamento são os erros experimentais na obtenção dos RDCs referentes às ligações  $^1\text{H}_\text{N}$ - $^{13}\text{CO}$ ,  $^{15}\text{N}$ - $^{13}\text{CO}$  e  $^{13}\text{C}_\alpha$ - $^{13}\text{CO}$ . Os erros dessas medidas são grandes uma vez que a relação sinal/ruído dos espectros é pequena. Esse fato é ainda agravado com a propagação dos erros pela normalização dos RDCs com relação a ligação  $^1\text{H}_\text{N}$ - $^{15}\text{N}$  (fator descrito na Tabela 2.1 – Capítulo 2).

Nesta tese foi apresentada a estrutura em solução da proteína CP, a primeira estrutura de um membro da família Cerato-Platanina. A proteína é composta por duas pequenas  $\alpha$ -hélices, uma hélice-3<sub>10</sub> e sete fitas- $\beta$ , sendo duas paralelas e cinco antiparalelas, formando uma única folha- $\beta$ . De modo geral, as medidas de relaxação indicam que a CP possui uma estrutura rígida, compacta e que não apresenta mudanças conformacionais em escalas de tempo de pico a nano segundo. Apenas as porções N- e C- terminais e a região entre os resíduos Ser<sup>99</sup>-Arg<sup>104</sup>, que formam o *loop* L7, apresentam uma maior flexibilidade, refletida pelos baixos valores de  $^1\text{H}$ - $^{15}\text{N}$ -NOE e valores de  $T_2$  maiores que a média. Mesmo os longos *loops* L1 e L4 não são flexíveis como esperado, provavelmente em consequência da ligação dissulfeto entre Cys<sup>23</sup>-Cys<sup>60</sup> que conecta esses segmentos. Além disso, a CP possui 8 resíduos de prolina em sua sequência que estão localizados principalmente no início ou no final das estruturas secundárias (Figura 1 – Artigo 3). Os resíduos de prolina, devido as suas propriedades cíclicas, tendem a interromper os elementos de estrutura secundária e diminuir a flexibilidade dos *loops*.

É interessante notar que a estrutura 3D da CP não apresentou semelhança conformacional a nenhuma estrutura das proteínas que constituem o clan das DPBB glucanases, nem mesmo à proteína barwin que, entre as proteínas dessas famílias, apresenta a maior identidade (20%) a proteína CP (Figura D.1). Talvez esse fato não seja



chaperona, por exemplo, podem ser fatores determinantes para a proteína manter sua estrutura e, conseqüentemente, sua função uma vez secretada pelo patógeno.

Também foi demonstrado que a CP possui uma alta estabilidade térmica. A relativa hidrofobicidade, a pouca acessibilidade ao solvente, constatada pelos experimentos de CLEANEX (Figura 3.4 – Capítulo 3) e baixa velocidade de troca H/D (Figura 4 - Artigo 3), e a pouca flexibilidade da CP são os principais fatores que determinam a sua estabilização térmica. Somado a isso, as duas ligações dissulfetos e a presença de oito resíduos de prolina também podem estabilizar a proteína pela diminuição da entropia do estado desenovelado (Vieille & Zeikus, 2001). Porém, apesar da alta estabilidade térmica, a CP não retorna completamente a sua conformação inicial no processo de renaturação. Isso pode ocorrer devido a dois fatores: ou a proteína sofre alterações irreversíveis à sua conformação original, ou a proteína se agrega após a desnaturação térmica e o sinal observado é menos intenso devido à presença de um menor número de moléculas enoveladas no meio. Entretanto, esse fenômeno também pode estar relacionado com a cinética de reenovelamento. Ou seja, devido ao alto número de prolinas na proteína, o caminho de reenovelamento possui um passo limitante de isomerização lenta ao redor de uma ou mais ligações peptídicas envolvendo esses aminoácidos. Isso poderia resultar em uma fração de proteínas na amostra que não conseguiria retornar ao estado nativo durante o tempo de experimento (Pertinhez et al., 2000).

Na tentativa de se descobrir qual é a função da CP, foram testadas as hipóteses da proteína apresentar atividade glicolítica assim como de possuir habilidade de ligar tetrâmeros de N-acetilglucosamina ou íons  $\text{Cu}^{2+}$ . Como todas as tentativas de identificar a função da proteína CP não foram bem sucedidas, partiu-se para a busca de proteínas com enovelamentos similares ao da proteína de estudo, já que esses resultados também podem dar indicações de possíveis funções da proteína. Assim, com o objetivo de encontrar domínios ou outras proteínas com enovelamento similar ao da CP, as 20 conformações que representam a família de estruturas da CP foram submetidas a um alinhamento de estrutura 3D utilizando-se a ferramenta de bioinformática *ProFunc* (Laskowski et al., 2005). A estrutura da CP não possui semelhança conformacional a nenhuma outra estrutura de proteína envolvida em avirulência de microorganismos ou a proteínas elicitoras de resposta

de defesa em plantas. Porém, surpreendentemente, os alinhamentos estruturais mostraram que a CP possui um enovelamento similar a moléculas pertencentes a família das quimiocinas e do MHC-I. De fato, a forma dimérica da IL-8, NAP-2 e os domínios  $\alpha 1, \alpha 2$  do MHC-I são as estruturas que melhor se sobrepõem a conformação de menor energia da CP.

Ambas famílias de proteínas estão relacionadas com a resposta do sistema imune inato de animais contra vários microorganismos. IL-8 e NAP-2 são citocinas bem caracterizadas que estimulam o movimento de leucócitos e regulam a sua migração do sangue para os tecidos. Os vários membros caracterizados dessa família diferem na sua extensão de oligomerização (monômeros, dímeros ou tetrâmeros) e/ou na superfície de interação usada para a oligomerização. Todos os monômeros são constituídos de 3 fitas- $\beta$  e uma  $\alpha$ -hélice (Skelton et al., 1999; Malkowski et al., 1997; Gerber et al., 2000). Já as moléculas de MHC-I mediam respostas imunes antivirais através da ligação de fragmentos de peptídeos de vírus dentro da célula infectada e do transporte do complexo resultante para a superfície da célula e subsequente reconhecimento das células T. A estrutura geral do MHC-I compreende três domínios ( $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$ ) que são não-covalentemente associados a um domínio  $\beta_2$ -microglobulina. Os domínios  $\alpha 1$  e  $\alpha 2$ , que são os domínios similares à CP, são regiões de reconhecimento e são constituídos de 2  $\alpha$ -hélices antiparalelas posicionadas sobre 8 fitas- $\beta$ . Entre esses domínios pode-se observar a presença de uma profunda cavidade na qual o antígeno se liga (Hickman-Miller et al., 2005).

Dessa forma, uma característica comum entre as proteínas com conformações similares ao da CP, IL-8, NAP-2 e MHC-I, é a sua função sinalizadora. O processo de sinalização ocorre através do reconhecimento da molécula por receptores de membrana específicos e relacionados com respostas de defesa contra patógenos. Esse fato sugere que a CP também seja uma molécula sinalizadora e que ela provavelmente tenha um receptor particular em plantas.

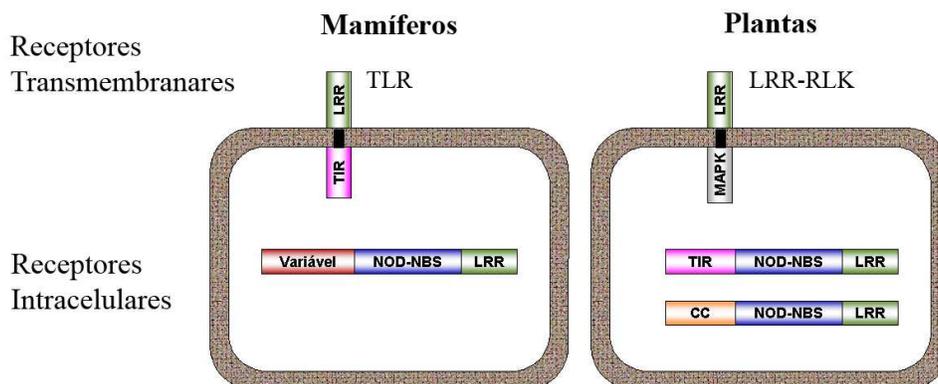
Os receptores das quimiocinas IL-8 e NAP-2, *CXCR1* e *CXCR2*, pertencem à classe de receptores acoplados a proteínas-G e têm uma importante função na patofisiologia de numerosas doenças auto-imunes e inflamatórias e na propagação de câncer e metástase.

Apesar da estrutura completa de receptores de quimiocinas não ser conhecida, análises de sequência predizem que eles são constituídos por um domínio N-terminal extracelular, um segmento transmembranar 7 TM (sete  $\alpha$ -hélices transmembranares), três *loops* extracelulares e um segmento C-terminal. Apesar da pouca de informação estrutural, são propostos dois sítios de interação receptor-ligante: um *loop* na região N-terminal do ligante com o domínio N-terminal do receptor (sítio 1) e os resíduos N-terminais Glu-Leu-Arg do ligante com um *loop* extracelular (sítio 2). Esses dois sítios de ligação podem modular a afinidade e seletividade da ligação receptor/ligante. A CP não possui a sequência Glu-Leu-Arg presente na região N-terminal desordenada da IL-8 que ativa o receptor. Por outro lado, estudos consistentes mostram que os resíduos centrais das quimiocinas, também teriam relevância na afinidade da proteína com o seu receptor (Rajagopalan & Rajarathnam, 2006; Skelton et al., 1999).

No caso da proteína MHC-I, a orquestração dos eventos de reconhecimento e sinalização requer a atividade de vários receptores associados a células T (*T cell receptors* ou TCR), incluindo co-receptores como o CD8. De maneira geral, TCRs são heterodímeros presentes na superfície celular compostos por domínios variáveis de imunoglobulina, seguido por um domínio transmembranar e uma pequena calda citoplasmática. Porém, a impressionante diversidade em que os TCRs podem reconhecer antígenos dificultam a elucidação das bases estruturais que definem os modos de ligação e sinalização dos complexos TCR/MHC (Rudolph et al., 2006).

Buscas no genoma de *Arabidopsis* feitas com as sequências de receptores de IL-8, NAP-2 e MHC-I, descritos acima, retornaram alguns possíveis alvos da proteína CP em planta. Ensaio de *pull-down* e espectrometria de massa poderiam confirmar a interação da CP com receptores ou possibilitariam o isolamento e identificação de parceiros de interação da proteína. Embora não tenham sido encontrados em *Arabidopsis* receptores com alta homologia de sequência aos receptores de mamíferos, sabe-se que a homologia estrutural é bem conservada. Portanto, mesmo que esses receptores de animais não possuam homologia de sequência direta em plantas, não se pode descartar a possibilidade de existir homólogos estruturais entre esses organismos, que corresponderiam a possíveis parceiros de interação com a CP.

Essa idéia é reforçada através de algumas características comuns já descritas entre a imunidade inata de animais e plantas. Em animais, uma família de proteínas conservadas que funciona diretamente ou indiretamente como PRRs são os receptores transmembranares *Toll-like* (TLRs). TLRs são caracterizados por um domínio extracelular rico em leucina (LRR) e um domínio receptor intracelular *Tool-interleucine-1* (TIR) de interação proteína-proteína. LRR são encontrados tanto em plantas quanto em animais. Porém, o domínio citoplasmático não é conservado e a sinalização intracelular em plantas é modulada através de proteínas quinases (LRR *receptor-like kinases*). Já no caso de receptores intracelulares a classe de proteínas predominante contém um sítio de ligação a nucleotídeo (*nucleotide-binding site* ou NBS), que é requerido para a ligação com ATP ou ADP, e um domínio LRR. Em plantas, as proteínas NBS-LRR podem ser divididas nas que carregam um N-terminal homólogo aos receptores TIR, como TIR-NBS-LRR, ou naquelas que carregam estrutura em *coiled-coil* (CC-NBS-LRR). Essas proteínas possuem a mesma arquitetura de proteínas Nod de mamíferos, que estão envolvidas no reconhecimento intracelular de infecção. Porém, em animais o domínio N-terminal pode ser variável (Figura D.2) (Ausubel, 2005; Gómez-Gómez, 2004).



**Figura D.2.** PRRs extracelulares e intracelulares de plantas e animais (Ausubel, 2005).

O fato da proteína CP possuir um enovelamento semelhante ao das citocinas IL-8 e NAP-2 não parece ser fortuito quando se analisa com atenção estudos já realizados com fungos patógenos de humano que secretam proteínas altamente similares a CP. Esse é o

caso do fungo *Aspergillus fumigatus* que secreta uma proteína com 83% de similaridade a CP e é responsável por causar uma variedade de doenças, incluindo micotoxicoses, reações alérgicas e doenças sistêmicas, como a aspergilose invasiva que possui altas taxas de mortalidade em pacientes imunocomprometidos. Estudos demonstram que células monocíticas de humanos após exposição ao *A. fumigatus* apresentam resposta pró-inflamatória ativada por receptores TLR-2 e TLR-4 e seguida pela tradução de sinal via NFκB e MAPKs (Shalit et al., 2006; Balloy et al., 2008). A pronunciada resposta pró-inflamatória de *A. fumigatus* está associada ao mesmo tempo à proteção contra a infecção e ao aumento da patogenicidade desse fungo. Embora tenha sido demonstrado que o extrato de proteínas secretadas pelo fungo mediam essa resposta, ainda não foi comprovado qual o componente que ativa a cascata de sinalização (Rementeria et al., 2005). Poderia-se especular que a proteína similar CP secretada pelo *A. fumigatus* estaria envolvida na ativação de respostas de defesa ao fungo. Também é interessante destacar que os receptores TLRs que estão envolvidos nos mecanismos de defesa em humanos são conservados plantas, como descritos anteriormente, sugerindo uma possível via de sinalização comum em plantas e animais na interação hospedeiro-patógeno.

Mais interessante ainda é notar que o fungo *A. fumigatus* estimula a síntese de IL-8 e do fator de necrose tumoral- $\alpha$  (*tumor necrosis factor alpha* ou TNF- $\alpha$ ) (Shalit et al., 2006). Sabe-se que a migração de leucócitos durante condições de infecção é um processo regulado pelo aumento da secreção de quimiocinas pró-inflamatórias, como IL-8 e NAP-2, em resposta a sinalização de moléculas como TNF- $\alpha$  e interferon- $\gamma$  (IFN- $\gamma$ ) (Rajagopalan & Rajarathnam, 2006). Assim, não parece ser acidental que uma proteína com enovelamento semelhante a IL-8 possa estar envolvida na resposta inflamatória, agindo como um mimético de IL-8 e favorecendo a invasão do patógeno. Essa hipótese é ainda mais reforçada pela determinação da atividade imunomodulatória da proteína ACA-1, que possui 62% de similaridade a CP. ACA-1 é uma proteína secretada pelo fungo *Antrodia camphorata*, usado tradicionalmente na medicina Chinesa para o tratamento de câncer, e que induz a produção de TNF- $\alpha$  e IFN- $\gamma$  (Sheu et al., 2005), ambos fatores envolvidos no estímulo de secreção de IL-8 e recrutamento de linfócitos (Hasan et al., 2004).

Embora esses resultados sejam muito promissores, ensaios funcionais para testar se a proteína CP age realmente como um mimético de IL-8/NAP-2, ativando ou bloqueando as vias de sinalização de atividade inflamatória, seriam essenciais para testar as hipóteses formuladas e poderiam explicar possíveis modos de ação para essa classe de elicitores fúngicos.

## CONCLUSÕES

Através dos estudos estruturais e dinâmicos da proteína elicitora de respostas de defesa em plantas CP foi possível demonstrar que:

- A estrutura tridimensional da CP é formada por duas pequenas  $\alpha$ -hélices, uma hélice-3<sub>10</sub>, duas fitas- $\beta$  paralelas e cinco fitas- $\beta$  antiparalelas, constituindo uma única folha- $\beta$ .
- A estrutura da CP possui pouca flexibilidade e não apresenta mudanças conformacionais em escalas de tempo de pico a nano segundo.
- A CP possui uma alta estabilidade frente a variação de temperatura, pH e presença de uréia.
- A CP não tem habilidade de ligar N-acetilglicosamina ou a íons de  $\text{Cu}^{2+}$ .
- A CP possui um enovelamento similar a forma dimérica das quimiomicinas IL-8 e NAP-2 e ao domínio  $\alpha 1$  e  $\alpha 2$  do MHC-I e portanto a nenhuma outra família de proteínas envolvidas em avirulência de microorganismos ou a proteínas elicitoras de resposta de defesa em plantas.

**REFERÊNCIAS BIBLIOGRÁFICAS**

- Alami I., Mari S., Clériveret A. (1998) A glycoprotein from *Ceratocystis fimbriata* f. sp. *platani* triggers phytoalexin synthesis in *Platanus x Acerifolia* cell suspension cultures. *Phytochemistry*, 48: 771-776.
- Ausubel F.M. (2005) Are innate immune signalling pathways in plants and animals conserved? *Nat Immunol*, 6: 973-979.
- Baker D. & Sali A. (2001) Protein structure prediction and structural genomics. *Science*, 294: 93-96.
- Baker C.J., Harrington T.C., Krauss U., Alfenas A. (2003) Genetic variability and host specialization in the Latin American clade of *Ceratocystis fimbriata*. *Phytopathology*, 93: 1274-1284.
- Balloy V., Sallenave J.M., Wu Y., Touqui L., Latgé J.P., Si-Tahar M., Chignard M. (2008) *Aspergillus fumigatus*-induced interleukin-8 synthesis by respiratory epithelial cells is controlled by the phosphatidylinositol 3-kinase, p38 MAPK, and ERK1/2 pathways and not by the Toll-like Receptor-MyD88 pathway. *J Biol Chem*, 283: 30513-30521.
- Bax A., Kontaxis G., Tjandra N. (2001) Dipolar couplings in macromolecular structure determination. *Methods Enzymol*, 339: 127-174.
- Boddi S., Comparini C., Calamassi R., Pazzagli L., Cappugi G., Scala A. (2004) Cerato-Platanin protein is located in the cell walls of ascospores, conidia and hyphae of *Ceratocystis fimbriata* f. sp. *platani*. *FEMS Microbiol Lett*, 233: 341-346.
- Boissy G., Fortelle E.L., Kahn R., Huet J., Bricogne G., Pernollet J., Brunie S. (1996) Crystal structure of a fungal elicitor secreted by *Phytophthora cryptogea*, a member of a novel class of plant necrotic proteins. *Structure*, 4: 1429-1439.
- Buchanan B.B., Gruissem W., Jones R.L. (2000) *Biochemistry and Molecular Biology of Plants*, 1st ed., Ed. American Society of Plant Physiologist, Rockville, MD., USA.
- Carresi L., Pantera B., Zoppi C., Cappugi G., Oliveira A.L., Pertinhez T.A., Spisni A., Scala A., Pazzagli L. (2006) Cerato-platanin, a phytotoxic protein from *Ceratocystis fimbriata*: expression in *Pichia pastoris*, purification and characterization. *Protein Expr Purif*, 49:159-167.
- Castillo R.M., Mizuguchi K., Dhanaraj V., Albert A., Blundell T.L., Murzin A.G. (1999) A six-stranded double-psi beta barrel is shared by several protein superfamilies. *Structure*, 15: 227-236.

- Chisholm S.T., Coaker G., Day B., Staskawicz B.J. (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*, 124: 803-814.
- Cicero D.O., Contessa G.M., Paci M., Bazzo R. (2006) HACACO revisited: residual dipolar coupling measurements and resonance assignments in proteins. *J Magn Reson*, 180: 222-228.
- Claridge T.D.W. (1999) High-resolution NMR techniques in organic chemistry, 1st ed., Ed. Pergamon, Oxford.
- Cole G.T., Zhu S.W., Pan S.C., Yuan L., Kruse D., Sun S.H. (1989) Isolation of antigens with proteolytic activity from *Coccidioides immitis*. *Infect Immun*, 57: 1524-1534.
- Cornileuscu G., Delaglio F., Bax A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, 13: 289-302.
- Cozzini P., Kellogg G.E., Spyraakis F., Abraham D.J., Costantino G., Emerson A., Fanelli F., Gohlke H., Kuhn L.A., Morris G.M., Orozco M., Pertinhez T.A., Rizzi M., Sotriffer C.A. (2008) Target Flexibility: an emerging consideration in drug discovery and design. *J Med Chem*, 51: 6237-6255.
- Dangl J.L. & Jones J.D. (2001) Plant pathogens and integrated defence responses to infection. *Nature*, 411: 826-833.
- de Alba E. & Tjandra N. (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. *Prog Nucl Magn Reson Spectrosc*, 40: 175-197.
- Delaglio F., Grzesiek S., Vuister G.W., Zhu G., Pfeifer J., Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 6: 277-293.
- Delaglio F., Torchia D.A., Bax A. (1991) Measurement of  $^{15}\text{N}$ - $^{13}\text{C}$  J coupling in staphylococcal nuclease. *J Biomol NMR*, 1: 439-446.
- Djonović S., Pozo M.J., Dangott L.J., Howell C.R., Kenerley C.M. (2006) Sm1, a proteinaceous elicitor secreted by the biocontrol fungus *Trichoderma virens* induces plant defense responses and systemic resistance. *Mol Plant Microbe Interact*, 19: 838-853.
- Djonović S., Vargas W.A., Kolomiets M.V., Horndeski M., Wiest A., Kenerley C.M. (2007) A proteinaceous elicitor Sm1 from the beneficial fungus *Trichoderma virens* is required for induced systemic resistance in maize. *Plant Physiol*, 145: 875-889.
- Ellis J., Catanzariti A.M., Dodds P. (2006) The problem of how fungal and oomycete avirulence proteins enter plant cells. *Trends Plant Sci*, 11: 61-63.

- Engelbrecht C.J.B., Harrington T.C., Steimel J., Capretti P. (2004) Genetic variation in eastern North American and putatively introduced populations of *Ceratocystis fimbriata* f. *platani*. *Mol Ecology*, 13: 2995-3005.
- Fefe S., Bouaziz S., Huet J., Pernollet J., Guittet E. (1997) Three-dimensional solution structure of  $\beta$ -cryptogein, a  $\beta$ -elicitor secreted by a phytopathogenic fungus *Phytophthora cryptogea*. *Protein Sci*, 6: 2279-2284.
- Ferentz A.E. & Wagner G. (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q Rev Biophys*, 33: 29-65.
- Ferreira F.A., Maffia L.A., Ferreira E.A. (2005) Detecção rápida de *Ceratocystis fimbriata* em lenho infectado de eucalipto, mangueira e outros hospedeiros lenhosos. *Fitopatol Bras*, 30: 543-545.
- Ferreira F.A., Maffia L.A., Barreto R.W., Demuner N.L., Pigatto S. (2006) Sintomologia da murcha de *Ceratocystis fimbriata* em eucalipto. *R. Árvore*, 30: 155-162.
- Finn R.D., Tate J., Misty J., Coghill P.C., Sammut J.S., Hotz H.R., Ceric G., Forslund K., Eddy S.R., Sonnhammer E.L., Bateman A. (2008) The Pfam protein families database. *Nucleic Acids Res Database*, 36: D281-D288.
- Flor H.H. (1971) Current status of gene-for-gene concept. *Annu Rev Phytopathol*, 9: 275-296.
- Fontana F., Santini A., Salvini M., Pazzagli L., Capuggi G., Scala A., Durante M., Bernardi R. (2008) Cerato-platanin treated plane leaves restrict *Ceratocystis platani* growth and overexpress defence-related genes. *J Plant Pathol*, 90: 293-304.
- Freialdenhoven A., Scherag B., Hollricher K., Collinge D.B., Thordal-Christensen H., Schulze-Lefert P. (1994) Nar-1 and Nar-2, two loci required for Mla12-specified race-specific resistance to powdery mildew in barley. *Plant Cell*, 6: 983-994.
- Gan H.H., Perlow R.A., Roy S., Ko J., Wu M., Huang J., Yan S., Nicoletta A., Vafai J., Sun D., Wang L., Noah J.E., Pasquali S., Schlick T. (2002) Analysis of protein sequence/structure similarity relationships. *Biophysical J*, 83: 2781-2791.
- Gerber N., Lowman H., Artis D.R., Eigenbrot C. (2000) Receptor-binding conformation of the "ELR" motif of IL-8: X-ray structure of the L5C/H33C variant at 2.35 Å resolution. *Proteins*, 38: 361-367.
- Gómez-Gómez L. (2004) Plant perception systems for pathogen recognition and defence. *Mol Immunol*, 41: 1055-1062.
- Grzesiek S. & Bax A. (1992a) Improved 3D triple resonance NMR techniques applied to a 31KDa protein. *J Magn Reson*, 96: 432-440.

- Grzesiek S. & Bax A. (1992b) Correlation backbone amide and side chain resonance in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*, 114: 6291-6293.
- Hakanpää J., Paananen A., Askolin S., Nakari-Setälä T., Parkkinen T., Penttilä M., Linder M.B., Rouvinen J. (2004) Atomic resolution structure of the HFBII hydrophobin, a self-assembling amphiphile. *J Biol Chem*, 279: 534-539.
- Hans M.R., Mueller L., Pardi A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol*, 5: 1065-1074.
- Harrington, T.C. (2000) Host specialization and speciation in the american wilt pathogen *Ceratocystis fimbriata*. *Fitopatol Bras*, 25: 262-263.
- Hasan Z., Mahmood A., Zafar S., Khan A.A., Hussain R. (2004) Leprosy patients with lepromatous disease have an up-regulated IL-8 response that is unlinked to TNF-alpha responses. *Int J Lepr Other Mycobact Dis*, 72: 35-44.
- Henrissat B & Bairoch A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, 293: 781-8.
- Hickman-Miller H.D., Bardet W., Gilb A., Luis A.D., Jackson K.W., Watkins D.I., Hildebrand W.H. (2005) Rhesus macaque MHC class I molecules present HLA-B-like peptides. *J Immunol*, 175: 367-375.
- Hwang T.L, van Zijl P.C., Moris S. (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN Chemical EXchange (CLEANEX-PM) approach with a fast-HSQC (FHSQC) detection scheme. *J Biomol NMR*, 11: 221-226.
- Ikura M., Kay L.E., Bax A. (1990) A novel approach for sequential assignment of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, 29: 4659-4667.
- Ishima R. & Torchia D.A. (2000) Protein dynamics from NMR. *Nature Struc Biol*, 7: 740-743.
- Jeong J.S., Mitchell T.K., Dean R.A. (2007) The *Magnaporthe grisea* snodprot1 homolog, MSP1, is required for virulence. *FEMS Microbiol Lett*, 273: 157-165.
- Johnson B. & Blevins R.A. (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J Biomol NMR*, 4: 603-614.
- Johnson J.A., Harrington T.C., Engelbrecht C.J. (2005) Phylogeny and taxonomy of the North American clade of the *Ceratocystis fimbriata* complex. *Mycologia*, 97: 1067-1092.

- Jones J.A., Wilkins D.K., Smith L.J., Dobson C.M. (1997) Characterization of protein unfolding by NMR diffusion measurements. *J Biomol NMR*, 10: 199-203.
- Jones J.D.G. & Dangl J.L. (2006) The plant immune system. *Nature*, 444: 323-329.
- Juge N. (2006) Plant protein inhibitors of cell wall degrading enzymes. *Trends Plant Sci*, 11: 359-367.
- Kay L.E. (1997) NMR methods for the study of protein structure and dynamics. *Biochem. Cell Biol.*, 75: 1-15.
- Kay L.E., Keifer P., Saarinen T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc*, 114: 10663-10665.
- Kay L.E., Xu G.Y., Singer A.U., Muhandiram D.R., Forman-Kay J.D. (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain  $^1\text{H}$  and  $^{13}\text{C}$  correlations in  $\text{H}_2\text{O}$  samples of proteins. *J Magn Reson*, B 101: 333-337.
- Kwan A.H.Y., Winefield R.D., Sunde M., Matthews J.M., Haverkamp R.G., Templeton M.D., Mackay J.P. (2006) Structural basis for rodlet assembly in fungal hydrophobins. *Proc Natl Acad Sci USA*, 103: 3621-3626.
- Laskowski R.A., Watson J.D., Thornton J.M. (2005) ProFunc: a server for prediction protein function from 3d structure. *Nucleic Acids Res*, 33: 89-93.
- Luderer R., Rivas S., Nürnberger T., Mattei B., van den Hooven H.W., van der Hoorn R.A., Romeis T., Wehrfritz J.M., Blume B., Nennstiel D., Zuidema D., Vervoort J., de Lorenzo G., Jones J.D., De Wit P.J., Joosten M.H. (2001) No evidence for binding between resistance gene product Cf-9 of tomato and avirulence gene product AVR9 of *Cladosporium fulvum*. *Mol Plant Microbe Interact*, 14: 867-876.
- Ludvigsen S. & Poulsen F.M. (1992) Three-dimensional structure in solution of barwin, a protein from barley seed. *Biochemistry*, 31: 8783-8789.
- Ma J.K., Bishop G.R., Davidson V.L. (2005) The ligand geometry of copper determines the stability of amicyanin. *Arch Biochem Biophys*, 444: 27-33.
- Malkowski M.G., Lazar J.B., Johnson P.H., Edwards B.F.P. (1997) The amino-terminal residues in the crystal structure of connective tissue activating peptide-III (des10) block the ELR chemotactic sequence. *J Mol Biol*, 266: 367-380.
- McFeeters R.L., Fowler A., Gaponenko V.V., Byrd A. (2005) efficient and precise measurement of  $\text{H}\alpha\text{-C}\alpha$ ,  $\text{C}\alpha\text{-C}'$ ,  $\text{C}\alpha\text{-C}\beta$  and  $\text{HN-N}$  residual dipolar couplings from 2D  $\text{HN-N}$  correlation spectra. *J Biomol NMR*, 31: 35-47.

- Mehta A., Brasileiro A.C., Souza D.S., Romano E., Campos M.A., Grossi-de-Sá M.F., Silva M.S., Franco O.L., Fragoso R.R., Bevitori R., Rocha T.L. (2008) Plant-pathogen interactions: what is proteomics telling us? *FEBS J*, 275: 3731-3746.
- Montelione G.T., Zheng D., Huang Y.J., Gunsalus K.C., Szyperski T. (2000) Protein NMR spectroscopy in structural genomics. *Nat Struct Biol*, 7, Suppl: 982-985.
- Neuhaus J.M., Flores S., Keefe D., Ahl-Goy P., Meins F.Jr. (1992) The function of vacuolar beta-1,3-glucanase investigated by antisense transformation. Susceptibility of transgenic *Nicotiana glauca* plants to *Cercospora nicotianae* infection. *Plant Mol Biol*, 19: 803-813.
- O'Connell R.J. & Panstruga R. (2006) Tête à tête inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes. *New Phytol*, 171: 699-718.
- Ottiger M. & Bax A. (1999) Bicelle-based liquid crystals for NMR measurement of dipolar couplings at acidic and basic pH values. *J Mol Biol*, 13: 187-191.
- Pascal S.M., Muhandiram D.R., Yamazaki T., Formankay J.D., Kay L.E. (1994) Simultaneous acquisition of  $^{15}\text{N}$ - and  $^{13}\text{C}$ -edited NOE spectra of proteins dissolved in  $\text{H}_2\text{O}$ . *J Magn Reson*, B 103: 197-201.
- Patt S.L. (1992) Single- and multiple-frequency-shifted laminar pulses. *J Magn Reson*, 96: 94-102.
- Pazzagli L., Cappugi G., Manao G., Camici G., Santini A., Scala, A. (1999) Purification, characterization and amino acid sequence of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata* f. sp. *platanii*. *J. Biol. Chem.*, 274: 24959-24964.
- Pertinhez T.A., Hamada D., Smith L.J., Chiti F., Taddei N., Stefani M., Dobson C.M. (2000) Initial denaturation conditions influence the slow folding phase of acylphosphatase associated with proline isomerization. *Protein Science*, 9: 1466-1473.
- Prestegard J.H., Mayer K.L., Valafar H., Benison G.C. (2005) Determination of protein backbone structures from residual dipolar couplings. *Methods Enzymol*, 394: 175-209.
- Rajagopalan L. & Rajarathnam K. (2006) Structural basis of chemokine receptor function-- a model for binding affinity and ligand selectivity. *Biosci Rep*, 26: 325-339.
- Rementeria A., López-Molina N., Ludwig A., Vivanco A.B., Bikandi J., Pontón J., Garaizar J. (2005) Genes and molecules involved in *Aspergillus fumigatus* virulence. *Rev Iberoam Micol*, 22: 1-23.

- Rodrigues M.L., Archer M., Martel P., Jacquet A., Cravador A., Carrondo M.A. (2002) Structure of  $\beta$ -cinnamoin, a protein toxic to some plant species. *Acta Cryst*, D158: 1314-1321.
- Rooney H.C., van't Klooster J.W., van der Hoorn R.A., Joosten M.H., Jones J.D., de Wit P.J. (2005) Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science*, 308: 1783-1786.
- Rudolph M.G., Stanfield R.L., Wilson I.A. (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol*, 24: 419-466.
- Santos A.F. & Ferreira F.A. (2003) Murcha-de-Ceratocystis em acácia-negra no Brasil. *Fitopatol Bras*, 28: 325.
- Saupe A. (1968) Recent results in the field of liquid crystals. *Angew Chem Int Ed Engl*, 7: 97-112.
- Sbrana F., Bongii L., Cappugi G., Fanelli D., Guarino A., Pazzagli L., Scala A., Vassalli M., Zoppi C., Tiribilli B. (2007) Atomic force microscopy images suggest aggregation mechanism in cerato-platanin. *Eur Biophys J*, 36: 727-732.
- Scala A., Pazzagli L., Comparini C., Santini A., Tegli S., Cappugi G. (2004) Cerato-platanin, an early-produced protein by *Ceratocystis fimbriata* f. sp. *Platani*, elicits phytoalexin synthesis in host and non-host plants. *J Plant Pathol*, 86: 23-29.
- Seidl V., Marchetti M., Schandl R., Allmaier G., Kubicek C.P. (2006) Ep11, the major secreted protein of *Hypocrea atroviridis* on glucose, is a member of a strongly conserved protein family comprising plant defense response elicitors. *FEBS J*, 273: 4346-4359.
- Shalit I., Halperin D., Haite D., Levitov A., Romano J., Osherov N., Fabian I. (2006) Anti-inflammatory effects of moxifloxacin on IL-8, IL-1beta and TNF-alpha secretion and NFkappaB and MAP-kinase activation in human monocytes stimulated with *Aspergillus fumigatus*. *J Antimicrob Chemother*, 57: 230-235.
- Sheu F., Hsieh K., Chien P., Tsao, C., Chin K. (2005) Protein ACA1 of *Antrrodia camphorata*. *United States Patent Application Publication*, Pub. No.: US 2005/0164931 A1.
- Silva S.D.V.M., Paim M.C., Castro W.M. (2004) Cacau "Jaca" resistente a *Ceratocystis fimbriata* na região cacaeira da Bahia, Brasil. *Fitopatol Bras*, 29: 538-540.
- Skelton N.J., Quan C., Reily D., Lowman H. (1999) Structure of a CXC chemokine-receptor fragment in complex with interleukin-8. *Structure*, 7: 157-168.

- Spronk C.A.E.M., Nabuurs S.B., Krieger E., Vriend G., Vuister G.W. (2004) Validation of protein structures derived by NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc*, 45: 315-337.
- Surridge C. (2001) Plant defence. *Nature*, 411: 825.
- Svensson B., Svendsen I., Højrup P., Roepstorff P., Ludvigsen S., Poulsen F.M. (1992) Primary structure of barwin: a barley seed protein closely related to the C-terminal domain of proteins encoded by wound-induced plant genes. *Biochemistry*, 31: 8767-8770.
- van Straaten K.E., Dijkstra B.W., Vollmer W., Thunnissen A.M. (2005) Crystal structure of MltA from *Escherichia coli* reveals a unique lytic transglycosylase fold. *J Mol Biol*, 352: 1068-1080.
- van't Slot K.A.E., van der Burg A.A., Kloks C.P.A.M, Hilbers C.H., Knogge W., Papavoine C.H.M. (2003) Solution structure of the plant disease resistance-triggering protein NIP1 from the fungus *Rhynchosporin secalis* shows a novel  $\beta$ -sheet fold. *J Biol Chem*, 278: 45730-45736.
- van't Slot K.A., Gierlich A., Knogge W. (2007) A single binding site mediates resistance- and disease-associated activities of the effector protein NIP1 from the barley pathogen *Rhynchosporium secalis*. *Plant Physiol*, 144, 1654-1666.
- Vargas W.A., Djonović S., Sukno S.A., Kenerley C.M. (2008) Dimerization controls the activity of fungal elicitors that trigger systemic resistance in plants. *J Biol Chem*, 11: 19804-19815.
- Vieille C. & Zeikus G.J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*, 65: 1-43.
- Vuister G.W. & Bax A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond  $J(\text{H}^{\text{N}}\text{H}^{\alpha})$  coupling constants in  $^{15}\text{N}$ -enriched proteins. *J Am Chem Soc*, 115: 7772-7777.
- Wang C.A., Guncar G., Forwood J.K., The T., Catanzariti A., Lawrence G.J., Loughlin F.E., Schirra H.J., Anderson P.A., Ellis J.G., Dodds P.N., Kobe B. (2007) Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell*, 19: 2898-2912.
- Wider G. (2000) Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy. *Biotechniques*, 29: 1278-1294.
- Wilkins D.K., Grimshaw S.B., Receveur V., Dobson C.M., Jones J.A., Smith L.J (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, 38: 16424-16431.

- Wilson L.M., Idnurm A., Howlett B. (2002) Characterization of a gene (sp1) encoding a secreted protein from *Leptosphaeria maculans*, the blackleg pathogen of *Brassica napus*. *Mol Plant Pathol*, 3: 487-493.
- Wishart D.S & Sykes B.D. (1994). The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J Biomol NMR*, 4: 171-180.
- Wishart D.S. & Nip A.M. (1998) Protein chemical shift analysis: a practical guide. *Biochem Cell Biol*, 76: 153-163
- Wittekind M. & Mueller L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson*, B 101: 201-205.
- Wu D.H., Chen A., Johnson C.S. (1995) Flow imaging by means of 1d pulsed-field-gradient NMR with application to electro osmotic flow. *J Magn Reson*, A 115: 123-126.
- Wüthrich K. (1986) NMR of proteins and nucleic acids, 1st ed., Ed. Wiley-Interscience, New York.
- Yip C.K. & Strynadka N.C. (2006) New structural insights into the bacterial type III secretion system. *Trends Biochem Sci*, 31: 223-230.
- Zaparoli G., Cabrera O.G., Medrano F.J., Tiburcio R., Lacerda G., Pereira G.G. (2008) Identification of a second family of genes in *Moniliophthora perniciosa*, the causal agent of witches' broom disease in cacao, encoding necrosis-inducing proteins similar to cerato-platanins. *Mycol Res*, in press (doi:10.1016/j.mycres.2008.08.004).

## DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha tese de Doutorado intitulada “Estudos estruturais e dinâmicos da proteína elicitora Cerato-Platanina”:

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

( x ) está inserido no **Projeto CIBio** (Protocolo nº CEB 5.1), intitulado Resolução da estrutura tridimensional da proteína elicitora de resposta de defesa, Cerato-platanina;

( ) tem autorização da **Comissão de Ética em Experimentação Animal/IB/UNICAMP** (Protocolo nº \_\_\_\_\_);

( ) tem autorização do **Comitê de Ética para Pesquisa com Seres Humanos/FCM/UNICAMP** (Protocolo nº \_\_\_\_\_);

( ) tem autorização de comissão de bioética ou biossegurança externa à UNICAMP.  
Especificar: \_\_\_\_\_

*Aline Lima de Oliveira*

Aluno: Aline Lima de Oliveira

*Celso Eduardo Benedetti*

Co-orientador: Celso Eduardo Benedetti

Para uso da Comissão ou Comitê pertinente:

(  ) Deferido ( ) Indeferido

*Helena Coutinho F. de Oliveira*

Nome:

Função:

Profa. Dra. HELENA COUTINHO F. DE OLIVEIRA  
Presidente  
Comissão Interna de Biossegurança  
CIBio/IB - UNICAMP

CEB 5.1

Uso exclusivo da CIBio:

**Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança da ABTLuS - Associação Brasileira de Tecnologia de Luz Síncrotron**

Título do projeto: Resolução da estrutura tridimensional da proteína elicitora de resposta de defesa, Cerato-platanina

Pesquisador responsável: Thelma Aguiar Pertinhez / Celso Eduardo Benedetti

Experimentador: Aline Lima de Oliveira

Nível do treinamento do experimentador:  -Iniciação científica,  -mestrado,  -doutorado,  -doutorado direto,  -pós-doutorado,  -nível técnico,  -outro, especifique: \_\_\_\_\_

Resumo do projeto:

O presente projeto tem como objetivo central a resolução da estrutura atômica da proteína Cerato-platanina (CP), um elicitor de resposta de defesa em plantas. A proteína CP do fungo *Ceratocystis platani* será produzida em células de *Pichia pastoris* e purificada por cromatografia de fase reversa. A proteína será marcada com isótopos  $^{15}\text{N}$  e  $^{13}\text{C}$  para estudos estruturais por ressonância magnética nuclear. Além disso, serão estudadas as propriedades de dinâmica da proteína em solução.

A CIBio analisou este projeto em reunião realizada no dia: 30.10.2008.

Parecer final:  -projeto aprovado,  -projeto recusado,  -projeto com deficiências, favor comentários abaixo:

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Jörg Kobarg  
Presidente de CIBio - ABTLuS  
Prof. Dr. Jörg Kobarg

Celso Eduardo Benedetti  
Membro da CIBio - ABTLuS  
Prof. Dr. Celso Eduardo Benedetti

Nilson Ivo Tonin Zanchin  
Membro da CIBio - ABTLuS  
Prof. Dr. Nilson Ivo Tonin Zanchin

**- ANEXO 1 -**

**Tabela.** Deslocamentos químicos ( $\delta$ , ppm) dos átomos de  $^{15}\text{N}$ ,  $^{13}\text{C}$  e  $^1\text{H}$  para os resíduos de amino ácido da proteína CP à concentração 0,7 mM em tampão fosfato 10 mM, pH 5,8 e à 20°C.

Resíduo	N	H <sup>N</sup>	C <sup>α</sup>	H <sup>α</sup>	C <sup>β</sup>	H <sup>β</sup>	Outros	C <sup>γ</sup>
<b>1 Glu</b>	-	-	54,50	4,29	28,26	2,00/2,05	C <sup>γ</sup> 34,28; H <sup>γ</sup> 2,25	176,06
<b>2 Lys</b>	124,85	8,38	54,12	4,40	31,33	1,78	C <sup>γ</sup> 22,70; H <sup>γ</sup> 1,49; C <sup>δ</sup> 30,97; H <sup>δ</sup> 1,94; C <sup>ε</sup> 39,87; H <sup>ε</sup> 3,10	175,78
<b>3 Arg</b>	124,56	8,40	53,37	4,94	30,34	1,87	C <sup>γ</sup> 24,97; H <sup>γ</sup> 1,63; C <sup>δ</sup> 41,35; H <sup>δ</sup> 3,23/3,13	176,00
<b>4 Val</b>	121,26	8,74	58,72	4,74	33,21	2,48	C <sup>γ</sup> 20,05; H <sup>γ</sup> 0,98	173,68
<b>5 Ser</b>	118,45	8,40	55,24	5,58	64,15	4,00/3,80	-	173,14
<b>6 Ile</b>	120,20	9,19	56,35	5,76	40,45	1,45	C <sup>γ</sup> 14,35; H <sup>γ</sup> 1,01; C <sup>δ</sup> 12,66; H <sup>δ</sup> 0,65/0,87	171,86
<b>7Ser</b>	123,24	8,66	54,77	5,02	63,78	3,93/3,78	-	172,54
<b>8 Tyr</b>	118,88	7,58	52,90	5,75	39,35	2,65/2,29	C <sup>δ</sup> 133,22; H <sup>δ</sup> 6,14; C <sup>ε</sup> 120,06; H <sup>ε</sup> 6,72	174,00
<b>9 Asp</b>	121,30	8,40	48,69	4,99	42,58	2,87/2,63	-	-
<b>10 Pro</b>	-	-	61,06	4,35	30,62	2,48/2,18	C <sup>γ</sup> 24,83; H <sup>γ</sup> 2,12; C <sup>δ</sup> 50,05; H <sup>δ</sup> 4,13/3,68	176,81
<b>11 Ile</b>	122,01	8,64	64,24	4,00	35,75	1,62	C <sup>γ1</sup> 26,29; H <sup>γ1</sup> 1,05/1,06; C <sup>δ</sup> 11,64; H <sup>δ</sup> 0,73; C <sup>γ2</sup> 14,27; H <sup>γ2</sup> 0,10	177,59
<b>12 Tyr</b>	121,97	6,91	56,16	4,62	34,15	3,34/2,73	C <sup>δ</sup> 133,22; H <sup>δ</sup> 6,14; C <sup>ε</sup> 120,06; H <sup>ε</sup> 6,72	173,44
<b>13 Ala</b>	128,44	8,14	48,27	4,79	20,28	1,66	-	177,28
<b>14 Ala</b>	121,39	8,97	52,11	4,01	16,77	1,38	-	177,80
<b>15 Asp</b>	111,85	8,03	50,69	4,49	38,00	2,89/2,66	-	175,69
<b>16 Leu</b>	124,33	7,49	54,09	4,17	40,71	1,80/2,13	C <sup>γ</sup> 24,54; H <sup>γ</sup> 0,99	176,63
<b>17 Ser</b>	122,08	8,77	56,24	4,49	61,80	3,88	-	177,45
<b>18 Met</b>	128,93	8,55	52,90	4,87	26,85	2,04	C <sup>γ</sup> 31,64; H <sup>γ</sup> 3,12;	175,98
<b>19 Gly</b>	105,64	8,42	44,46	3,87/4,43	-	-	-	175,99
<b>20 Ser</b>	116,22	8,06	56,89	4,74	62,44	4,20	-	173,44
<b>21 Val</b>	111,46	6,85	57,11	4,76	33,68	2,67	C <sup>γ1</sup> 20,34; H <sup>γ1</sup> 1,06; C <sup>γ2</sup> 17,29; H <sup>γ2</sup> 1,062	176,00
<b>22 Ala</b>	121,35	8,76	52,90	4,57	16,94	1,55	-	178,78
<b>23 Cys</b>	116,39	6,91	50,15	5,24	35,08	3,18	-	174,65
<b>24 Ser</b>	117,63	7,81	54,77	4,86	62,21	3,73	-	175,31

<b>25 Asn</b>	123,86	9,65	50,25	4,99	36,80	3,36/3,06	-	174,68
<b>26 Gly</b>	109,32	8,52	42,62	3,74/4,24	-	-	-	175,08
<b>27 Asp</b>	123,52	9,10	54,95	4,24	38,83	2,17/2,02	-	177,75
<b>28 His</b>	119,28	10,29	52,41	4,74	27,10	2,86	C <sup>ε</sup> 138,57; H <sup>ε</sup> 7,70	174,48
<b>29 Gly</b>	107,40	7,19	43,41	3,62	-	-	-	173,54
<b>30 Leu</b>	119,95	7,79	53,36	4,79	44,37	2,06	C <sup>γ</sup> 23,96; H <sup>γ</sup> 1,17; C <sup>δ</sup> 24,06; H <sup>δ</sup> 1,08	177,18
<b>31Met</b>	121,13	8,47	55,95	4,31	31,96	2,02/1,69	C <sup>γ</sup> 31,79; H <sup>γ</sup> 2,51/2,26	175,66
<b>32 Ala</b>	122,39	8,62	52,43	3,10	15,86	1,27	-	179,06
<b>33 Gln</b>	115,66	7,22	55,17	3,97	28,27	1,27/0,96	C <sup>γ</sup> 32,35; H <sup>γ</sup> 2,02/1,92	175,40
<b>34 Tyr</b>	120,33	7,97	50,58	5,11	37,65	2,77/3,09	C <sup>δ1</sup> 134,64; H <sup>δ1</sup> 7,23; C <sup>δ2</sup> 119,14; H <sup>δ2</sup> 6,85; C <sup>ε1</sup> 119,14; H <sup>ε1</sup> 6,85; C <sup>ε1</sup> 134,64; H <sup>ε1</sup> 7,20	-
<b>35 Pro</b>	-	-	63,42	4,43	30,86	2,49/2,10	C <sup>γ</sup> 26,28; H <sup>γ</sup> 2,30/1,91; C <sup>δ</sup> 49,03; H <sup>δ</sup> 3,94/4,20	178,77
<b>36 Thr</b>	107,61	7,39	56,81	5,24	70,34	4,40	C <sup>γ</sup> 19,76; H <sup>γ</sup> 1,16	174,82
<b>37 Leu</b>	125,27	8,90	55,16	4,40	39,51	2,39/1,60	C <sup>γ</sup> 22,86; H <sup>γ</sup> 1,29; C <sup>δ1</sup> 21,50; H <sup>δ1</sup> 0,88; C <sup>δ1</sup> 22,76; H <sup>δ1</sup> 0,91	177,91
<b>38 Gly</b>	105,43	8,99	43,05	3,12/2,12	-	-	-	174,63
<b>39 Glu</b>	118,88	7,46	54,07	4,15	29,93	2,60/2,22	C <sup>γ</sup> 35,99; H <sup>γ</sup> 2,20	175,69
<b>40 Val</b>	124,27	7,81	58,86	2,99	30,58	2,46	C <sup>γ1</sup> 21,24; H <sup>γ1</sup> 1,23; C <sup>γ2</sup> 18,10; H <sup>γ2</sup> 0,68	176,00
<b>41 Pro</b>	-	-	62,20	4,23	30,33	2,42/1,57	-	175,69
<b>42 Gly</b>	115,78	9,00	42,11	4,42	-	-	-	174,75
<b>43 Phe</b>	125,69	8,53	57,58	3,89	37,90	3,55/3,06	-	-
<b>44 Pro</b>	-	-	61,27	4,43	31,79	2,35/1,96	C <sup>γ</sup> 25,41; H <sup>γ</sup> 2,12; C <sup>δ</sup> 49,03; H <sup>δ</sup> 3,73/3,95	174,37
<b>45 Asn</b>	132,29	8,94	50,08	4,87	33,68	3,52/3,18	-	174,92
<b>46 Val</b>	113,68	8,28	55,69	5,74	35,51	2,17	C <sup>γ1</sup> 16,85; H <sup>γ2</sup> 0,87 C <sup>γ2</sup> 18,64; H <sup>γ2</sup> 0,87	174,37
<b>47 Gly</b>	103,03	7,24	42,66	4,30	-	-	-	172,02
<b>48 Gly</b>	112,52	9,26	42,62	4,44/3,24	-	-	-	173,19
<b>49 Ile</b>	113,40	7,15	55,47	4,31	35,55	0,975	C <sup>γ</sup> 14,97; H <sup>γ</sup> 0,63	-
<b>50 Pro</b>	-	-	63,21	3,65	29,18	1,71/1,44	C <sup>γ</sup> 25,26; H <sup>γ</sup> 0,82/0,51; C <sup>δ</sup> 47,00; H <sup>δ</sup> 2,17/1,04	178,06
<b>51 Asp</b>	113,51	6,83	54,25	4,24	38,83	2,79	-	174,93
<b>52 Ile</b>	119,40	6,56	59,03	4,27	36,02	1,89	C <sup>γ1</sup> 25,27; H <sup>γ1</sup> 1,51; C <sup>δ</sup> 25,20; H <sup>δ</sup> 1,24; C <sup>γ2</sup> 15,41; H <sup>γ2</sup> 0,91	174,26
<b>53 Ala</b>	131,46	8,61	50,22	4,49	17,86	1,43	-	176,81
<b>54 Gly</b>	105,69	7,24	42,58	4,10	-	-	-	170,80
<b>55 Trp</b>	121,53	8,62	57,11	4,65	26,65	3,47/3,28	C <sup>δ</sup> 128,98; H <sup>δ</sup> 7,67; C <sup>ε2</sup> 115,75; H <sup>ε2</sup> 7,36; C <sup>η</sup> 125,73; H <sup>η</sup> 7,27;	177,31

							C <sup>ζ3</sup> 123,51; H <sup>ζ3</sup> 7,34; C <sup>ε3</sup> 120,37; H <sup>ε3</sup> 7,61; N <sup>ε1</sup> 131,20; H <sup>ε1</sup> 9,82	
<b>56 Asp</b>	126,19	8,14	52,90	3,99	37,90	2,30	-	174,68
<b>57 Ser</b>	111,46	6,84	53,97	4,72	63,38	4,19/3,84	-	-
<b>58 Pro</b>	-	-	61,74	-	29,86	-	-	177,38
<b>59 Ser</b>	113,54	8,25	57,05	4,68	60,64	3,90/ 3,68	-	175,88
<b>60 Cys</b>	126,19	8,44	58,30	4,24	39,49	3,37/2,77	-	175,38
<b>61 Gly</b>	111,91	8,94	43,70	4,74/3,74	-	-	-	174,43
<b>62 Thr</b>	111,39	7,71	60,40	4,37	69,30	4,41	C <sup>γ</sup> 22,73; H <sup>γ</sup> 1,46	172,16
<b>63 Cys</b>	123,13	9,56	52,90	5,52	42,24	3,10/2,79	-	171,80
<b>64 Trp</b>	130,83	9,37	54,30	4,99	29,65	3,26/3,03	C <sup>δ1</sup> 128,10; H <sup>δ</sup> 7,38; C <sup>ζ2</sup> 115,00; H <sup>ζ2</sup> 7,38; C <sup>η</sup> 126,73; H <sup>η</sup> 7,35; C <sup>ζ3</sup> 123,36; H <sup>ζ3</sup> 7,25; C <sup>ε3</sup> 122,19; H <sup>ε3</sup> 7,63; N <sup>ε1</sup> 134,83; H <sup>ε1</sup> 11,07	174,56
<b>65 Lys</b>	126,93	9,68	53,77	4,74	32,38	1,98/1,65	C <sup>γ</sup> 23,09; H <sup>γ</sup> 0,90/0,60; C <sup>δ</sup> 27,29; H <sup>δ</sup> 1,56; C <sup>ε</sup> 39,61; H <sup>ε</sup> 2,70	178,77
<b>66 Val</b>	132,95	9,00	59,27	4,62	31,33	1,56	C <sup>γ1</sup> 18,80; H <sup>γ1</sup> 0,31; C <sup>γ2</sup> 18,41; H <sup>γ2</sup> 0,07	174,34
<b>67 Thr</b>	123,58	8,80	59,16	5,12	69,30	3,77	C <sup>γ</sup> 18,80; H <sup>γ</sup> 1,20	175,14
<b>68 Ile</b>	122,99	8,33	57,95	4,69	34,63	2,20	C <sup>γ1</sup> 17,44; H <sup>γ1</sup> 0,68; C <sup>δ</sup> 16,55; H <sup>δ</sup> 0,84; C <sup>γ2</sup> 18,51; H <sup>γ2</sup> 1,05	-
<b>69 Pro</b>	-	-	64,31	4,31	29,86	1,94	C <sup>γ</sup> 25,37; H <sup>γ</sup> 2,01; C <sup>δ</sup> 48,45; H <sup>δ</sup> 3,93/3,71	177,14
<b>70 Asn</b>	115,54	8,15	51,49	4,49	34,91	3,23/2,87	-	176,46
<b>71 Gly</b>	110,76	8,70	42,69	3,61/4,37	-	-	-	174,37
<b>72 Asn</b>	124,14	8,36	51,44	4,62	36,29	2,90	-	173,39
<b>73 Ser</b>	114,54	7,97	54,51	5,74	65,55	3,47/3,33	-	172,69
<b>74 Ile</b>	120,74	8,37	58,21	4,25	40,24	1,66	C <sup>γ1</sup> 25,99; H <sup>γ1</sup> 1,05; C <sup>δ</sup> 13,67; H <sup>δ</sup> 0,80; C <sup>γ2</sup> 14,69; H <sup>γ2</sup> 0,68	171,37
<b>75 Phe</b>	123,92	8,18	54,62	5,69	37,71	3,12/2,95	-	175,10
<b>76 Ile</b>	124,19	9,63	54,32	4,49	38,42	1,90	C <sup>γ1</sup> 22,95; H <sup>γ1</sup> 1,26; C <sup>δ</sup> 20,19; H <sup>δ</sup> 0,98; C <sup>γ2</sup> 15,05; H <sup>γ2</sup> 0,63	174,30
<b>77 Arg</b>	130,04	8,99	52,36	5,62	30,64	1,93/1,70	H <sup>γ</sup> 3,25	176,35
<b>78 Gly</b>	114,59	9,76	45,63	3,37/4,58	-	-	-	174,00
<b>79 Val</b>	116,33	8,54	58,05	4,72	32,32	2,55	C <sup>γ1</sup> 18,75; H <sup>γ1</sup> 0,91; C <sup>γ2</sup> 14,66; H <sup>γ2</sup> 0,69	171,94
<b>80 Asp</b>	117,25	7,50	49,44	5,24	42,58	2,71/2,46	-	173,67
<b>81 Ser</b>	118,57	8,08	55,29	4,87	61,80	3,72	-	175,94
<b>82 Gly</b>	117,15	8,07	42,47	3,75	-	-	-	173,28
<b>83 Arg</b>	121,72	8,66	54,55	3,87	25,51	1,83/1,46	C <sup>γ</sup> 35,64; H <sup>γ</sup> 0,23; C <sup>δ</sup> 41,35; H <sup>δ</sup> 3,25/3,16	176,79
<b>84 Gly</b>	106,23	8,49	42,76	3,57/4,24	-	-	-	173,64
<b>85 Gly</b>	109,53	7,11	42,10	4,19	-	-	-	170,60

<b>86 Phe</b>	115,43	8,26	55,01	4,97	40,26	2,86/2,65	-	175,69
<b>87 Asn</b>	127,40	8,54	50,26	5,37	38,36	2,65/2,98	-	174,55
<b>88 Val</b>	115,04	8,99	57,29	5,32	34,77	2,51	C <sup>γ1</sup> 16,30; H <sup>γ1</sup> 1,24; C <sup>γ2</sup> 20,05; H <sup>γ2</sup> 0,97	174,38
<b>89 Asn</b>	121,56	7,79	50,55	5,12	39,25	3,43/3,20	-	-
<b>90 Pro</b>	-	-	63,92	4,27	30,47	2,14	-	178,13
<b>91 Thr</b>	119,96	8,61	64,54	4,00	65,75	4,39	C <sup>γ</sup> 19,58; H <sup>γ1</sup> 1,27	176,44
<b>92 Ala</b>	127,81	7,87	53,16	3,99	16,62	1,62	-	178,21
<b>93 Phe</b>	118,21	8,14	60,40	3,66	38,40	3,19/2,89	C <sup>δ</sup> 133,80; H <sup>δ</sup> 7,29; C <sup>ε</sup> 133,72; H <sup>ε</sup> 7,62; C <sup>ζ</sup> 130,93; H <sup>ζ</sup> 7,10	176,44
<b>94 Thr</b>	116,83	8,43	65,08	3,56	66,71	4,30	C <sup>γ</sup> 19,19; H <sup>γ</sup> 1,18	176,63
<b>95 Lys</b>	124,72	7,94	57,00	3,99	29,93	1,94	C <sup>γ</sup> 22,86; H <sup>γ</sup> 1,55; C <sup>δ</sup> 26,91; H <sup>δ</sup> 1,74; C <sup>ε</sup> 39,95; H <sup>ε</sup> 2,98	177,60
<b>96 Leu</b>	119,80	7,51	55,24	4,04	40,27	1,71/1,27	C <sup>γ</sup> 19,47; H <sup>γ</sup> 1,27; C <sup>δ</sup> 23,90; H <sup>δ</sup> 0,57/0,66	175,99
<b>97 Val</b>	115,42	8,27	61,52	3,62	31,52	1,60	C <sup>γ1</sup> 19,97; H <sup>γ1</sup> 0,78; C <sup>γ2</sup> 19,50; H <sup>γ2</sup> 0,16	177,00
<b>98 Gly</b>	111,67	8,12	43,52	4,48/3,74	-	-	-	173,73
<b>99 Ser</b>	112,90	7,41	55,24	4,73	61,93	4,19/3,85	-	174,25
<b>100 Thr</b>	114,35	8,49	60,25	4,37	66,49	4,10	C <sup>γ</sup> 19,06; H <sup>γ</sup> 0,78	175,88
<b>101 Glu</b>	124,99	8,53	56,64	4,12	27,79	2,06	C <sup>γ</sup> 34,25; H <sup>γ</sup> 2,33	176,83
<b>102 Ala</b>	122,49	8,11	51,02	4,37	17,87	1,52	-	178,93
<b>103 Gly</b>	107,45	8,06	44,80	3,87/4,12	-	-	-	172,31
<b>104 Arg</b>	121,12	8,03	53,36	5,24	31,33	1,82	C <sup>γ</sup> 25,20; H <sup>γ</sup> 1,50; C <sup>δ</sup> 41,27; H <sup>δ</sup> 3,19	174,88
<b>105 Val</b>	126,33	9,06	59,91	4,37	32,01	1,93	C <sup>γ1</sup> 19,63; H <sup>γ1</sup> 1,10; C <sup>γ2</sup> 15,88; H <sup>γ2</sup> 0,80	174,51
<b>106 Asp</b>	127,65	8,56	52,11	4,89	41,18	2,77/2,55	-	175,04
<b>107 Asn</b>	115,78	8,51	52,39	4,39	35,35	3,03/2,85	-	174,09
<b>108 Val</b>	116,59	8,85	59,43	4,37	30,40	2,10	C <sup>γ1</sup> 18,95; H <sup>γ1</sup> 0,81	175,32
<b>109 Asn</b>	122,63	8,53	50,08	5,37	39,77	2,81/2,64	-	173,96
<b>110 Tyr</b>	119,69	8,59	53,65	6,12	39,84	3,16/2,95	C <sup>δ</sup> 134,07; H <sup>δ</sup> 7,08; C <sup>ε</sup> 119,50; H <sup>ε</sup> 6,81	173,21
<b>111 Val</b>	119,16	8,46	58,17	4,63	33,68	2,19	C <sup>γ1</sup> 18,49; H <sup>γ1</sup> 1,05; C <sup>γ2</sup> 18,17; H <sup>γ2</sup> 0,93	172,12
<b>112 Gln</b>	132,30	8,68	54,23	3,49	24,89	1,27/0,49	C <sup>γ</sup> 31,37; H <sup>γ</sup> 2,05/1,48	175,69
<b>113 Val</b>	120,33	8,39	56,62	4,71	34,61	2,24	C <sup>γ1</sup> 18,25; H <sup>γ1</sup> 1,01; C <sup>γ2</sup> 14,82; H <sup>γ2</sup> 0,68	174,57
<b>114 Asp</b>	120,19	8,24	53,35	4,56	40,71	2,53	-	177,75
<b>115 Leu</b>	127,40	8,87	56,18	3,97	40,30	1,77	C <sup>γ</sup> 23,24; H <sup>γ</sup> 1,06	177,48
<b>116 Ser</b>	116,33	8,75	59,48	3,89	60,02	-	-	175,76
<b>117 Asn</b>	119,41	7,46	54,51	3,87	35,81	2,01	-	175,70

<b>118 Cys</b>	117,96	7,59	61,48	4,50	37,43	3,35	-	175,07
<b>119 Ile</b>	117,41	7,28	61,80	3,73	36,23	1,83	C <sup>γ1</sup> 12,66; H <sup>γ1</sup> 0,85; C <sup>δ</sup> 11,79; H <sup>δ</sup> 0,40; C <sup>γ2</sup> 15,29; H <sup>γ2</sup> 0,8	176,88
<b>120 Asn</b>	120,08	8,23	52,20	4,68	36,49	2,85	-	176,35
<b>121 Gly</b>	109,55	7,83	43,71	4,11	-	-	-	173,50
<b>122 Ala</b>	124,78	7,65	50,55	4,40	17,74	1,44	-	176,25
<b>123 Asn</b>	124,17	7,79	52,59	4,53	38,83	2,84/2,73	-	

- ANEXO 2 -

Letter to the Editor:  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonance assignments of cerato-platanin, a phytotoxic protein from *Ceratocystis fimbriata*.

Aline L. Oliveira<sup>a,b</sup>, Luigia Pazzagli<sup>c</sup>, Barbara Pantera<sup>c</sup>, Gianni Cappugi<sup>c</sup>, Celso E. Benedetti<sup>b</sup>,  
Alberto Spisni<sup>a</sup> & Thelma A. Pertinhez<sup>a\*</sup>

<sup>a</sup>*Department of Experimental Medicine, University of Parma, Parma, Italy.*  
<sup>b</sup>*Center for Molecular Structural Biology, LNLS, Campinas, Brazil*  
<sup>c</sup>*Department of Biochemical Sciences, University of Florence, Florence, Italy*

\*To whom correspondence should be addressed. E-mail: [thelma@unipr.it](mailto:thelma@unipr.it)

Address:  
Dipartimento di Medicina Sperimentale  
Sez. "Chimica e Strutturistica Biochimica"  
Università di Parma  
Via Volturmo, 39  
43100, Parma  
Italy  
phone: ++39 0521 033825  
fax: ++39 0521 033802

*Keywords:* Cerato-platanin, phytotoxin

## Biological context

Cerato-platanin (CP) is produced by the ascomycete fungus *Ceratocystis fimbriata*, which is the causative agent of canker stain, a severe disease with incidence in a great number of plants, such as *Platanus acerifolia*, *Theobroma cacao*, *Coffea arabica*, among others. The protein secreted in the medium by cultured *Ceratocystis fimbriata* is able to elicit phytoalexin synthesis in *Platanus acerifolia* leaves and to enhance cell death (Pazzagli *et al.*, 1999). Besides that, CP is known to accumulate in the mycelium and to be located in the cell walls of the fungus ascospores, hyphae and conidia, suggesting a role in forming the fungal cell wall (Boddi *et al.*, 2004).

CP primary sequence is formed by 120 amino acids, 40% of which are hydrophobic. According to the EMBL data bank, CP is the reference protein of a new family: “the cerato-platanin family” (EBI-InterPro IPR010829) that is constituted by six other fungal secreted proteins characterized by a high amino acid sequence similarity and involved in phytopathological phenomena and/or immunological reactions such as the infection of wheat leaves, the blackleg disease of *Brassica napus*, the allergic reactions against *A. fumigatus* in human and the antigenic responses in a human respiratory disease caused by *C. immitis* (Kurup *et al.*, 2000; Pan & Cole, 1995).

While the functions of this protein remain speculative, the determination of its structure is of obvious interest. In fact, considering that CP does not show significant structural homology with any protein in the PDB, the resolution of the 3D solution structure is expected to provide new clues on its biological functions and may also disclose some new protein fold. In this respect, the observation that the recombinant CP can be expressed with good yield in *E. coli*, together with its limited molecular size, high solubility and remarkable stability (Carresi *et al.*, 2006) clearly make this protein appropriate for being studied by NMR. In this paper we report the CP backbone and side-chain assignments.

## Methods and experiments

Recombinant CP was obtained by cloning and expressing *cp* gene in the yeast *Pichia pastoris* (GS115 strain) using pPIC9 extracellular expression vector containing a signal sequence that enables the secretion of the cloned protein (Carresi *et al.*, 2006). The uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled protein was obtained by growing the yeast in MGY (3.4 g/L YNB, 1%  $\text{NH}_4\text{SO}_4$ , 10% glycerol, 0.02% biotin) containing  $^{15}\text{NH}_4\text{SO}_4$  and  $^{13}\text{C}$ -glycerol, at 30 °C until an  $\text{OD}_{600\text{nm}} = 7$ . The cells were then re-suspended in MM (3.4g/L YNB, 1%  $\text{NH}_4\text{SO}_4$ , 5% methanol, 0.02% biotin). To maintain CP induction aliquots of 100%  $^{13}\text{C}$ - methanol were added every 24h for 10 days. The  $^{15}\text{N}$ ,  $^{13}\text{C}$  CP was purified from culture filtrate by RP-HPLC (C-18 Vydac) chromatography.

NMR samples contained 0.7 mM of  $^{13}\text{C}$ ,  $^{15}\text{N}$  CP in 10 mM phosphate buffer pH 5.8, 0.05%  $\text{NaN}_3$  (using either 100% or 5%  $\text{D}_2\text{O}$ ). NMR spectra were collected at 293 K on a Varian Inova AS 500 MHz and 600 MHz spectrometers.

The  $^{15}\text{N}$  edited HSQC spectrum and the HNCA, HN(CO)CA, HNCO, HNCACB and CBCACONH 3D triple resonance experiments and TOCSY-NHSQC were obtained using the 5%  $\text{D}_2\text{O}$  sample for the backbone assignment process. For the side-chain assignments, we employed NMR data from HCCH-TOCSY, HCCH-COSY, NOESY-CHSQC and CHSQC experiments, collected with the 100%  $\text{D}_2\text{O}$  sample. Spectra were processed using the NMRPipe software package (Delaglio *et al.*, 1995) and analyzed with NMRView (Johnson & Blevins, 1994).

## Extent of assignments and data deposition

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC of CP is shown in Figure 1. 100% of the backbone atoms have been assigned. For the side chains, 92% of all  $^{13}\text{C}$  chemical shifts and 95% of the proton chemical shifts have been assigned. A list of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts have been deposited in the BioMagRes Bank (<http://www.bmrb.wisc.edu>) under accession number 6990.

Based on the Chemical Shift Index (Wishart *et al.*, 1992) for C $\alpha$ , H $\alpha$  and C $\beta$  atoms the secondary structure of CP is characterized by the presence of seven  $\beta$ -strands and two  $\alpha$ -helices (data not shown), in agreement with the computational secondary structure prediction analyses.

### **Acknowledgement**

This work was supported by FAPESP, Brazil and by MIUR-COFIN, Italy. A.L.O. is a recipient of a FAPESP doctoral fellowship. T.A.P. is the recipient of a Marie Curie Incoming International Fellowship (MIF1-CT-2005-022050).

### **References**

- Boddi, S., Comparini, C., Calamassi, R., Pazzagli, L., Cappugi, G. and Scala, A. (2004) *FEMS Microbiol. Letters*, **233**, 341-346.
- Carresi, L., Pantera, B., Zoppi, C., Cappugi, G., Oliveira, A.L., Pertinhez, T.A., Spisni, A., Scala, A. and Pazzagli, L. (2006) *Protein Expr. Purif.*, accepted for publication.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277-293.
- Kurup, V.P., Banerjee, B., Hemmann, S., Greenberger, P.A., Blaser, K. and Crateri, R. (2000). *Clin. Experim. Allergy*, **30**, 988-993.
- Johnson, B. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603-614.
- Pan S. and Cole G.T. (1995). *Infect. Immun.*, **63**, 3994-4002.
- Pazzagli, L., Cappugi, G., Manao, G., Camici, G., Santini, A. and Scala, A. (1999) *J. Biol. Chem.*, **274**, 24959-24964.
- Wishart, D.S., Sykes, B. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647-1651.

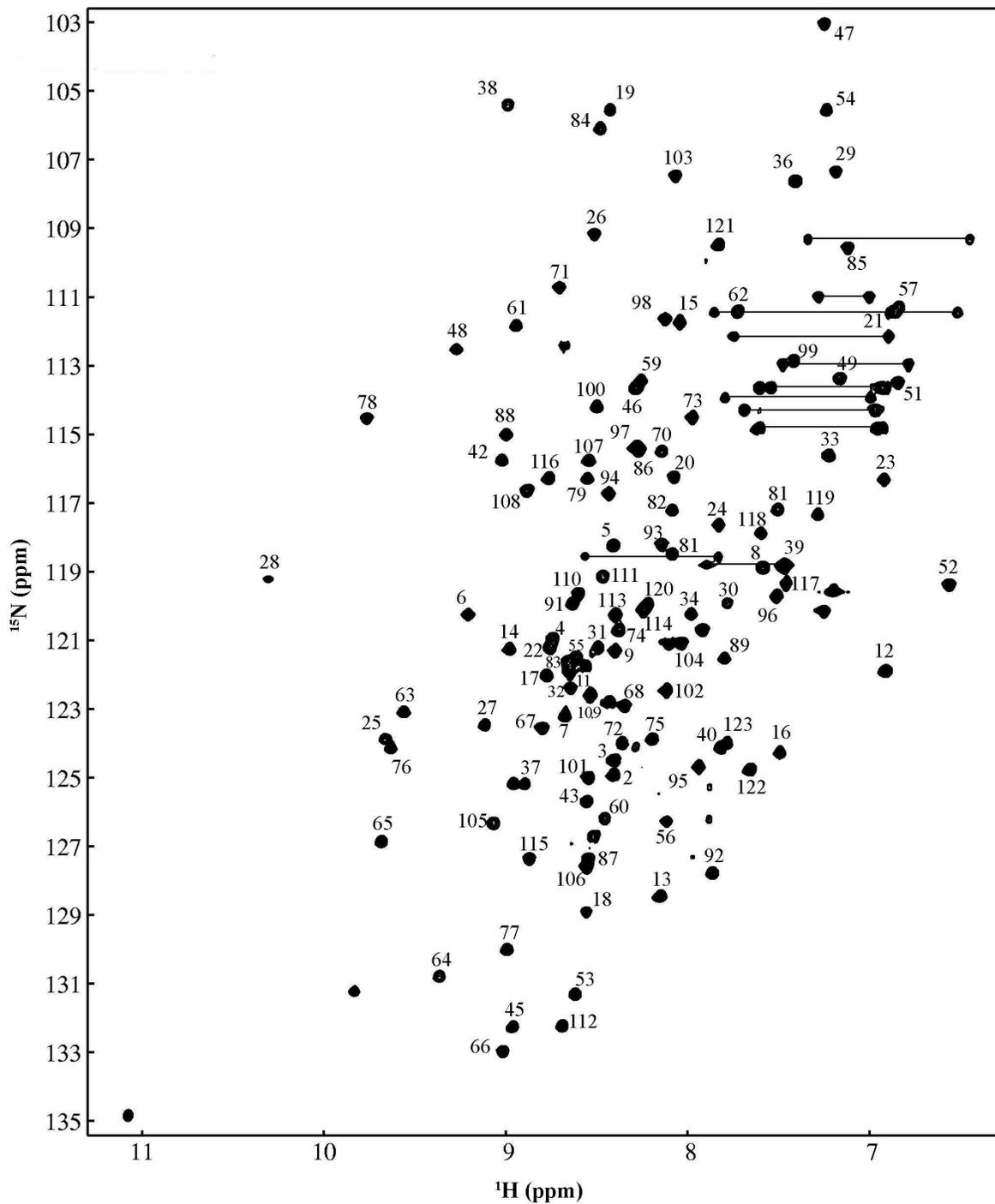
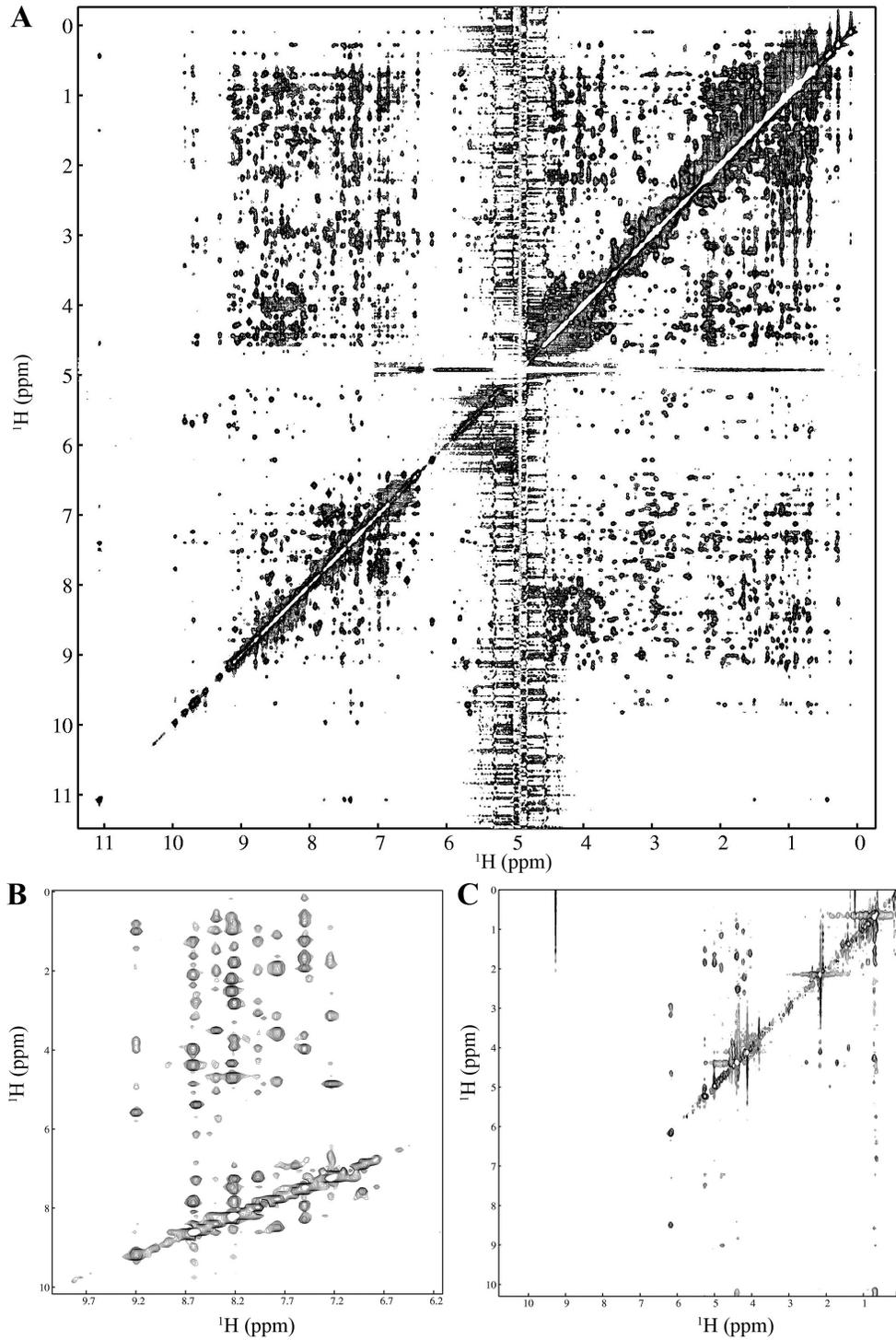


Figure 1.  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of CP. The spectrum was acquired at 293K, pH 5.8 with 128 t1 increments and 16 scans on a Varian Inova 600 MHz. Backbone amide groups is indicated by residue number. Side-chain amide cross peaks are connected by lines and the other side-chain correlations are not indicated.

- ANEXO 3 -



**Figura.** Espectro bidimensional  $^1\text{H}$ -NOESY (A) e planos dos espectros tridimensionais  $^{15}\text{N}$ -NOESY-HSQC (B) e  $^{13}\text{C}$ -NOESY-HSQC (C) obtidos para 0,7 mM da proteína  $^{13}\text{C}/^{15}\text{N}$ -CP, 10 mM tampão fosfato, pH 5,8 a 20°C.

**- ANEXO 4 -**

**Tabela.** Ângulos dihedros  $\phi$  e  $\psi$  derivados da análise do programa TALOS obtidos para a proteína CP. A avaliação de cada ângulo refere-se a qualidade da predição.

<b>Resíduo</b>	<b><math>\phi</math></b>	<b>Erro <math>\phi</math></b>	<b><math>\psi</math></b>	<b>Erro <math>\psi</math></b>	<b>Avaliação</b>
<b>1 Glu</b>	-79,230	16,170	140,130	18,850	Novo
<b>2 Lys</b>	-89,620	17,820	128,790	15,430	Novo
<b>3 Arg</b>	-98,810	15,120	138,610	18,740	Bom
<b>4 Val</b>	-124,390	15,160	143,370	18,280	Bom
<b>5 Ser</b>	-107,330	26,190	136,610	27,230	Bom
<b>6 Ile</b>	-124,990	8,860	139,370	23,150	Bom
<b>7Ser</b>	-123,350	17,440	131,850	31,130	Bom
<b>8 Tyr</b>	-122,350	15,410	148,450	27,290	Bom
<b>9 Asp</b>	-117,510	20,480	145,740	38,840	Novo
<b>10 Pro</b>	-81,790	29,880	133,600	35,280	Bom
<b>11 Ile</b>	-90,490	20,140	132,830	38,470	Novo
<b>12 Tyr</b>	-85,900	10,580	126,440	32,660	Bom
<b>13 Ala</b>	-106,630	18,300	152,970	30,460	Novo
<b>14 Ala</b>	-70,100	9,120	-18,360	28,090	Novo
<b>15 Asp</b>	-101,190	18,950	2,700	22,570	Novo
<b>16 Leu</b>	-74,120	12,080	126,700	27,300	Bom
<b>17 Ser</b>	-88,290	29,320	126,670	40,340	Novo
<b>18 Met</b>	-72,190	7,630	-20,310	54,350	Novo
<b>19 Gly</b>	81,530	13,100	11,120	48,200	Novo
<b>20 Ser</b>	-88,120	14,540	153,320	31,200	Novo
<b>21 Val</b>	-134,030	21,970	158,980	32,010	Novo
<b>22 Ala</b>	-86,360	25,960	139,030	69,340	Bom
<b>23 Cys</b>	-137,390	16,170	152,450	65,300	Novo
<b>24 Ser</b>	-106,660	12,740	122,040	51,470	Bom
<b>25 Asn</b>	-88,640	15,650	132,670	28,360	Novo
<b>26 Gly</b>	-82,230	21,010	140,930	31,560	Novo
<b>27 Asp</b>	-80,470	17,080	134,380	30,320	Novo
<b>28 His</b>	-88,420	15,330	5,050	23,100	Bom
<b>29 Gly</b>	89,840	8,800	6,710	26,590	Novo
<b>30 Leu</b>	-99,880	27,480	145,520	31,110	Novo
<b>31Met</b>	-68,250	23,620	155,040	36,250	Novo
<b>32 Ala</b>	-63,650	8,340	-31,720	28,930	Bom
<b>33 Gln</b>	-83,510	10,270	-26,760	48,880	Bom
<b>34 Tyr</b>	-116,880	23,380	120,480	47,980	Novo
<b>35 Pro</b>	-64,370	10,820	-20,840	64,470	Novo
<b>36 Thr</b>	-126,140	23,560	151,800	39,620	Novo
<b>37 Leu</b>	-90,210	26,580	0,110	50,000	Novo
<b>38 Gly</b>	-3,700	87,830	-9,530	47,450	Novo
<b>39 Glu</b>	-122,250	31,040	157,460	60,710	Novo
<b>40 Val</b>	-87,650	18,160	125,310	47,240	Novo

<b>41 Pro</b>	-79,960	20,590	131,740	60,410	Novo
<b>42 Gly</b>	-104,390	18,140	119,370	47,420	Novo
<b>43 Phe</b>	-86,450	16,880	133,240	44,950	Novo
<b>44 Pro</b>	-85,790	16,120	119,690	55,670	Novo
<b>45 Asn</b>	-92,610	16,580	132,540	65,960	Novo
<b>46 Val</b>	-130,920	16,580	154,750	79,310	Bom
<b>47 Gly</b>	-127,730	19,500	145,040	60,550	Bom
<b>48 Gly</b>	-121,850	29,610	142,860	55,980	Novo
<b>49 Ile</b>	-92,110	18,580	148,770	59,800	Novo
<b>50 Pro</b>	-61,130	5,220	-24,070	48,800	Novo
<b>51 Asp</b>	-89,990	12,330	-11,950	36,810	Novo
<b>52 Ile</b>	-106,840	31,740	129,720	24,130	Bom
<b>53 Ala</b>	-96,420	24,420	125,770	29,910	Novo
<b>54 Gly</b>	173,880	89,370	124,150	42,890	Novo
<b>55 Trp</b>	-82,990	11,770	130,110	41,520	Novo
<b>56 Asp</b>	-85,740	27,240	-37,540	35,180	Novo
<b>57 Ser</b>	-122,910	18,360	159,430	48,660	Bom
<b>58 Pro</b>	-61,530	4,920	-16,210	41,170	Novo
<b>59 Ser</b>	-90,270	5,640	9,300	46,910	Novo
<b>60 Cys</b>	56,580	17,940	37,480	43,860	Novo
<b>61 Gly</b>	-98,720	3,320	3,910	51,010	Novo
<b>62 Thr</b>	-121,670	22,070	154,350	58,540	Novo
<b>63 Cys</b>	-111,610	14,540	144,420	60,770	Novo
<b>64 Trp</b>	-96,820	19,250	126,190	55,010	Bom
<b>65 Lys</b>	-106,600	20,800	121,160	32,320	Bom
<b>66 Val</b>	-113,040	22,770	122,610	27,100	Bom
<b>67 Thr</b>	-108,080	24,350	131,390	34,070	Novo
<b>68 Ile</b>	-83,950	15,400	133,170	46,250	Novo
<b>69 Pro</b>	-68,380	13,990	-32,620	41,950	Bom
<b>70 Asn</b>	-91,470	14,510	9,040	32,270	Bom
<b>71 Gly</b>	86,660	13,110	0,640	28,120	Novo
<b>72 Asn</b>	-79,730	14,580	128,880	32,890	Novo
<b>73 Ser</b>	-125,780	17,040	149,690	36,690	Bom
<b>74 Ile</b>	-140,620	18,900	147,040	30,460	Novo
<b>75 Phe</b>	-103,760	20,630	126,440	32,690	Bom
<b>76 Ile</b>	-128,740	18,140	138,220	35,130	Bom
<b>77 Arg</b>	-102,170	15,490	118,340	44,740	Bom
<b>78 Gly</b>	-83,490	21,720	133,700	53,690	Novo
<b>79 Val</b>	-125,210	16,230	154,060	38,660	Bom
<b>80 Asp</b>	-130,150	10,680	137,850	32,020	Bom
<b>81 Ser</b>	-102,400	25,250	121,470	37,250	Bom
<b>82 Gly</b>	-97,730	15,390	134,140	43,840	Novo
<b>83 Arg</b>	-78,540	22,150	-24,910	46,020	Novo
<b>84 Gly</b>	85,460	11,930	5,790	43,120	Novo
<b>85 Gly</b>	-110,380	18,170	140,670	43,450	Novo
<b>86 Phe</b>	-116,290	22,560	128,130	35,760	Bom
<b>87 Asn</b>	-117,480	20,280	138,990	32,400	Bom

<b>88 Val</b>	-130,070	17,310	150,070	33,390	Novo
<b>89 Asn</b>	-119,960	29,660	149,940	47,760	Bom
<b>90 Pro</b>	-72,220	17,170	124,350	54,100	Novo
<b>91 Thr</b>	-69,640	6,160	-34,420	34,720	Bom
<b>92 Ala</b>	-65,820	4,640	-40,600	24,520	Bom
<b>93 Phe</b>	-64,340	6,890	-44,910	19,960	Bom
<b>94 Thr</b>	-62,370	5,530	-37,080	24,780	Bom
<b>95 Lys</b>	-72,650	7,650	-31,790	21,340	Bom
<b>96 Leu</b>	-72,030	15,210	-25,080	22,180	Novo
<b>97 Val</b>	-92,730	15,790	-24,930	27,310	Novo
<b>98 Gly</b>	88,110	11,920	0,600	29,170	Novo
<b>99 Ser</b>	-96,360	13,110	138,950	28,760	Novo
<b>100 Thr</b>	-72,620	17,570	136,860	21,190	Novo
<b>101 Glu</b>	-74,510	10,710	-14,140	20,580	Novo
<b>102 Ala</b>	-98,700	13,880	8,930	20,740	Bom
<b>103 Gly</b>	84,710	13,910	8,440	25,410	Novo
<b>104 Arg</b>	-116,480	18,250	127,370	26,580	Novo
<b>105 Val</b>	-113,350	18,340	129,220	19,720	Bom
<b>106 Asp</b>	-91,830	8,710	150,950	24,090	Novo
<b>107 Asn</b>	-81,900	11,630	147,990	24,740	Novo
<b>108 Val</b>	-96,680	20,890	125,930	24,480	Novo
<b>109 Asn</b>	-114,670	22,900	141,410	29,420	Bom
<b>110 Tyr</b>	-124,110	11,630	141,510	30,190	Bom
<b>111 Val</b>	-128,910	14,880	124,210	41,390	Bom
<b>112 Gln</b>	-73,720	14,170	130,120	56,100	Novo
<b>113 Val</b>	-120,950	24,430	159,650	48,320	Novo
<b>114 Asp</b>	-84,130	29,240	121,360	29,590	Bom
<b>115 Leu</b>	-65,330	11,960	-44,370	26,350	Bom
<b>116 Ser</b>	-66,660	6,020	-39,090	25,660	Bom
<b>117 Asn</b>	-70,840	14,530	-31,740	46,000	Bom
<b>118 Cys</b>	-70,790	5,690	-37,530	52,960	Bom
<b>119 Ile</b>	-72,230	12,420	-24,490	38,430	Bom
<b>120 Asn</b>	-93,540	13,240	1,780	21,380	Bom
<b>121 Gly</b>	82,150	12,550	7,170	22,120	Novo
<b>122 Ala</b>	-90,280	21,190	130,590	42,190	Novo
<b>123 Asn</b>	***	***	***	***	None

**- ANEXO 5 -**

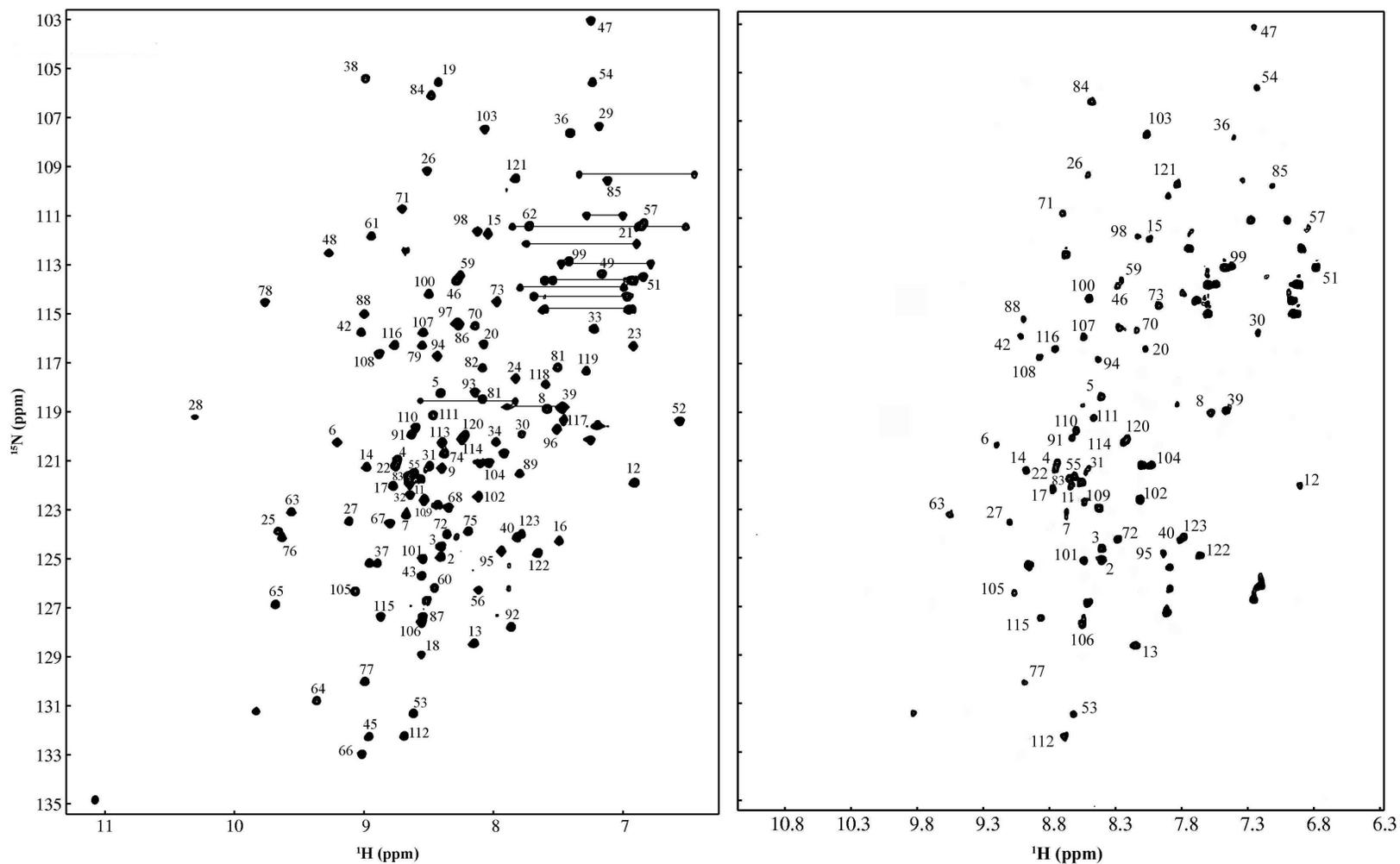
**Tabela.** RDCs das ligações  $^1\text{H}_\text{N}$ - $^{15}\text{N}$ ,  $^1\text{H}_\text{N}$ - $^{13}\text{CO}$ ,  $^{15}\text{N}$ - $^{13}\text{CO}$ ,  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  e  $^{13}\text{C}_\alpha$ - $^{13}\text{CO}$  obtidos para a proteína CP.

<b>Resíduo</b>	$^1\text{H}_\text{N}$ - $^{15}\text{N}$	$^1\text{H}_\text{N}$ - $^{13}\text{CO}$	$^{15}\text{N}$ - $^{13}\text{CO}$	$^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$	$^{13}\text{C}_\alpha$ - $^{13}\text{CO}$
<b>1 Glu</b>	1,29	6,61	3,15	0,38	2,06
<b>2 Lys</b>	5,28	-1,38	-8,11	7,02	-0,89
<b>3 Arg</b>	6,21	9,97	5,59	7,72	3,54
<b>4 Val</b>	**	**	**	8,79	2,39
<b>5 Ser</b>	4,73	7,51	14,25	0,65	10,84
<b>6 Ile</b>	-4,53	11,77	11,61	-5,97	**
<b>7Ser</b>	-1,28	4,44	3,60	-3,97	7,52
<b>8 Tyr</b>	5,02	**	**	0,71	-5,15
<b>9 Asp</b>	8,65	**	**	1,20	**
<b>10 Pro</b>	**	**	**	11,84	7,46
<b>11 Ile</b>	**	**	**	-3,49	-7,65
<b>12 Tyr</b>	-6,12	**	**	-10,91	-0,03
<b>13 Ala</b>	8,28	**	**	4,22	3,68
<b>14 Ala</b>	0,46	4,18	13,73	-2,52	0,61
<b>15 Asp</b>	-0,66	-1,79		-0,59	-4,23
<b>16 Leu</b>	1,80	-12,96	-21,45	11,34	3,40
<b>17 Ser</b>	8,71	2,99	1,422	3,22	5,00
<b>18 Met</b>	-2,14	-2,76	-7,92	-12,23	**
<b>19 Gly</b>	-3,0	1,69	4,92	-6,95	2,44
<b>20 Ser</b>	6,45	-8,13	-10,78	**	6,73
<b>21 Val</b>	-1,12	**	**	9,01	**
<b>22 Ala</b>	**	**	**	-3,01	-13,37
<b>23 Cys</b>	-1,70	3,23	15,03	6,69	8,94
<b>24 Ser</b>	3,88	8,19	7,54	-3,44	2,67
<b>25 Asn</b>	7,51	15,79	-6,30	5,79	**
<b>26 Gly</b>	-4,74	-6,85	-3,50	8,36	-11,60
<b>27 Asp</b>	2,94	14,82	-1,31	**	-1,69
<b>28 His</b>	9,14	8,34	-8,06	-7,42	**
<b>29 Gly</b>	0,55	-0,02	-4,41	20,64	-5,26
<b>30 Leu</b>	4,89	-3,46	7,77	**	**
<b>31Met</b>	**	0,19	-8,89	6,72	**
<b>32 Ala</b>	2,74	**	-0,84	2,57	**
<b>33 Gln</b>	-4,60	-16,28	3,99	2,19	0,05
<b>34 Tyr</b>	1,01	-3,11	-9,16	-8,80	**
<b>35 Pro</b>	**	**	**	**	6,04
<b>36 Thr</b>	10,02	9,53	0,12	4,38	1,14
<b>37 Leu</b>	**	18,99	-12,57	8,31	-9,88
<b>38 Gly</b>	6,50	**	14,73	0,05	-5,33
<b>39 Glu</b>	11,73	**	**	-9,03	0,74
<b>40 Val</b>	**	**	**	-12,82	**

<b>41 Pro</b>	**	**	**	8,41	-11,91
<b>42 Gly</b>	-0,22	-3,90	-2,49	-7,22	**
<b>43 Phe</b>	-6,28	-8,42	-14,68	-4,36	**
<b>44 Pro</b>	**	**	**	0,60	**
<b>45 Asn</b>	3,44	**	**	4,84	-3,38
<b>46 Val</b>	6,36	**	**	-0,50	0,34
<b>47 Gly</b>	1,94	-0,09	10,69	3,80	-0,27
<b>48 Gly</b>	-0,30	**	**	**	2,31
<b>49 Ile</b>	-2,02	**	**	-0,10	**
<b>50 Pro</b>	**	**	**	**	-1,01
<b>51 Asp</b>	7,51	8,52	5,98	0,47	10,42
<b>52 Ile</b>	-0,77	0,55	-0,53	**	**
<b>53 Ala</b>	3,40	3,67	6,51	-9,88	1,94
<b>54 Gly</b>	9,57	-4,16	-16,91	8,19	3,66
<b>55 Trp</b>	62,69	-9,61	-12,02	1,26	**
<b>56 Asp</b>	-2,22	-10,86	-6,77	2,00	3,56
<b>57 Ser</b>	**	**	**	-4,11	**
<b>58 Pro</b>	**	**	**	5,04	-9,18
<b>59 Ser</b>	11,31	-7,13	-13,43	-1,60	-10,75
<b>60 Cys</b>	3,21	5,52	8,75	-3,40	13,92
<b>61 Gly</b>	4,66	**	**	**	-6,51
<b>62 Thr</b>	6,48	11,10	-6,68	**	2,44
<b>63 Cys</b>	3,58	2,10	5,90	9,37	12,44
<b>64 Trp</b>	1,03	-1,13	7,01	-2,85	1,66
<b>65 Lys</b>	-1,53	-15,08	-10,26	-21,60	-2,88
<b>66 Val</b>	-1,03	-11,85	-24,97	2,81	2,47
<b>67 Thr</b>	3,79	-8,84	-17,44	6,34	-0,06
<b>68 Ile</b>	7,37	**	**	0,62	**
<b>69 Pro</b>	**	**	**	-3,27	**
<b>70 Asn</b>	1,42	-4,51	-13,79	-9,07	-11,29
<b>71 Gly</b>	8,30	-2,45	-20,12	0,02	8,31
<b>72 Asn</b>	5,50	6,99	-8,15	3,18	0,91
<b>73 Ser</b>	7,23	-6,16	-13,40	6,21	11,12
<b>74 Ile</b>	5,12	3,66	-10,30	2,25	12,25
<b>75 Phe</b>	0,29	-5,90	-35,57	-7,21	1,98
<b>76 Ile</b>	-5,17	7,53	0,23	4,94	8,28
<b>77 Arg</b>	-1,45	-12,99	-4,10	4,53	-11,61
<b>78 Gly</b>	5,59	**	**	10,44	10,27
<b>79 Val</b>	-1,93	1,29	-3,82	11,74	-11,93
<b>80 Asp</b>	-2,01	**	**	7,99	0,90
<b>81 Ser</b>	1,32	6,94	5,41	-6,70	**
<b>82 Gly</b>	2,27	-7,11	-10,64	-4,88	-3,36
<b>83 Arg</b>	-6,14	-10,64	12,24	-1,23	8,70
<b>84 Gly</b>	-2,13	-1,33	1,14	-4,08	-3,86
<b>85 Gly</b>	-4,70	-6,84	-5,97	-5,53	6,87
<b>86 Phe</b>	-7,32	-2,37	5,51	-3,26	**
<b>87 Asn</b>	-1,60	10,56	-2,07	-0,97	**

<b>88 Val</b>	-5,47	0,01	-2,39	**	**
<b>89 Asn</b>	3,03	-9,43	-11,08	0,62	**
<b>90 Pro</b>	**	**	**	1,52	-9,45
<b>91 Thr</b>	-4,21	9,66	1,05	2,50	4,31
<b>92 Ala</b>	2,76	-3,77	-12,99	-1,73	8,20
<b>93 Phe</b>	0,57	-0,83	17,90	-6,33	-0,16
<b>94 Thr</b>	-4,24	-9,67	-3,19	-7,17	-8,70
<b>95 Lys</b>	-2,72	-5,13	-21,00	-8,26	8,00
<b>96 Leu</b>	3,70	4,64	0,65	-7,34	3,00
<b>97 Val</b>	-0,71	-11,27	-11,85	**	2,46
<b>98 Gly</b>	-5,57	-14,41	-8,02	-8,98	-3,78
<b>99 Ser</b>	0,45	-1,064	25,35	-3,56	-3,82
<b>100 Thr</b>	-1,25	-5,281	-0,38	10,97	**
<b>101 Glu</b>	6,48	-4,08	-22,61	-4,35	-0,30
<b>102 Ala</b>	8,07	4,46	6,55	-6,54	**
<b>103 Gly</b>	0,11	-2,80	-6,89	1,75	0,79
<b>104 Arg</b>	-0,48	-9,98	-9,73	12,03	1,06
<b>105 Val</b>	-4,74	-0,72	5,18	-3,79	3,14
<b>106 Asp</b>	-0,83	**	**	-8,64	0,09
<b>107 Asn</b>	8,73	-2,27	-7,08	-8,50	-4,18
<b>108 Val</b>	11,67	-5,94	-19,90	-5,82	1,68
<b>109 Asn</b>	2,98	5,62	12,09	-3,66	**
<b>110 Tyr</b>	-4,44	-10,42	-0,54	2,47	0,99
<b>111 Val</b>	-0,99	-5,20	3,32	0,60	13,64
<b>112 Gln</b>	-1,14	-6,16	-2,16	2,33	**
<b>113 Val</b>	2,17	-12,60	-30,72	0,88	10,72
<b>114 Asp</b>	**	**	**	-5,03	3,28
<b>115 Leu</b>	-0,74	-6,22	-19,60	-6,25	-3,49
<b>116 Ser</b>	1,20	5,18	17,14	-6,64	**
<b>117 Asn</b>	-0,47	1,04	-10,12	-1,34	**
<b>118 Cys</b>	-1,25	-13,29	4,83	1,92	3,46
<b>119 Ile</b>	-1,29	-0,12	-21,13	4,46	**
<b>120 Asn</b>	**	**	**	2,18	0,64
<b>121 Gly</b>	1,54	**	0,06	5,41	-1,47
<b>122 Ala</b>	-0,31	-3,37	-3,50	-1,39	**
<b>123 Asn</b>	2,82	3,29	1,67	-1,52	0,35

- ANEXO 6 -



**Figura.** Espectro  $^1\text{H}$ - $^{15}\text{N}$ -HSQC (esquerda) e CLEANEX da região de  $^1\text{H}$ - $^{15}\text{N}$  (direita) obtidos para 0,7 mM da proteína  $^{15}\text{N}$ -CP, 10 mM tampão fosfato, pH 5,8 a 20°C.