



Joice Helena Paiva

**Estrutura, função e estabilidade de hidrolases glicosídicas pertencentes
à família GH5 com potencial aplicação na conversão de biomassa
lignocelulósica em açúcares fermentáveis**

**Structure, function and Stability of the glycosides hydrolases that
belong to GH5 family with potential application in the conversion of
lignocellulosic biomass in fermentable sugars**

Campinas

2013



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

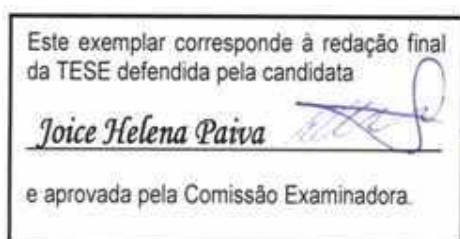
JOICE HELENA PAIVA

**“Estrutura, função e estabilidade de hidrolases glicosídicas
pertencentes à família GH5 com potencial aplicação na conversão de
biomassa lignocelulósica em açúcares fermentáveis”**

**“Structure, function and Stability of the glycosides hydrolases that
belong to GH5 family with potential application in the conversion of
lignocellulosic biomass in fermentable sugars”**

Tese apresentada ao Instituto de Biologia da
UNICAMP para obtenção do Título de
Doutora em Biologia Funcional e Molecular
na área de BIOQUÍMICA.

Thesis presented to the Institute Biology of
the UNICAMP to obtain the title of PhD in
Molecular Biology and Functional in the area
of BIOCHEMISTRY



Orientador: Dr. Mário Tyago Murakami

Campinas

2013

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

P166e Paiva, Joice Helena, 1985-
Estrutura, função e estabilidade de hidrolases glicosídicas pertencentes à família GH5 com potencial aplicação na conversão de biomassa lignocelulósica em açúcares fermentescíveis / Joice Helena Paiva. – Campinas, SP : [s.n.], 2013.

Orientador: Mário Tyago Murakami.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Glicosídeo hidrolases. 2. Enzimas - Biotecnologia. 3. Cristalografia de proteína. 4. Estabilidade térmica. 5. Dicroísmo circular. I. Murakami, Mário Tyago. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Structure, function and stability of the glycosyl hidrolases that belong to GH5 family with potencial application in the conversion of lignocellulosic biomass in fermentable sugars

Palavras-chave em inglês:

Glycosyl hydrolases
Enzymes - Biotechnology
Protein crystallography
Thermal stability
Circular dichroism

Área de concentração: Bioquímica

Titulação: Doutora em Biologia Funcional e Molecular

Banca examinadora:

Mário Tyago Murakami [Orientador]
Sandro Roberto Marana
Gonçalo Amarante Guimarães Pereira
Marcio Vinicius Bertacine Dias
Adriana Franco Paes Leme

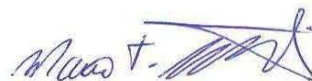
Data de defesa: 16-08-2013

Programa de Pós-Graduação: Biologia Funcional e Molecular

Campinas, 16 de agosto de 2013

BANCA EXAMINADORA

Dr. Mário Tyago Murakami (orientador)



Assinatura

Prof. Dr. Sandro Roberto Marana



Assinatura

Prof. Dr. Gonçalo Amarante Guimarães Pereira



Assinatura

Dr. Marcio Vinicius Bertacine Dias



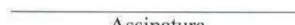
Assinatura

Dra. Adriana Franco Paes Leme



Assinatura

Prof. Dr. Claudio Chrysostomo Werneck



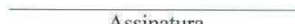
Assinatura

Dr. Daniel Maragno Trindade



Assinatura

Dr. Iuri Estrada Gouvea



Assinatura

Resumo

O material lignocelulósico é considerado uma promissora fonte renovável de energia para substituição da atual matriz energética baseada no petróleo. Além disso, diversos ramos industriais empregam a biomassa vegetal como matéria-prima, sendo assim biotecnologias visando sua degradação em açúcares fermentescíveis ou alteração das propriedades físico-químicas tem despertado interesse tanto do domínio acadêmico como industrial. Um dos desafios em bioprocessos envolvendo biomassa lignocelulósica é seu caráter extremamente recalcitrante a hidrólise enzimática o que requer um arsenal de enzimático que incluem hemicelulases, celulasas, esterases e outras enzimas acessórias, e os altos custos com enzimas representam um gargalo econômico para uso industrial. Assim, esclarecer as bases mecanísticas do funcionamento destas enzimas bem como a racionalização e discretização de estratégias moleculares que visam o aumento da eficiência, estabilidade, modulação da função e seletividade, requisitos que são de suma importância para o aprimoramento do uso dessas enzimas em bioprocessos com escala industrial.

Nesta tese, os estudos foram direcionados às hidrolases glicosídicas (GH) pertencentes à família 5 (GH5), que possuem enovelamento barril TIM (α/β)₈ e apesar da alta conservação estrutural apresentam grande diversidade na função. As enzimas alvo deste estudo incluem a celulase 5A de *Bacillus subtilis*; a celulase 5 de *Xanthomonas axonopodis* pv. *citri*; uma nova endocelulase GH5 oriunda de uma biblioteca metagenômica do solo de cana-de-açúcar; e a β -mananase de *Thermotoga petrophila*. Tais enzimas tiveram sua estrutura cristalina determinada tanto na forma nativa como complexada com ligantes. Também foram realizados ensaios funcionais combinados com mutagênese deletéria que contribuíram para o entendimento da relação estrutura-função-estabilidade dessa classe de enzima. Estudos em solução por espalhamento de raios X a baixos ângulos (SAXS) e outras ferramentas biofísicas foram empregadas para dar suporte as análises estruturais e de estabilidade.

Como principais resultados destaca-se o efeito termo-estabilizante do íon manganês sobre o domínio catalítico da celulase 5A de *Bacillus subtilis* (BsCel5A) com um ganho de estabilidade de 20 °C. A estrutura do complexo indicou que o íon se liga ao bolsão negativamente carregado, localizado na região oposta à interface catalítica. A remoção do domínio acessório (CBM – módulo de ligação a carboidrato) mostrou diminuição da eficiência catalítica. A estrutura resolvida por ressonância magnética nuclear (RMN) em colaboração com o grupo da Dra. Ana Carolina Zeri revelou ser um CBM3 que não possui o sítio de ligação ao cálcio. Interessantemente, o domínio acessório, presente na β -mananase de *Thermotoga petrophila* (CBM27), teve um papel distinto daquele observado na BsCel5A, contribuindo para a termo-estabilidade ao invés do aumento da eficiência catalítica. Além disso, sua estrutura tridimensional foi elucidada por SIRAS com iodo e diversos complexos foram obtidos contribuindo para o entendimento molecular da preferência de glicomananas como substrato. Outra enzima ortóloga a BsCel5A, de *Xanthomonas axonopodis* pv. *citri*, também demonstrou ganho de estabilidade na presença do íon manganês. Por fim, foi determinada a primeira estrutura de uma celulase 5 proveniente de uma biblioteca metagenômica de DNA do solo de cana-de-açúcar (CelE1) que irá ajudar na compreensão de enzimas especializadas a degradação de biomassa.

Abstract

Cellulose is the most abundant component in the lignocellulosic biomass, and is considered a promising renewable energy source to replace the current petroleum-based energy sources. Moreover, a number of industrial sectors also employ plant biomass as feedstock. Thus, biotechnologies for degradation of plant biomass into simple sugars have attracted much attention from both academic and industrial domain. However, one of the economical challenges in converting the lignocellulosic material into fermentable sugars by enzymatic routes is the high costs with enzymes. In addition, due to its inherently recalcitrant characteristic, the plant biomass requires a complex enzymatic cocktail including endoglucanases, exoglucanases, glucosidases, and a variety of hemicellulases and accessory proteins. Therefore, the understanding of mechanisms by which these enzymes hydrolyze glycosidic bonds, and their molecular strategies to modulate catalytic efficiency, stability and specificity are critical to improve their use in bioprocesses at industrial scale.

In this work, we have investigated four glycosyl hydrolases that belong to family 5 (GH5) including the cellulase 5A from *Bacillus subtilis*, cellulase 5 from *Xanthomonas axonopodis* pv. *citri*, cellulase 5 from a sugarcane soil metagenomic library and β -mannanase from *Thermotoga petrophila*. All enzymes had their three-dimensional structures determined by either X-ray crystallography or NMR. Enzymatic characterization and extensive mutagenesis studies contributed to shed light on the structure-function-stability relationships of these modular enzymes.

Among the several findings, we highlight the thermostabilizing effect of the manganese ion on the catalytic core of some specific cellulases 5. In BsCel5A, the ion binds to a negatively-charged pocket at the opposite face of the active site, which could serve as a structural motif to be engineered aiming at higher thermal stability. Moreover, the deletion of the carbohydrate binding module (CBM) reduced the catalytic efficiency and also affected positively the conformational stability. The NMR structure of the CBM3 solved in cooperation with Dr. Ana Carolina Zeri, revealed that CBM3 did not have a structural calcium, which explains the lower stability of this accessory domain in relation to its parental catalytic domain. On the other hand, the CBM27 of β -mannanase from *Thermotoga petrophila* conferred higher thermal stability and did not affect the enzyme function. Moreover, its three-dimensional structure was elucidated by SIRAS with iodine derivative, and several complexes with glucose, maltose and structurally related ligands were solved at high-resolution depicting a wide-range representation of the subsites forming the active-site cleft. The BsCel5A ortholog enzyme from *Xanthomonas axonopodis* pv. *citri*, also showed a gain in the thermal stability in presence of the manganese ion; however the BsCel5A manganese-binding site is not conserved. It suggests a distinct binding site and structural studies are currently in progress to elucidate the structural basis. Finally, the structure of a cellulase 5 recovered from a sugarcane soil metagenomic library was characterized and its structure determined, contributing to a better understanding of enzymes specialized in plant biomass degradation.

Sumário

Resumo.....	vii
Abstract	viii
Agradecimentos	xi
Lista de Figuras	xiii
Lista de tabela	xv
Lista de abreviaturas	xvii
I Introdução	1
I. 1 O material lignocelulósico e sua arquitetura estrutural	1
I. 2 Enzimas envolvidas na degradação do material lignocelulósico.....	5
I.3 Utilização das Hidrolases Glicosídicas	8
I. 4 Aspectos estruturais das celulasas.....	9
I. 5 As enzimas da família GH5	12
I. 6 Mecanismo de catálise das enzimas GH5	14
I. 7 Organismos produtores de celulasas	16
II. Objetivo	21
III. Material e métodos.....	23
III. 1 Clonagem.....	23
III. 2 Expressão das proteínas recombinantes.....	23
III. 3 Purificação das proteínas recombinantes	24
III. 4 Espalhamento de Luz Dinâmico - DLS	26
III. 5 Dicroísmo circular - CD	26
III. 6 Proteólise limitada	27
III. 7 Ensaios de cristalização	27
III. 8 Coleta e processamento de dados da difração	28
III. 9 Resolução de Estrutura e Refinamento.....	29
III. 10 Espalhamento de raios X a baixos ângulos - SAXS	30
III. 11 Caracterização bioquímico-cinética	31
III. 11. a. Atividade sobre diferentes substratos	31
III. 11. b. Cinética	31
IV. Resultados e discussão.....	33
IV. 1 Celulase 5A de <i>Bacillus subtilis</i> – BsCel5A.....	33
IV. 2 Celulase 5 de <i>Xanthomonas axonopodis</i> pv. <i>citri</i> – XacCel5	49
IV. 3 Celulase 5 de biblioteca de metagenômica do solo de cana-de-açúcar – CelE1	59

IV. 4 GH5-CBM27/ endo-1,4- β -D-mananase de <i>Thermotoga petrophila</i>	69
V. conclusões.....	83
VI. Referencias bibliográficas	85
Anexo I.....	97
A cristalografia como ferramenta para elucidação de estruturas tridimensionais	97
Anexo 2 Artigos publicados.....	107
Dissecting structure–function–stability relationships of a thermostable GH5-CBM3 cellulase from <i>Bacillus subtilis</i> 168.	107
Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- β -D-mannanase	123
The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction	139

Agradecimentos

“Agradeço primeiramente” (...), costumeiramente é assim que se iniciam os agradecimentos, mas é difícil elenca-los, pois muitos que passaram em minha vida durante esses cinco anos, de uma maneira ou de outra contribuíram para que eu chegasse até essa etapa da minha carreira acadêmica. Tenho absoluta certeza de que em todo esse tempo minha vida sempre foi guiada por Deus e pela minha intercessora, a Virgem Maria, passando na frente.

Acho justo inicialmente mencionar meus familiares. Com todos os desafios que vivemos, são meus sustento e fortaleza e me proporcionaram as oportunidades que busquei desde o dia em que deixei nossa casa para iniciar o percurso da minha carreira como bióloga. Nunca deixaram de acreditar no meu sonho ou disseram que ele não seria possível. Mas sempre afirmaram com gestos e palavras: “*Joice, você é capaz*”. Um eterno obrigado a meus pais, José Carlos e Luzia Paiva, e as minhas irmãs, meu porto seguro sempre, Ana Paula e Cristina. Vocês me ensinaram a procurar, buscar e realizar o que eu realmente quis.

Obrigada ao meu orientador, Mário Murakami, que mesmo quando eu não acreditava na minha capacidade, ele sempre esteve dizendo que era possível, que é “*questão de se familiarizar*” ... Obrigada por me integrar ao grupo, me ensinar, confiar no meu trabalho.

Agradeço as minhas amigas, que me acompanham desde ensino fundamental, que mesmo passando os anos, não importando distância ou quanto tempo se passa, estamos sempre juntas... Angélica Macorin, Jéssica Bortolassi, Marisa Macorin... Obrigada.

As minhas amigas de universidade, a conclusão da minha graduação foi graças à ajuda e o apoio que recebi de vocês... Pois no momento em que o desânimo batia, o cansaço me prostava e a saudade dos meus falava fundo, elas me sustentaram e sempre me impulsionavam a estudar um pouco mais... Além de momentos inesquecíveis que passei ao lado delas, histórias que contarei aos meus filhos. Bárbara Bonine, Bruna Guido, Lilian Carolina e Livia Botta: carreguei vocês no meu coração onde é que eu vá.

Aos amigos que título de Doutorado me proporcionou: Aline Sampaio, Andrey Nascimento, Carla Polo, Cristiane Tambascia, e Fernanda Basei... Se esse doutorado se concretiza, é graças à amizade, companheirismo, apoio, ouvidos, esclarecimentos de vocês... Que não importando o turno, tornaram meu trabalho mais alegre, meus dias leves, a jornada mais fácil. Que compartilham comigo momentos de diversão, aventuras, e muitas lutas. São os amigos que dizer o até breve será extremamente difícil, mas será necessário quando nossas carreiras se

consolidarem, e cada um seguir o destino de sucesso, que certamente está reservado para cada um de nós.

Ao meu conselheiro, amigo, Daniel MT sempre disposto a ajudar, me escutar... Obrigada Dan.

A Celisa, Givanil, Renata, Teresa, e aos demais colaboradores do LNBio, que sempre estiveram dispostas a compartilhar o conhecimento, e consolidaram meu aprendizado.

Aos companheiros e amigos do grupo BBE, Mariana, Leandro, Priscila, Rosa, Vanessa Peixoto, Camila e Plínio. Obrigada por toda a informação partilhada, por todo ensinamento, por toda disposição.

Aos amigos que o laboratório me proporcionou: André, Tiago, Alisson, Germana, Patrícia, Tatiana, Bárbara, Rafael Canevarolo, Américo, Micheli, Gustavo, Kelvin, Rafael, Vanessa Pegos, João. Obrigada pela disposição em ajudar, em escutar, em esclarecer dúvidas, e proporcionar momentos que levarei comigo.

Agradeço a Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), por financiar a minha pesquisa, e tornar real o sonho. Ao LNBio, por toda infraestrutura.

Lista de Figuras

Figura 1: Esquematização dos componentes da estrutura recalcitrante do material lignocelulósico.

Figura 2: Representação do componente da celulose.

Figura 3: Exemplos de monômeros responsáveis pela composição da hemicelulose.

Figura 4: Estruturas dos compostos fenólicos precursores da lignina.

Figura 5: Representação sistemática da ação sinérgica de exocelulase, endocelulase e β -glicosidase na hidrólise da celulose amorfa e microcristalina.

Figura 6: Representação em *cartoon* das endocelulases presentes nos clãs A, B e M.

Figura 7: Representação da topologia barril TIM/enovelamento $(\alpha/\beta)_8$.

Figura 8: Mecanismo enzimático de hidrólise da ligação glicosídica utilizando a via de inversão ou retenção do carbono anomérico.

Figura 9: Representação da arquitetura dos domínios presentes na Celulase 5A de *Bacillus subtilis*.

Figura 10: SDS-PAGE 13% da cromatografia de exclusão molecular da enzima BsCel5A e dos seus domínios Catalítico e CBM3.

Figura 11: Espectros de UV distante das diferentes construções da celulase 5A de *Bacillus subtilis*.

Figura 12: Desnaturalizações térmicas da proteína BsCel5A, e dos seus domínios CC_BsCel5A e CBM_BsCel5A.

Figura 13: Estrutura do CBM_BsCel5A determinada por RMN, comparada com o CBM3 de *Clostridium thermocellum*.

Figura 14: Efeito do pH e da temperatura sobre a atividade catalítica da BsCel5A e do Domínio Catalítico.

Figura 15: Atividade relativa da BsCel5A e do domínio catalítico em diferentes substratos.

Figura 16: Cristais do domínio catalítico da celulase 5A de *Bacillus subtilis*.

Figura 17: Estrutura do domínio catalítico da celulase 5A de *Bacillus subtilis*.

Figura 18: Estrutura do domínio catalítico da Celulase 5A de *Bacillus subtilis* em complexo com o íon manganês.

Figura 19: Curvas de espalhamento e distribuição de distâncias (Pr) de SAXS dos domínios da Celulase 5A de *Bacillus subtilis* e da enzima selvagem.

Figura 20: Envelope molecular da Celulase 5A de *Bacillus subtilis* obtidos por SAXS.

Figura 21: SDS-PAGE 13% das etapas de purificação da celulase 5 de *Xanthomonas axonopodis* pv. *citri*.

Figura 22: Cristais da XacCel5 obtidos durante os ensaios iniciais de cristalização.

Figura 23: Cristais de XacCel5 adequados para os experimentos de difração de raio X.

Figura 24: Espectro de dicroísmo circular no UV distante da celulase 5 de *Xanthomonas axonopodis* pv. *citri* - XacCel5.

Figura 25: Curvas de desnaturações térmicas da XacCel5 com distintos íons .

Figura 26: Espectro de dicroísmo circular na faixa do UV distante da XacCel5 após ensaios de desnaturação térmica.

Figura 27: Espectros de dicroísmo circular na faixa do UV distante da XacCel5 antes e pós desnaturação térmica e na presença e ausência do íon manganês.

Figura 28: Alinhamento das sequências das celulases de *Bacillus subtilis* (BsCel5A) e *Xanthomonas axonopodis* pv. *citri* (XacCel5).

Figura 29: SDS-PAGE 13% das etapas de purificação da celulase 5 oriunda de uma biblioteca de metagenômica do solo de cana-de-açúcar.

Figura 30: Espectro de dicroísmo circular no UV distante CelE1.

Figura 31: Curvas de Desnaturação térmica da CelE1 em pH 7,4, com $MnCl_2$ e $CaCl_2$.

Figura 32: Cristais obtidos da celulase 5 oriunda da biblioteca de metagenômica do solo de cana-de-açúcar.

Figura 33: Representação da estrutura tridimensional da celulase 5 oriunda da biblioteca de metagenômica do solo de cana-de-açúcar.

Figura 34: Representação da superfície molecular e da estrutura tridimensional da CelE1.

Figura 35: Destaque para os resíduos que formam a fenda catalítica da CelE1.

Figura 36: Comparação estrutural entre quatro celulases da família GH5 com a CelE1.

Figura 37: Representação da arquitetura dos domínios pertencentes à endo-1,4- β -D-mananase de *Thermotoga petrophila* - TpMan.

Figura 38: SDS-PAGE 13% da cromatografia de afinidade da enzima TpMan e da DC_TpMan.

Figura 39: Caracterização biofísica por dicroísmo circular da enzima TpMan.

Figura 40: Alinhamento das β -mananases da família GH5 com estrutura tridimensional determinada.

Figura 41: Estrutura do domínio catalítico da TpMan.

Figura 42: Localização dos ligantes na estrutura da DC_TpMan.

Figura 43: Interação dos ligantes nos subsítios que compõem a fenda catalítica da DC_TpMan.

Figura 44: Detalhe dos resíduos que compõem o subsítio -1 e +1 da TpMan.

Lista de tabela

Tabela 1: Condição para a produção das proteínas heterólogas.

Tabela 2: Soluções tamponantes utilizadas na purificação das proteínas recombinantes.

Tabela 3: Temperatura de transição (T_M) das desnaturações térmicas das BsCel5A, CC_BsCel5A e CBM_BsCel5A.

Tabela 4: Parâmetros cinéticos da celulase 5A de *Bacillus subtilis*.

Tabela 5: Condições de cristalização do domínio catalítico da celulase 5A de *Bacillus subtilis*.

Tabela 6: Estatísticas da coleta de dados dos cristais do Domínio Catalítico BsCel5A, e do refinamento da estrutura.

Tabela 7: Condições iniciais de cristalização da enzima XacCel5.

Tabela 8: Condições do refinamento da cristalização da celulase 5 de *Xanthomonas axonopodis* pv. *citri*.

Tabela 9: Temperatura de transição (T_M) das desnaturações térmicas da XacCel5.

Tabela 10: Temperaturas de transições (T_M) das desnaturações térmicas da CelE1.

Tabela 11: Condições propícias para obter cristais da nova celulase oriunda de biblioteca de metagenômica do solo de cana-de-açúcar pertencente à família GH5.

Tabela 12: Estatística da coleta de dados e do refinamento da CelE1.

Tabela 13: Condições de cristalização do domínio catalítico da β -mananase de *Thermotoga petrophila*.

Tabela 14: Estatística da coleta de dados e do refinamento da estrutura da DC_TpMan.

Tabela 15: Estatísticas da coleta de dados e do refinamento das estruturas de DC_TpMan com ligantes.

Lista de abreviaturas

BLAST: *Basic Local Alignment Search Tool*

BsCel5A: celulase da família GH5 pertencente a *Bacillus subtilis*

CBM_BsCel5A: módulo de ligação a carboidrato da endocelulase da família GH5 de *Bacillus subtilis*.

CC_BsCel5A: domínio catalítico da celulase da família GH5 de *Bacillus subtilis*.

CCP4: *Collaborative Computational Project number 3*

CD: *circular dichroism* (dicroísmo circular)

CelE1: endocelulase oriunda da biblioteca de metagenômica do solo de cana-de-açúcar

C-terminal: carboxi terminal

DC_TpMan: domínio catalítico da endo-1,4- β -manosidase pertencente a família GH5 da *Thermotoga petrophila*.

DLS: *dynamic light scattering* (espalhamento dinâmico de luz)

DNA: ácido desoxirribonucleico

dNTP: desoxinucleotídeo fosfatado

EDTA: ácido etilenodiamino tetra-acético

FPLC: *fast performance liquid chromatography*

GH: *Glycosides Hydrolases* – Hidrolases Glicosídicas

IMAC: *immobilized metal ion affinity chromatography* (cromatografia por afinidade a ion metálico imobilizado)

IPTG: isopropil β -D-1-tiogalactopiranosídeo

N-terminal: amino-terminal

PCR: *polimerase chain reaction* (reação em cadeia da polimerase)

PDB: *Protein Data Bank*

PEG: Polietilenoglicol

PMSF: fluoreto de fenilmetilsulfonil

SDS-PAGE: eletroforese em gel de poliacrilamida contendo dodecil sulfato de sódio

SEC: *size exclusion chromatography* (cromatografia por exclusão molecular)

SIRAS: *Single Isomorphous Replacement with Anomalous Scattering*

T_M: temperatura de transição / *melting temperature*

TpMan: manana endo-1,4-β-D-manosidase, pertencente a família GH5 de *Thermotoga petrophila*.

R_h: raio hidrodinâmico

R_g: raio de giro

XacCel5: domínio catalítico da endocelulase da família GH5 de *Xanthomonas axonopodis* pv. *citri*.

SAXS: *Small Angle X-ray Scattering* - Espalhamento de Raios X a baixos ângulos

X. citri: *Xanthomonas axonopodis* pv. *citri*.

I Introdução

I. 1 O material lignocelulósico e sua arquitetura estrutural

As matérias-primas lignocelulósicas são as fontes renováveis de energia mais comuns na natureza, compreendidas pelos materiais agroindustriais, pelos resíduos urbanos e pelas madeiras de angiospermas e gimnospermas (Coughlan e McHale, 1988). Os resíduos das agroindústrias vêm se destacando como importante fonte de energia renovável de baixo custo, fácil acesso e disponível em diversas regiões. Esses resíduos são utilizados na produção biológica de indústrias químicas como ácido lático, ácido acético, ácido propinóico e combustíveis (Wyman et al., 1992)

A biomassa lignocelulósica é constituída por três componentes principais: celulose, hemicelulose e lignina; em menores proporções por proteínas, lipídios, pectinas, açúcares solúveis e minerais. Para exemplificação a cana-de-açúcar é composta por 3% de celulose, 23% de hemicelulose e 22% de lignina (Dawson, 2005). As fibras da fração celulósica encontram-se dispostas como espirais, conferindo força e flexibilidade ao material e está envolta pela lignina, cuja função é aumentar a resistência da estrutura a ataques químicos e enzimáticos. Já a hemicelulose atua como um elo químico entre a celulose e a lignina (figura 1) (Jeffreis, 1990).

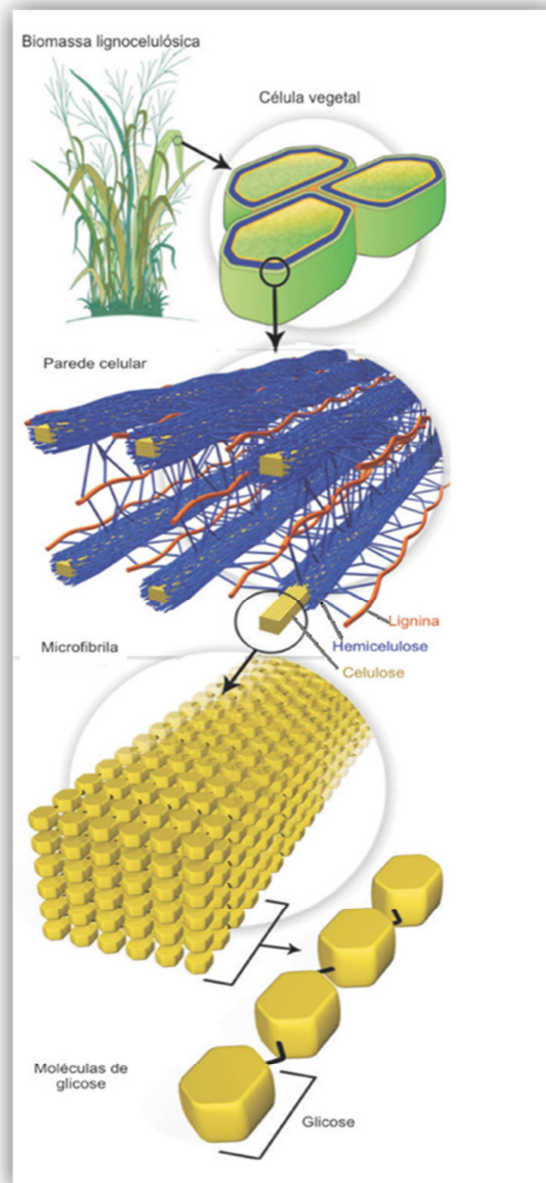


Figura 1: Esquematisação dos componentes da estrutura recalcitrante do material lignocelulósico, sendo composta principalmente de celulose, hemicelulose e lignina.

A celulose, constituinte mais abundante do material lignocelulósico, é um homopolissacarídeo constituído por unidades de D-glicose (figura 2A) unidas entre si por ligações glicosídicas $\beta(1-4)$ que apresenta um grau de polimerização de até 15000 unidades (Nelson e Cox, 2011). A estrutura linear (figura 2B), obtida pela configuração das ligações glicosídicas, possibilita a formação de ligações de hidrogênio intra e intermoleculares, o que acarreta na agregação das cadeias celulósicas em “fibrilas elementares”. As fibrilas são constituídas de duas regiões uma ordenada e outra desordenada. A região regularmente empacotada forma uma rede cristalina típica que podem conter configuração α

(predominante em algas) e $\text{I}\beta$ (predominante em plantas). A segunda região, a amorfa, tem as fibrilas desordenadas. Essa conformação estrutural é necessária para proporcionar rigidez e flexibilidade, para a expansão, contração e relaxamento existente na parede celular vegetal (Matthews et al. 2006; Mittal et al., 2011).

A conformação da celulose confere elevada resistência à tensão, que consequentemente a deixa insolúvel em inúmeros solventes, explicando em parte, a resistência do homopolímero à degradação microbiana (Gardner e Blackwell, 1974; Kolpak e Blackwell, 1976; Ding e Himmel 2006; Matthews et al., 2006). A vantagem na degradação da celulose está na capacidade de ter como produto final apenas a glicose. Tal açúcar é fonte abundante de energia estocadas entre as ligações C-H e C-C (Matthews et al., 2006), que pode ser utilizada na composição dos biocombustíveis, em diferentes ramos industriais, além de aplicações na medicina (Gil e Ferreira, 2006).

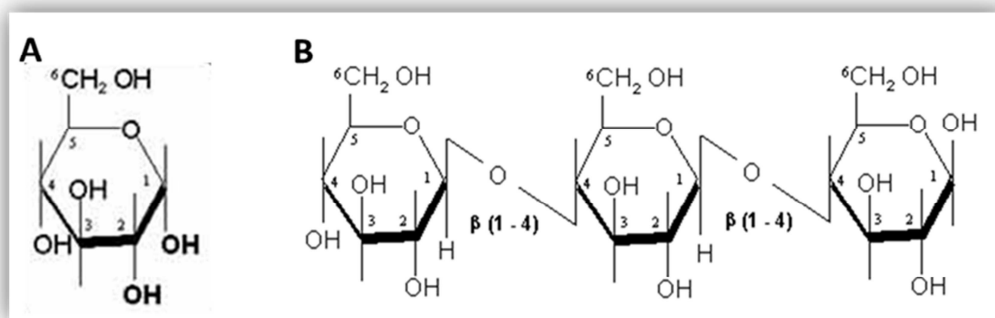


Figura 2: Representação do componente da celulose. **(A)** D-Glicose (monômero). **(B)** Celulose (homopolímero).

Outro componente do material lignocelulósico, a hemicelulose, nome utilizado para vários polissacarídeos, tem sua natureza heteropolissacarídica ramificada formada, principalmente, por blocos de pentoses (xilose, raminose e arabinose), hexoses (glicose, manose e galactose), ácidos urônicos (ácido 4-O-metilglucurônico e ácido galacturônico) e radicais acetilas. A composição da hemicelulose é determinada por aspectos ambientais, teciduais, tipos de plantas e idade da mesma. Por exemplo, em gramíneas a principal composição da hemicelulose é a xilana, composta por monômeros de xilose (figura 3A), em gimnospermas o componente fundamental é manana que tem como principal componente a mananose (figura 3B), enquanto que a hemicelulose de muitas angiospermas é formada por xiloglucana (glicose, xilose e galactose) (figura 3C) (Horn et al., 2012). Acredita-se que as

hemiceluloses estejam envolvidas na regulação da elongação e modificação da parede celular vegetal.

A hemicelulose apresenta seu grau de polimerização inferior a 200 unidades, e podem existir na forma acetilada (Ramos, 2003). De maneira geral, as hemicelulases são mais propícias à ação das enzimas aptas à degradação, porém há certas estruturas que são recalcitrantes (difícil degradação) devido ao processo de ramificação e acetilação (Agger et al., 2010). A degradação da hemicelulose tem como produto final uma combinação de diferentes açúcares, o que pode dificultar a fermentação (Horn et al., 2012). Indícios revelam que as interações entre hemiceluloses e as microfibrilas de celulose sejam mais fortes do que as interações entre as hemiceluloses e ligninas.

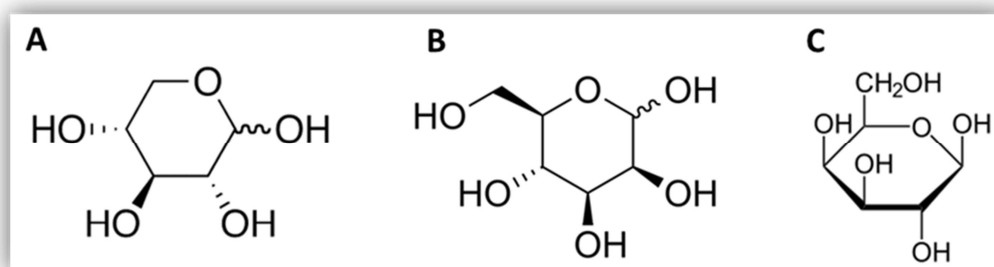


Figura 3: Exemplos de monômeros responsáveis pela composição da hemicelulose. (A) Xilose. (B) Mananose. (C) Galactopiranosose.

A lignina é um biopolímero rígido derivado de tirosina e fenilalanina (Nelson e Cox, 2011), sendo uma macromolécula complexa, formada pela polimerização radicalar de unidades fenil-propano (álcool p-cumarílico, álcool coniferílico e álcool sinapílico) (figura 4). O caráter hidrofóbico diminui significativamente a entrada de água na biomassa vegetal contribuindo, para a baixa taxa de degradação por coquetéis enzimáticos. A lignina é constituída da fração não polissacarídica mais abundante da parede celular vegetal, e muitas vezes se apresenta covalentemente ligada à fração hemicelulósica (Kuhad et al., 1996; Lawoko et al., 2006). A presença da lignina na biomassa lignocelulósica proporciona uma barreira física à ação das enzimas hidrolíticas. Já que as enzimas até então descritas capazes de degradarem a lignina são comumente dependentes de oxiredutases, logo sua aplicação em escala industrial fica prejudicada devido aos altos custos da produção (Martinez et al., 2005).

As ligninas, quando degradadas a frações menores, podem ser utilizadas na fabricação de espumas de poliuretanas, resinas fenólicas e epóxi, como fontes de fenol e etileno e podendo ser convertidas em fibras de carbono (Lora e Wolfgana, 2002).

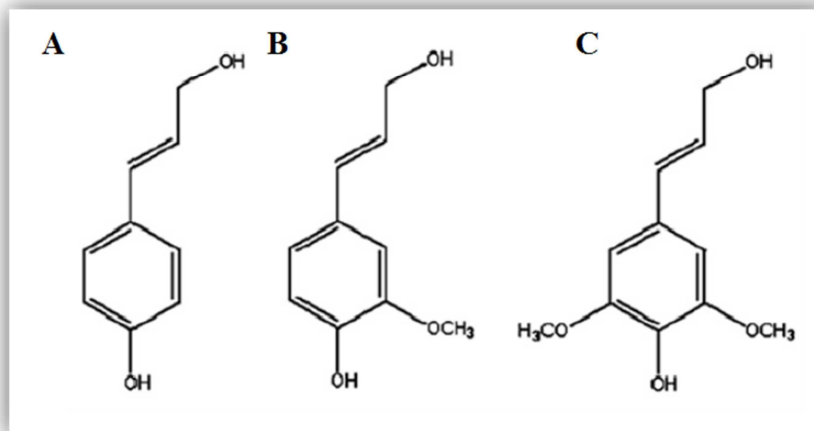


Figura 4: Estruturas dos compostos fenólicos precursores de lignina. (A) cumaril. (B) coniferil. (C) sinapil.

I. 2 Enzimas envolvidas na degradação do material lignocelulósico

Para que ocorra a liberação dos monômeros de açúcares são necessárias enzimas capazes de degradar o complexo lignocelulolítico. Essas enzimas incluem majoritariamente as hidrolases glicosídicas (*Glycosides Hydrolases* - GHs), juntamente com as carboidratos esterases (*Carbohydrate Esterases* - CE) e polissacarídeos liases (*polysaccharide lyases* - PL). Tais enzimas são formadas por domínios catalíticos independentes ou podem estar em fusão com módulos de ligação a carboidratos (*Carbohydrate Binding Module* - CBM), ou outros domínios acessórios como fibrinectina-símile ou imunoglobulina-símile. Recentemente, foram descobertas novas enzimas auxiliares na degradação de biomassa vegetal que são as mono-oxigenases e enzimas ligninolíticas (Levasseur et al., 2013).

As GHs constituem uma superfamília, composta por cerca de 130 famílias já identificadas. Essas enzimas são classificadas pela *Enzyme Commission* (EC) com a codificação 3.2.1.x, onde o x corresponde sua especificidade ao substrato e ao mecanismo molecular (Henrissat, 1991). A classificação das GHs está padronizada de acordo com seu local de atuação no substrato e são denominadas como endocelulases, exocelulases e β -glucosidades.

As *endocelulases* clivam randomicamente as ligações internas da região amorfa, liberando oligossacarídeos com terminações redutoras e não redutoras livres. Com a ação dessa enzima há uma rápida diminuição da viscosidade, tendo como produtos resultantes glicose, celobiose e celodextrinas de vários tamanhos. As *exocelulases* atuam nas extremidades das cadeias de celulose. Estas são subdivididas em celobiohidrolases – responsáveis pela hidrólise dos terminais redutores e não redutores, e glucanohidrolases –

capazes de liberar moléculas de glicose diretamente dos terminais do polímero. As β -*glucosidases* hidrolisam celobiose e os oligossacarídeos solúveis de baixo grau de polimerização (menor que sete) em glicose (figura 5) (Lynd et al., 2002).

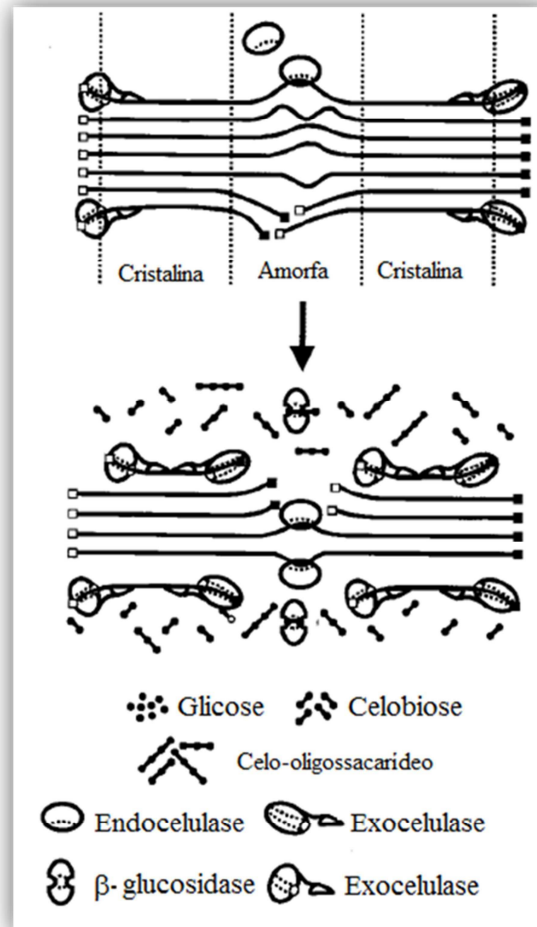


Figura 5: Representação sistemática da ação sinérgica de exocelulase, endocelulase e β -glucosidase na hidrólise da celulose amorfa e microcristalina (adaptação de Lynd et al., 2002).

Esse conjunto de enzimas forma o sistema celulolítico cuja característica é o sinergismo, tendo sua ação de forma coordenada favorecendo maior eficiência na hidrólise (Lynd et al., 2002). Há quatro formas de sinergismos entre as enzimas celulolíticas que são descritas em duetos: I sinergia - endocelulase e exocelulase; II sinergia - exocelulase e exocelulase, uma exocelulase atua na extremidade da cadeia reduzida e a outra na extremidade não reduzida; III sinergia - exocelulase e β -glucosidase, clivagem da celobiose em glicose; IV sinergia - domínio catalítico e CBM, uma sinergia intramolecular para aumento de eficiência catalítica (Blanco et al., 1994; Teeri 1997).

O módulo de ligação a carboidratos, CBM (carbohydrate-binding module), tem como função primordial aumentar a afinidade do domínio catalítico com o substrato (Tomme et al., 1995). Esse domínio acessório tem ação diferenciada quando a enzima atua em substratos cristalinos, característica não observada em regiões que a celulose está amorfa (Jalak e Våljamäe, 2010). Além de aumentar a eficiência catalítica, outras funções também são conhecidas para o CBM, como maior proximidade da enzima com o substrato; seletividade pelo substrato (Arante e Saddler, 2010); e maior estabilidade térmica (Santos et al., 2012).

Historicamente, as GHs começaram a ser estudadas durante a Segunda Guerra Mundial. A deterioração de fardas, barracas, bolsas, e outros objetos de algodão, intrigaram os soldados das forças armadas norte-americanas. Algumas organizações juntamente com as forças armadas montaram um laboratório para achar explicações cabíveis para o fenômeno da degradação. Os estudos detectaram organismos capazes de inutilizar os materiais de algodão, também encontraram o mecanismo de ação e os métodos eficazes de combatê-los. Até 1953, a pesquisa tinha determinado que as enzimas naturais, chamadas de celulasas, constituíam complexos de diversas moléculas com distintas habilidades de degradar o substrato. Mas, em 1956, o foco da pesquisa com as celulasas passou a ser otimizado na produção dessas enzimas para inúmeras aplicações industriais (Castro e Pereira, 2010).

A crise do petróleo deflagrada na década 70 acarretou em prejuízos marcantes à economia global (Canilha et al., 2009). A eminente escassez das reservas de petróleo, principal fonte energética mundial, juntamente com a preocupação com a preservação ambiental, foram os principais impulsos que levaram os governos a buscarem estratégias para uma maior produção e consumo de combustíveis que sejam renováveis e sustentáveis (Khesghi et al., 2000; Himmel et al., 2007). No Brasil, o Programa Nacional do Álcool (Proálcool), instituído pelo Governo Federal em 1975, fomentou o uso do etanol, produzido a partir do caldo da cana-de-açúcar como combustível veicular, até tal momento a produção de biocombustível era praticamente inexistente no país.

Embora sejam notórios os avanços conseguidos pelo Brasil na produção de etanol de primeira geração, a partir do caldo de cana-de-açúcar (sacarinas), pouco se investiu no desenvolvimento de tecnologias para a produção de etanol a partir do bagaço da cana-de-açúcar (etanol de segunda geração). Entretanto, há um grande esforço da comunidade científica para o desenvolvimento de novos processos, economicamente viáveis, para o aproveitamento da fração lignocelulósica da biomassa dos resíduos agrícolas (palha e bagaço de cana-de-açúcar, palha de trigo e resíduos de milho), resíduos florestais (pó e restos de madeira), e o capim elefante (Zheng et al., 2009); todas frações de biomassa que não são utilizadas com fins alimentícios.

O emprego da conversão bioquímica na produção de etanol de segunda geração é vantajoso já que preserva a estrutura original dos carboidratos, os açúcares monoméricos. Com isso a degradação biomassa por enzimas hidrolíticas é considerada como uma tecnologia sustentável para a sacarificação.

1.3 Utilização das Hidrolases Glicosídicas

O emprego das GHs não está restrito apenas para geração de biocombustíveis de fontes renováveis, se estende a aplicações terapêuticas (Hughes e Rudge, 1994) e nas indústrias de alimentos, fármacos, cosméticos, detergentes e tecidos (Singhania et al., 2010). Além disso, essa superfamília de enzimas possuem uma diversidade de papéis biológicos, como defesa contra patógenos, penetração de patógenos na célula, e reciclagem dos carboidratos (Henrissat et al., 1995).

Um dado econômico relevante é que as celulasas representam aproximadamente 20% do mercado mundial de enzimas, ocupando a terceira posição do ranking dos negócios nesta área (Singhania et al., 2010). No âmbito nacional, considerando-se as importações e exportações no ano de 2008, as celulasas movimentaram USD 1,35 milhões (Castro e Pereira, 2010). Com a perspectiva de aplicação dessas enzimas na produção de etanol ou outros derivados de processos fermentativos, podem transforma-las nas principais enzimas de uso industrial (Wilson, 2009).

O início da aplicação de celulasas no ramo industrial ocorreu por volta de 1980 na indústria de alimentos como um aditivo nas rações de ruminantes, visto que permitia o aumento da digestibilidade desses animais. Como a celulose é uma fonte de carboidratos, o uso das celulasas pode aumentar a biodisponibilidade desses carboidratos no trato digestivo, agindo sinergicamente com outras enzimas endógenas e exógenas, já existentes nesses animais.

Na indústria alimentícia, as celulasas juntamente com as hemicelulasas e pectinases são utilizadas no processo de maceração enzimática, o que permite maior rendimento dos frutos na produção de sucos, deixando-os mais nutritivos, além de auxiliar na extração e clarificação dos mesmos. Também são utilizados na fermentação de cerveja e produção de vinho; produção de néctares de frutas e purês; alteração das propriedades sensoriais das frutas e vegetais; e extração do óleo de oliva (Bhat, 2000). As xilanases são utilizadas para o aperfeiçoamento da qualidade de massas, pães, biscoitos e outros produtos de padaria. Essas

enzimas são capazes de hidrolisar arabinoxilana, substância presente em compostos utilizados na panificação. A quebra da arabinoxilana facilita a redistribuição de água nas massas e tal modificação facilita o manuseio das mesmas (Flander et al., 2008).

Na indústria têxtil, os tecidos de jeans ou brim alcançaram uma popularidade marcante nas últimas décadas. Nas décadas de 1970 e 1980 as indústrias desenvolveram uma metodologia de lavagem em tecidos para conferir maciez, e consequentemente maior conforto. Uma das possibilidades dessa lavagem é utilização de um coquetel de celulasas microbianas, processo denominado *bio-stoning*. Durante o processo, as celulasas atuam degradando as fibras da superfície do tecido (fibras soltas e microfibrilas), proporcionando a perda do índigo, tornando assim os tecidos mais lisos e macios (Bhat, 2000).

Nas indústrias de papel e celulose, as xilanases são as enzimas de maior destaque. Estas atuam no processo de branqueamento do papel, auxiliando na retirada da lignina, tornando o material mais apto ao tratamento químico, reduzindo assim os custos de produção. Se concomitante com a xilanase, houver a utilização da mananase, há uma maior eficiência no “biobranqueamento” do material (Bhat, 2000).

I. 4 Aspectos estruturais das celulasas

As celulasas têm sido alvo de muitos estudos tanto da academia como da indústria visto que hoje representam um dos principais gargalos do processo de degradação da biomassa vegetal por rotas enzimáticas. Estas apresentam baixa eficiência catalítica e, comumente, os principais organismos produtores de GHs não possuem um repertório balanceado de enzimas necessárias para completa redução do material lignocelulósico.

Análises de estrutura primária e terciária mostraram que as celulasas são enzimas modulares podendo ser formadas por domínios conectados por segmentos polipeptídicos flexíveis. O principal domínio é o catalítico, presente obrigatoriamente em todas as enzimas (Ferreira et al., 1990; Gilbert et al., 1990; Tomme et al., 1995). Os domínios adicionais, também conhecidos como acessórios, são responsáveis pelo reconhecimento e ligação ao substrato, modulando a ação da enzima em substratos insolúveis ou em estruturas cristalinas de celulose, proporcionando aumento da eficiência catalítica (Bolam et al., 1998; Boraston et al., 2003; Boraston et al., 2004).

Sabe-se que as celulasas exibem diferenças em suas propriedades estruturais e catalíticas, sendo unidas de acordo com a similaridade da sequência primária, enovelamento

estrutural, especificidade do substrato e mecanismo de ação (Henrissat, 1991; Henrissat e Bairoch, 1993; Henrissat et al., 1995; Henrissat e Davies, 1997). Atualmente as celulasas ou endocelulasas são classificadas em mais de 20 famílias (*Carbohydrate Active enZYme database* – CAZy, Cantarel et al., 2009). Muitas dessas famílias são estruturalmente relacionadas, existindo uma conservação dos resíduos e da maquinaria catalítica.

A maioria das enzimas com função de endocelulase pertence ao clã GH-A, formado pelas famílias 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 72, 79, 86, 113 e 128. Os membros desse clã hidrolisam β -glicosídeos através do mecanismo de retenção do carbono anomérico. O domínio catalítico presente nessas enzimas tem o enovelamento $(\alpha/\beta)_8$, com o próton doador e o nucleófilo localizados no C-terminal do *loop* 4 e da folha β 7, respectivamente, sendo o nome alternativo para esse clã 4/7 (figura 6A) (Ly e Wither, 1999).

Além do GH-A, há os grupos GH-B e GH-C, que também possuem endocelulasas, com o enovelamento característico *β -jelly roll* (barril beta), que pertencem às famílias 7 e 16 (GH-B) (figura 6C) e 11 e 12 (GH-C). Estas também clivam as ligações glicosídicas via retenção do carbono anomérico. Outro clã, o GH-M, com enovelamento característico $(\alpha/\alpha)_6$ também com enzimas de ação endocelulase, porém o mecanismo de reação é pela via de inversão do carbono anomérico (figura 6B). Apesar das diferenças estruturais marcantes entre os distintos clãs, tais enzimas compartilham da mesma função, a de endocelulasas (CAZy - Carbohydrate-Active enZYmes Database).

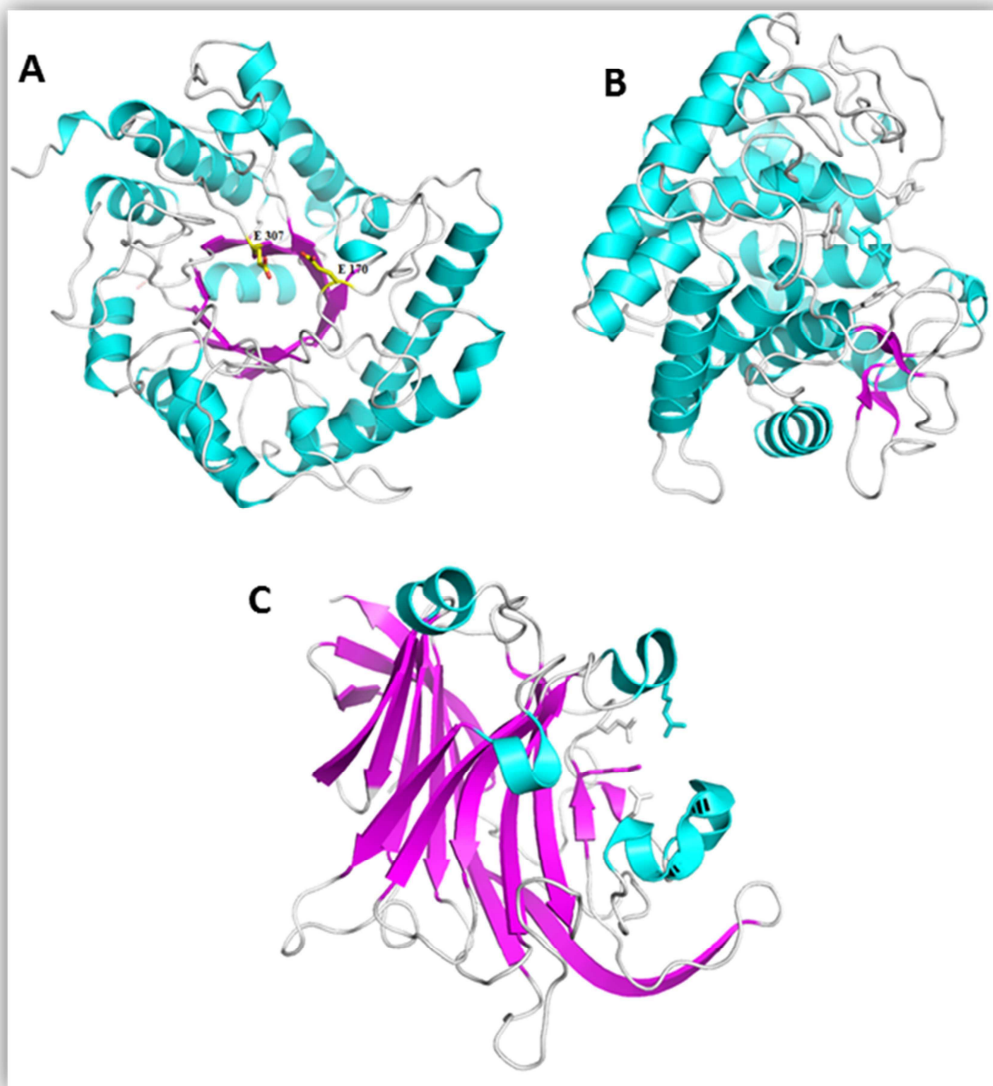


Figura 6: Representação em *cartoon* das endocelulases presentes nos clãs A, B e M. **(A)** Enzima pertencente ao clã GH-A (α/β)₈, CelCCA de *Clostridium cellulosyticum*, família GH5. Em amarelo os principais resíduos da catálise enzimática (PDBID 1EDG). **(B)** Enzima pertencente ao clã GH-M (α/α)₆, família 8, CelA de *Clostridium thermocellum* (PDBID 1CEM). **(C)** Enzima pertencente ao clã GH-B, família 6 com enovelamento β jelly roll, pLamA de *Pyrococcus furiosus* (PDBID 2VY0).

I. 5 As enzimas da família GH5

As enzimas pertencentes à família GH5 são encontradas principalmente em fungos e bactérias, dentre os membros há inúmeras celulasas. As enzimas dessa família tem a topologia estrutural $(\alpha/\beta)_8$ (figura 7), na qual oito fitas β , alternam-se sequencialmente com oito hélices α , arranjadas de tal modo que as fitas β são circundadas pelas hélices α . A primeira descrição dessa estrutura ocorreu em 1975, por Banner e colaboradores, e foi encontrada no músculo do coelho, estrutura que ficou reconhecida como barril TIM (Banner et al., 1975). Esse enovelamento pode ser encontrado em outras famílias de proteínas, com a grande maioria tendo o sítio ativo localizado na extremidade carboxi-terminal do barril (figura 7A) (Lesk et al., 1989; Forouhar et al., 2006; Hidaka et al., 2009; Repo et al., 2012).

Juntamente com a localização do sítio catalítico na extremidade C-terminal é característico dessas enzimas a presença de dois resíduos de ácido glutâmicos. Tais resíduos são bastante conservados e se localizam no C-terminal do loop 4 e da fita $\beta 7$ (figura 7B), identificados como doador de próton e nucleófilo, respectivamente, sendo essenciais para a catálise enzimática (Søgaard et al., 1993; McCarter e Withers, 1994). Além dos resíduos ácidos, os aromáticos e os polares tem atividade crucial na ação enzimática, sendo responsáveis por formar a fenda catalítica, onde há a acomodação do substrato para a clivagem. Os resíduos ácidos, aromáticos e polares interagem entre si e com o substrato formando uma rede de ligações de hidrogênio (Chapon et al. 2001; Dominguez et al. 1995; Ducros et al. 1995, Sakon et al. 1996).

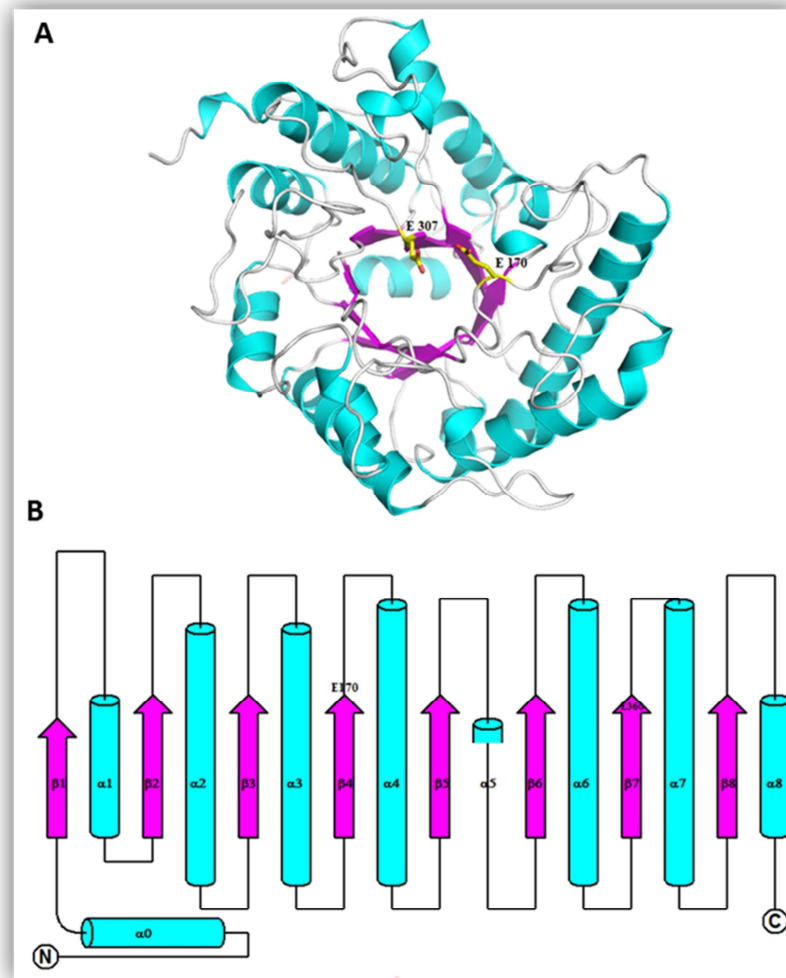


Figura 7: Representação da topologia barril TIM/enovelamento $(\alpha/\beta)_8$, característico nos membros da família GH5. Estrutura resolvida por cristalografia de raios X do domínio catalítico da CelCCA, celulase pertencente à bactéria gram-positiva *Clostridium cellulosum* (A) Representação das estruturas secundárias por *ribon*, em amarelo estão os resíduos Glu170 e Glu307, essenciais para a catálise. (B) Topologia dos elementos de estrutura secundária da CelCCA com 8 hélices α demonstradas por cilindros e 8 fitas β representadas por setas. (adaptação do Ducros et al., 1995).

Há uma variedade de atividades catalíticas presentes na família GH5, podendo ser de endocelulase, mananase, manosidase, xilanase, quitosanase, glicosidase, liquenase, fucosidase, galactomanase, endoglicoceramidase e manana-transglicosidase (Fukumori et al., 1986; Han et al., 2010; Dodd et al., 2010; Tanabe et al., 2003; Cutfield et al., 1999; Jakiviak et al., 2011; Yoshida et al., 2011; Sakamoto et al., 2007; Caines et al., 2007; Bourgault et al., 2005). A diversidade funcional dos membros dessa família é intrigante do ponto de vista estrutural uma vez que o enovelamento proteico $(\alpha/\beta)_8$ se mantém em todos os membros classificados nessa família. Apesar da baixa similaridade na sequência das enzimas da família

GH5, oito resíduos são imprescindíveis e invariáveis, desses sete são responsáveis pela interação com o substrato (Davies et al., 1998).

A variedade funcional indica que a família GH5 tem um papel chave no metabolismo dos carboidratos presentes na parede celular vegetal o que as tornam candidatas atrativas para a produção de biocombustíveis de segunda geração e para outras aplicações industriais (Park et al., 2012). Essa diversidade permitiu a divisão em subfamílias de acordo com a similaridade de sequência e o tipo de maquinaria catalítica. A classificação dos membros da família GH5 em subfamílias incide em uma divergência evolutiva dentro da família, indicando que as enzimas não compartilham de um ancestral comum quando se analisa o sítio ativo; mas há uma conservação evolutiva relacionada ao enovelamento das mesmas. Os aspectos para a classificação em subfamília ocorrem a partir de análises das regiões de reconhecimento do substrato (subsítios +1/-2). Os subsítios fornecem evidências diretas para evolução convergente quando se analisa os resíduos responsáveis pela ligação ao anel piranosídico (Wang et al., 1993).

Os membros da família GH5 podem ser divididos em quatro subfamílias. Na subfamília 5-1, o subsítio +1 é formado pelo Trp213 (numeração baseada na estrutura com código PDB: 1ECE) que está posicionado no *loop* entre fita β 5 e a hélice α 5; tal região é formada por um longo *loop* responsável por compor a superfície de ligação do substrato. Na subfamília 5-2, o Trp178 (nomeação baseada na estrutura PDB: 1EGZ) forma a base do subsítio (+1), através de um empilhamento do anel aromático com as faces hidrofóbicas dos anéis piranosídicos. O Trp178 é conservado em todas as sequências da subfamília 5-2, localizado na fita β 5 que precede uma curta hélice α (Davies et al., 1998). Na subfamília 5-3, o resíduo doador de próton está no *loop* entre a fita β 4 e a hélice α 4, e não na fita β 4; outra característica é que o anel aromático que contribui para a formação do subsítio -1 provem da Tyr176 (nomeação baseada na estrutura PDB: 1CEC), e não mais do Trp (Dominguez et al., 1995; Dominguez et al., 1996). Na subfamília 5-4, o Trp180 que caracteriza o grupo está localizado no *loop* entre a fita β 4 e a hélice α 4 (PDB: 1EDG) (Ducros et al., 1995).

I. 6 Mecanismo de catálise das enzimas GH5

As enzimas pertencentes às GHs podem hidrolisar as ligações glicosídicas por dois mecanismos; pela via de retenção ou de inversão do carbono anomérico (figura 8) (Koshland, 1953; Sinnott, 1990). A família GH5 utiliza a primeira via. Esses mecanismos operam de maneiras distintas; enquanto a via que retém o carbono utiliza um mecanismo de duplo

deslocamento que envolve a formação de um intermediário enzima-glicose, mantendo assim a configuração β original da glicose. A via que inverte o carbono anomérico age através de um deslocamento do grupo com a participação de uma molécula de água que efetua a quebra da ligação glicosídica, concomitante com a inversão da configuração do substrato, passando a ser α -glicose (McCarter e Withers, 1994). Ambos os mecanismos de catálise ácida ocorrem com a participação dos resíduos carboxilados: o nucleófilo e o doador de prótons (ácido/base).

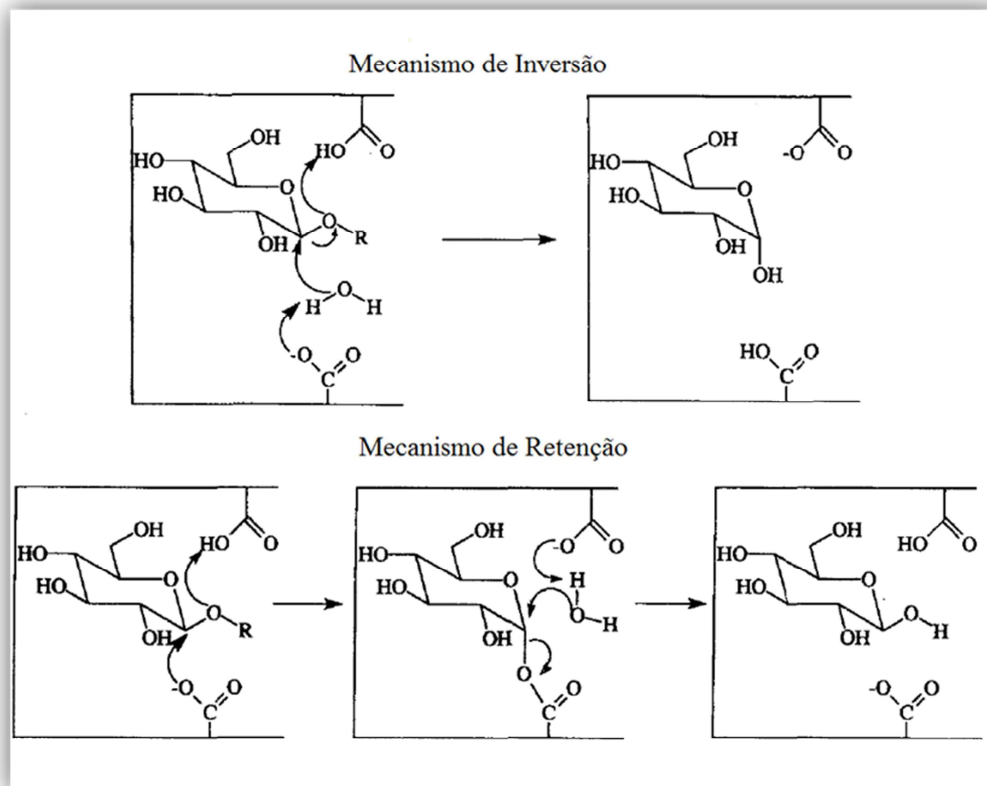


Figura 8: Mecanismo enzimático de hidrólise da ligação glicosídica utilizando a via de inversão ou retenção do carbono anomérico (adaptação McCarter e Withers, 1994).

A arquitetura dos sítios catalíticos dessas enzimas apresenta o resíduo nucleofílico e o resíduo doador de prótons em lados opostos no sítio catalítico. Para a via com retenção, a distância média entre os resíduos é de aproximadamente 5 Å já que a base do ataque nucleófilo encontra-se na vizinhança próxima ao carbono anomérico. A via de inversão tem uma distância média de 10 Å para que ocorra a acomodação da água e do substrato entre os grupos carboxilas (McCarter e Withers, 1994; Wang et al., 1994). Como demonstrado na figura 8, na via de inversão o carbono é atacado pelo oxigênio da água (devidamente polarizado pelo nucleófilo da enzima) essa molécula se posiciona no interior do sítio ativo. Já no mecanismo de

retenção o carbono anomérico é atacado pelo próprio grupo nucleófilo da enzima (Davies e Henrissat, 1995).

A quebra da ligação glicosídica pela via de retenção ocorre através de dois passos com duplo deslocamento, que envolve a formação e a hidrólise de um intermediário covalente enzima-glicose. Os passos procedem através do estado de transição do íon *oxocarbenium*. No sítio catalítico, os ácidos carboxílicos têm papéis distintos: de ação como nucleófilo, atacando o centro anomérico do açúcar para formar a espécie enzima-glicose; e atuando como catalisador ácido/base. O oxigênio glicosídico do primeiro passo faz a protonação (catálise ácida) e o segundo passo, a desprotonação da água (catálise básica) (Ly e Withers, 1999).

O trabalho de Davies e colaboradores (1998) mostra, através de resolução da estrutura cristalina, a reação enzimática de retenção do carbono anomérico da família GH5. O estudo apresenta quatro estruturas: a nativa; em complexo com substrato; com o intermediário e com o produto. Estudos anteriores demonstraram que o açúcar no subsítio -1 tem uma distorção chamada padrão 4C_1 , conformação que permite um estado de transição planar, aceito para as enzimas que promovem catálise ácido/base (Sulzenbacher et al. 1996; Tews et al., 1996). A molécula de água está situada na posição para um ataque nucleofílico, no centro anomérico acima do anel piranosídeo e da ligação de hidrogênio dos resíduos catalíticos ácido/base. A densidade eletrônica observada no produto na posição do subsítio -1 foi difusa, consistindo com o tipo de ligação fraca que é esperada entre o produto e a enzima (Davies et al., 1998).

A compreensão do mecanismo de ação dessas enzimas permite buscar melhorias na sua ação enzimática, modulando-as a agirem em condições adversas, como diferentes pHs, ampla faixa de temperatura e diferentes concentrações salinas.

I. 7 Organismos produtores de celulasas

Microrganismos celulolíticos são os responsáveis pela degradação do material presente em plantas e solos. Devido à eficiência na decomposição de materiais lignocelulósicos, os microrganismos que produzem enzimas do complexo celulolítico desempenham um importante papel no meio ambiente estabelecendo um elo chave no ciclo do carbono. A maioria desses organismos secretam as celulasas fora da sua célula, pois são incapazes de transportar materiais insolúveis (b Wilson, 2009).

Os fungos e bactérias aptos a produzirem enzimas capazes de degradar a celulose, são comuns em habitats como solo, compostos ou rúmen de animais (Hilge et al., 1998). Dentre os

organismos produtores, os que são comumente utilizados para a produção de celulases são os fungos *Trichoderma sp.*, *Aspergillus sp.*, *Penicillium sp.*, e bactérias do gênero *Clostridium*, *Bacillus* entre outros. Nesse trabalho, os alvos são enzimas oriundas das bactérias do gênero termofílicas *Bacillus*, e *Thermotoga*, a mesofílica *Xanthomonas* e uma enzima oriunda da biblioteca de metagenômica do solo de cana-de-açúcar, sem organismo indentificado.

As espécies do gênero *Bacillus* como *B. cereus*, *B. licheniformis*, *B. polymyxa* e *B. subtilis* são capazes de produzir enzimas que degradam a biomassa lignocelulósica. Entre as enzimas produzidas estão as endocelulases (Sharma et al., 1990; Han et al., 1995). Uma das enzimas discutidas nesse trabalho, a BsCel5A é uma endocelulase de *Bacillus subtilis*, uma bactéria gram-positiva, não patogênica, saprófita e termofílica. Utilizada como probiótico e amplamente empregada no setor industrial para produção de enzimas e moléculas de interesse médico e biotecnológico (Sonenshien et al., 2002). A cepa *B.subtilis* está correlacionada com diferentes linhas de pesquisa, sendo considerado um organismo modelo, o que facilita os estudos com as suas celulases, devido ao conhecimento biológico já adquirido desse microrganismo.

Também foi estudada uma celulase da família GH5 pertencente à bactéria gram-negativa *Xanthomonas axonopodis* pv. *citri* (*X.citri*), tal organismo pertence à classe gamma de Proteobactérias. A bactéria é o agente causador do cancro cítrico, uma das mais importantes doenças que acomete a citricultura no país (Rodrigues Neto e Ribeiro, 200). A *X.citri* é a responsável pelo cancro cítrico na sua forma mais agressiva, com ampla ocorrência e afetando praticamente todas as variedades de citros comerciais (Brunings e Gabriel, 2003; Schubert e Sun, 1996; Barbosa-Mendes et al., 2009). Esta penetra no hospedeiro através dos estômatos ou pequenos ferimentos na epiderme das folhas, frutos e ramos. A disseminação local também ocorre através de podas com ferramentas contaminadas e pode ser facilitada por insetos herbívoros, como a larva minadora dos citros (*Phyllocnistis citrella*), a qual gera galerias nas folhas (Brown, 2001).

Devido à sua importância econômica a *X.citri* teve seu genoma completamente sequenciado (da Silva et al., 2002), e foram identificados várias ORFs (*open reading frames*) associadas ao processo de infecção, sobrevivência, patogenicidade, e sequências relacionadas com enzimas de degradação da parede celular da planta hospedeira. Estão presentes no genoma de *X.citri* mais de 100 hidrolases glicosídicas que podem causar desordem na parede celular vegetal (CAZy). Estudos com outros fitopatógenos demonstraram que degradação da parede celular de plantas hospedeiras parece, em muitos casos, ser um dos principais

mecanismos de interação com a planta e, conseqüentemente, com a colonização do tecido (Toth et al., 2003).

A relevância dos estudos de celulasas pertencentes a *X.citri*, está além de suas aplicações biotecnológicas visto que, do ponto de vista biológico, os fitopatógenos em geral contém enzimas altamente específicas para interação com seu hospedeiro. Adicionalmente, a caracterização estrutural dos componentes desta bactéria trará contribuição para comunidade científica e para os estudos da patogenicidade bacteriana.

A abordagem metagenômica é uma ferramenta poderosa para prospecção de novas moléculas com diversas funções e características. Uma pequena amostra de solo primitivo pode conter dez mil diferentes tipos de bactérias, mais de um milhão de sequências proteicas, sendo muitas dessas com características exclusivas (Lorenz e Eck, 2005). Uma das enzimas caracterizadas nesse trabalho é oriunda de uma biblioteca metagenômica de DNA do solo de cana-de-açúcar. Os propósitos da metagenômica incluem o entendimento dos processos ecológicos, e a busca de novos genes, conseqüentemente de proteínas de interesse a partir de organismos não cultivados, que podem ser biotecnologicamente relevantes (Watanabe e Kassai, 2008). A abordagem metagenômica realizada por Alvarez TM (2012, dados não publicados) no solo de cana-de-açúcar, localizou uma nova celulase pertencente à família GH5, a CelE1, demonstrando o potencial da mesma para a localização de enzimas aptas a degradar biomassa vegetal, e para aplicação nos diversos ramos industriais.

A bactéria hipertermófila *Thermotoga petrophila*, é capaz de produzir enzimas que degradam a biomassa vegetal, as quais mais de 50 enzimas são classificadas como hidrolases glicosídicas pertencentes as mais diversas famílias. Entre as GHs encontra-se a manana endo-1,4- β -D-manosidase (EC 3.2.1.78, TpMan), pertencente à família GH5. Esta bactéria foi isolada a partir de um reservatório de óleo subterrâneo de Nigata, Japão. A *T.petrophila* pode ser encontrada em uma faixa de temperatura de 47 a 88 °C com crescimento ótimo em torno de 80 °C (Takahata et al., 2001). Outras bactérias pertencentes ao gênero *Thermotoga*, tem a condição ótima para o crescimento as temperatura entre 66 - 70 °C (Fardeau et al., 1997; Jeanthon et al., 1995; Ravot et al., 1995). Nos últimos anos, os estudos com as enzimas termoestáveis vêm crescendo devido ao interesse industrial relacionado à elaboração de um coquetel enzimático termofílico compatível com condições atuais de fermentação da cana-de-açúcar.

As pesquisas relacionadas ao alto rendimento na degradação de celulose estão distantes de atingir a saturação, todas as informações, funcionais ou estruturais, relacionadas a tais enzimas é de suma importância para concretizar esse ramo industrial. Para máxima eficiência dessas enzimas é necessário superar obstáculos, e adquirir conhecimentos de novas

celulases que sejam capazes de realizar a hidrólise mesmo em condições enzimáticas desfavoráveis, como altas temperaturas, concentração de sais e variações de pH.

II. Objetivo

Contribuir na elucidação das bases moleculares envolvidas na correlação estrutura-função-estabilidade das enzimas GH5 com potenciais aplicações biotecnológicas e desvendar possíveis novas estratégias para termoestabilização baseado no enovelamento barril (α/β)₈. Para isso uma abordagem multidisciplinar explorando ferramentas biofísicas e a determinação de estruturas tridimensionais foi empregada obtendo desta forma informações sobre a termoestabilidade e especificidade de enzimas da família GH5.

III. Material e métodos

III. 1 Clonagem

Os clones para expressão das enzimas selvagens de BsCel5A, TpMan e CelE1, clonadas em pET-28a, foram cedidos pelos pesquisadores Prof. Dr Fábio Squina e Prof. Dr Rolf Prade, ambos vinculados ao Laboratório Nacional de Ciências e Tecnologia do Bioetanol –CTBE.

A enzima XacCel5 foi amplificada do DNA genômico da *Xanthomonas axonopodis pv. citri*; para tanto foram desenhados oligonucleotídeos iniciadores correspondentes a sequência nucleotídica do fragmento alvo capazes de hibridizar com a extremidade 5' (CATATGCTGAAGTATGTTGGCGTCAATC) e 3' (GTCGACTTATGCGTACTTGCTCAGGATC). Para as reações de amplificação do fragmento de interesse, utilizou-se a técnica de reação em cadeia da polimerase (PCR - *Polymerase Chain Reaction*) contendo PCR Rxn Buffer invitrogen®; 1,5 mM MgCl₂; 0,2 mM dNTP mix Fermentas®; oligonucleotídeos específicos com concentração final de 0,5 mM; 20 ng de DNA fita molde e 2,5U de enzima Taq DNA polimerase *high fidelity*. Inicialmente a XacCel5 foi clonada no vetor pGEM-T; após confirmação do fragmento de interesse por sequenciamento de DNA, o inserto foi digerido com as enzimas Nde I e Sal I, finalmente, subclonado em vetor de expressão pET-28a.

III. 2 Expressão das proteínas recombinantes

Para a produção das proteínas recombinantes utilizou-se meio LB (Luria-Bertani)-canamicina. A tabela 1 sumariza condições de expressão utilizadas para proteínas recombinantes, incluindo cepas, temperatura, e tempo de indução.

Tabela 1: Condição para a produção das proteínas heterólogas.

Proteínas	Cepas	Temperatura °C	Tempo de indução
BsCel5A	BL21(BE3) Δ slyD	37	4 hrs
CC_BsCel5A	BL21(BE3) Δ slyD	37	4 hrs
CBM_BsCel5A	BL21(BE3) Δ slyD	37	4 hrs
TpMan	BL21(BE3) Δ slyD	37	4 hrs
DC_TpMan	BL21(BE3) Δ slyD	37	4 hrs
XacCel5	BL21(DE3)	37/20	16 hrs
CelE1	Roseta- gami™(DE3)pLys-S	37	4 hrs

III. 3 Purificação das proteínas recombinantes

Para que as proteínas de interesse pudessem ser purificadas do extrato bruto bacteriano, fez-se a lise das células em diferentes soluções tamponantes (tabela 2), através de choque mecânico por ultrassom (Vibra Cell® Sonics, 40% de amplitude). O lisado bacteriano foi centrifugado durante 1 hora a 4 °C a uma rotação de 20000 xg. Todas as proteínas possuíam cauda de histidina fusionada e, portanto, o primeiro passo cromatográfico foi por afinidade a níquel imobilizado (IMAC), em coluna pré-empacotada HiTrap® Chelatinhg HP (GE), acoplada ao sistema AKTA-FPLC (*Fast Performance Liquid Chromatography* – Amersham Biosciences). O sistema foi previamente equilibrado com solução tamponante A (tabela 2). A eluição das proteínas de interesse foi realizada através de gradiente não linear de solução tamponante A contendo imidazol (solução B) e monitorada em comprimento de onda 280 nm. As amostras coletadas foram analisadas em SDS-PAGE 13% (Laemmli, 1970) corado com azul de Coomassie.

As proteínas apresentaram elevado grau de pureza, sendo necessário apenas mais uma etapa cromatográfica, a exclusão molecular (SEC – *size exclusion chromatography*), para realizar um polimento da amostra, garantindo assim homogeneidade estrutural. Foi utilizada a coluna pré-empacotada HiLoad Superdex 16/60 (GE) acoplada ao sistema AKTA-FPLC, (tabela 2). A enzima BsCel5A foi uma exceção, sendo necessário um passo de troca catiônica entre a IMAC e a SEC. Para tal foi utilizada a coluna Hitrap HP SP (GE) acoplada ao sistema AKTA-FPLC.

Tabela 2: Soluções tamponantes utilizadas na purificação das proteínas recombinantes.

Solução	Composição
Lise celular BsCel5A	50 mM fosfato pH 7,2; 100 mM de NaCl; 1 mM PMSF; 5 mM de benzamidina
IMAC A BsCel5A	50 mM fosfato de Sódio pH 7,2; 100 mM NaCl; 5 mM benzamidina
IMAC B BsCel5A	50 mM fosfato de Sódio pH 7,2; 100 mM NaCl e 500 mM imidazol
TC A BsCel5A	50 mM fosfato pH 7,2; 5 mM benzamidina
TC B BsCel5A	50 mM fosfato pH 7,2; 1 M de NaCl
SEC BsCel5A	50 mM fosfato pH 7,2; 250 mM de NaCl; 5 mM benzamidina
Lise celular CC_BsCel5A/CBM_BsCel5A	50 mM fosfato de Sódio pH 7,2; 100 mM NaCl; 1 mM PMSF; 5 mM benzamidina e 0,03 mg/ml lisozima
IMAC A CC_BsCel5A/CBM_BsCel5A	50 mM fosfato de Sódio pH 7,2; 100 mM NaCl
IMAC B CC_BsCel5A/CBM_BsCel5A	50 mM fosfato de Sódio pH 7,2; 100 mM NaCl e 500 mM imidazol
SEC CC_BsCel5A/CBM_BsCel5A	20 mM tampão fosfato de Sódio pH 7,2; 150 mM NaCl
Lise celular TpMan/DC_TPman	20 mM fosfato de Sódio pH 7,4; 50 mM NaCl; 5 mM imidazol; 1 mM PMSF e 0,03 mg/ml lisozima
IMAC A TpMan/DC_TPman	20 mM fosfato de Sódio pH 7,4; 50 mM NaCl; 5 mM imidazol
IMAC B TpMan/DC_TPman	20 mM fosfato de Sódio pH 7,4; 50 mM NaCl; 500 mM imidazol
SEC TpMan/DC_TPman	20 mM fosfato de Sódio pH 7,4; 150 mM NaCl
Lise celular XacCel5	20 mM fosfato de Sódio pH 7,4; 500 mM NaCl; 5% glicerol; 1 mM PMSF; 5 mM Benzamidina e 0,03 mg/ml de lisozima
IMAC A XacCel5	20 mM fosfato de Sódio pH 7,4; 500 mM NaCl e 5% glicerol
IMAC B XacCel5	20 mM fosfato de Sódio pH 7,4; 500 mM NaCl; 5% glicerol e 500 mM imidazol
SEC XacCel5	20 mM fosfato de Sódio pH 7,4 e 200 mM NaCl
Lise celular CelE1	20 mM fosfato de Sódio pH 7,4; 100 mM NaCl; 1 mM PMSF; 5 mM Benzamidina e 0,03 mg/mL lisozima
IMAC A CelE1	20 mM fosfato de Sódio pH 7,4; 100 mM NaCl
IMAC B CelE1	20 mM fosfato de Sódio pH 7,4; 100 mM NaCl e 500 mM imidazol
SEC CelE1	50 mM fosfato de Sódio pH 7,4 e 150 mM NaCl

III. 4 Espalhamento de Luz Dinâmico - DLS

Experimentos de espalhamento de luz dinâmico - DLS (*Dynamic light scattering*) foram realizados com as proteínas recombinantes purificadas, com objetivo de verificar a monodispersidade populacional em solução (homogeneidade estrutural). A medida de distribuição populacional tem-se demonstrado essencial para o sucesso dos ensaios de cristalização e realização de análises biofísicas (Wårdell et al., 1993; Goldberg, 1994). Além da polidispersidade, também se obtém o raio hidrodinâmico (R_h), que depende do tamanho físico e forma da molécula e do ambiente (difusão, viscosidade) onde esta se encontra. Com o R_h foi possível inferir os estados de oligomerização da proteína em solução.

Para realizar as medidas de DLS, as amostras foram previamente centrifugadas por 20 minutos a 4 °C com uma rotação de 20000 xg. O número médio de aquisições para cada amostra foi 100. Os experimentos foram realizados no equipamento DynaPRO 180 (*Protein Solutions*) a 18 °C, com temperatura controlada por um sistema eletrônico *Peltier*. Os parâmetros hidrodinâmicos foram calculados utilizando o software DYNAMICS v.6.10.1.2.

III. 5 Dicroísmo circular - CD

O espectro do dicroísmo circular (CD – *Circular Dichroism*) fornece principalmente características da estrutura secundária das proteínas, e as mudanças no espectro de CD pode refletir a perturbação na estrutura tridimensional. Por esta técnica pode-se realizar estudos de desnaturação térmica e resposta a ligação da proteína ao seu ligante, monitorando o espectro da mesma (Vedadi et al., 2010).

Os espectros de ultravioleta distante (190 – 260 nm) do CD foram coletados no espectropolarímetro Jasco J-180 (JASCO) equipado com módulo *Peltier* para o controle da temperatura. As medidas foram realizadas em uma cubeta de quartzo de 1 mm de caminho ótico. Os dados gerados permitiram analisar a composição da estrutura secundária, assim como a estabilidade da proteína alvo.

Os experimentos foram realizados com as amostras na concentração de 10 μ M. Os dados foram coletados com uma varredura de 50 nm/min, com tempo de resposta 4 segundos e 20 acumulações com intervalo de 60 segundos. O espectro final foi obtido pela acumulação das varreduras corrigidas pela subtração do espectro do tampão correspondente. Os dados gerados com a subtração dos espectros foram convertidos em elipticidade molar residual $[\Theta]$, obtida pela equação:

$$[\Theta] = \Theta / 10 \times l \times C \times N$$

na qual Θ é a elipticidade observada (grau), l é o caminho ótico em centímetros (cm), C é a concentração da proteína (molar) e N é o número de resíduos da proteína (Adler et al., 1973). Os dados convertidos foram tratados utilizando o programa ORIGIN 8.

Para avaliar a estabilidade proteica, foram realizados experimentos de desnaturação térmica, monitorados no comprimento de onda de 222 nm, já que há o pico negativo característico das hélices α , estrutura secundária presente na família GH5 das celulasas. Já para a CBM_BsCel5A, o comprimento de onda utilizado foi o 230 nm, que é um comprimento de onda típico para monitoramento dos resíduos aromáticos como o triptofano. As desnaturações foram realizadas em diferentes pHs e na presença de sais divalentes. Os dados obtidos da curva de desnaturação térmica foram transformados para frações desenoveladas (fD) de acordo com a seguinte equação:

$$fD = (X - X_N) / (X_D - X_N)$$

na qual fD é a fração desenovelada, X_N é a proteína nativa, X_D a proteína desenovelada e X é o valor observado (Greenfield, 2007).

Para observar a estabilidade de proteína utiliza-se o T_M , que é o ponto médio da curva de desnaturação, conhecido com temperatura de transição (*melting temperature*) (Privalov, 1979; Schellman, 1997; Vedadi et al., 2006).

III. 6 Proteólise limitada

A proteólise limitada dos domínios da celulase 5A de *Bacillus subtilis* foi realizada com enzima tripsina (SIGMA), na proporção de 1:100 (enzima: proteína) (Dong et al., 2007). Para analisar o tempo ideal de clivagem realizou-se um teste em pequena escala, onde foram coletadas as amostras nos tempos 0, 10, 20 e 30 minutos e 16 horas. Os testes foram realizados a temperatura ambiente e a 4°C. As melhores condições para a clivagem foram visualizadas através de SDS-PAGE 13% corado com azul de Coomassie.

III. 7 Ensaio de cristalização

A proteína a qual deseja cristalizar geralmente deve estar isolada por técnicas de purificação, homogênea, e biologicamente ativa (Gilliland, 1988). Atualmente sabe-se que todos os passos da cristalização de proteína estão interligados, e ocorrem em continua

evolução. A nucleação e o crescimento procede após a total supersaturação da solução diminuir (Boistelle e Astier, 1988).

A tarefa de cristalizar uma macromolécula biológica é complexa, já que envolve a interferência de variáveis como: estabilidade, grau de pureza, pH, temperatura, agentes precipitantes, força iônica, entre outras. Diante de tantas variáveis torna-se difícil fazer a predição de condições específicas para a cristalização da proteína, fazendo com que isso se torne um processo amplamente dependente de tentativas experimentais (Bergford, 1999).

A técnica de cristalização de macromoléculas biológicas utilizada foi a difusão de vapor, que consiste no equilíbrio entre duas soluções através da fase de vapor um sistema isolado. Essa técnica pode ser aplicada de diferentes formas: gota suspensa (*hanging drop*), gota sentada (*sitting drop*) e gota sanduíche (*sandwich drop*) (Bergford, 1999). Nesse trabalho foram utilizados dois métodos, a gota suspensa e a gota sentada.

Inicialmente empregou-se ensaios em gotas sentadas (0,5 μ L da solução de proteína com 0,5 μ L da solução condição) em placas de 96 poços, preparadas pelo robô *Honeybee* disponível no Laboratório Robotizado de Cristalização de Proteínas (Robolab-LNBio-CNPq). Para testes iniciais foram utilizados os *kits* comerciais *Hampton Screen* I e II, *Wizard* I e II, *PACT*, *JCSG+*, *SaltRx* e *Precipitant Synergy* os quais abrangem uma ampla faixa de pH, concentração de sal e precipitantes. As placas foram preparadas e mantidas a 20 °C.

Para melhorar de cristais obtidos nos testes iniciais, preparou-se o refinamento das condições onde sempre uma variável foi modificada enquanto, outros componentes da condição foram mantidos. Num primeiro momento manteve-se o pH e variou-se o precipitante, em pequenas escalas (0,5 a 1% de volume). Posteriormente, fez-se uma varredura em diferentes pHs, e, por fim foram testados aditivos, tais como glicerol e dioxano, até obtenção de um cristal com forma e tamanho adequados para experimentos de difração.

III. 8 Coleta e processamento de dados da difração

As coletas de dados de difração dos monocristais foram realizadas na linha de luz do Laboratório Nacional de Luz Síncrotron (LNLS), Campinas – SP que possui um detector marmosaic 225 CCD (Mar Research) e uma mesa experimental da NatXray/Irelec. Para diminuir o dano à radiação que os cristais sofrem devido à formação de radicais livres, utilizou-se da coleta sob condições criogênicas a 100 K com fluxo de nitrogênio gasoso (Oxford Nitrogen

Cryojet XL) (Garman e Schneider, 1997). Os dados foram indexados, integrados e escalonados utilizando o pacote HKL2000 (Otwinowski e Minor, 1997).

Para a CC_BsCel5A empregou-se uma oscilação de 1° para uma faixa de 180° . Antes do resfriamento os cristais passaram por uma solução crio-protetora contendo 20% de glicerol. Para a coleta do cristal com cloreto de manganês, o cristal passou por uma solução crio-protetora contendo também 50 mM de MnCl_2 , por 5 minutos. O comprimento de onda utilizado foi 1,45 Å. Houve a coleta de 3 conjuntos de dados com resoluções limites foram 2,90; 2,10 e 1,97 Å.

Para a DC_TpMan, o cristal do domínio catalítico foi difratado sem passar por crioproteção, pois a condição já apresenta ação crio-protetora. Os cristais com derivados de iodo foram preparados de acordo com o método de *quick cryo-soaking* de Dauter e colaboradores, 2000. Tal método usa-se a solução crio-protetora com iodo para fazer a incorporação de íons na região ordenada do solvente que fica em volta das moléculas de proteínas. Tais cristais foram mergulhados na solução crioproteção com 12,5% de glicerol e 500 mM de iodeto de sódio, por 1 minuto. Para resolução da estrutura com glicose, também foi realizado *soaking*, com a solução do cristalizante com 500 mM de glicose, por 5 minutos. Assim como o complexo com maltose, que foi obtido por adição de 100 mM de maltose na solução cristalizante. O conjunto derivado com iodo foi coletado a energia de 7797 eV, para que o sinal anômalo fosse acentuado, enquanto que os complexos com açúcares foram coletados a 8500 eV. Para tais cristais, um total de 360 imagens foi coletado, usando uma oscilação de 1° , com uma exposição de 30 segundos por imagem.

Cristais da CelE1 também não necessitaram de adição de solução crioprotetora, já que a solução cristalizante continha 4% de glicerol. O comprimento de onda utilizado foi 1,46 Å com tempo de exposição de 20 segundos e oscilação de 1° . Cristais de CelE1 difratam a resolução de 1,78 Å.

III. 9 Resolução de Estrutura e Refinamento

Para resolver os problemas das fases foram utilizados dois métodos substituição molecular e SIRAS (*single isomorphous replacement with anomalous scattering*).

Para as celulasas, CC_BsCel5A e CelE1, foi utilizado o método de substituição molecular, com o auxílio do programa MOLREP (Vagin e Teplyakov, 2010), que faz parte do programa CCP4 (*Colaborative Computational Project Number 4*).

Como modelo de busca para a BsCel5A, utilizou-se a celulase 5A de *Alkalophilic Bacillus* sp. (PDB ID: 7A3H) que apresenta uma identidade da estrutura primária de 70%. Para a CelE1, utilizou-se as coordenadas do domínio catalítico da Celulase 5G de *Pseudoalteromonas haloplanktis* (código do pdb: 1TVP) com uma identidade de sequência de 67%.

Já para a mananase, o problema das fases foi resolvido com o método SIRAS, *quick cryo-soaking*, utilizando o iodo. As posições dos átomos de iodo com ocupância acima de 0,4 foi obtida utilizando o programa SHELXD (Schneider e Sheldrick, 2002). Para estimar as fases, foi utilizado o programa SHELXE, o qual também emprega rotinas de modificação de densidade para melhorar a qualidade das fases estimadas experimentalmente (Sheldrick, 2002). A maior parte do molelo, 90%, foi construído automaticamente com o programa ARP/wARP e concluído manualmente no programa COOT.

Todas as estruturas foram refinadas com o programa REFMAC com ciclos de intercalados de inspeção manual no programa COOT. As moléculas de água e os ligantes foram adicionados após refinamento das cadeias polipeptídicas. Para dados abaixo de 2 Å foi empregado o refinamento com restrições geométricas e isotrópico, enquanto para dados de alta resolução foi utilizado refinamento anisotrópico.

III. 10 Espalhamento de raios X a baixos ângulos - SAXS

Os experimentos de SAXS (*Small Angle X-ray Scattering*) foram realizados na linha de luz D02A-SAXS2 (LNLS-CNPq Campinas – SP), equipado com detector MarCCD 165 mm. Os dados bidimensionais gerados foram analisados e a integração das contagens sobre a região detectada para obtenção da curva de espalhamento foi realizada no programa FIT2D (Hammersley, copyright 1987 – 2005).

A celulase BsCel5A e os seus domínios BsCel5A foram medidos na concentração de 2, 4 e 8 mg/mL em 10 mM de tampão fosfato pH 7,0 com temperatura controlada a 18 °C, utilizando uma célula de mica de 1 mm e um feixe de raios-X monocromático ($\lambda = 1,488 \text{ Å}$).

Antes do experimento de espalhamento, as amostras foram centrifugadas durante 15 minutos a 20000 xg, a 4 °C. As imagens coletadas foram de 300 segundos, sendo o dado relativo ao tampão subtraído da medida da amostra. As curvas de intensidade foram individualmente corrigidas para resposta do detector, escalonadas pela intensidade do feixe incidente e absorção da amostra.

O raio de giro (R_g) foi elucidado, primeiramente, usando a aproximação Guinier (Guinier e Fournet, 1955, Glatter e Kratky, 1982; Feigin e Svergun, 1987). Em seguida o R_g

também foi calculado a partir da função de distribuição de distâncias, $p(r)$, pela transformada de Fourier indireta da curva de intensidade utilizando o programa GNOM (Svergun, 1992) implantado no programa PRIMUS (Konarev et al., 2003). A função $p(r)$ também fornece a dimensão máxima (D_{max}) da molécula, além disso, a representação de Kratky da curva de intensidade ($q^2 I(q)$ vs. q) foi utilizada para analisar a compactação/flexibilidade da conformação da proteína (Glatter e Kratky, 1982; Feigin e Svergun, 1987). O envelope da baixa resolução dos domínios da BsCel5A foram modelados usando o método *ab initio* no programa DAMMIN (Svergun, 1999), o qual contém uma configuração capaz de ajustar os modelos experimentais, garantindo a compactação e a conectividade da solução. O modelo de baixa resolução e a estrutura modelada foram sobrepostos usando o programa SUPCOMB (Kozin e Svergun, 2001).

III. 11 Caracterização bioquímico-cinética

III. 11. a. Atividade sobre diferentes substratos

Os testes de atividade foram baseados na detecção de açúcares redutores pelo método do DNS (ácido 3,5-dinitrosalicílico) (Miller, 1959). A análise do perfil de atividade da celulase BsCel5A em diferentes substratos foi feita em reação contendo 40 μL de solução tamponante, 50 μL de substrato e 10 μL de amostra das enzimas a 100 ng/ μL . Após o tempo de 10 min de incubação a 60 °C, adicionou-se a 100 μL do reagente de DNS para interromper a reação através do aquecimento a 95 °C por 5 min, resfriada em banho de gelo por 2 min. A atividade foi quantificada pela leitura de absorbância de 100 μL de reação em leitor de microplacas Infinite 200 (TECAN®) a 540nm. O valor da absorbância foi correlacionado a uma curva padrão de açúcar redutor.

Os substratos testados foram β -glucana, carboximetil-celulose (CMC), arabinogalactomanana, xiloglucano, arabinano desramificado, 1-4- β -D-manana, arabinoxilano, xilano de madeira de faia, pectina e liquenano. A unidade de enzima foi definida como a quantidade de enzima necessária para liberação de 1 μmol de produto por minuto.

A glicose foi utilizada para montar a curva padrão. Para determinar o pH e a temperatura ótima, as reações enzimáticas foram realizadas em diferentes tampões, no sistema adaptado de McIlvaine.

III. 11. b. Cinética

Os ensaios de cinética para as enzimas foram realizados com CMC para a celulase. Foram testadas 12 concentrações de substratos, com adição da enzima a 100 ng/ μL em pH 6,5

a 60 °C. A análise dos resultados foi feita no software GraphPad Prism 5 ajustando os dados a uma regressão não-linear segundo o modelo cinético de Michaelis-Menten.

IV. Resultados e discussão

IV. 1 Celulase 5A de *Bacillus subtilis* – BsCel5A

Nesta parte do trabalho, investigamos as propriedades estruturais e de estabilidade da celulase 5A de *Bacillus subtilis* (BsCel5A), bem como os efeitos do domínio acessório na função do domínio catalítico parental e os modos de reconhecimento e interação ao substrato. A BsCel5A apresenta um peso molecular de aproximadamente 55 kDa, formada por um domínio catalítico com enovelamento barril (α/β)₈ associado ao domínio acessório (*Carbohydrate-binding module*, CBM) da família CBM3 na sua extremidade C-terminal, unidos por uma junção de 50 resíduos. Essa enzima pode ter aplicações em tecnologias e processos que envolvem a degradação de biomassa vegetal, pois atua no mais abundante polímero da parede celular de plantas, a celulose. A BsCel5A é uma endocelulase e portanto cliva ligações internas do tipo $\beta(1\rightarrow4)$ nas cadeias celulósicas. Sinergisticamente com celobiohidrolases e β -glucanases formam o complexo enzimático fundamental para redução da celulose em açúcares fermentescíveis (glicose).

Para o entendimento do papel de cada domínio na função e estabilidade da enzima três construções foram desenhadas: a proteína selvagem (BsCel5A), o domínio catalítico parental (CC_BsCel5A) e domínio acessório CBM3 (CBM_BsCel5A) (figura 9). A expressão da enzima ocorreu a 37 °C, com indução de 0,5 mM de IPTG por 4 horas. Para a purificação da proteína selvagem foram utilizados três passos cromatográficos incluindo afinidade ao níquel, troca catiônica e exclusão molecular. Para a CC_BsCel5A e o CBM_BsCel5A utilizaram-se duas etapas de cromatografia líquida, afinidade em coluna carregada de níquel, e exclusão molecular. As análises das purificações foram realizadas por eletroforese em gel de poliacrilamida sob condições desnaturantes (SDS-PAGE), indicando um elevado grau de pureza (> 95%) (figura 10).

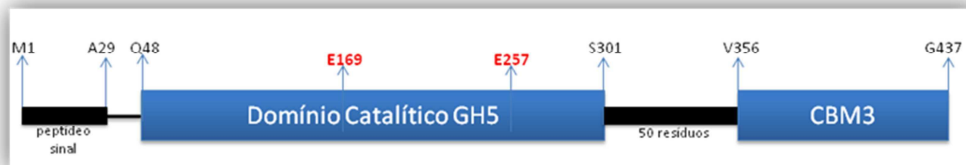


Figura 9: Representação da arquitetura dos domínios presentes na Celulase 5A de *Bacillus subtilis*.

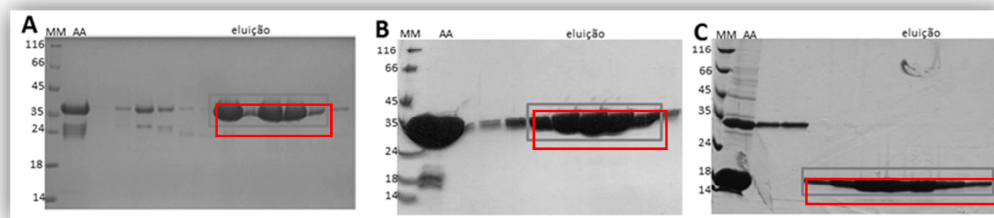


Figura 10: SDS-PAGE 13% da cromatografia de exclusão molecular da enzima BsCel5A e dos seus domínios catalítico e CBM3. **(A)** Cromatografia de exclusão molecular da proteína selvagem BsCel5A. **(B)** Cromatografia de exclusão molecular do domínio catalítico (CC_BsCel5A). **(C)** Cromatografia de exclusão molecular do domínio acessório (CBM_BsCel5A). MM marcador molecular; AA amostra aplicada na cromatografia de exclusão molecular e originada da concentração das amostras oriundas da afinidade ou da troca catiônica. Em destaque, no quadro vermelho, temos as amostras que foram utilizadas nos testes biofísicos e ensaios de cristalização.

A homogeneidade estrutural das amostras também foi analisada com intuito de inferir o estado oligomérico das construções e adequabilidade para estudos biofísicos e estruturais. Experimentos de espalhamento dinâmico de luz (DLS) indicaram que todas as proteínas heterologicamente expressas apresentaram baixa polidispersividade ($< 20\%$) e raios hidrodinâmicos de 3,3; 2,1 e 1,9 nm para a selvagem, o domínio catalítico e o CBM3, respectivamente, sugerindo que a BsCel5A e suas construções se comportam como monômeros em solução e estão em conformidade aos requerimentos para cristalização e outras análises espectroscópicas.

Estudos por espectroscopia de dicroísmo circular no UV distante mostram que a CBM_BsCel5A têm predominantemente elementos estruturais tipo β , enquanto que o domínio catalítico tem uma composição mista de hélices α e fitas β , compatível com seu enovelamento barril TIM. Enquanto que no espectro da proteína selvagem, pode-se observar uma composição equivalente das características dos domínios que a compõem (figura 11). Além de fornecer noções básicas do tipo de enovelamento estrutural, esses resultados corroboram os estudos por DLS, indicando que as proteínas recombinantes preservam suas estruturas nativas.

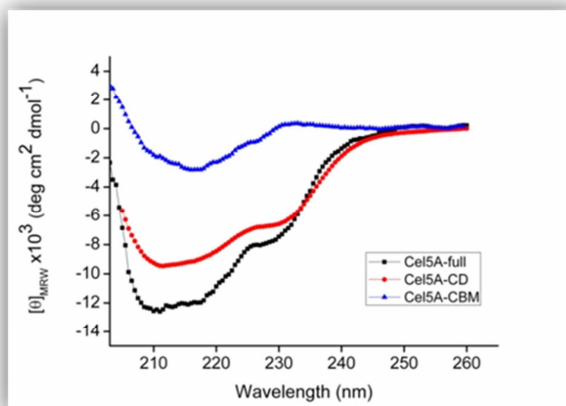


Figura 11: Espectros no UV distante das diferentes construções da celulase 5A de *Bacillus subtilis*. Observa-se um pico negativo em 216 nm da CBM_BsCel5A, característico de proteínas com predominância em estrutura β . No espectro da CC_BsCel5A tem-se os valores mínimos nos comprimentos de onda 222 e 208 nm, típicos de hélices α . Já na proteína selvagem observa-se uma composição das características presentes nas duas curvas anteriores.

Sabendo quais as composições das estruturas secundárias predominantes em cada domínio, diferentes comprimentos de onda foram utilizados para monitorar a estabilidade térmica. Para a enzima selvagem e para o domínio catalítico o comprimento de onda adequado foi 222 nm, devido à presença de elementos de estrutura secundária do tipo hélice α . Para a CBM_BsCel5A o comprimento de onda escolhido para os experimentos de desnaturação térmica foi o 230 nm. O sinal do espectro em 230 nm é característico do resíduo aromático triptofano e de pontes dissulfeto (Haas et al., 1998; Hider et al., 1988), e pode ser empregado para monitorar o desdobramento de proteínas ricas em elementos β , quando comportamentos anômalos são observados no comprimento de onda 216 nm.

Foram realizadas desnaturações térmicas em distintos pHs, e na presença de 10 mM EDTA e 10 mM MnCl_2 (figura 12). Os dados de desnaturação revelaram que tanto os domínios quanto a proteína selvagem apresentam apenas um estado de transição na desnaturação. Apesar da BsCel5A selvagem ser composta por dois domínios distintos, ligados por 50 resíduos a desnaturação ocorreu por uma via de transição única, sugerindo interações intermoleculares entre os domínios.

Outra característica importante é o ganho da estabilidade em torno de 20 °C, quando o domínio catalítico e a proteína selvagem estão em pH 5,5 com 10 mM MnCl_2 (figura 12). Outros íons divalentes como Ni^{2+} , Ca^{2+} e Mg^{2+} , mas não apresentaram resultados relevantes o que indica um efeito específico do íon manganês na estabilidade da BsCel5A. Pode-se observar

também que os íons não alteraram a estabilidade do CBM_BsCel5A, como observado na figura 12, mostrando que o efeito de termoestabilização é direcionado ao domínio catalítico.

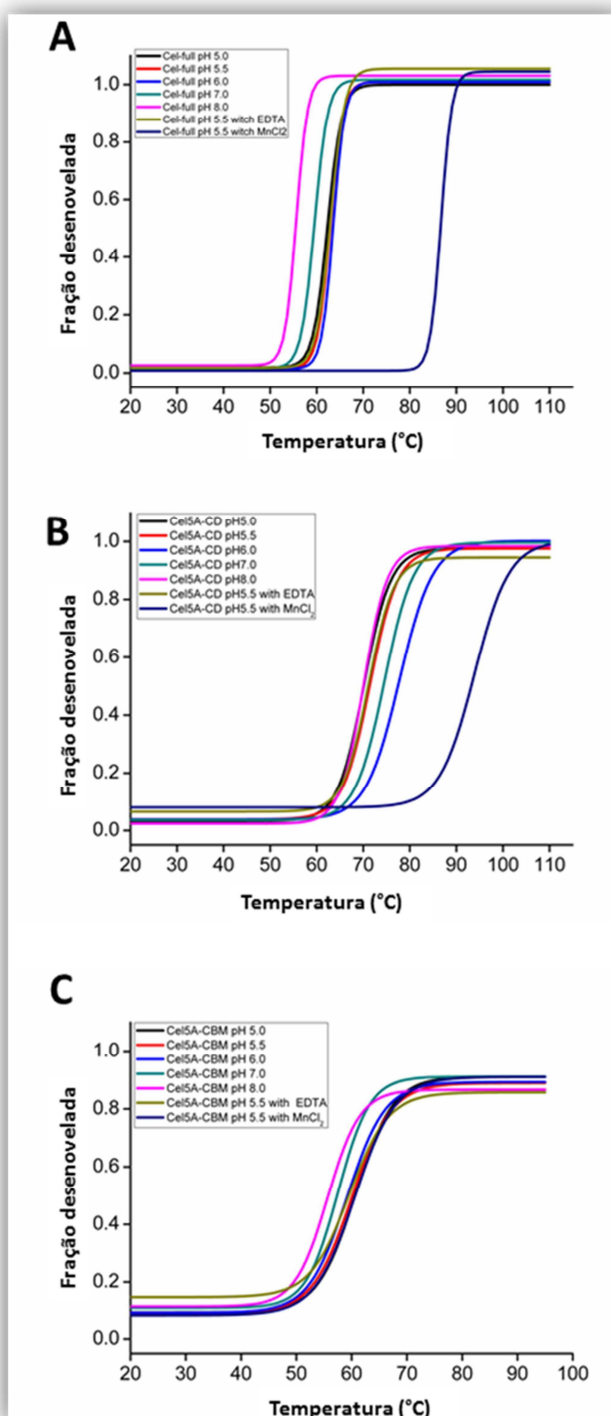


Figura 12: Desnaturalizações térmicas da proteína BsCel5A, e dos seus domínios CC_BsCel5A e CBM_BsCel5A, as condições de desnaturação variam em pH 5,0 a 8,0, e na presença de EDTA e $MnCl_2$. (A) Perfil da desnaturação da BsCel5A. (B) Desnaturação CC_BsCel5A. (C) Desnaturação CBM_BsCel5A.

Analisando as temperaturas de transição (do inglês, *melting temperature* - T_M) (tabela 3), pode-se concluir que a estabilidade é maior para o domínio catalítico, do que para a proteína selvagem e para o módulo de ligação a carboidrato. Por meio das informações obtidas pela estrutura do CBM_BsCel5A (RMN – colaboração com a Profa. Dra. Ana Zeri) foi revelado que a baixa estabilidade deste domínio é devido a ausência do canônico sítio de ligação ao íon cálcio (figura 13). O CBM_BsCel5A, faz parte da família 3 dos CBMs, com enovelamento sanduíche- β , comum entre os módulos de ligação a carboidratos e classicamente membros da família 3 possuem um cálcio estrutural (Yaniv et al., 2012) que não é conservado no CBM_BsCel5A, indicando a existência de uma subclasse de CBM3s que não ligam cálcio.

Tabela 3: Temperatura de transição (T_M) das desnaturações térmicas de BsCel5A, CC_BsCel5A e CBM_BsCel5A.

	BsCel5A	CC_BsCel5A	CBM_BsCel5A
pH	T_M (°C)	T_M (°C)	T_M (°C)
5,5	62,9	62,8	59,6
5,5 +EDTA	62,9	62,0	59,5
5,5 + $MnCl_2$	80,0	95,0*	60,7
6,0	63,5	64,9	59,4
7,0	59,6	62,4	57,5
8,0	55,0	58,9	55,6

* valor estimado

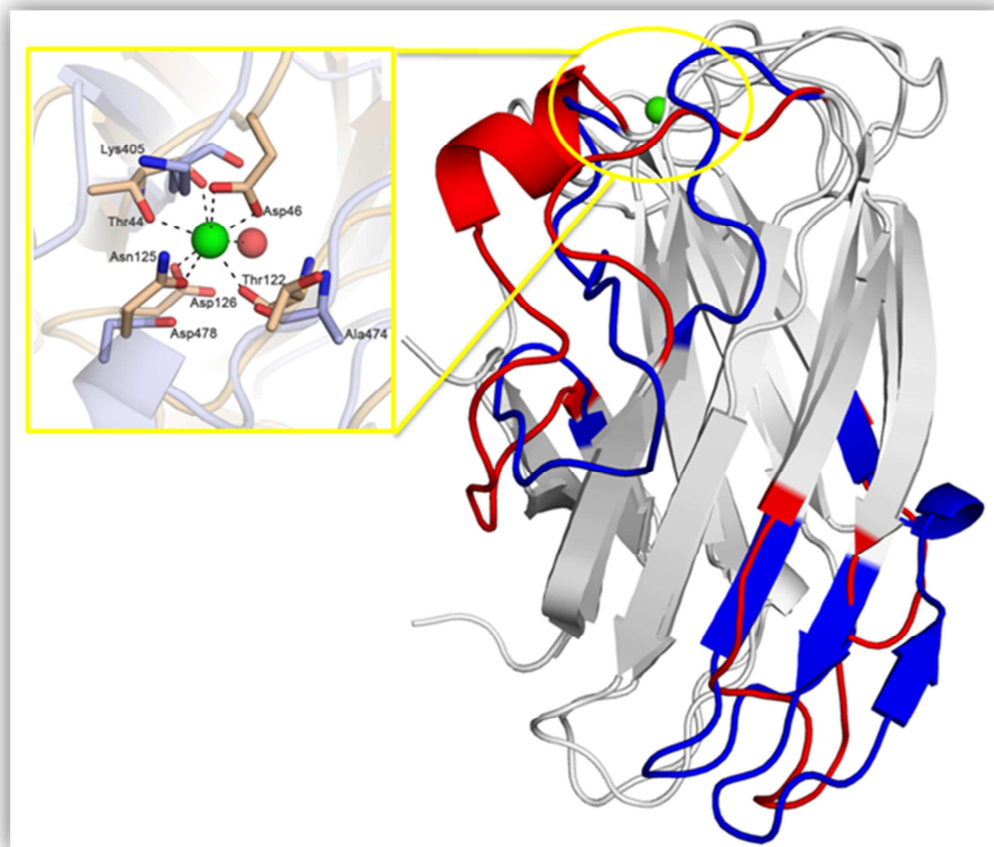


Figura 13: Estrutura do CBM_BsCel5A determinada por RMN, comparada com o CBM3 de *Clostridium thermocellum* (CBM_Ct), que mantém o sítio de ligação de cálcio. As diferenças entre ambos estão coloridas em azul para o CBM_BsCel5A e vermelho para CBM_Ct, o íon cálcio (esfera verde) está na estrutura do CBM_Ct. No zoom destaca-se a região de ligação do íon, com os resíduos de ligação do CBM_ct em bege, e os resíduos da mesma região do CBM_BsCel5A em azul.

Experimentos funcionais com a enzima BsCel5A e seu domínio catalítico CC_BsCel5A foram realizados, em colaboração com os pesquisadores Dr. Roberto Ruller e Dr. Fabio M. Squina (Laboratório Nacional de Ciência e Tecnologia do Bioetanol –CTBE), para determinação das condições de catálise (pH e temperatura) e a influência do íon sobre a atividade. A BsCel5A teve sua melhor atividade catalítica em pH 6,0 e na faixa de temperatura em torno de 60 °C (figura 14). Ainda a presença do íon Mn^{2+} não alterou o perfil funcional da enzima, sendo mantido a temperatura e o pH ótimo, o que pode representar uma vantagem, pois geralmente ganhos de estabilidade estão acoplados com alterações das condições físico-químicas ideais das enzimas.

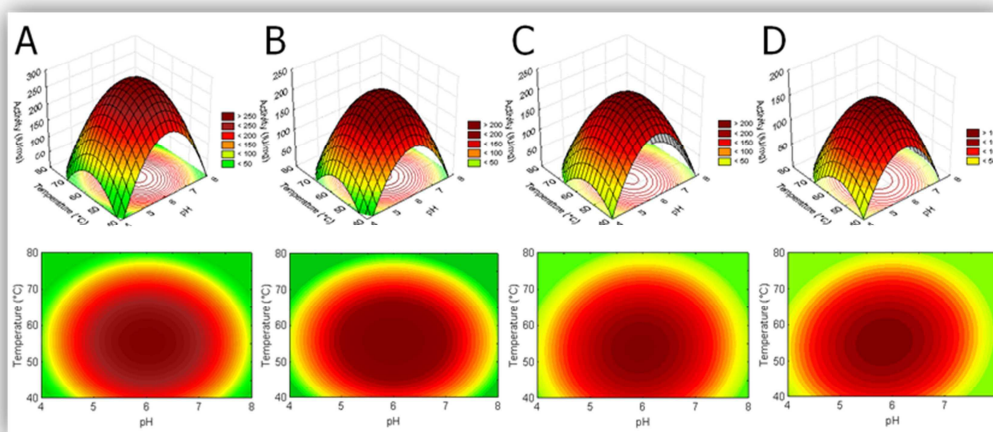


Figura 14: Efeito do pH e da temperatura sobre a atividade catalítica da BsCel5A e do Domínio Catalítico. (A) Proteína selvagem sem Mn^{2+} . (B) Proteína selvagem com 10 mM de Mn^{2+} . (C) Domínio catalítico sem Mn^{2+} . (D) Domínio catalítico com 10 mM de Mn^{2+} .

Experimentos de atividade sobre diferentes substratos (β -glucano, carboximetilcelulose e liquenana) revelaram que na presença do íon a atividade enzimática teve pequena variação (< 30%) em contrapartida ao grande ganho de estabilidade, que é interessante do ponto de vista de aplicações industriais (figura 15). Outros estudos mostraram o efeito do íon manganês na função de hidrolases glicosídicas como na xilanase de *Cellulosimicrobium* sp. HY-12 (Oh et al., 2008) e na celulase GH5 derivada de uma biblioteca metagenômica de solo (Voget et al., 2006) tiveram um aumento na atividade relativa na presença do íon. Tais estudos indicam o íon manganês como um potencial aditivo para melhoramento de desempenho de hidrolases glicosídicas seja por ganho de estabilidade ou aumento de atividade catalítica para as diversas aplicações biotecnológicas.

Parâmetros cinéticos mostram que a presença do íon diminuiu a velocidade de reação V_{max} e não teve efeitos significativos no K_m (tabela 4). Tal efeito foi inicialmente atribuído a um possível mecanismo competitivo no sítio ativo, como observado para a celobiohidrolase de *Humicola insolens*, que teve sua estrutura resolvida com o íon cádmio ligado aos resíduos catalíticos (Varrot et al., 2002). Porém, dados estruturais não indicam a presença do íon manganês na interface catalítica da BsCel5A, descartando assim esta hipótese. Desta forma, assumimos que a alteração negativa do V_{max} esteja relacionada à da rigidez global adquirida pela enzima na presença do íon, impedindo pequenas modificações conformacionais necessárias para condições ideais para a catálise.

A presença do módulo acessório CBM proporcionou uma melhor eficiência catalítica (K_{cat}/K_m) para a enzima, e uma maior afinidade da enzima pelo substrato. Por exemplos, em ensaios realizados com celulase cristalina, como o AVICEL®, só houve atividade catalítica na presença da enzima selvagem, confirmando a importância do CBM para reconhecimento do substrato.

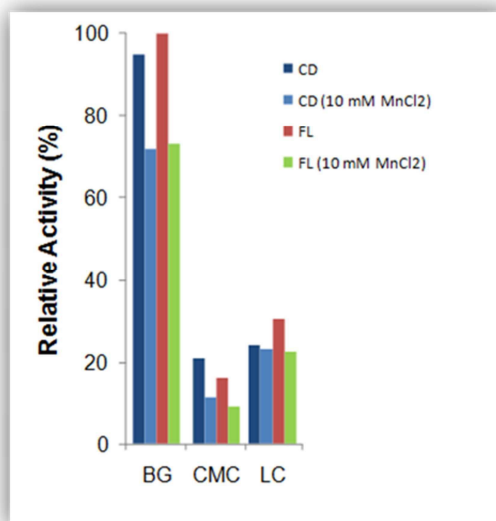


Figura 15: Atividade relativa da BsCel5A e do domínio catalítico em diferentes substratos. FL proteína selvagem, e CD core catalítico. Os substratos utilizados foram β-glucano (BG), carboximetilcelulose (CMC) e liquenana (LG).

Tabela 4: Parâmetros cinéticos da celulase 5A de *Bacillus subtilis*.

Parâmetros	BsCel5A		CC_BsCel5A	
	0 mM Mn ²⁺	10 mM Mn ²⁺	0 mM Mn ²⁺	10 mM Mn ²⁺
V _{max} (IU/mg)	46,2±3,1	35,7±1,8	143,7±6,6	92,3±5,9
K _m (mg/mL)	0,32±0,2	0,5±0,2	1,4±0,2	1,6±0,3
K _{cat} (S ⁻¹)	40,4	31,2	78,7	50,5
K _{cat} /K _m	127,8	58,8	56,2	31,6

Para elucidar os determinantes estruturais do mecanismo de termoestabilização pelo íon manganês no domínio catalítico foi determinada a estrutura da CC_BsCel5A no seu estado nativo e complexada com o íon manganês. O domínio catalítico se apresentou muito solúvel e alta concentração da amostra foi necessária para atingir a saturação nos experimentos de

cristalização (40 a 75 mg/ml). Cristais com distintas morfologias foram obtidos (figura 16) após diversas etapas de refinamento das condições iniciais de cristalização.

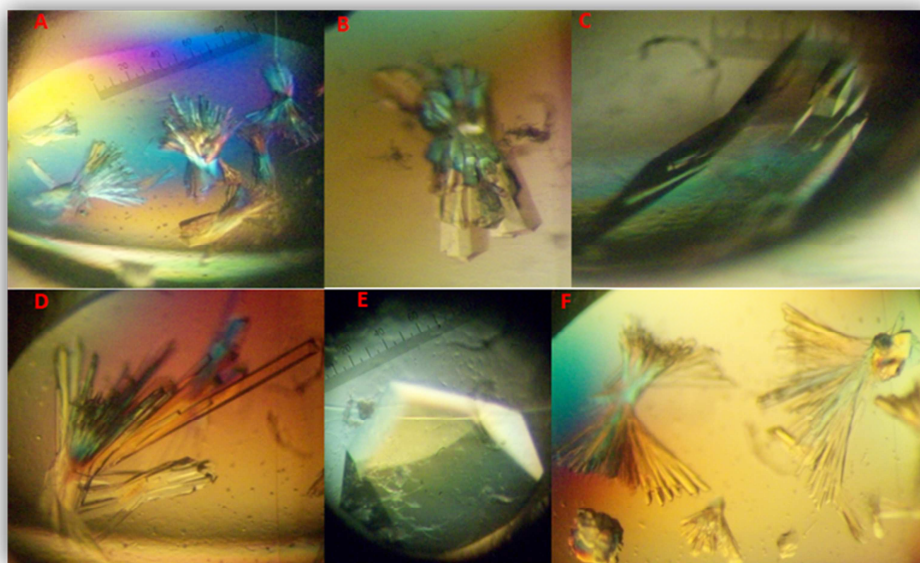


Figura 16: Cristais do domínio catalítico da celulase 5A de *Bacillus subtilis*. As letras correspondem a condições de cristalização presentes na tabela 5.

Tabela 5: Condições de cristalização do domínio catalítico da celulase 5A de *Bacillus subtilis*.

Condição	Tampão	Precipitante	Sal
A	100 mM NaAc/HCl pH 4,6	8% PEG 4000	
B	100 mM SPG* pH 6,0	25% PEG 1500	
C	100 mM Bis-Tris-propano pH 6,5	20% PEG 3350	200 mM nitrato de sódio
D	100 mM Tris/HCl pH 7,0	10% PEG 8000	200 mM cloreto de magnésio
E	200 mM Formaito de potássio pH 7,3	20% PEG 3350	
F	100 Tris/HCl pH 8,5	20% etanol	

* ácido succínico, di-hidrogeno fosfato de sódio mono-hidratado, glicina.

Os cristais obtidos foram utilizados em experimentos de difração de raios X na linha de luz MX2 do LNLS, Campinas - SP. Antes da coleta, os cristais passaram por uma solução crioprotetora com glicerol. Foram coletadas duas formas cristalinas (forma I e forma II), que

difrataram a 2,90 e 2,10 Å, respectivamente. Sendo que a forma I a proteína cristalizou na condição C (tabela 5), enquanto que a forma II a condição utilizada foi a B (tabela 5). Com o cristal II fizemos *soaking* com manganês (forma II*). A difração do cristal com manganês foi de 1,90 Å. O processamento dos dados de difração mostraram que os cristais referentes as formas II e II* pertencem ao grupo espacial $P2_12_12_1$ enquanto para a forma I o grupo espacial é o C2. Através do programa *BLAST* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) realizou-se busca por proteínas que apresentassem identidade com a CC_BsCel5A, e tivessem sua estrutura tridimensional resolvida e depositada no *Protein Data Bank* (PDB). A proteína selecionada para a substituição molecular foi a celulase 5A de *Alkalophilic Bacillus* (código PDB: 1LF1) que possui identidade sequencial de 70% com a CC_BsCel5A. Na tabela 6 estão sumarizadas as estatísticas das coletas, processamento de dados e refinamento da CC_BsCel5A.

A estrutura do domínio catalítico da celulase 5A tem um enovelamento barril $(\alpha/\beta)_8$ (figura 17A) que é clássico na família GH5. O sítio catalítico localiza-se no C-terminal da enzima, tendo os resíduos catalíticos ácidos localizados no loop entre a folha β_4 e a hélice α_5 , e na extremidade da fita β_7 , Glu169 e Glu257, respectivamente. O Glu169 é o doador de prótons, enquanto que o Glu257 é o nucleófilo que inicia a reação atacando a ligação glicosídica cindível (figura 17C). Outros resíduos contribuem para a formação do sítio catalítico incluindo Arg92, Trp207, His235, Thr256, e Trp291 (figura 17B). Os resíduos aromáticos Trp207 e Trp291 formam a entrada do sítio catalítico e os resíduos polares His235, Arg92 e Thr256 complementam o sítio ativo (figura 17D). A treonina 256 é considerada o terceiro elemento da catálise com participação efetiva da hidrólise ajudando a estabilizar o íon carbônio intermediário resultante da hidrólise a reação.

Tabela 6: Estatística da coleta de dados dos cristais do CC_BsCel5A, e do refinamento da estrutura.

Parâmetros	Forma I	Forma II	Forma II*
Dados da coleta			
Grupo Espacial	C2	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Parâmetros da célula unitária (Å, °)	a=155,36 b=81,08 c=114,27	a=50,24 b=110,76 c=121,38	a=49,53 b=110,62 c=122,26
Resolução	50,00-2,90 (3,00-2,90) *	50,00-2,10 (2,18-2,10)	50,00-1,97 (2,04-1,97)
Reflexões únicas	30407 (2607)	39069 (3393)	47366 (4163)
Multiplicidade	3,5 (2,7)	6,9 (6,2)	4,8 (4,3)
Completeza (%)	97,1 (83,7)	97,2 (85,9)	97,6 (87,8)
R _{merge}	8,3 (43,6)	11,6 (31,9)	8,4 (37,3)
I/ σ (I)	11,3 (2,0)	13,5 (5,0)	14,7 (3,8)
Refinamento			
R _{work} /R _{free}	23,0/29,6	17,1/22,1	16,4/20,6
Moléculas na unidade assimétrica	4	2	2
Moléculas de água	30	330	452
Média do fator B	54,5	32,7	29,1
RMSD ⁺ do comprimento das ligações (Å)	0,006	0,023	0,025
RMSD dos ângulos de ligação (°)	1,261	1,903	1,938
Ramachandran			
Região favorável (%)	90,63	96,44	97,09
Regiões permitidas	8,02	3,39	2,91
Outras regiões	1,35	0,17	0,00
PDB	3PZV	3PZU	3PZT

* Os valores em parênteses são relativos da última camada.

⁺ Root Mean Square Deviation

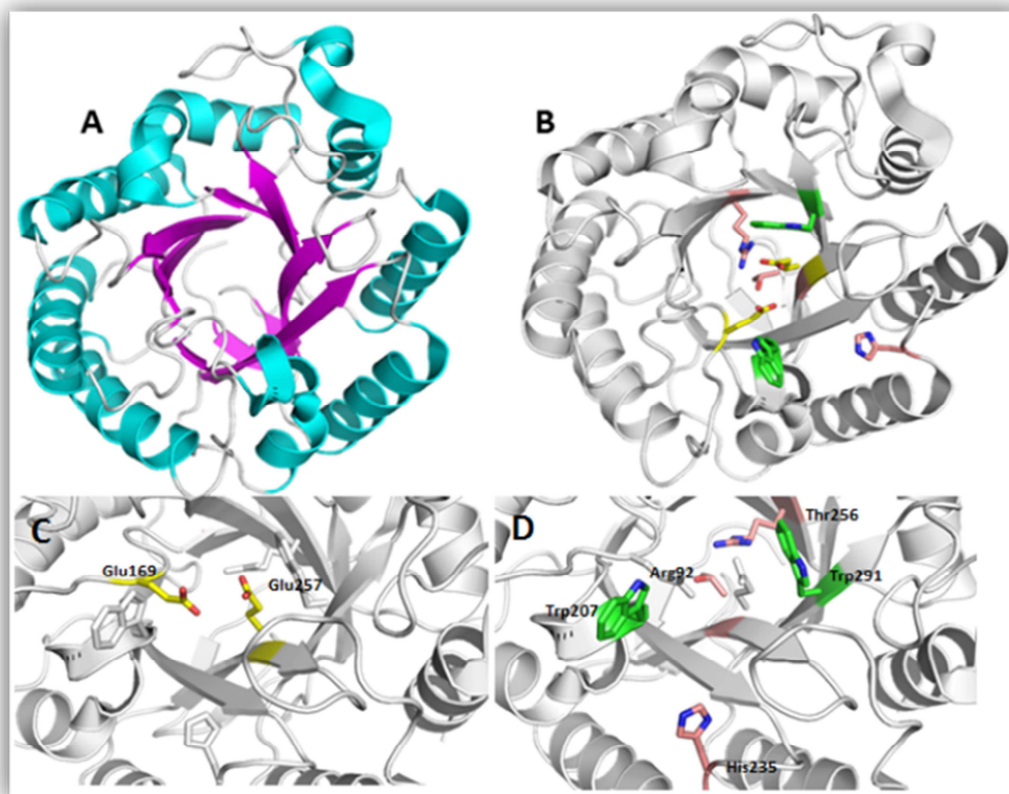


Figura 17: Estrutura do domínio catalítico da Celulase 5A de *Bacillus subtilis*. (A) Representação da estrutura proteica, mostrando o enovelamento barril (α/β)₈. (B) Detalhes do sítio ativo da BsCel5A, em amarelos são os resíduos catalíticos essenciais, em verde tem os resíduos responsáveis pela interação com o substrato, e em rosa os resíduos que ajudam a manter a ligação enzima/substrato (C) Destaque para os resíduos primordiais para a ação catalítica da enzima. (D) Representação dos resíduos acessórios responsáveis pela interação enzima-substrato.

Para compreender o ganho de estabilidade proporcionado pelo íon manganês, a estrutura da CC_BsCel5A com o Mn^{2+} foi elucidada. Para que o íon fosse incorporado na estrutura, fez-se *soaking* com o mesmo. Com a resolução da estrutura, observou-se que o íon localiza-se em um sítio carregado negativamente, que está posicionado na face oposta do sítio catalítico, próximo ao motivo β -hairpin no N-terminal (figura 18A). O íon é hexa-coordenado por duas moléculas de água, e pelas cadeias laterais dos resíduos Asp195, Asp197 e Asn198, e pelo oxigênio da carbonila do resíduo Gly157 (figura 18C). O sítio de ligação do manganês é estrategicamente localizado próximo à região N-terminal e do β -hairpin (figura 18B) permitindo interações intermediadas por moléculas de água com porções proteicas importantes para estabilidade estrutural, aumentando a rigidez, e consequentemente sua estabilidade. Comparando as formas II e II*, sem e com o manganês respectivamente, pode-se observar que o íon substitui uma molécula de água no sítio carregado negativamente. Entretanto, a localização do íon Mn^{2+} não afeta a arquitetura da fenda catalítica, não havendo modificações

no estado conformacional dos resíduos que são essenciais para formação do centro catalítico. O íon tem uma distância média de 30 Å do sítio catalítico.

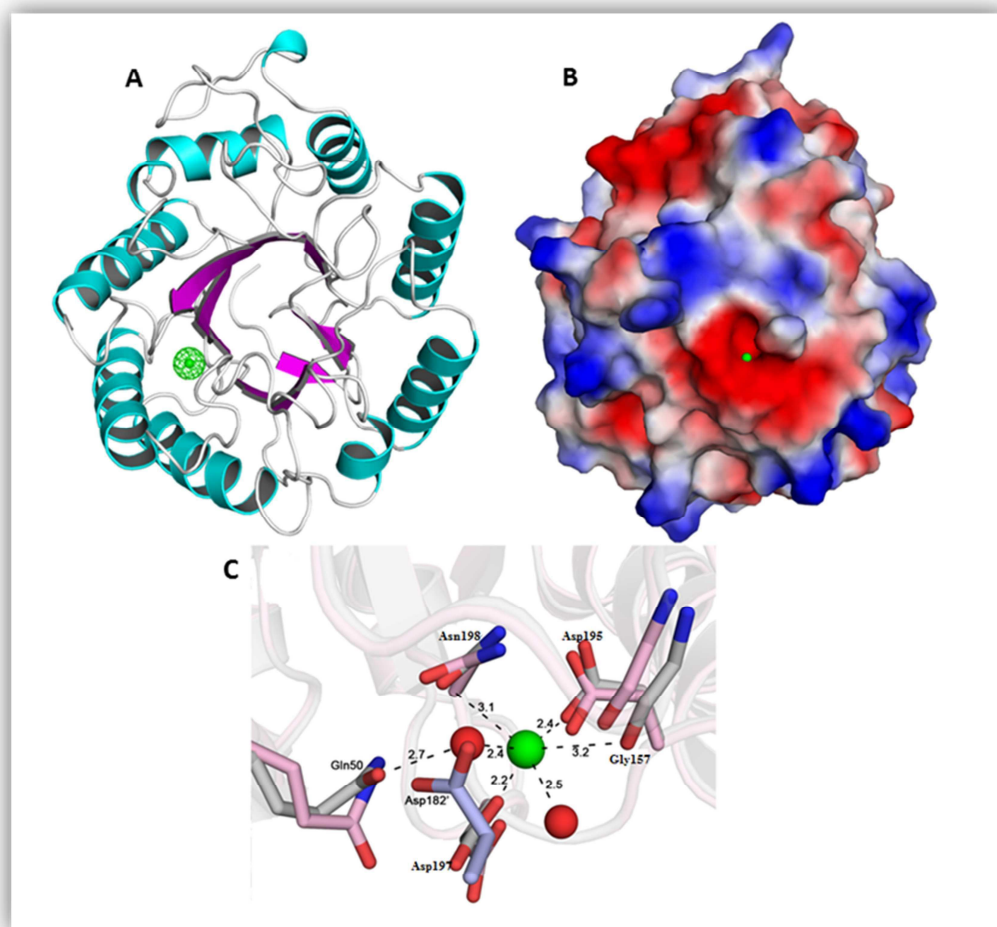


Figura 18: Estrutura do domínio catalítico da Celulase 5A de *Bacillus subtilis* em complexo com o íon manganês. (A) Representação da estrutura proteica com destaque para a localização do íon manganês (esfera verde), próximo ao N-terminal da proteína. (B) Representação da superfície eletrostática do domínio colorida por carga, vermelho (-), azul (+), indicando a presença do manganês no bolsão carregado negativamente. (C) Interações dos resíduos da proteína em cinza, com moléculas de H₂O (esferas vermelhas) e o íon manganês (esfera verde), nas cores rosa e cinza tem o deslocamento dos resíduos na presença do íon. O resíduo em azul é o aspartato que na ausência do íon fica deslocado.

Para obter a estrutura da enzima selvagem e observar as interações entre os domínios que a compõem, exaustivas tentativas de cristalização foram realizadas, mas sem sucesso, provavelmente devido a flexibilidade conferida pela junção entre os domínios que gera condições entropicamente desfavoráveis para formar os contatos cristalinos. Então, a técnica SAXS foi empregada para fornecer dados estruturais da proteína selvagem, bem como de seus domínios em solução. Os valores obtidos para o raio de giro (R_g) foram 1,75; 1,63 e 3,04 nm

(figura 19), e as dimensões máximas foram 4,65; 4,8 e 9,4 nm para o domínio catalítico, CBM3 e proteína selvagem, respectivamente, comprovando que as proteínas são monoméricas em solução.

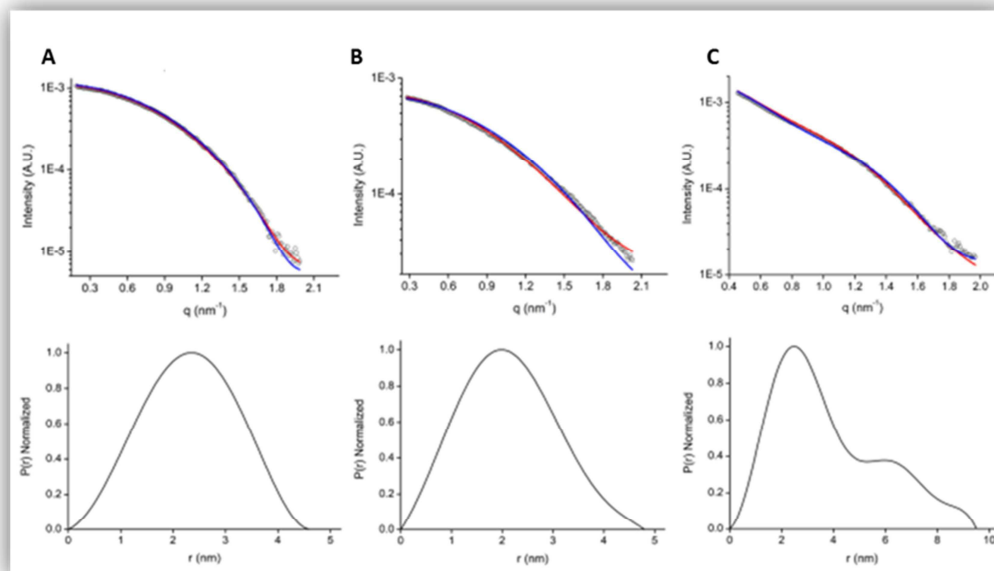


Figura 19: Curvas de espalhamento e distribuição de distâncias ($P(r)$) de SAXS dos domínios da Celulase 5A de *Bacillus subtilis* e da enzima selvagem. **A** domínio catalítico (CC_BsCel5A) e **B** módulo de ligação a carboidrato (CBM_BsCel5A), **C** Celulase 5A (BsCel5A). Os gráficos superiores mostram as curvas de espalhamento experimental (círculos abertos), calculado pelo envelope do SAXS (linha vermelha) e da estrutura de alta resolução (linha azul). Nos gráficos inferiores tem a função distribuição de distâncias calculada com os dados experimentais.

Com os dados obtidos de SAXS foi possível prever a organização do domínio catalítico e do módulo de ligação a carboidrato na enzima selvagem. Os envelopes moleculares construídos para o domínio catalítico (figura 20A), e o CBM_BsCel5A (figura 20B), foram sobrepostos com as coordenadas atômicas dos domínios. A sobreposição demonstrou um excelente ajuste permitindo no caso da CC_BsCel5A observar a cavidade catalítica. O modelo da proteína selvagem baixa resolução (figura 20C) indicou resultados similares ao obtido para a celulase Cel5G da bactéria psicrotrófica (organismo que cresce em baixas temperaturas) *Pseudoalteromonas haloplanktis* A23, que possui um $D_{\max}$ de 13 ± 5 nm e uma função $P(r)$ semelhante à da enzima BsCel5A (Sonan et al., 2007).

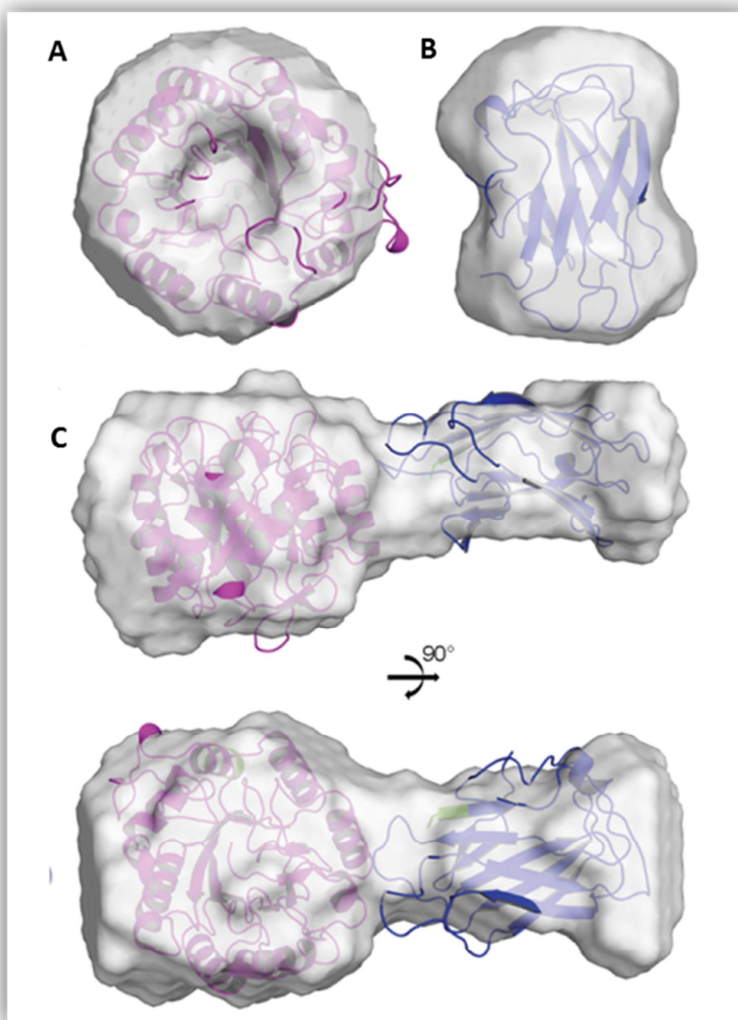


Figura 20: Envelope molecular da Celulase 5A de *Bacillus subtilis* obtidos por SAXS. **(A)** envelope de SAXS do domínio catalítico com a sobreposição da estrutura cristalográfica. **(B)** envelope de SAXS do CBM3, com a estrutura tridimensional. **(C)** Envelope da Celulase 5A, com os domínios modelados.

O envelope molecular da celulase 5A permitiu evidenciar uma superfície plana e contínua, formada pela interface do sítio ativo da fração catalítica alinhada com a face do CBM3 que reconhece a celulose, e ocorre a interação com o carboidrato. Essa superfície é essencial para que a celulose possa se acomodar e ser clivada. A organização estrutural entre o domínio catalítico e o módulo de ligação a carboidrato já foi observado em outras celulases, como a endoglucanase de *Clostridium cellulolyticum*, a Cel9G e a celulase E4 de *Thermobifida fusca* (Mandelman et al., 2003; Sakon et al., 1997). Esta arquitetura molecular corrobora os dados funcionais sugerindo que o CBM ajuda no reconhecimento do substrato, o que reflete no aumento da eficiência catalítica da enzima.

IV. 2 Celulase 5 de *Xanthomonas axonopodis* pv. *citri* – XacCel5

Xanthomonas axonopodis pv. *citri* (*X.citri*) é uma fitobactéria gram-negativa altamente adaptada ao ambiente de plantas, sendo raramente encontrada livre no solo (Swings e Civerolo, 1993). Essa bactéria é o agente causal do cancro cítrico, uma doença que resulta em enormes perdas econômicas na cultura de citros. Embora os sintomas causados por essa bactéria não estejam tipicamente associados à degradação massiva da parede celular da planta, *X.citri* apresenta vários genes que codificam enzimas com esse tipo de função. Quando cultivada em meio de cultura que simula o espaço intercelular na planta, a bactéria expressa vários desses genes (Astua-Monge et al., 2005). Interessantemente, a mutação no gene *xpsD*, que codifica um componente do sistema de secreção do tipo II, afeta a degradação de celulose pela bactéria e também sua virulência (Baptista et al., 2010), indicando uma possível relação entre a expressão de celulasas e hemicelulasas e a infecção por *X.citri*. Desta forma, a prospecção de novas enzimas glicosídicas desses microorganismos pode contribuir para a descoberta ou desenvolvimento de enzimas mais eficazes e especializadas a degradação de biomassa vegetal.

Uma dessas enzimas é a celulase GH5 (XacCel5) que tem homologia com o domínio catalítico da BsCel5A e ainda não foi caracterizada funcional e estruturalmente. A proteína recombinante tem 36 kDa e foi subclonada no vetor de expressão pET28a. Sua expressão ocorreu em BL21(DE3), com indução de 0,5 mM de IPTG, a 20 °C por 18 horas. No processo de extração a proteína apresentou-se na fração solúvel e dois passos cromatográficos foram utilizados, afinidade e exclusão molecular. O grau de pureza da amostra foi monitorado por SDS-PAGE 13% (figura 21), que foi superior a 95%.

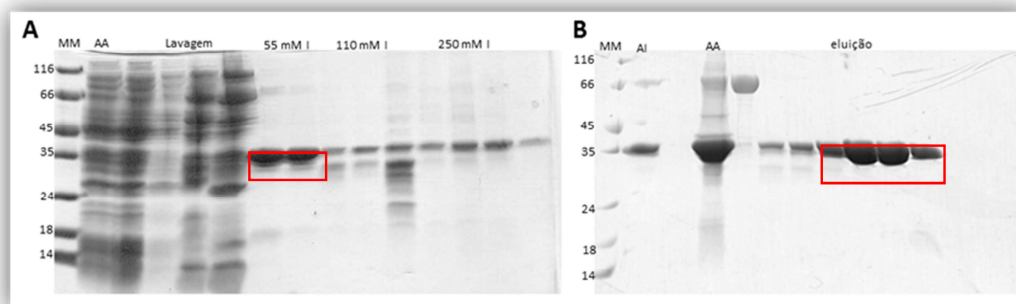


Figura 21: SDS-PAGE 13% das etapas de purificação da celulase 5 de *Xanthomonas axonopodis* pv. *citri*. Em destaque as amostras utilizadas. (A) cromatografia de afinidade a níquel imobilizado. (B) cromatografia de exclusão molecular. MM marcador molecular; AA amostra aplicada do extrato bruto de 1 litro de expressão; 55 mM I, 110 mM I e 250 mM I são as eluições com imidazol; AI amostra inicial, frações unidas da afinidade antes da concentração, AA amostra concentrada aplicada na filtração em gel.

Após a cromatografia de exclusão molecular, a proteína foi concentrada até 7,5 mg/mL e submetida aos ensaios de cristalização pela técnica de difusão de vapor em gota sentada. Os ensaios foram realizados com o sistema robotizado *Honeybee* e a análise dos experimentos remotamente pelo sistema Rock Imager (Formulatrix). O volume da solução proteica e da solução do reservatório foi 0,5 μ L. Algumas condições dos ensaios iniciais de cristalização resultaram em micro ou pequenos cristais de proteína (figura 22), porém ainda não adequados à difração de raios X.

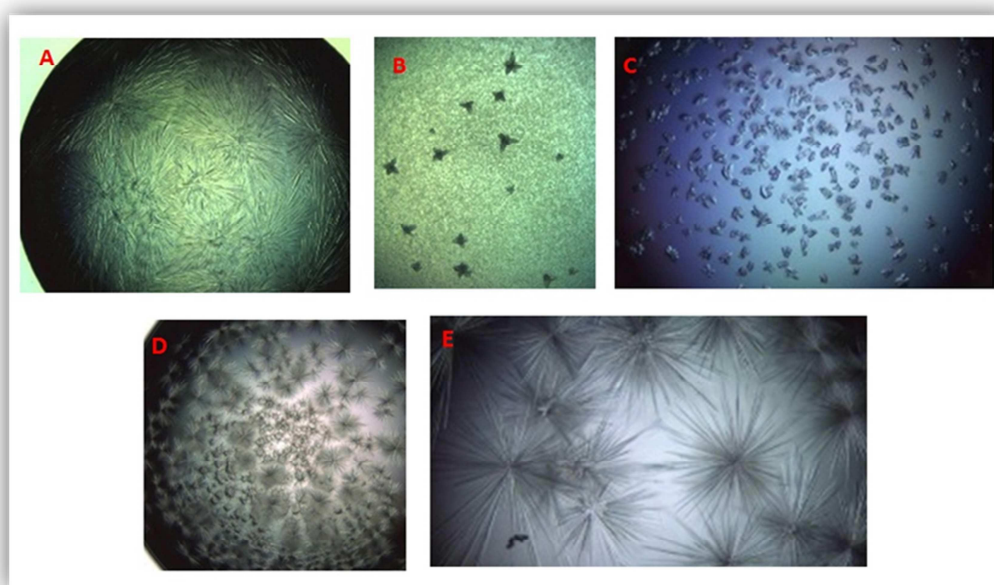


Figura 22: Cristais da XacCel5 obtidos durante os ensaios iniciais de cristalização com a proteína a 7,5 mg/mL. Na tabela 7 estão as condições de cristalização referentes às imagens.

Tabela 7: Condições iniciais de cristalização da enzima XacCel5.

Condição	Tampão	Precipitante	Sal
A	100 mM HEPES pH 7,5	30% PEG 400	200 mM cloreto de magnésio
B	100 mM TRIS-HCl pH 8,5	20% PEG-1000	
C	100 mM Na3 citrato/HCl pH 5,6	30% PEG 4000	200 mM acetato de amônia
D	100 mM Bis-Tris pH 5,5	25% PEG-3350	200 mM sulfato de amônia
E	100 mM HEPES pH 7,5	45% MPD	400 mM acetato de amônia

Para obter cristais adequados para os experimentos de difração inúmeros refinamentos foram realizados, variando a concentração de precipitantes, gradiente salino, solução tamponante e temperatura (figura 23). Tais cristais apresentam morfologia e dimensões adequadas para a atual configuração da linha de luz MX2 disponível no LNLS.

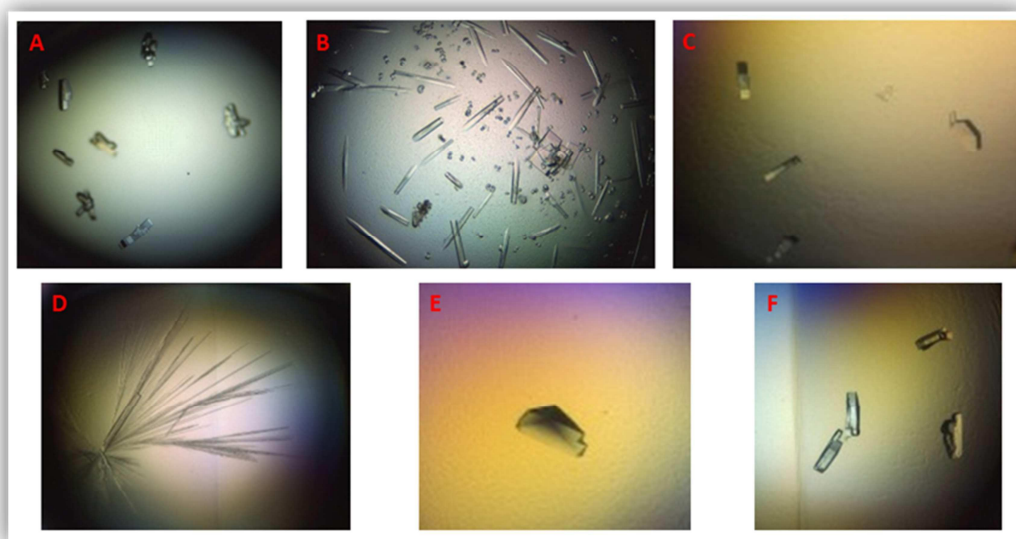
**Figura 23:** Cristais de XacCel5 adequados para os experimentos de difração de raio X. As condições de cristalização estão listadas na tabela 8.

Tabela 8: Condições do refinamento da cristalização da celulase 5 de *Xanthomonas axonopodis* pv. *citri*.

Condição	Tampão	Precipitante	Sal
A	100 mM citrato de sódio pH 5,6	30% PEG 4000	200 mM acetato de amônio.
B	100 mM TRIS-HCl pH 8,5	27% PEG 4000 5% glicerol	200 mM sulfato de lítio
C	100 mM citrato de sódio pH 5,6	30% PEG 4000	200 mM acetato de amônia
D	100 mM citrato de Sódio pH 5,6	29% PEG-4000 10% glicerol	200 mM sulfato de amônia
E	100 mM cacodilato de Sódio pH 6,5	29% PEG-8000 5% glicerol	200 mM acetato de sódio
F	100 mM citrato de Sódio pH 5,6	29% PEG-4000	200 mM acetato de sódio

Concomitante aos ensaios de cristalização foi realizada a caracterização biofísica da proteína por espectroscopia de dicroísmo circular. Primeiramente foram coletados espectros na presença de diferentes íons que na literatura demonstraram afetar a estabilidade ou função de celulasas GH5 (figura 24) e como é possível observar, não houve alterações no espectro do UV distante da XacCel5 na presença destes íons.

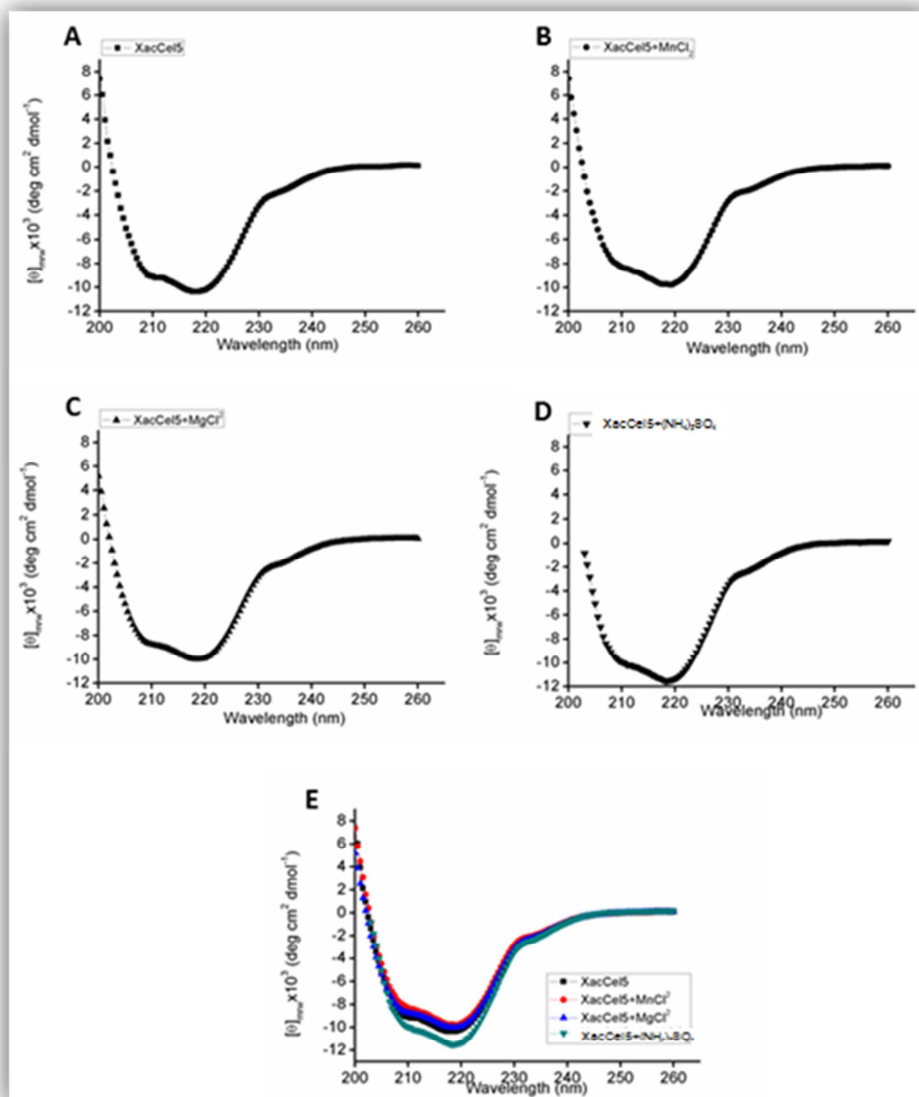


Figura 24: Espectro de dicroísmo circular no UV distante da celulase 5 de *Xanthomonas axonopodis* pv. *citri* - XacCel5. (A) XacCel5 em pH 7,4. (B) pH 7,4 e 10 mM MnCl₂. (C) pH 7,4 e 10 mM MgCl₂. (D) pH 7,4 e 10 mM (NH₄)₂SO₄. (E) Sobreposição dos espectros no UV distante da XacCel5.

Para avaliar a estabilidade térmica da proteína perante estes íons foi realizada a desnaturação (figura 25). O íon magnésio foi escolhido para os testes de desnaturação já que o mesmo está descrito em estruturas de endocelulases da família GH5, como na celulase 5 de *Bacillus agaradhaerens* e na de *Erwinia chrysanthemi* (Varrot et al., 2000; Chapon et al., 2001). Enquanto que o íon manganês teve um papel termoestabilizante na Cel5A de *Bacillus subtilis* (bSantos et al., 2012). A desnaturação da XacCel5 revelou uma T_M média de 47 °C na ausência de íons, e tal temperatura se manteve para $MgCl_2$ e $(NH_4)SO_2$ (tabela 9). Interessamente, o íon Mn^{2+} teve um efeito hiper-estabilizante na XacCel5 (figura 26B), assim como observado na celulase 5 de *Bacillus subtilis* (bSantos et al., 2012).

Tabela 9: Temperatura de transição (T_M) das desnaturações térmicas da XacCel5.

Condição	T_M
pH 7,4	47,55
pH 7,4 + $MgCl_2$	49,73
pH 7,4 + $MnCl_2$	80,0*
pH 7,4 + $(NH_4)_2SO_4$	49,62

*Valor estimado

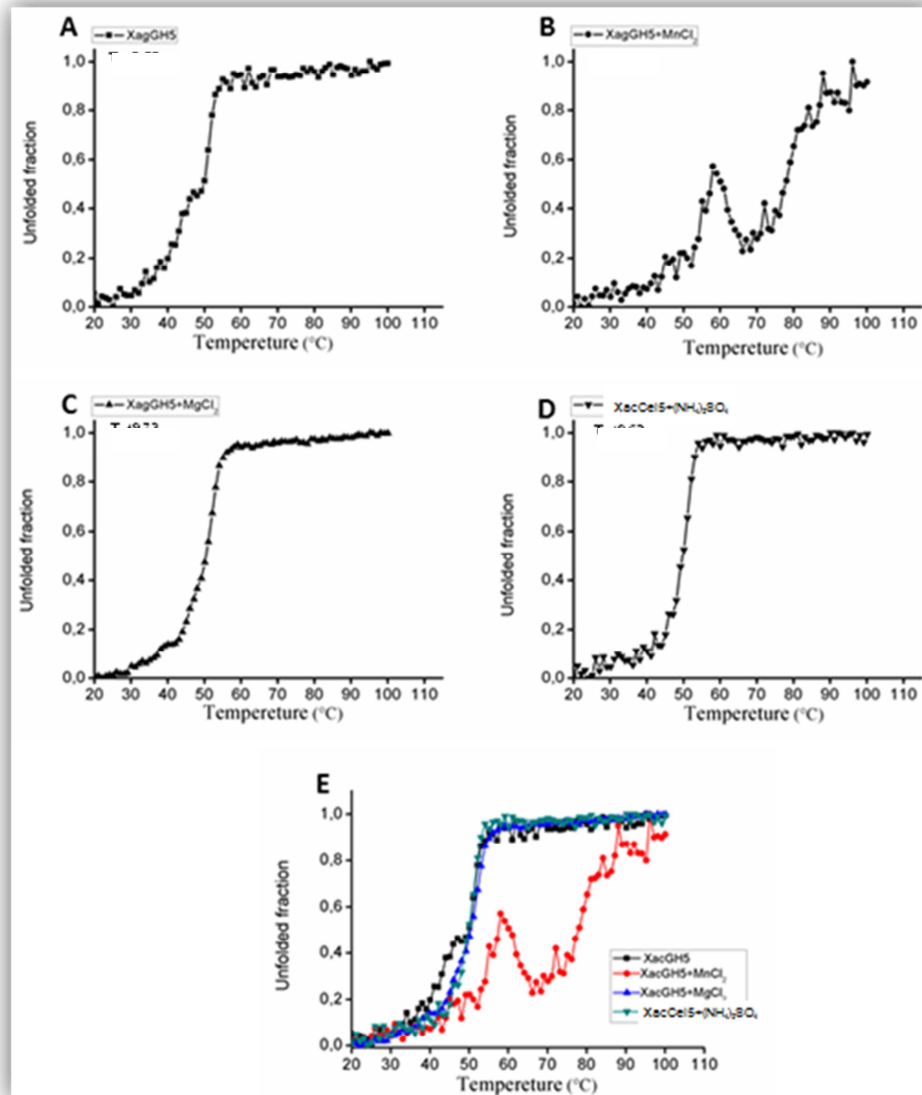


Figura 25: Curvas de desnaturações térmicas da XacCel5 com distintos íons . (A) XacCel5 em pH 7,4. (B) pH 7,4 e 10 mM MnCl_2 . (C) pH 7,4 e 10 mM MgCl_2 . (D) pH 7,4 e 10 mM $(\text{NH}_4)_2\text{SO}_4$. (E) comparação das desnaturações.

O espectro da proteína após passar por desnaturação foi observado (figura 26), mostrando que o processo é irreversível na ausência de íons ou presença de sulfato ou magnésio. Porém, o espectro da proteína com manganês manteve-se com os picos característicos de hélices α e folhas β (pico negativo em 208 e 222 para hélices α , e pico 216 para folhas β), corroborando o efeito termo-estabilizante deste íon na XacCel5 (figura 27).

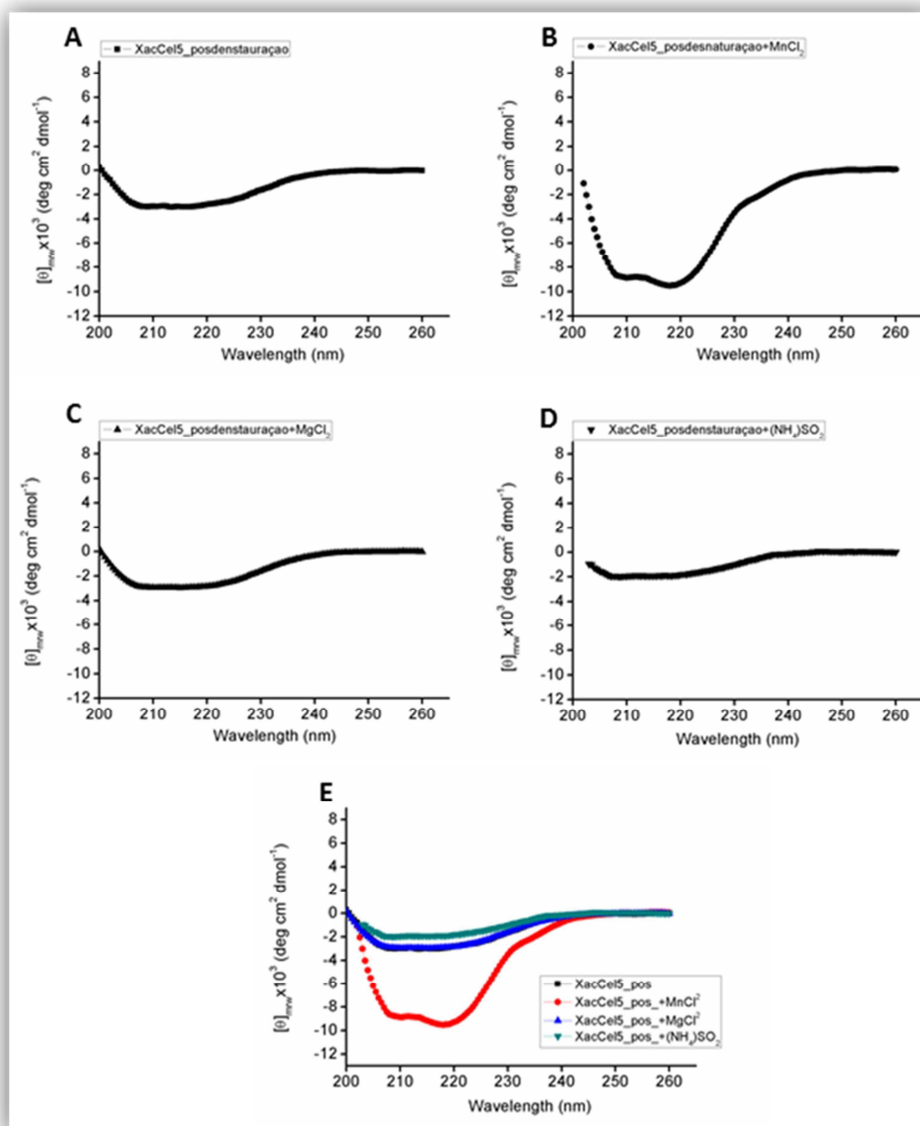


Figura 26: Espectro de dicroísmo circular na faixa do UV distante da XacCel5 após ensaios de desnaturação térmica. (A) XacCel5 em pH 7,4. (B) pH 7,4 e 10 mM MnCl_2 . (C) pH 7,4 e 10 mM MgCl_2 . (D) pH 7,4 e 10 mM $(\text{NH}_4)_2\text{SO}_4$. (E) Sobreposição dos espectros da XacCel5 em diferentes condições.

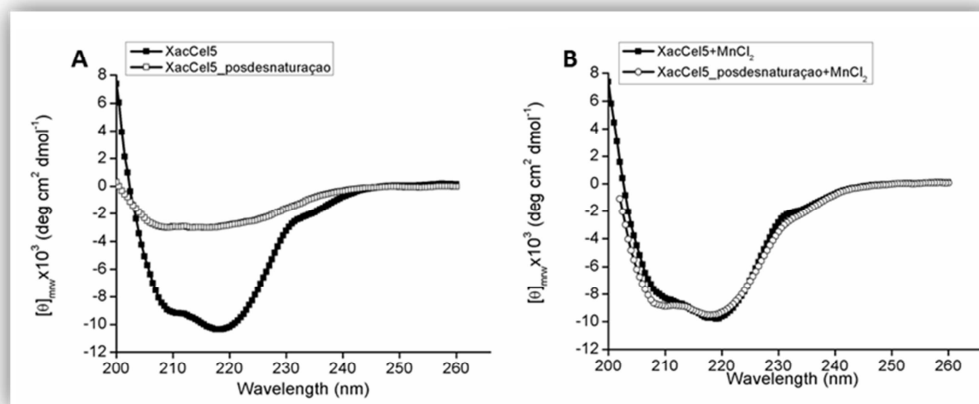


Figura 27: Espectros de dicroísmo circular na faixa do UV distante da XacCel5 antes e pós desnaturação térmica e na presença e ausência do íon manganês. **(A)** XacCel5 pH 7,4. **(B)** XacCel5 pH 7,4 e 10 MnCl_2 .

A capacidade do íon Mn^{2+} de estabilizar celulasas pertencentes à família GH5 não fica restrita a esse caso. Como mencionado anteriormente, a celulase 5A de *Bacillus subtilis* (BsCel5A) também apresentou ganho de estabilidade na presença do íon, com aumento da T_M em torno de 20 °C. A estrutura tridimensional do domínio catalítico da BsCel5A foi revolvida com o íon Mn, revelando a localização do mesmo na face oposta ao sítio catalítico, sendo hexa-coordenado pelas cadeias laterais dos resíduos Asp195, Asp197, Asn198, pelo oxigênio da carbonila da Gly157 e por duas moléculas de água. A posição estratégica do íon permitiu interações que aumentassem a rigidez da BsCel5A, e consequentemente a sua estabilidade (Santos et al., 2012).

O alinhamento das sequências da XacCel5 e BsCel5A (figura 28), apresentou uma identidade de apenas 20%, e os resíduos que coordenam o Mn^{+2} apresentam algumas diferenças, sendo apenas o Asn198 conservado em ambas as enzimas. Enquanto que os Asp195 e Asp197 são substituídos por Gly e Lys, respectivamente, e Gly157 da BsCel5 é substituída por outra Lys na XacCel5.

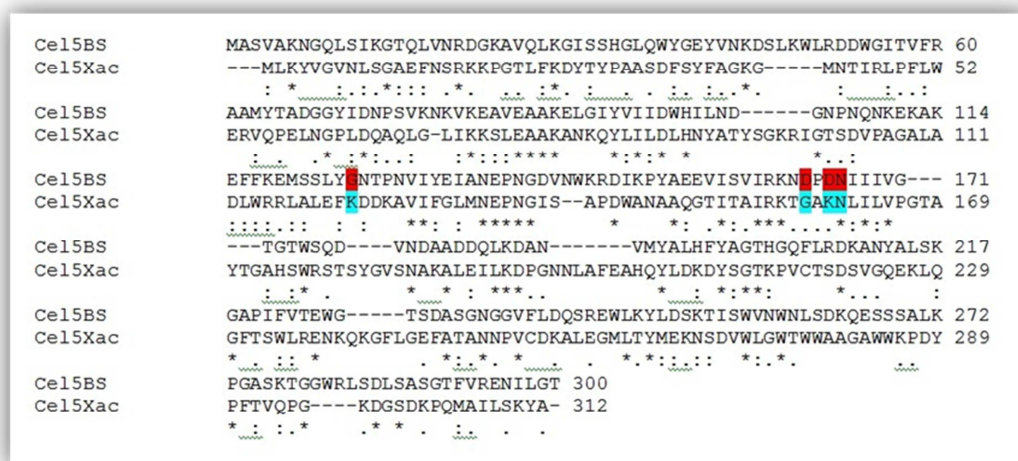


Figura 28: Alinhamento das sequências das celulases de *Bacillus subtilis* (BsCel5A) e *Xanthomonas axonopodis* pv. *citri* (XacCel5). Em destaque os resíduos responsáveis pela coordenação do íon Mn²⁺ na BsCel5, em azul têm os resíduos equivalentes na XacCel5. O alinhamento foi realizado com o programa clustal 2.1.

As próximas etapas da caracterização da XacCel5 incluem a resolução da estrutura tridimensional da enzima, selvagem e ligada ao íon Mn²⁺ para desvendar a base molecular do efeito do íon que indica ser distinto daquela para BsCel5A e conduzir estudos funcionais para avaliação do seu potencial biotecnológico em processos envolvendo a degradação de biomassa lignocelulósica.

IV. 3 Celulase 5 de biblioteca de metagenômica do solo de cana-de-açúcar – CelE1

A abordagem metagenômica é uma ferramenta poderosa para prospecção de novas moléculas com diferentes funções e características. Uma nova celulase GH5 foi descoberta a partir de uma biblioteca de metagenômica do solo de cana-de-açúcar pelo grupo Prof. Fabio M. Squina e em colaboração realizamos a caracterização estrutural e biofísica desta nova celulase GH5, aqui denominada CelE1.

A CelE1 é uma proteína recombinante com 33 kDa, que foi subclonada no vetor de expressão pET28a. A expressão da mesma ocorreu em Roseta-gami™(DE3)pLys-S, em meio LB contendo o antibiótico canamicina a 37 °C e indutor IPTG (0,5 mM). A enzima apresentou-se na fração solúvel e sua purificação foi realizada em dois passos cromatográficos: afinidade por íon imobilizado e exclusão molecular. A amostra purificada apresentou grau de pureza superior a 95% de acordo com análise por SDS-PAGE 13% (figura 29).

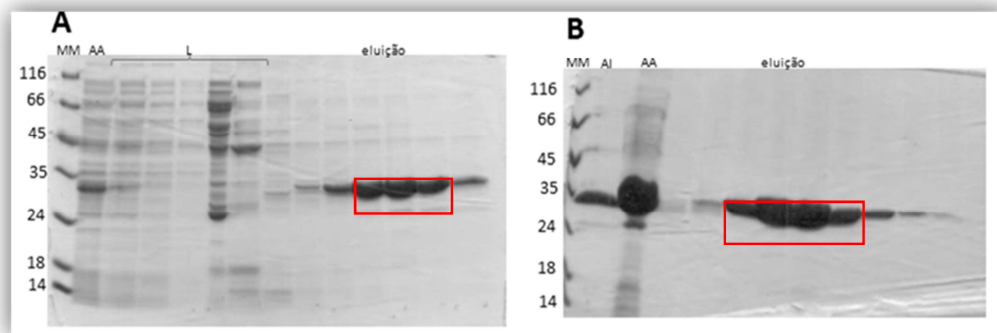


Figura 29: SDS-PAGE 13% das etapas de purificação da celulase 5 oriunda de uma biblioteca de metagenômica do solo de cana-de-açúcar. Em destaque as amostras selecionadas para os ensaios biofísicos. **(A)** Cromatografia de afinidade por níquel da CelE1. **(B)** Cromatografia de exclusão molecular da CelE1. MM marcador molecular; AA amostra aplicada na coluna de afinidade oriunda da lise do extrato de 1 litro de expressão; L lavagem; AI amostra inicial – amostras unidas da afinidade; AA amostra aplicada na filtração em gel.

O espectro de dicroísmo circular da enzima no UV distante indicou um perfil canônico para enzimas com enovelamento barril TIM (figura 30), que sugere o estado nativo da enzima recombinante.

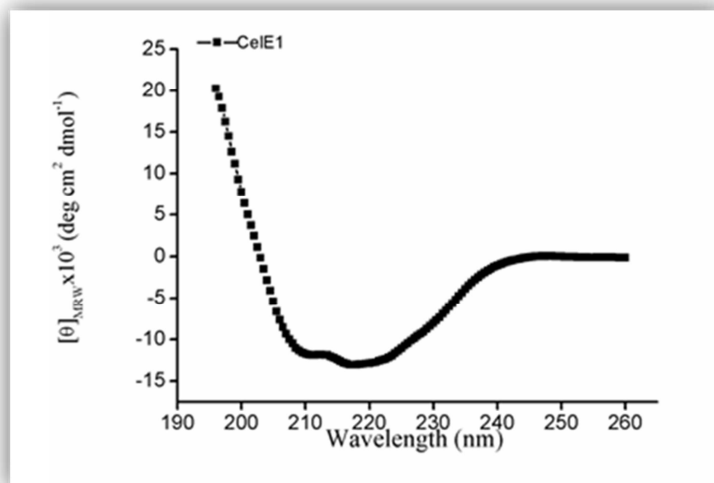


Figura 30: Espectro de dicroísmo circular no UV distante da CelE1.

Para observar a estabilidade da CelE1, ensaios de desnaturação térmica foram realizados com a proteína nativa e com 10 mM de MnCl_2 e CaCl_2 (figura 31). Entretanto, não houve alteração no perfil da desnaturação como observado para a BsCel5A e XacCel5, apresentando um T_M em torno de 55 °C (tabela 10). As temperaturas de transição encontradas para as outras enzimas caracterizadas foram de 64 e 50 °C para BsCel5A e XacCel5, respectivamente. A enzima de *B.subtilis* é de organismo termofílico, o que explica a maior estabilidade térmica em relação a enzima de *X. citri* e da biblioteca de metagenômica do solo de cana-de-açúcar.

Tabela 10: Temperaturas de transições (T_M) das desnaturações térmicas da CelE1.

Condição	T_M
pH 7,4	55,3
pH 7,4 + CaCl_2	55,0
pH 7,4 + MnCl_2	54,6

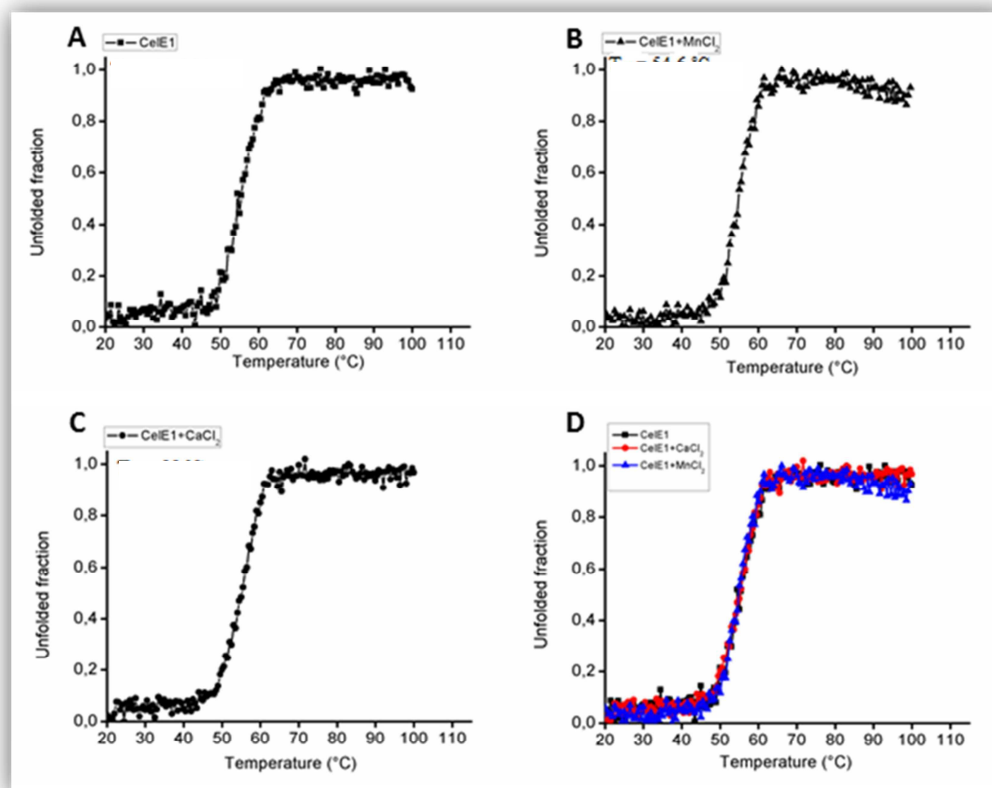


Figura 31: Curvas de Desnaturação térmica da CelE1 em pH 7,4, com MnCl₂ e CaCl₂, coletados a 222 nm. (A) desnaturação térmica da CelE1 pH 7,4. (B) desnaturação térmica da CelE1 com 10 mM cloreto de cálcio. (C) desnaturação da CelE1 com 10 mM de cloreto de manganês. (D) comparação das curvas de desnaturação da CelE1.

Ensaio cristalográficos foram realizados com a enzima a 17 mg/mL, utilizando a técnica difusão de vapor com gota sentada. O volume de solução proteica e da solução cristalizante foi de 0,5 µL cada. Em algumas condições foi possível observar cristais, que foram aprimorados variando a concentração de precipitante (figura 32). Após obter os cristais adequados para os experimentos de difração, esses foram levados para a linha de luz MX2 (LNLS- Campinas, SP) para análise por difração de raios X.

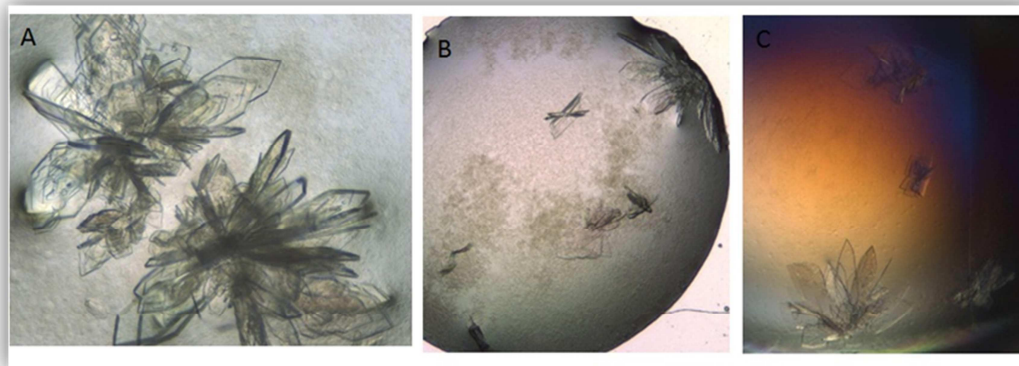


Figura 32: Cristais obtidos da celulase 5 oriunda da biblioteca de metagenômica do solo de cana-de-açúcar, a enzima estava a 17 mg/ml. As condições propícias para crescimento dos cristais estão listadas na tabela 11.

Tabela 11: Condições propícias para obter cristais da celulase oriunda da biblioteca de metagenômica do solo de cana-de-açúcar pertencente à família GH5.

Condição	Tampão	Precipitante
A	100 mM Tris pH 8,5	31% PEG 3350 30% isopropanol
B	100 mM TRIS-HCl pH 8,5	30% PEG 3350, 30% isopropanol, 5,5% glicerol
C	100 mM TRIS-HCl pH 8,5	30% PEG 3350 30% isopropanol

Os dados coletados tiveram resolução máxima de 1,8 Å. As reflexões foram indexadas no sistema cristalino monoclinico, com os parâmetros de cela unitária $a = 41,89 \text{ Å}$; $b = 87,51 \text{ Å}$; $c = 66,63 \text{ Å}$; e $\beta = 98,55^\circ$, pertencente ao grupo espacial $P2_1$. Duas moléculas foram encontradas na unidade assimétrica. A estrutura da proteína foi resolvida por substituição molecular, tendo como molécula molde uma celulase 5 de *Pseudoalteromonas haloplanktis* (PDBID: 1TVP), com similaridade de 63%. Na tabela 12 tem-se as estatísticas do processamento de dados e refinamento cristalográfico da CelE1.

Tabela 12: Estatística de coleta de dados e refinamento da CelE1.

Parâmetros	Resultados
Dados da coleta	
Grupo espacial	$P2_1$
Parâmetros da cela unitária (Å, °)	$a = 41,88$, $b = 87,51$, $c = 66,63$, $\beta = 98,55$
Resolução, Å	50,00 – 1,80 (1,86-1,80)
Multiplicidade	6,5 (3,8)
Completeza (%)	93,5 (80,4)
R_{merge} (%)	9,6 (30,2)
$I/\sigma(I)$	16,9 (3,9)
Refinamento	
Resolução	30,08 – 1,80
Número de reflexões	39062
$R_{\text{work}}/R_{\text{free}}$	0,133/0,173
Número de átomos	4895
Ligante/íon	2
Moléculas na unidade assimétrica	2
Moléculas de água	379
Média do fator B	9,226
RMSD do comprimento das ligações (Å)	0,0090
RMSD dos ângulos de ligação (°)	0,979
Ramachandran	
Região favorável (%)	94,05%
Regiões permitidas	5,10%
Outras regiões	0,85%

A análise da estrutura de CelE1 revela uma proteína com enovelamento $(\alpha/\beta)_8$, assim como as outras estruturalmente caracterizadas neste trabalho (figura 33), que é ubíquo da família GH5. O barril é formado por oito fitas β circundadas por oito hélices α unidas por longos segmentos de resíduos que na extremidade C-terminal do barril formam a interface catalítica.

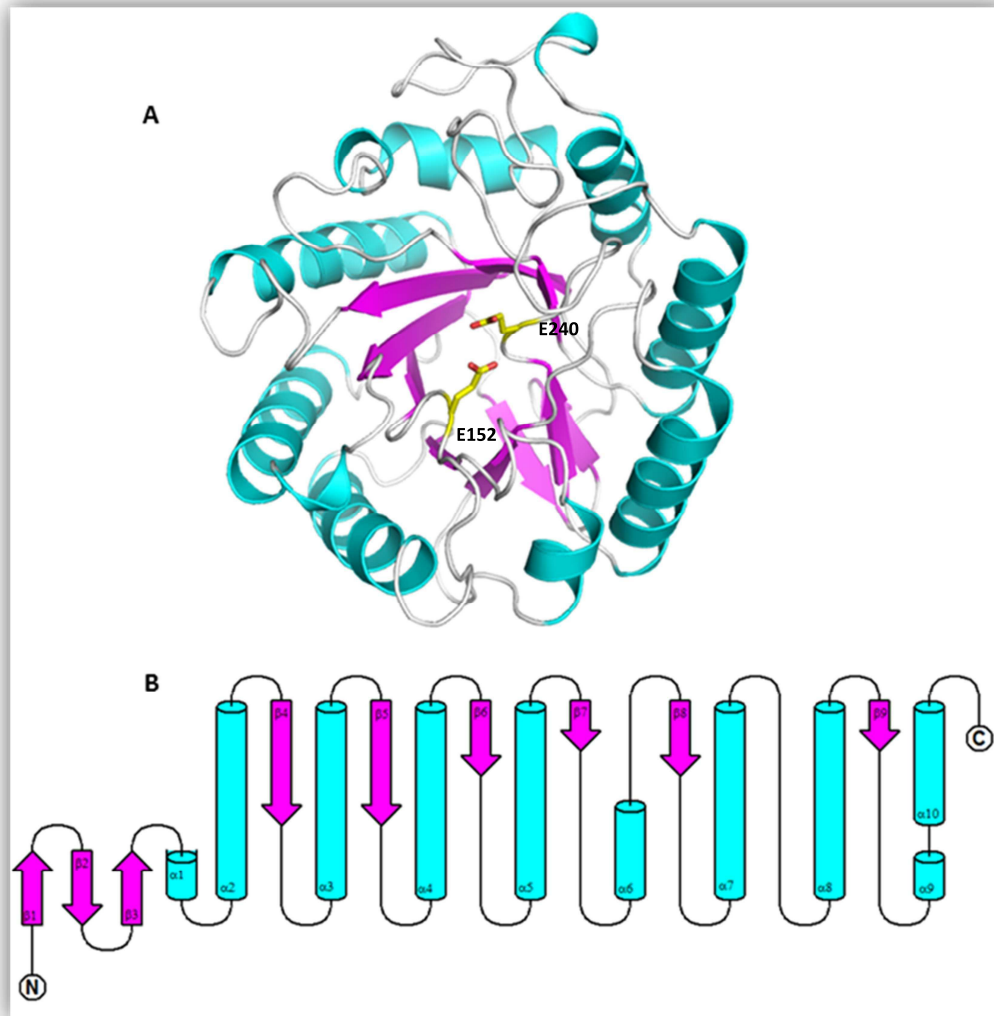


Figura 33: Representação da estrutura tridimensional da celulase 5 oriunda de biblioteca de metagenômica do solo de cana-de-açúcar, demonstrando o enovelamento característico da família GH5. **(A)** Estrutura tridimensional da CelE1 em *cartoon*. Em amarelo os dois resíduos ácidos catalíticos. **(B)** Topologia da estrutura da CelE1 (topdraw).

Esse enovelamento proteico, com formato globular, contém uma fenda catalítica (figura 34A) que é carregada negativamente, e localizada no C-terminal. A CelE1 contém um motivo estrutural extra no N-terminal que é um β -*harpin* (figura 34B). Este motivo também está presente em outros membros da família GH5, como na Cel5 (EGZ) de *Erwinia chrysanthemi* e na Cel5 de *Bacillus agaradherans*. Na Celulase 5A de *Bacillus subtilis*, esse elemento extra é responsável por interações com o íon manganês que proporcionou alta estabilidade proteica (Davies et al., 1998; Chapon et al., 2001; Shaw et al., 2002; Santos et al., 2012).

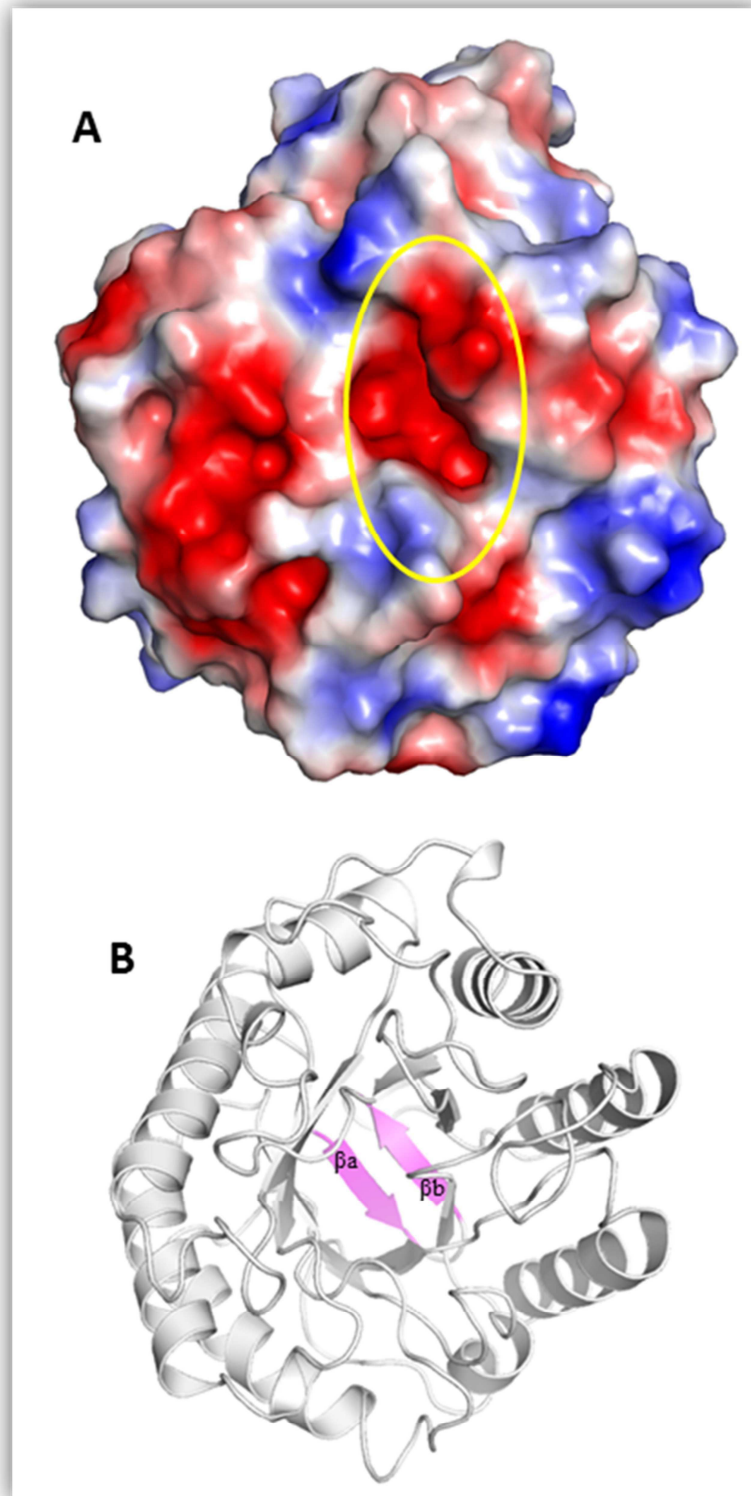


Figura 34: Representação da superfície molecular e da estrutura tridimensional da CelE1. **A** Superfície da molécula da CelE1, com destaque para a fenda catalítica localizada em uma região carregada negativamente (azul +, vermelho -). **B** Estrutura terciária da CelE1, evidenciando as duas folhas β anti-paralelas que formam o β -harpin no N-terminal da enzima.

Estudo estrutural comparativo da CelE1 com outras celulasas da família GH5 mostrou que os principais resíduos da catálise, Glu152 e Glu249, são conservados, sendo o primeiro o doador de prótons e o segundo o nucleófilo (figura 35B). Para que ocorra a catálise enzimática é necessária a participação de outros resíduos, como His127, His212, Trp189 e Trp241, que são conservados em todos os membros da família GH5 (figura 35A) (Chapon et al., 2001; Dominguez et al., 1995; Ducros et al., 1995). Os resíduos Trp189 e Trp241 são considerados essenciais pela ligação ao substrato através de interações com anéis glicopiranosos (Vyas, 1991) e estão a uma distância de 16,2 Å (figura 35C). Os resíduos His127 e His212 também contribuem para a formação da superfície de interação com o substrato, tendo uma distância de 11,9 Å. Tais resíduos compõem a superfície hidrofóbica responsável pela interação com substrato.

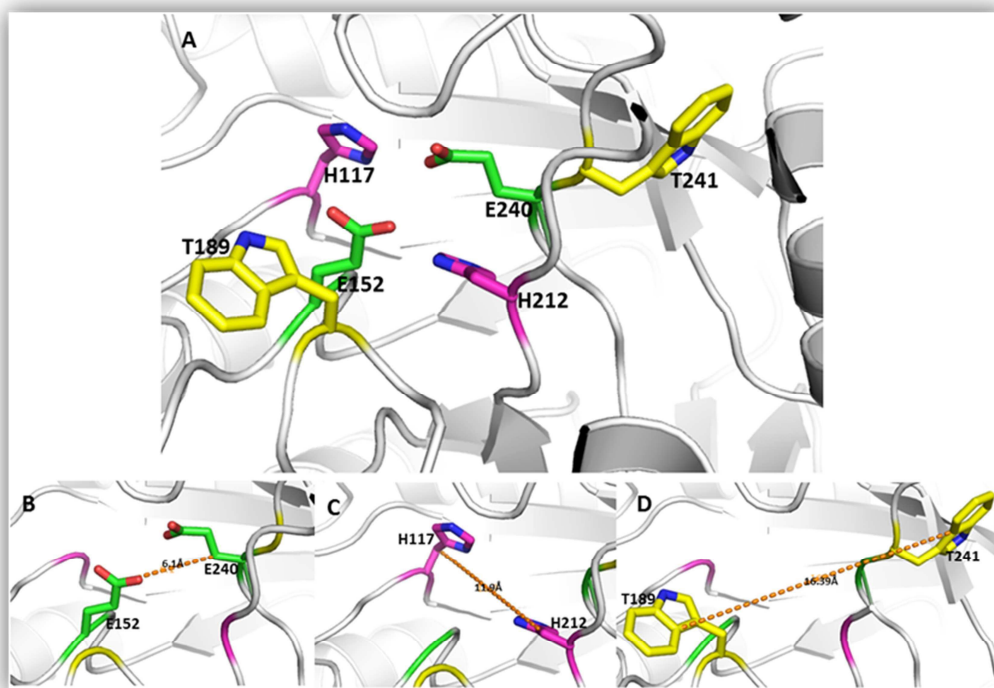


Figura 35: Destaque para os resíduos que formam a fenda catalítica da CelE1. (A) Destaque dos resíduos responsáveis pela hidrólise da celulose na CelE1. (B) Distância dos resíduos ácidos glutâmicos, chaves na catálise. (C) Distância das histidinas que fazem parte do bolsão catalítico. (D) Distância entre os resíduos dos triptofanos que interagem com a celulose na interface catalítica.

Fazendo um alinhamento entre alguns membros da família GH5, foi possível observar uma conservação estrutural entre eles. O alinhamento estrutural, mostra que a celulase 5 de *Erwinia chrysanthemi* tem um RMSD de 0.39 Å para os átomos Ca comparado com a CelE1. As celulasas de *Pseudoalteromonas haloplanktis*, *Bacillus agaradhaerens* e *Bacillus subtilis* resultaram em RMSDs de 0,43; 0,56 e 0,64 Å, respectivamente (Chapon et al. 2001; Violot et al. 2005; Varrot et al. 2000; b Santos et al. 2012). Pequenas diferenças estruturais podem ser observadas entre essas GH5, principalmente nos *loops* que conectam as hélices α com as fitas β (figura 36) e que estão voltados para a interface catalítica.

Comparando a CelE1 com a Cel5 de *Erwinia chrysanthemi* (PDBID: 1EGZ) observamos uma diferença estrutural significativa no loop $\beta 8-\alpha 8$ com uma inserção de 8 resíduos (²⁹²ANGGWTSS²⁹⁹) na CelE1. Esta inserção favorece novos contatos intramoleculares do loop $\beta 8-\alpha 8$ com a hélice $\alpha 1$ que pode refletir na estabilidade conformacional da proteína (figura 36A) (Chapon et al., 2001). A celulase 5G de *Pseudoalteromonas haloplanktis* (PDBID: 1TVN) também apresenta esta inserção, que proporciona essa interação, juntamente com a deleção de três resíduos no loop $\beta 2-\alpha 2$ na CelE1 (figura 36B) (Violot et al., 2005). As enzimas oriundas de organismos capazes de sobreviver em baixas temperaturas, não apresentam o longo loop $\beta 8-\alpha 8$, e conseqüentemente não há a interação com a hélice $\alpha 1$.

A celulase 5A de *B. subtilis* (PDBID: 3PZU) é a enzima com o maior número de inserções nos loops interfaciais incluindo os loops $\beta 1-\alpha 1$ (WYGE), $\beta 3-\alpha 3$ (NPN) e $\beta 7-\alpha 7$; a enzima de *B. subtilis* mantém o longo loop $\beta 8-\alpha 8$ compartilhando interações intramoleculares com a hélice $\alpha 1$ (figura 36C) (b Santos et al., 2012). A Cel5A de *B. agaradhaerens* (PDBID: 1QHZ) apresenta uma maior variabilidade estrutural na interface catalítica que inclui todas as inserções nos loops interfaciais descritas para a Cel5A de *B. subtilis* e adicionalmente alterações conformacionais no loop $\beta 4-\alpha 4$ (figura 36D) (Davies et al., 1998). As enzimas oriundas dos organismos mesofílicos e termofílicos apresentam um longo loop $\beta 8-\alpha 8$, que com as interações intramoleculares podem auxiliar na maior estabilidade global dessas enzimas, sendo um importante mecanismo de estabilização.

De forma geral, celulasas pertencentes à família GH5 apresentam um enovelamento conservado com inserções ou deleções em loops interfaciais que podem estar relacionados a estabilidade e eficiência enzimática. Então, conclui-se desse trabalho que a nova celulase 5 de uma biblioteca metagenômica do solo de cana-de-açúcar possui propriedades estruturais extremamente conservadas incluindo todos os resíduos cataliticamente relevantes e

diferenças pontuais residem nos loops interfaciais, estudos funcionais serão conduzidos pelos pesquisadores do CTBE com intuito de avaliar o potencial biotecnológico dessa enzima.

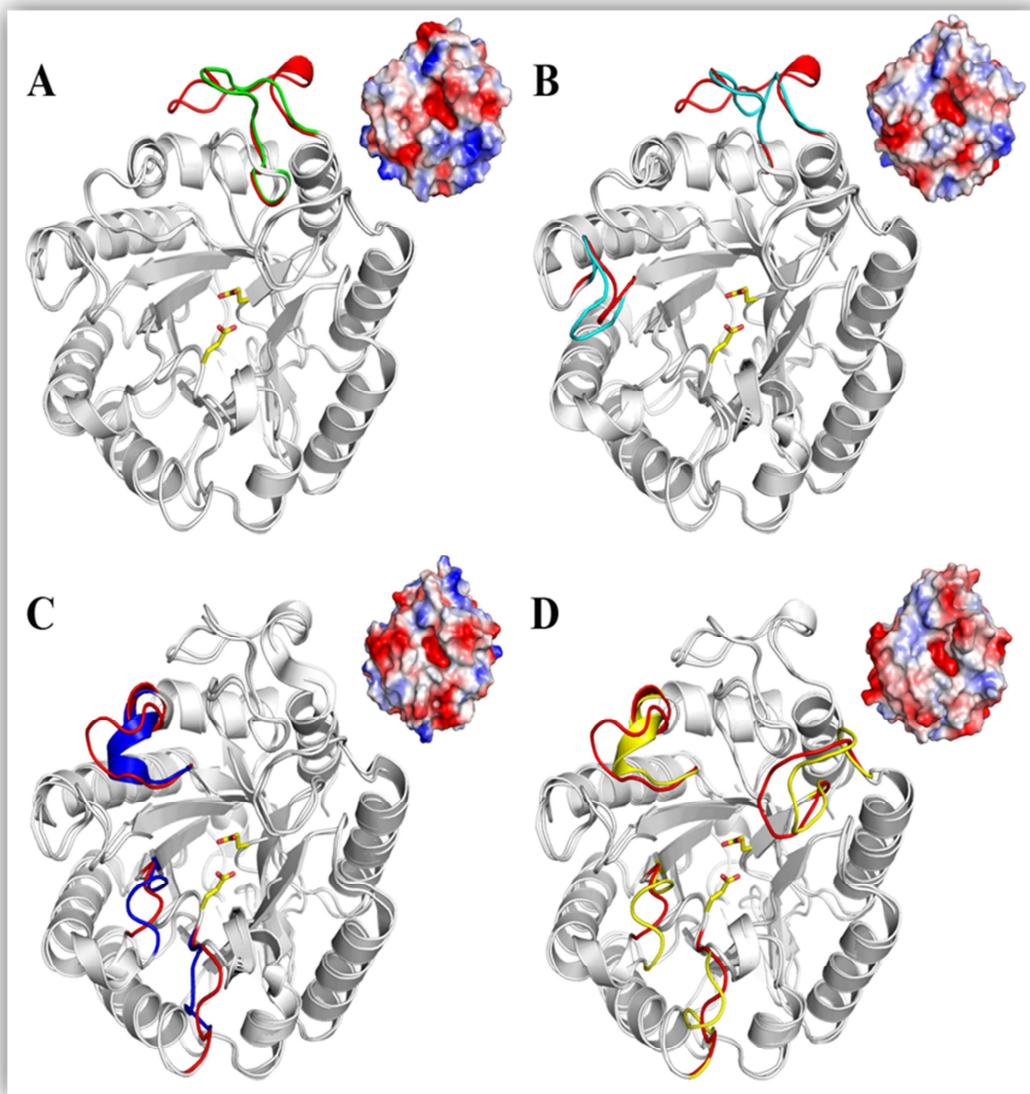


Figura 36: Comparação estrutural entre quatro celulasas da família GH5 com a CelE1. Em vermelho a CelE1; verde a Cel5 de *Erwinia chrysanthemi*; ciano a Cel5G de *Pseudoalteromonas haloplanktis*; azul a Cel5A de *Bacillus subtilis*; amarelo a Cel5 de *Bacillus agaradhaerens* (A) Cel5 de *Erwinia chrysanthemi* comparada com a CelE1. (B) Comparando a Cel5G de *Pseudoalteromonas haloplanktis* com a CelE1. (C) Comparação da Cel5A de *Bacillus subtilis* com a CelE1. (D) Comparação da Cel5A de *Bacillus agaradhaerens* com a CelE1.

IV. 4 GH5-CBM27/ endo-1,4- β -D-mananase de *Thermotoga petrophila*

Dentro da família GH5 encontra-se uma diversidade de atividades incluindo quitosanase (EC 3.2.1.132); β -manosidase (EC 3.2.1.25); endo- β -1,4-glicanase / celulase (EC 3.2.1.4); β -1,3-glicosidase (EC 3.2.1.58); liqueninase (EC 3.2.1.73); endo-1,6- β -glicosidase (EC 3.2.1.75); endo- β -1,4-manosidase (EC 3.2.1.78); endo- β -1,4-xilanase (EC 3.2.1.8); β -1,4-cellobiosidase (EC 3.2.1.91); β -1,3-mananase (EC 3.2.1.-); endo- β -1,4-glicanase xiloglicana-específica (EC 3.2.1.151); manana transglicosilase (EC 2.4.1.-); endo- β -1,6-galactanase (EC 3.2.1.164); endoglicoceramidase (EC 3.2.1.123); β -primeverosidase (EC 3.2.1.149); β -glicosilceramidase (EC 3.2.1.45); hesperidina 6-O- α -L-ramanosil- β -glicosidase (EC 3.2.1.168); e exo- β -1,4-glicanase / celodextrinase (EC 3.2.1.74). Dentre essas funções além das celulases, destacam-se as endo-1,4- β -D-mananases que possuem estrutura modular similar as celulases GH5 e apresentam grande potencial biotecnológico para degradação de biomassa vegetal, pois contribuem na degradação da fração hemicelulósica aumentando o rendimento de açúcares fermentescíveis. Os estudos comparativos irão fornecer os determinantes estruturais para diferenciação funcional, especificidade e os efeitos dos módulos acessórios na função e estabilidade dessas enzimas. Para isso, realizamos uma caracterização abrangente da endo-1,4- β -D-mananase do microorganismo hipertermofílico *Thermotoga petrophila*.

Duas construções foram empregadas neste estudo: a proteína selvagem (TpMan), com aproximadamente 77 kDa, formada por um domínio catalítico de aproximadamente 46 kDa, pertence à família GH5, seguido por um junção de 98 resíduos e no C-terminal um CBM pertencente a família 27, com aproximadamente 20 kDa (figura 37); e apenas o domínio catalítico da enzima, denominado DC_TpMan.

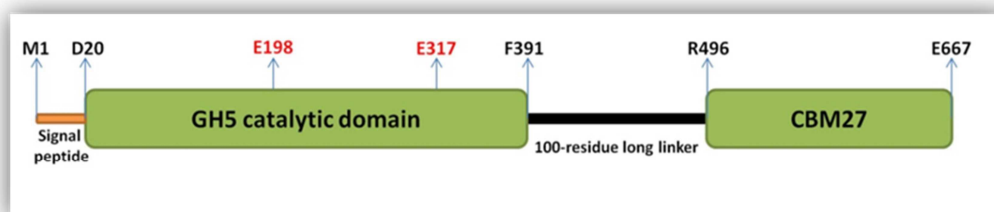


Figura 37: Representação da arquitetura dos domínios pertencentes à endo-1,4- β -D-mananase de *Thermotoga petrophila* - TpMan.

A expressão de ambas as construções foi em BL21(DE3) Δ SlyD a 37 °C com indução de 4 horas por 0,5 mM IPTG. Para realização dos estudos biofísicos, 1 L de expressão proporcionou quantidades suficientes da proteína recombinante pura após dois passos cromatográficos, afinidade em coluna de níquel imobilizado e exclusão molecular, com colunas pré-empacotadas. As análises das etapas de cromatografia foram analisadas por SDS-PAGE 13% (figura 38).

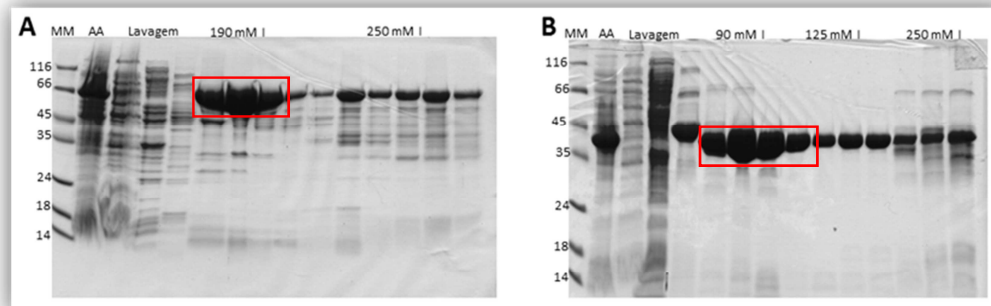


Figura 38: SDS-PAGE 13% da cromatografia de afinidade da enzima TpMan e da DC_TpMan. Em destaque as amostras utilizadas para os estudos posteriores. **(A)** cromatografia de afinidade da TpMan. **(B)** cromatografia de afinidade da DC_TpMan. MM marcador molecular; AA amostra aplicada da expressão; lavagem da coluna; 90 mM, 125 mM, 190 mM e 250 mM correspondem à concentração do imidazol.

Estudos por espectroscopia de dicroísmo circular mostraram um espectro na faixa do UV distante típico da família GH5 de forma análoga ao observado para as celulasas 5 (figuras 39A e 39B). As enzimas oriundas de seres termofílicos, têm demonstrado um papel de destaque para possíveis aplicações biotecnológicas, devido à capacidade de manter-se funcional mesmo em condições adversas. Para observar a estabilidade da TpMan em altas temperaturas, ensaios de desnaturação térmica foram realizados (figuras 39C e 39D). A enzima selvagem demonstrou ser mais termoestável que o domínio catalítico isolado, com T_M de 100 e 88 °C, respectivamente. Tal estabilidade pode ser originária da presença do CBM27, já que o mesmo efeito foi descrito na β -mananase-26 da bactéria termofílica *Caldicellulosiruptor* cepa Rt8B.4, que contem dois CBM27 no N-terminal (Roske et al., 2004).

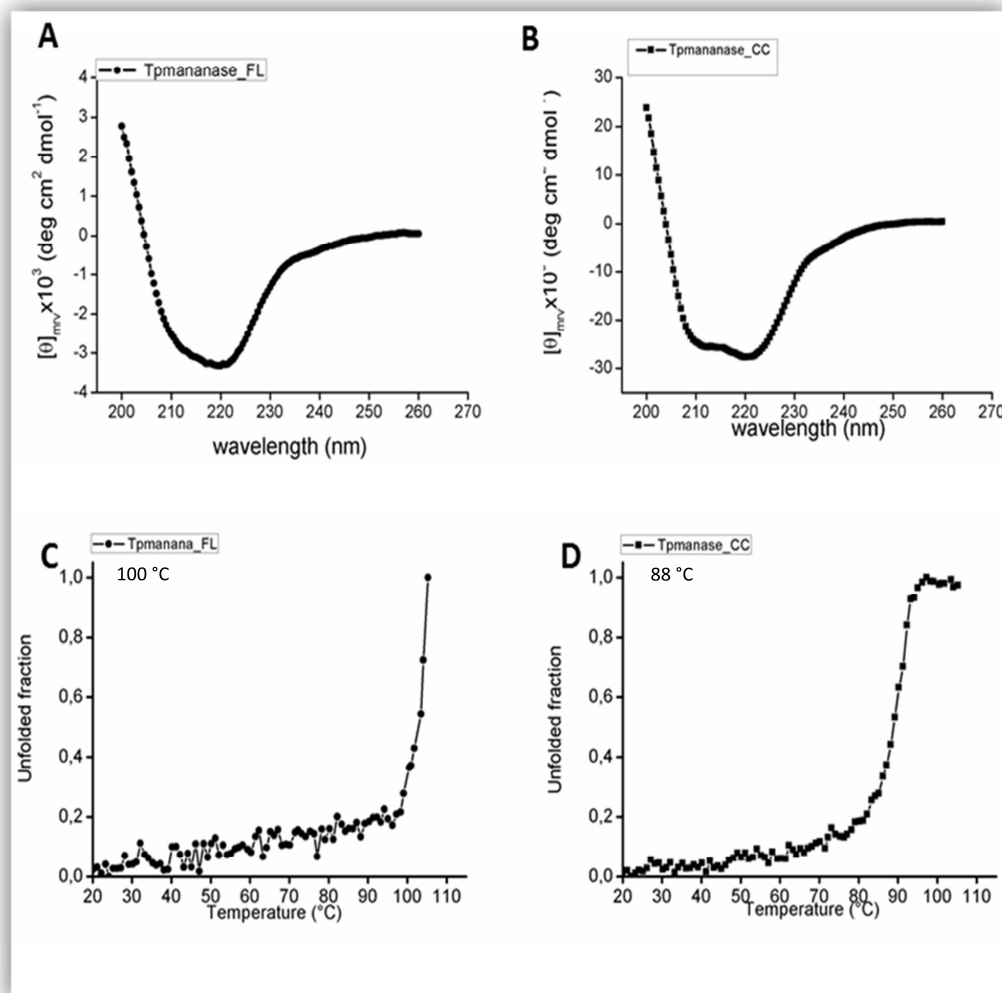


Figura 39: Caracterização biofísica por dicroísmo circular da enzima Tpman. **(A)** Espectro do UV distante da enzima selvagem. **(B)** espectro do UV distante do domínio catalítico. **(C)** Curva da desnaturação térmica enzima selvagem. **(D)** Curva da desnaturação térmica do domínio catalítico.

As enzimas aptas à degradação de manana da família GH5 que já tiveram suas estruturas tridimensionais elucidadas apresentam uma identidade de aproximadamente 30% com a TpMan. A β -mananase de *Lycopersicon Esculentum* (LeMan, PDBID: 1RH9) tem similaridade de 33%, mesmo percentual da β -mananase de *Trichoderma Reesei* (TrMan, PDBID: 1QNO). Enquanto que a exo-manosidase de *Cellivibrio mixtus* (CmMan, PDBID: 1UUQ) compartilha de uma identidade de 26%. O alinhamento das sequências dessas β -mananases confirma a baixa conservação entre os resíduos (figura 40). A comparação das sequências demonstra que doze resíduos são estritamente conservados em todas as β -mananases, e desses sete estão vinculados ao sítio catalítico (Hilge et al., 1998).

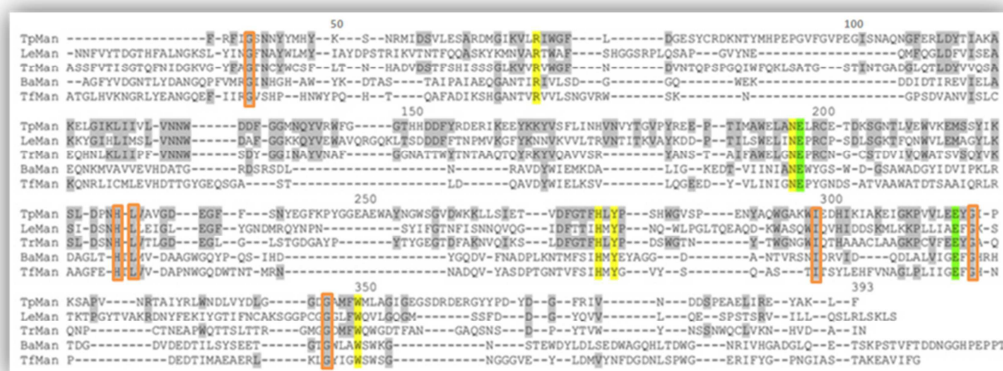


Figura 40: Alinhamento das β -mananases da família GH5 com estrutura tridimensional determinada. Os resíduos marcados em cinza são os resíduos idênticos entre TpMan e outras β -mananases. Os resíduos em verdes são os resíduos ácidos catalíticos. Os resíduos destacados em amarelo são os constituintes da fenda catalítica. LeMan, β -mananase de *Lycopersicon Esculentum*. TrMan, β -mananase de *Trichoderma Reesei*. BaMan, β -Mananase de *Bacillus alkaliphilic*. TfMan, β -Mananase de *Thermomonospora Fusca*.

Para uma melhor compreensão do mecanismo de ação da β -mananase de *T. petrophila*, realizou-se ensaios de cristalização com o domínio catalítico, já que a enzima selvagem não houve resultados satisfatórios. Os ensaios iniciais de cristalização com DC_TpMan foram realizados com a proteína 12 mg/ml, em 25 mM TRIS-HCl, pH 7,5. Na tabela 13, está relacionada as condições de cristalização do domínio catalítico da β -mananase de *Thermotoga petrophila*.

Tabela 13: Condições de cristalização do domínio catalítico da β -mananase de *Thermotoga petrophila*.

Condição	Tampão	Precipitante	Sal
A	100 mM citrato pH 5,5	10% glicerol	850 mM fosfato de amônia
B	100 mM TRIS-HCl pH 8,5	10% glicerol	600 mM citrato de sódio
C	100 mM fosfato pH 4,2	5% PEG 1000 36% etanol 10% glicerol	
D	100 mM citrato pH 5,5		1 M fosfato de amônia

Os cristais foram difratados na a linha de luz MX2 do LNLS, Campinas - SP. A estrutura da DC_TpMan foi resolvida por SIRAS (*single isomorphous replacement anomalous scattering*), em colaboração com a Dra Camila Ramos dos Santos, assistente de pesquisa do Laboratório Nacional de Biociências – LNBio, pois métodos de substituição molecular não resultaram em mapas de densidade eletrônica interpretáveis devido a baixa identidade das estruturas homólogas disponíveis no PDB. Na tabela 14 encontram-se as estatísticas da coleta de dados e do refinamento cristalográfico.

A estrutura tridimensional do domínio catalítico TpMan apresenta um enovelamento clássico da família GH5, o barril TIM (α/β)₈ (figura 42A), com o sítio catalítico localizado no C-terminal (figura 42C). Os principais resíduos que constituem a fenda catalítica são aromáticos e ácidos, dentre esses os Glu198 e Glu317 principais responsáveis pela catálise, sendo o primeiro doador de prótons e o segundo nucleofílico (figuras 42B e 42D). O Glu198 localiza-se no loop entre a fita β 4 e a hélice α 4, e o Glu317 no fita β 7. A composição estrutural do sítio ativo como uma fenda aberta ($\sim 1500 \text{ \AA}^3$) é típico das β -mananases, proporcionando uma melhor acomodação dos substratos ramificados (Bourgault et al., 2005).

Tabela 14: Estatística da coleta de dados e do refinamento da estrutura de DC_TpMan.

Parâmetros	Forma I	Forma II	Iodo
Dados da coleta			
Grupo Espacial	<i>I</i> 222	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Parâmetros da célula unitária (Å, °)	a=91,03 b=89,97 c= 97,90	a=55,46 b=83,35 c=92,18	a=55,55 b=83,63 c=92,66
Resolução (Å)	30,00-1,40 (1,45-1,40)	30,00-1,42 (1,47-1,42)	30-1,82 (1.89-1.82)
Reflexões únicas	76427 (6160)	80074 (6891)	39216 (3790)
Multiplicidade	12,1 (8,1)	9,1 (5,3)	13,6 (12,5)
Completeza (%)	96,5 (78,7)	98,4 (86,0)	99,9
R_{merge}	5,1 (33,7)	7,2 (56,8)	9,0 (31,1)
$I/\sigma(I)$	40,3 (4,9)	30,2 (2,2)	24,5 (8,9)
Refinamento			
$R_{\text{work}}/R_{\text{free}}$	15,4/18,4	14,5/17,8	-
Moléculas de água	319	417	-
Média do fator B	17,4	17,8	-
RMSD do comprimento das ligações (Å)	0,030	0,030	-
RMSD dos ângulos de ligação (°)	2,301	2,229	-
<i>Ramachandran</i>			-
Região favorável (%)	96,9	96,6	-
Regiões permitidas	3,1	3,4	-
PDB	3PZG	3PZ9	-

A rede de interações que subsidia o sítio ativo envolve contatos entre o Glu198 com a His278 e Trp134, enquanto que Glu317 é coordenado por Arg71 e a cadeia lateral da Tyr280 (figura 41B). Os resíduos Tyr45, Trp73, Asn197, Arg200, Glu235, Trp284 e Trp350 formam a interface de reconhecimento ao substrato, e são altamente conservados nas β -mananases.

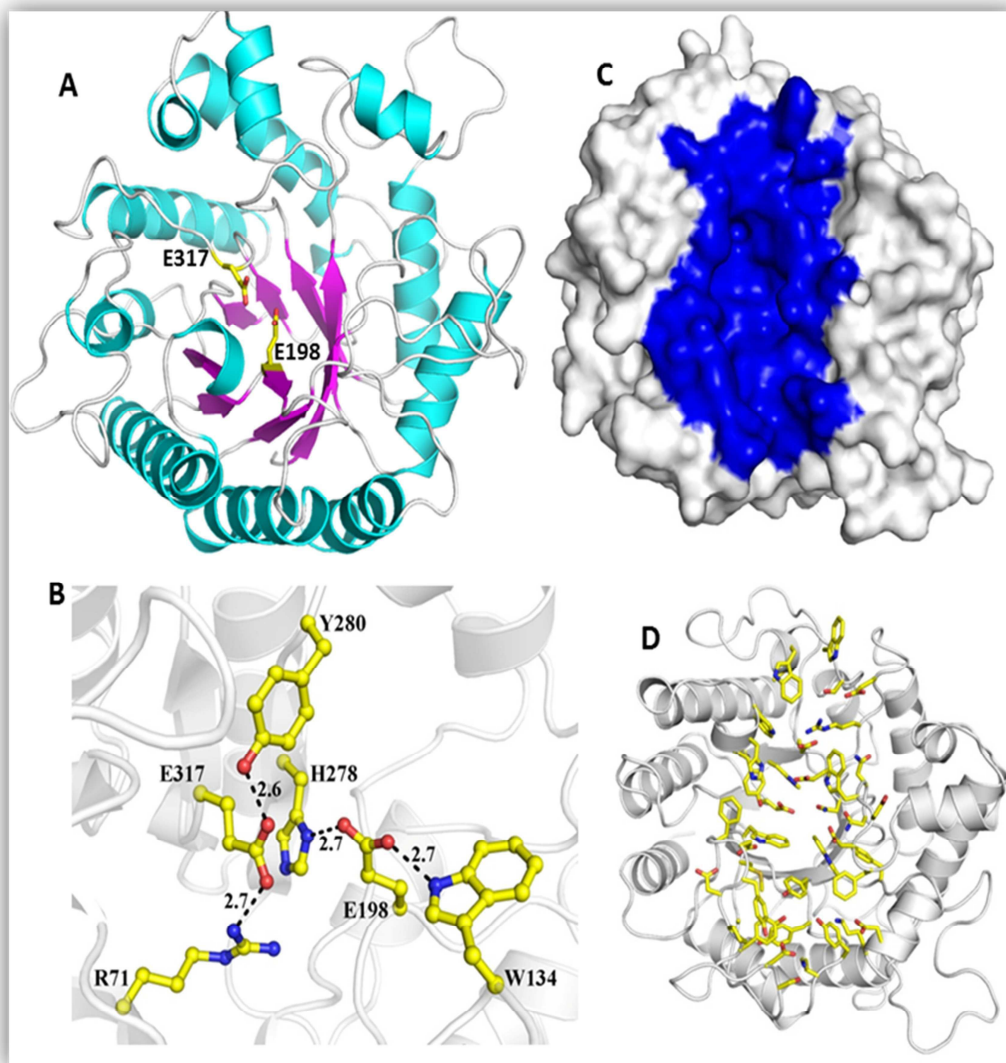


Figura 41: Estrutura do domínio catalítico da TpMan, destacando a fenda catalítica, juntamente com os principais resíduos da catálise. (A) estrutura tridimensional ilustrando o enovelamento TIM barril (α/β)₈ característico da família GH5. (B) rede de ligações de hidrogênio que mantem a conformação dos resíduos catalíticos Glu198 e Glu317. (C) perfil da superfície molecular do sítio ativo, em azul. (D) representação dos resíduos que formam o bolsão catalítico.

Há um interesse em compreender como as β -mananases interagem e clivam seus substratos, e quais são os subsítios utilizados na clivagem. Tal empenho para compreender o

mecanismo de ação existe devido à heterogeneidade dos polissacarídeos de glicomanana, que tem a cadeia principal formada por glicose e manose unidas por ligação β -1,4. Estudos realizados com outras enzimas revelaram que os subsítios em que há a interação do substrato são enumerados da seguinte forma: -4, -3, -2, -1, +1 e +2, sendo da extremidade não redutora para a extremidade redutora do polissacarídeo (Davies et al., 1997).

A TrMan, uma β -mananase de *Trichoderma Reesei* pertencente a família GH5, teve sua estrutura tridimensional resolvida com diferentes ligantes no sítio catalítico, o TRIS, o glicerol, e o dissacarídeo manobiose. As interações do TRIS - enzima e glicerol - enzima, incidiram nos subsítios -2, -1 e +1. Já a manobiose ocupou os subsítios +1 e +2 da Trman (Sabini et al., 2000).

Análises com sobreposições da região do subsítio -1 de diferentes β -mananases, revelam uma conservação dos resíduos que compõem esse local de interação. O subsítio -1 é formado por uma plataforma hidrofóbica integrada por três triptofanos, responsáveis por interações com o anel piranose. Já o subsítio -2 não é formado rigorosamente pelos mesmos resíduos, e estudos mostram que tal subsítio tem maior afinidade para ligações com a glicose do que com manose. Enquanto que o subsítio +1 não tem preferência nem por glicose ou manose, permitindo que as interações podem ocorrer com ambos. Diante de tais interações propostas, acredita-se que as β -mananases da família GH5 são aptas a interagir com glicose e manose, desta forma capazes de degradar a glicomanana (Tailford et al., 2009).

A β -mananase de *Thermomonospora fusca* teve sua estrutura tridimensional resolvida com dois ligantes, a manotriose e a manohexaose. Os açúcares ocupavam os subsítios -2, -3, e -4, similar à localização dos subsítios -2 e -3 da celulase 5A de *Bacillus agaradherans*. Apesar de não haver uma densidade eletrônica conclusiva para o subsítio -1, a sobreposição da estrutura nativa da β -mananase com as estruturas com manotriose e manohexaose, relevou um padrão de densidade eletrônica capaz de representar o subsítio -1 (Hilge et al., 1998).

Para observar o comportamento da ligação dos substratos na DC_TpMan, foram resolvidas as estruturas, por substituição molecular, da enzima com glicose, maltose e glicose + maltose. Tais açúcares foram incorporados utilizando *soaking*, com soluções de 0,1 M de maltose e 0,5 M de glicose. As estatísticas dos dados coletados e dos refinamentos das estruturas estão detalhados na tabela 15.

Tabela 15: Estatísticas das coletas de dados e dos refinamentos das estruturas de DC_TpMan com ligantes.

Parâmetros	Glicose	Maltose	Glicerol + Maltose
<i>Dados da coleta</i>			
Grupo Espacial	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Parâmetros da célula unitária	a=55,43	a=55,11	a=55,29
(Å, °)	b=83,23	b=83,19	b=83,30
	c= 91,92	c=91,52	c=92,08
Resolução (Å)	30,00-1,55	30,00-1,92	30,00 -1,55
	(1,61-1,55)	(1,99-1,92)	(1,61-1,55)
Reflexões únicas	62462 (6068)	32825 (32250)	61007 (5208)
Multiplicidade	6,1 (5,3)	10,0 (10,5)	7,1 (5,7)
Completeza (%)	99,9 (99,1)	99,6 (99,1)	97,7 (84,5)
R_{merge}	7,3 (52,2)	9,5 (50,9)	7,1 (56,0)
$I/\sigma(I)$	23,2 (2,9)	25,0 (5,0)	24,7 (2,4)
<i>Refinamento</i>			
$R_{\text{work}}/R_{\text{free}}$	14,0/18,3	16,3/21,0	17,1/20,2
Moléculas de água	370	265	291
Média do fator B	17	18,5	17,4
RMSD do comprimento das ligações (Å)	0,027	0,025	0,033
RMSD dos ângulos de ligação (°)	2,073	1,997	2,639
Ligantes	Glicose	Maltose	Glicerol + Maltose
<i>Ramachandran</i>			
Região favorável (%)	96,9	96,1	96,9
Regiões permitidas	3,1	3,9	3,1
PDB	3PZI	3PZO	3PZQ

Comparando os diferentes ligantes obtidos a alta resolução ($> 1,8 \text{ \AA}$) foi possível mapear a interação de cada substrato com a DC_TpMan (figura 42) e assim contribuir para entender a preferência desta enzima por glicomanana ao invés de manana linear. A estrutura da DC_Tpman com glicerol, tem-se o ligante localizado no subsítio +1, exibindo uma orientação usual, na qual o átomo O^1 é estabilizada por ligações de hidrogênio com os resíduos Glu198, Arg235, His283 e Trp284 (figura 42C). Já a maltose teve sua localização nos subsítios -3 e -2 (MAL1), e enquanto a segunda se fixou nos subsítios +1 e -1 (MAL2) (figura 42E). Ambas compartilham de interações com os resíduos catalíticos e resíduos aromáticos, que formam a fenda catalítica. Além disso, outros ligantes provenientes das condições de cristalização e que mimetizam as interações dos açúcares são observados na fenda catalítica, particularmente no subsítio +1. A capacidade de interação com diferentes ligantes observada para o subsítio +1 é compatível com especificidade demonstrada pela TpMan com os substratos manana e glicomanana. Os complexos ligantes-proteína aqui demonstrados revelaram interações importantes para o reconhecimento e interação da glicose com a TpMan (figura 43).

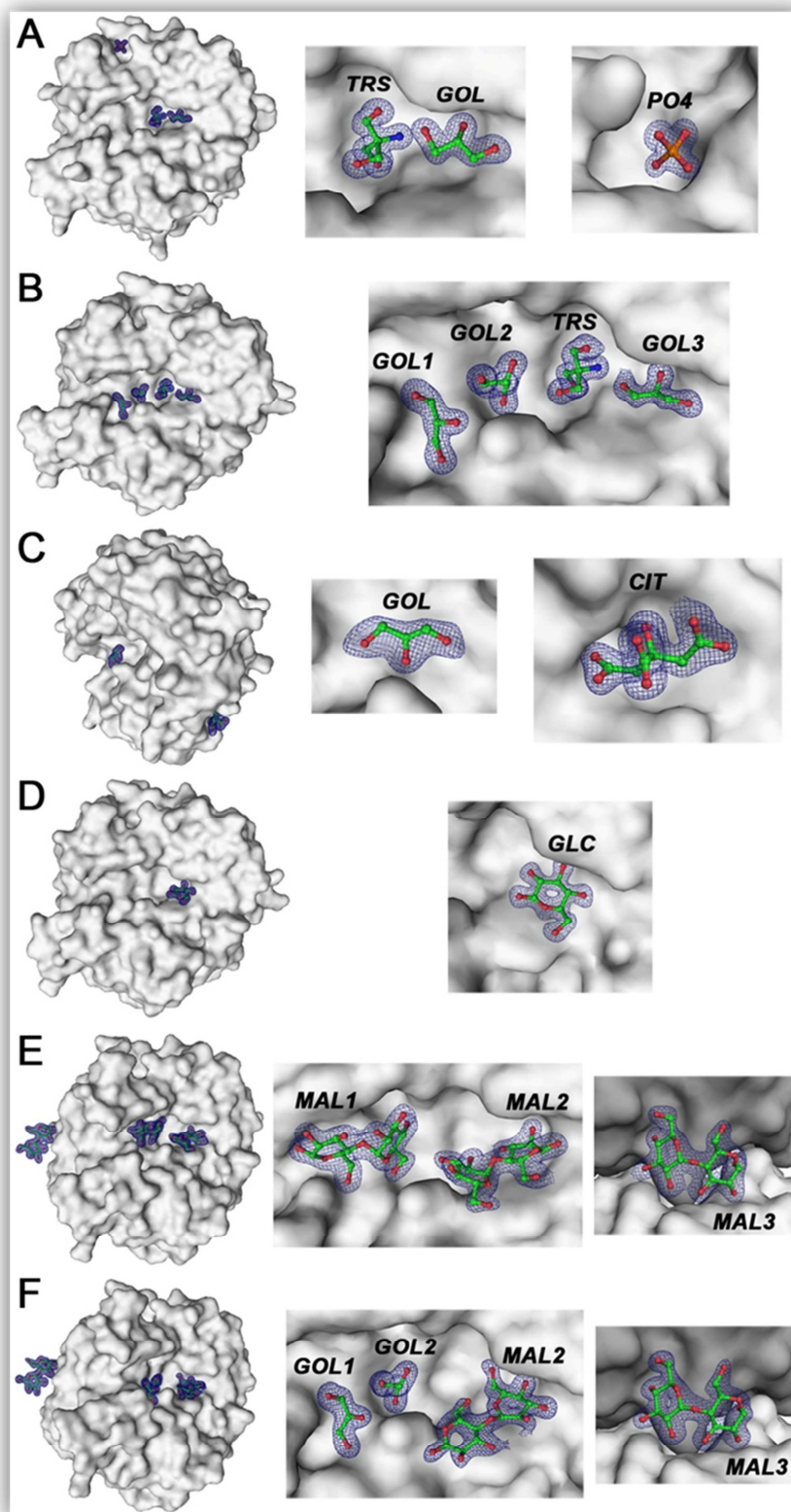


Figura 42: Localização dos ligantes na DC_TpMan, com o mapa de densidade eletrônica em 1.0 sigma. (A) Estrutura do grupo espacial I222. (B) Complexo de TRIS e glicerol. (C) Complexo de CIT + glicerol. (D) Complexo glicose. (E) Complexo de maltose. (F) Complexo de maltose + glicose.

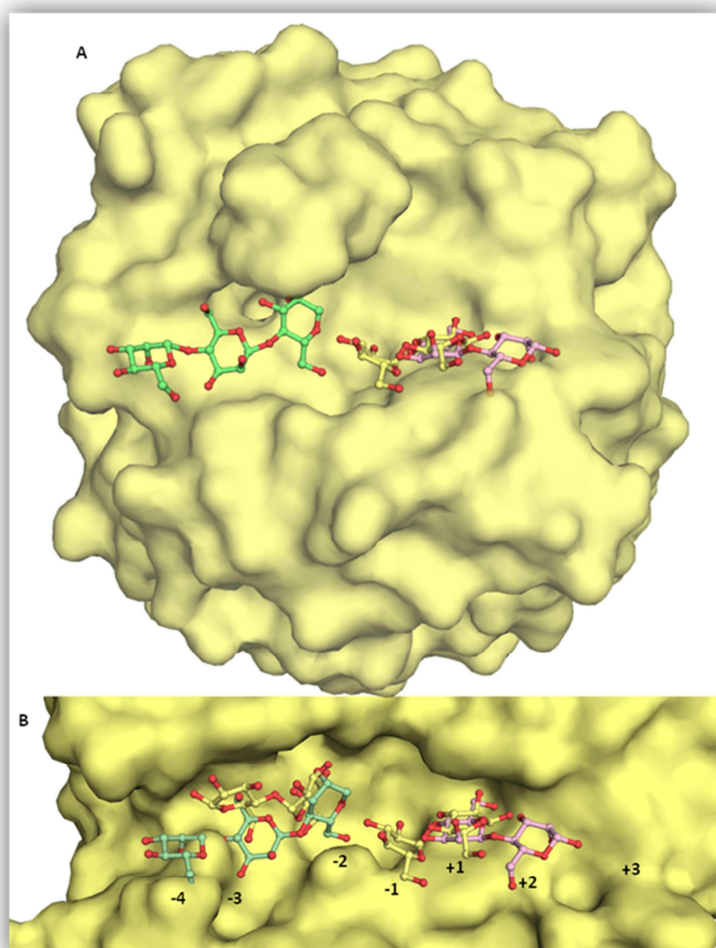


Figura 43: Intareção dos ligantes nos subsítios que compões a fenda catalítica da DC_TpMan. (A) representação da superfície molecular da TpMan com os substratos no sítio catalítico. (B) Detalhe da localização dos substratos nos seus respectivos subsítios.

Dos resíduos conservados entre as β -mananases da família GH5 com estrutura tridimensional conhecida, sete deles estão presentes na TpMananase, Arg71, Asn197, Glu198, His278, Tir280, Glu317 e Trp350, sendo todos relacionados na formação dos subsítios -1 e +1 (figura 44). O subsítio +1 da TpMan tem um resíduo de histidina (His283) não conservado, a qual é substituída por uma glutamina na β -mananase de *Lycopersicon esculentum*, e por uma serina na β -mananase de *Trichoderma reesei*. Na estrutura nativa, a His283 exibe uma dupla conformação, mas nas estruturas complexadas com sacarídeos, a His283 exibe uma conformação próxima aos resíduos aromáticos Trp238 e Phe373. Tal resíduo participa na

coordenação das moléculas de maltose e glicose, sugerindo um papel importante no reconhecimento e interação com a glicomanana.

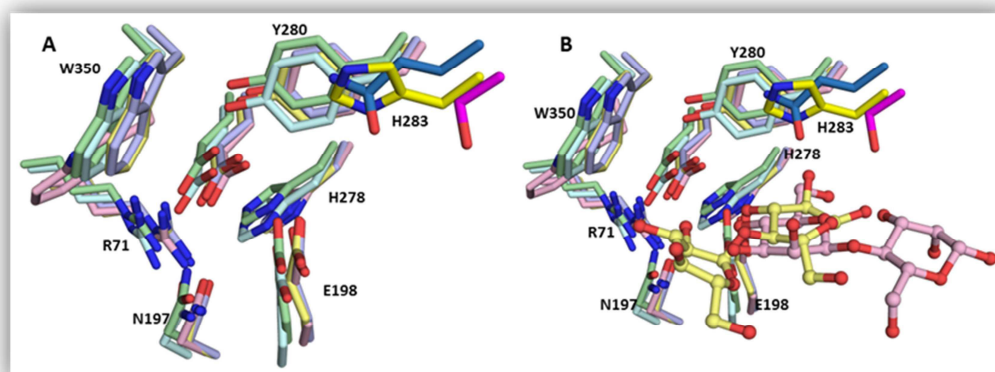


Figura 44: Detalhe dos resíduos que compõem o subsítio -1 e +1 da TpMan. Os átomos de carbono de cada resíduo e seu ligante foram coloridos de acordo com cada β -mananase: *Thermotoga petrofila* e a maltose amarelo; *Trichoderma reesei* e manobiose rosa; *Thermomonospora fusca* e manotriose verde; *Lycopersicon esculentum* azul; *Bacillus agaradherans* ciano. **A** apenas os resíduos. **B** os resíduos catalíticos e os ligantes.

A resolução da estrutura da TpMan revelou a topologia da interface catalítica, assim como que a resolução da proteína com ligantes permitiu definir os resíduos primordiais para a formação dos subsítios, consequentemente os resíduos responsáveis pela interação com o substrato. Além disso, estudos biofísicos revelaram o papel do domínio acessório na estabilidade térmica da enzima, indicando que o CBM27 pode ser uma ferramenta para aumento de termoestabilidade de enzimas pertencentes à família GH5. Tais dados podem contribuir para uma melhor compreensão da clivagem da glicomanana, e auxiliar o entendimento das bases estruturais das enzimas termofílicas.

V. conclusões

Os resultados desse trabalho resulta na caracterização de quatro enzimas pertencentes a família 5 das hidrolases glicosídicas com propriedades funcionais (celulase e mananase) e estruturais (termo-estabilidade) de interesse para tecnologias relacionadas a degradação de biomassa vegetal.

A BsCel5A, celulase 5A de *Bacillus subtilis*, teve suas propriedades funcionais e estruturais dissecadas, revelando um novo mecanismo de termo-estabilização baseado no íon manganês. Os ensaios bioquímicos mostraram que o manganês prolonga o tempo de vida da enzima em altas temperaturas, sem interferir nas condições ótimas da BsCel5A que é desejável para aplicações industriais. É importante ressaltar que tal íon localiza-se a 30 Å do sítio catalítico, o que permite concluir que não há qualquer intervenção direta do mesmo com os resíduos catalíticos. Dados cristalográficos mostraram que o manganês é hexacoordenado pelos resíduos Asp195, Asp197, Asn198 e Gly157, juntamente com mais duas moléculas de água. Além disso, dados estruturais por RMN mostraram a base estrutural da baixa estabilidade do CBM3 que não possui o clássico sítio de ligação ao cálcio estrutural. Por fim, dados de SAXS demonstraram a ampla e plana superfície formada pelos domínios da BsCel5A para acomodar seu substrato, e que subsidia os estudos de cinética.

Outras duas celulasas inéditas pertencentes a família GH5 isoladas de fontes adaptadas a degradação de parede celular vegetal foram investigadas. A XacCel5, celulase 5 da bactéria *Xanthomonas axonopodis* pv. *citri*, que apresentou uma termo-estabilidade aumentada na presença do íon manganês assim como a BsCel5A. Entretanto, o sítio de ligação ao manganês não é conservado e estudos cristalográficos estão em andamento para revelar a base molecular desta termo-estabilização. A outra celulase, CelE1, oriunda de metagenômica do solo de cana-de-açúcar, teve sua estrutura cristalina determinada indicando uma alta conservação estrutural com apenas deleções e inserções em laços interfaciais. Estudos funcionais estão sendo realizados para avaliação do seu potencial biotecnológico.

A TpMan, β -mananase da bactéria hipertermofílica *Thermotoga petrophila*, apresenta arquitetura modular com um domínio catalítico pertencente a família GH5, assim como a BsCel5A. Porém, neste caso o domínio acessório CBM27 mostrou um efeito termo-estabilizante em contraste com a BsCel5A, indicando que evolutivamente em organismos hipertermofílicos os domínios acessórios tem um papel chave na termoestabilidade. Devido a

baixa identidade sequencial a estrutura da TpMan foi determinada pela técnica SIRAS com derivado de iodo. Para compreender a interação enzima-substrato e a preferência da TpMan por glicomanana ao invés de manana linear, estudos estruturais foram realizados na presença de dois ligantes: glicose e maltose. Os subsítios foram assinalados, e resíduos essenciais para o reconhecimento da glicose foram determinados, explicando a maior eficiência na clivagem do substrato glicomanana.

VI. Referencias bibliográficas

- Adler AJ, Greenfield NJ, Fasman GD. (1973) Circular dichroism and optical rotatory dispersion of proteins and polypeptides. **Methods in Enzymology**. vol. 27, p. 675-735.
- Agger J, Vikso-Nielsen A, Meyer AS. (2010) Enzymatic xylose release from pretreated corn bran arabinoxylan: differential effects of deacetylation and deferuloylation on insoluble and soluble substrate fractions. **Journal of Agricultural and Food Chemistry**. vol. 58, p. 6141–6148.
- Arantes V, Saddler JN. (2010) Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. **Biotechnology for Biofuels**. vol. 3, p. 4-15.
- Astua-Monge GFAJ, Bacocina G, Roncoletta J, Carvalho SA, Machado MA. (2005) Expression profiling of virulence and pathogenicity genes of *Xanthomonas axonopodis* pv. *citri*. **Journal of Bacteriology**. vol. 187, p. 1201-1205.
- Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AJ, Milman JD, Offord RE, Priddle JD, Waley SG. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. **Nature**. vol. 255, p. 609–614. Genet Mol Biol. 2010 Jan-Mar; 33(1): 146–153.
- Baptista JC, Machado MA, Homem RA, Torres OS, Vojnov AA, Amaral AM. (2010) Mutation in the *xpsD* gene of *Xanthomonas axonopodis* pv. *citri* affects cellulose degradation and virulence. **Genetics and Molecular Biology**. vol.33, p.146-153.
- Barbosa-Mendes JM, Mourão-Filho AAF, Bergamin-Filho A, Harakava R, Beer VS, Mendes BMJ. (2009) Genetic transformation of *Citrus sinensis* cv. Hamlin with *hrpN* gene from *Erwinia amylovora* and evaluation of the transgenic lines for resistance to citrus canker. **Scientia Horticulturae**. vol. 122, p. 109–115.
- Bergford TM. (1999) Protein crystallization: techniques, strategies and tips – a laboratory manual. La Jolla: UIL Biotechnology Series.
- Bhat MK. (2000) Cellulases and related enzymes in biotechnology. **Biotechnology Advances**. vol. 18, p.355-383.
- Blanco P, Sieiro C, Diaz A, Villa TG. (1994) Production and partial characterization of an endopolygalacturonase from *Saccharomyces cerevisiae*. **Canadian Journal of Microbiology**. vol. 40, p. 974–977.
- Boistelle R, Astier JP. (1988) Crystallization mechanisms in solution. **Journal of Crystal Growth**. vol. 90, p. 14–30.
- Bolam DN, Ciruela A, McQueen-Mason S, Simpson P, Williamson MP, Rixon JE, Boraston A, Hazlewood GP, Gilbert HJ. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. **Biochemical Journal**. vol. 331, p. 775-781.
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. **Biochemical Journal**. vol. 382, p. 769-781.

- Boraston AB, Kwan E, Chiu P, Warren RA, Kilburn DG. (2003) Recognition and hydrolysis of noncrystalline cellulose. **The Journal of Biological Chemistry**. vol. 278, p. 6120-6127.
- Bourgault R, Oakley AJ, Bewley JD, Wilce MC. (2005) Three-dimensional structure of (1,4)-beta-D-mannan mannanohydrolase from tomato fruit. **Protein Science**. vol. 14, p. 1233-1241.
- Brown K. (2001) Florida fights to stop citrus canker. **Science**. vol. 292, p. 2275-2278.
- Brunings AM, Gabriel DW. (2003) *Xanthomonas citri*: breaking the surface. **Molecular Plant Pathology**. vol. 4, p. 141-157.
- Caines ME, Vaughan MD, Tarling CA, Hancock SM, Warren RA, Withers SG, Strynadka NC. (2007). Structural and mechanistic analyses of endoglycoceramidase II, a membrane-associated family 5 glycosidase in the Apo and GM3 ganglioside-bound forms. **The Journal of Biological Chemistry**. vol. 282, p. 14300-14308.
- Canilha L, Milagres AMF, Silva SS, Almeida e Silva JB, Felipe MGA, Rocha GJM, Ferraz A, Carvalho W. (2009) Sacarificação da biomassa lignocelulósica através de pré-hidrólise ácida seguida por hidrólise enzimática: uma estratégia de “deconstrução” da fibra vegetal. **Revista Analytica**. vol. 44, p. 48-54.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. **Nucleic Acids Research**. vol. 37, p. 233-238.
- Castro AM, Pereira NJr. (2010) Produção, propriedades e aplicação de celulasas na hidrólise de resíduos agroindustriais. **Química Nova**. vol. 33, p. 181-188.
- Chapon V, Czjzek M, El Hassouni M, Py B, Juy M, Barras F. (2001) Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase Cel5 (formerly EGZ) from *Erwinia chrysanthemi*. **Journal of Molecular Biology**. vol. 310, p. 1055-1066.
- Coughlan MP, McHale A. (1988) Purification of the P-D-glucoside glucohydrolases of *Talaromyces emersonii*. **Methods in Enzymology**. vol. 160, p. 437 - 443.
- Cutfield SM, Davies GJ, Murshudov G, Anderson BF, Moody PC, Sullivan PA, Cutfield JF. (1999) The structure of the exo- β -(1,3)-glucanase from *Candida albicans* in native and bound forms: relationship between a pocket and groove in family 5 glycosyl hydrolases. **Journal of Molecular Biology**. vol. 294, p. 771-783.
- da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM, do Amaral AM, Bertolini MC, Camargo LE, Camarotte G, Cannavan F, Cardozo J, Chambergio F, Ciapina LP, Cicarelli RM, Coutinho LL, Cursino-Santos JR, El-Dorri H, Faria JB, Ferreira AJ, Ferreira RC, Ferro MI, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EG, Lemos MV, Locali EC, Machado MA, Madeira AM, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CF, Miyaki CY, Moon DH, Moreira LM, Novo MT, Okura VK, Oliveira MC, Oliveira VR, Pereira HA, Rossi A, Sena JA, Silva C, de Souza RF, Spinola LA, Takita MA, Tamura RE, Teixeira EC, Tezza RI, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC, Kitajima JP. (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. **Nature**. vol. 417, p. 459-463.

- Dauter Z, Dauter M, Rajashankar KR. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. **Acta crystallographica Section D, Biological Crystallography**. vol. 56, p. 232–237.
- Davies GJ, Dauter M, Brzozowski AM, Bjørnvad ME, Andersen KV, Schülein M. (1998) Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. **Biochemistry**. vol. 37, p. 1926–1932.
- Davies GJ, Henrissat B. (1995) Structures and mechanisms of glycosyl hydrolases. **Structure**. vol. 3, p. 853–859.
- Davies GJ, Wilson KS, Henrissat B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. **Biochemical Journal**. vol. 321, p. 557–559.
- Dawson L (2005). Optimization of pretreatment conditions for high efficiency ethanol production from post-harvest sugarcane residue. M.S. Thesis submitted to Nicholls States University, Thibodaux, LA.
- Ding SY, Himmel ME. (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. **Journal of Agriculture and Food Chemistry**. vol. 54, p. 597–606.
- Dodd D, Moon YH, Swaminathan K, Mackie RI, Cann IK. (2010) Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic *Bacteroidetes*. **The Journal of Biological Chemistry**. vol. 285, p. 30261–30273.
- Dominguez R, Souchon H, Lascombe M, Alzari PM. (1996) The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. **Journal of Molecular Biology**. vol. 257, p. 1042–1051.
- Dominguez R, Souchon H, Spinelli S, Dauter ZN, Wilson KS, Chauvaux S, Beâguin P, Alzari PM. (1995) A common fold and similar active site in two distinct families of β -glycanases. **Nature Structural & Molecular Biology**. vol. 2, p. 569–576.
- Dong A, Xu X, Edwards AM; Midwest Center for Structural Genomics; Structural Genomics Consortium, Chang C, Chruszcz M, Cuff M, Cymborowski M, Di Leo R, Egorova O, Evdokimova E, Filippova E, Gu J, Guthrie J, Ignatchenko A, Joachimiak A, Klostermann N, Kim Y, Korniyenko Y, Minor W, Que Q, Savchenko A, Skarina T, Tan K, Yakunin A, Yee A, Yim V, Zhang R, Zheng H, Akutsu M, Arrowsmith C, Avvakumov GV, Bochkarev A, Dahlgren LG, Dhe-Paganon S, Dimov S, Dombrowski L, Finerty P Jr, Flodin S, Flores A, Gräslund S, Hammerström M, Herman MD, Hong BS, Hui R, Johansson I, Liu Y, Nilsson M, Nedyalkova L, Nordlund P, Nyman T, Min J, Ouyang H, Park HW, Qi C, Rabeh W, Shen L, Shen Y, Sukumard D, Tempel W, Tong Y, Tresagues L, Vedadi M, Walker JR, Weigelt J, Welin M, Wu H, Xiao T, Zeng H, Zhu H. (2007) In situ proteolysis for protein crystallization and structure determination. **Nature Methods**. vol. 12, p. 1019–1021.
- Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich JP, Davies GJ, Haser R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. **Structure**. vol. 15, p. 939–949.
- Fardeau ML, Ollivier B, Patel BKC, Magot M, Thomas P, Rimbault A, Rocchiccioli F, Garcia JL. (1997). *Thermotoga hypogea* sp. nov., a xylanolytic, thermophilic bacterium from an oil-producing well. **International Journal of Systematic and Evolutionary Microbiology**. vol. 47, p. 1013–1019.

- Feigin LA, Svergun DI. (1987) **Structure Analysis by Small-Angle X-Ray and Neutron Scattering**. Plenum Press, New York. pp335.
- Ferreira LMA, Durrant AJ, Hall J, Hazlewood GP, Gilbert HJ. (1990) Spatial separation of protein domains is not necessary for catalytic or substrate binding in a xylanase. **Biochemical Journal**. vol. 269, p. 261-264.
- Flander F, Rouau X, Morel MH, Autio K, Seppanen-Laakso T, Kruus k, Buchert J. (2008) Effects of laccase and xylanase on the chemical and rheological properties of oat and wheat doughs. **Journal of Agriculture and Food Chemistry**. vol. 56, p. 5732-5742.
- Forouhar F, Hussain M, Farid R, Benach J, Abashidze M, Edstrom WC, Vorobiev SM, Xiao R, Acton TB, Fu Z, Kim JJ, Miziorko HM, Montelione GT, Hunt JF. (2006) Crystal structures of two bacterial 3-hydroxy-3-methylglutaryl-CoA lyases suggest a common catalytic mechanism among a family of TIM barrel metalloenzymes cleaving carbon-carbon bonds. **The Journal of Biological Chemistry**. vol. 281, p. 7533-7545.
- Fukumori F, Kudo T, Narahashi Y, Horikoshi K. (1986). Molecular cloning and nucleotide sequence of the alkaline cellulase gene from the alkalophilic *Bacillus* sp. strain 1139. **Journal of general microbiology**. vol. 132, p. 2329 –2335.
- Garman EF, Schneider TR. (1997) Macromolecular cryocrystallography. **Journal of Applied Crystallography**. vol. 30, p. 211-237.
- Gardner KH, Blackwell J. (1974) The structure of native cellulose. **Biopolymers**. vol. 13, p. 1975-2001.
- Gil MH, Ferreira P. (2006) Polissacarídeos como biomateriais. **Química 100**. vol. 10, p. 72-74.
- Gilbert HJ, Hall J, Hazlewood GP, Ferreira LMA. (1990) The N-terminal region of an endoglucanase from *Pseudomonas fluorescens* subspecies *cellulosa* constitutes a cellulosebinding domain that is distinct from the catalytic centre. **Molecular Microbiology**. vol. 4, p. 759-767.
- Gilliland GL. (1988) A biological macromolecule crystallization database: A basis for a crystallization strategy. **Journal of Crystal Growth**. vol. 90, p. 51–59.
- Glatter O, Kratky O. (1982) **Small angle x-ray scattering**. Academic Press Pages, Londres. 515pp.
- Golburg WI. (1994) Dynamic Light Scattering. **American Journal of Physics**. vol. 67, p. 1152-1164.
- Greenfield NJ. (2007) Using circular dichroism spectra to estimate protein secondary structure. **Nature Protocols**. vol. 1, p. 2876-2890.
- Guinier A, Fournet G. (1955) Small-Angle Scattering of X-Rays, John Wiley and Sons, New York, 1955.
- Hammersley A. (Copyright 1987-2005), Program FIT2D v.12.077, ESRF, Grenoble, France, <http://www.esrf.eu/computing/scientific/FIT2D/>.
- Han SJ, Yoo YJ, Kang HS. (1995) Characterization of a bifunctional cellulase and its structural gene. The cell gene of *Bacillus* sp. D04 has exo- and endoglucanase activity. **The Journal of Biological Chemistry**. vol. 270, p. 26012-26019.

- Han Y, Dodd D, Hespen CW, Ohene-Adjei S, Schroeder CM, Mackie RI, Cann IK. (2010) Comparative analyses of two thermophilic enzymes exhibiting both β -1,4 mannosidic and β -1,4 glucosidic cleavage activities from *Caldanaerobius polysaccharolyticus*. **Journal of Bacteriology**. vol. 192, p. 4111-4121.
- Haas W, MacColl R, Banas JA. (1998) Circular dichroism analysis of the glucan binding domain of *Streptococcus mutans* glucan binding protein-A. **Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology**. vol. 1384, p. 112-120.
- Henrissat B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemical Journal**. vol. 280, p. 309-316.
- Henrissat B, Bairoch A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemical Journal**. vol. 293, p. 781-788.
- Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP and Davies G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. **PNAS, Proceedings of the National Academy of Sciences**. vol. 92, p. 7090-7094.
- Henrissat B, Davies G. (1997) Structural and sequence-based classification of glycoside hydrolases. **Current Opinion in Structural Biology**. vol. 7, p. 637-644.
- Hidaka M, Nishimoto M, Kitaoka M, Wakagi T, Shoun H, Fushinobu S. (2009) The crystal structure of galacto-N-biose/lacto-N-biose I phosphorylase: a large deformation of a TIM barrel scaffold. **The Journal of Biological Chemistry**. vol. 284, p. 7273-7283.
- Hider RC, Kupryszewski G, Rekowski P, Lammek B. (1988) Origin of the positive 225-230 nm circular dichroism band in proteins: Its application to conformational analysis. **Biophysical Chemistry**. vol. 31, p. 45-51.
- Hilge M, Gloor SM, Rypniewski W, Sauer O, Heightman TD, Zimmermann W, Winterhalter K, Piontek K. (1998) High-resolution native and complex structures of thermostable beta-mannanase from *Thermomonospora fusca* - substrate specificity in glycosyl hydrolase family 5. **Structure**. vol. 6, p. 1433-1444.
- Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. **Science**. vol. 315, p. 804-807.
- Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG (2012). Novel enzymes for the degradation of cellulose. **Biotechnology for Biofuels**. vol. 5, p. 45.
- Hughes AB, Rudge AJ. (1994) Deoxynojirimycin: synthesis and biological activity. **Journal of Natural Products**. vol. 11, p. 135-162.
- Iakiviak M, Mackie RI, Cann IK. (2011) Functional analyses of multiple lichenin-degrading enzymes from the rumen bacterium *Ruminococcus albus* 8. **Applied and Environmental Microbiology**. vol. 77, p. 7541-7550.
- Jalak J, Våljamäe P. (2010) Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. **Biotechnology and Bioengineering**. vol. 106, p. 871-883.
- Jeanthon C, Reysenbach AL, L'Haridon S, Gambacorta A, Pace NR, Glénat P, Prieur D. (1995) *Thermotoga subterranean* sp. nov., a new thermophilic bacterium isolated from a continental oil reservoir. **Archives of Microbiology**. vol. 164, p. 91-97.

- Jeffries TW. (1990) Biodegradation of lignin-carbohydrate complexes. **Biodegradation**. vol. 1, p. 163-176.
- Kheshgi HS, Prince RC, Marland G. (2000) The potential of biomass fuels in the context of global climate change. **Annual Review of Environment and Resources**. vol. 25, p. 199-244.
- Kolpak FJ, Blacwell J. (1976) Determination of the structure of cellulose II. **Macromolecules**. vol. 9, p. 273-278
- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. (2003) PRIMUS: a windows PCbased system for small-angle scattering data analysis. **Journal Applied Crystallography**. vol. 36, p. 1277-1282.
- Koshland DE (1953) Stereochemistry and the mechanism of enzymatic reactions. **Biological Reviews of the Cambridge Philosophical Society**. vol. 28, p. 416-438.
- Kozin MB, Svergun DI. (2001) Automated matching of high- and low-resolution structural models. **Journal Applied Crystallography**. vol. 34, p. 33-41.
- Kuhad RC, Singh A, Eriksson KE. (1996) Microorganisms and enzymes involved in the degradation of plant fiber cell walls. **Advances in Biochemical Engineering and Biotechnology**. vol. 57, p. 45-125.
- Laemmli UK. (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. **Nature**. vol. 227, p. 680-685.
- Lawoko M, Henriksson G, Gellerstedt G. (2006) Characterisation of lignin-carbohydrate complexes (LCCs) of spruce wood (*Picea abies* L.) isolated with two methods. **Holzforschung**. vol. 60, p. 156-61.
- Lesk AM, Brändén CI, Chothia C. (1989). Structural principles of α/β barrel proteins: the packing of the interior of the sheet. **Proteins: Structure, Function, and Genetics**. vol. 5, p. 139-148.
- Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. **Biotechnology for Biofuels**. vol. 6, p. 41.
- Lora JH, Wolfgang GG. (2002) Recent Industrial Applications of Lignin: A Sustainable Alternative to Nonrenewable Materials. **Journal of Polymers and the Environment**. vol. 10, p. 39-48.
- Lorenz P, Eck J. (2005) Metagenomics and industrial applications. **Nature Reviews Microbiology**. vol. 3, p. 510-516.
- Ly HD, Withers SG. (1999) Mutagenesis of glycosidases. **Biochemistry - Annual Review of Biochemistry**. vol. 68, p. 487-522.
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. (2002) Microbial cellulose utilization: fundamentals and biotechnology. **Microbiology and Molecular Biology Reviews**. vol. 66, p. 506-577.
- Mandelman D, Belaich A, Belaich JP, Aghajari N, Driguez H, Haser R. (2003) X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. **Journal of Bacteriology**. vol. 185, p. 4127-4135.

- Martinez AT, Speranza M, Ruiz-Duenas FJ, Ferreira P, Camarero S, Guillen F, Martinez MJ, Gutierrez A, del Rio JC. (2005) Biodegradation of lignocellulosics: microbial chemical, and enzymatic aspects of the fungal attack of lignin. **International Microbiology**. vol. 8, p. 195–204.
- Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J. (2006) Computer simulation studies of microcrystalline cellulose I β . **Carbohydrate Research**. vol. 341, p. 138–152.
- McCarter JD, Withers SG. (1994) Mechanisms of enzymatic glycoside hydrolysis. **Current Opinion in Structural Biology**. vol. 4, p. 885–892.
- Miller GL. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. **Analytical Chemistry**. vol. 31, p. 426-428.
- Mittal A, Katahira R, Himmel ME, Johnson DK. (2011) Effects of alkaline or liquid-ammonia treatment on crystalline cellulose: changes in crystalline structure and effects on enzymatic digestibility. **Biotechnology for Biofuels**. vol. 4, p. 41.
- Nelson DL, Cox MM. *Princípios de Bioquímica de Lehninger*. 5 edição. New York. W.H. Freeman and Company, 2011 - 1100p.
- Oh HW, Heo SY, Kim do Y, Park DS, Bae KS, Park HY. (2008) Biochemical characterization and sequence analysis of a xylanase produced by an exo-symbiotic bacterium of *Gryllotalpa orientalis*, *Cellulosimicrobium* sp. HY-12. **Antonie Van Leeuwenhoek**. vol. 93, p. 437-442.
- Otwinowski Z, Minor W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. **Methods Enzymol**. vol. 276, p. 307–326.
- Park IH, Chang J, Lee YS, Fang SJ, Choi YL. (2012) Gene cloning of endoglucanase Cel5A from cellulose-degrading *Paenibacillus xylanilyticus* KJ-03 and purification and characterization of the recombinant enzyme. **The Protein Journal**. vol. 31, p. 238-245.
- Privalov PL. (1979) Stability of proteins: small globular proteins. **Advances in Protein Chemistry**. vol. 33, p. 167-241.
- Ramos LP. (2003) The chemistry involved in the steam treatment of lignocellulosic materials. **Química Nova**. vol. 26, p. 863-871.
- Ravot G, Magot M, Fardeau ML, Patel BKC, Prensier G, Egan A, Garcia JL, Ollivier B. (1995) *Thermotoga elfii* sp. nov., a novel thermophilic bacterium from an African oil-producing well. **International Journal of Systematic and Evolutionary Microbiology**. vol. 45, p. 308-314.
- Repo H, Oeemig JS, Djupsjöbacka J, Iwai H, Heikinheimo P. (2012) A monomeric TIM-barrel structure from *Pyrococcus furiosus* is optimized for extreme temperatures. **Acta crystallographica Section D, Biological Crystallography**. vol. 68, p. 1479-1487.
- Rodrigues Neto J, Ribeiro JGB. (2002) Manual técnico de procedimentos do cancro cítrico. Brasília: MAPA/SDA/ DDIV. 66p.
- Roske Y, Sunna A, Pfeil W, Heinemann U. (2004) High-resolution crystal structures of *Caldicellulosiruptor* strain Rt8B.4 carbohydrate-binding module CBM27-1 and its complex with mannohexaose. **Journal of Molecular Biology**. vol. 340, p. 543-554.

- Sabini E, Schubert H, Murshudov G, Wilson KS, Siika-Aho M, Penttilä M. (2000) The three-dimensional structure of a *Trichoderma reesei* β -mannanase from glycoside hydrolase family 5. **Acta crystallographica Section D, Biological Crystallography**. vol. 56, p. 3-13.
- Sakamoto T, Taniguchi Y, Suzuki S, Ihara H, Kawasaki H. (2007) Characterization of *Fusarium oxysporum* β -1,6-galactanase, an enzyme that hydrolyzes larch wood arabinogalactan. **Applied and Environmental Microbiology**. vol. 73, p. 3109–3112.
- Sakon J, Adney WS, Himmel ME, Thomas SR, Karplus PA. (1996) Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. **Biochemistry**. vol.35, p.10648-10660.
- Sakon J, Irwin D, Wilson DB, Karplus PA. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. **Nature Structural e Molecular Biology**. vol. 4, p. 810-818.
- Santos CR, Paiva JH, Meza AN, Cota J, Alvarez TM, Ruller R, Prade RA, Squina FM, Murakami MT. (2012) Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- β -D-mannanase. **Journal of Structural Biology**. vol. 177, p. 469-476.
- b Santos CR, Paiva JH, Sforça ML, Neves JL, Navarro RZ, Cota J, Akao PK, Hoffmam ZB, Meza AN, Smetana JH, Nogueira ML, Polikarpov I, Xavier-Neto J, Squina FM, Ward RJ, Ruller R, Zeri AC, Murakami MT. (2012) Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. **Biochemical Journal**. vol. 441, p. 95-104.
- Schellman JA. (1997) Temperature, stability, and the hydrophobic interaction. **Biophysical Journal**. vol. 73, p. 2960-2964.
- Schneider TR, Sheldrick GM. (2002). Substructure Solution with SHELXD. **Acta crystallographica Section D, Biological Crystallography**. vol. 58, p. 1772–1779.
- Schubert TS, Sun X. (1996) Bacterial Citrus Canker. Plant Pathology Circular. 377 Florida Department of Agriculture and Consumer Services, Divisio of Plant Industry, May/June. (fifth revision January 2003).
- Sharma P, Gupta JK, Vadehra DV, Dube DK. (1990) Purification and properties of an endoglucanase from a *Bacillus* isolate. **Enzyme and Microbial Technology**. vol. 12, p. 132-137.
- Shaw A, Bott R, Vornrhein C, Bricogne G, Power S, Day AG. (2002) A novel combination of two classic catalytic schemes. **Journal Molecular Biology**. vol. 320, p. 303–309.
- Sheldrick GM. (2002) Macromolecular phasing with SHELXE. **Z Kristallogr**. vol. 217, p. 644–650.
- Singhania RR, Sukumaran RK, Patel AK, Larroche C, Pandey A. (2010) Advancement and comparative profiles in the production technologies using solid-state and submerged fermentation for microbial cellulases. **Enzyme and Micribial Technology**. vol. 46, p. 541-549.
- Sinnott ML. (1990) Catalytic mechanisms of enzymic glycosyl transfer. **Chemical Reviews**. vol. 90, p. 1171-1202.
- Søgaard M, Kadziola A, Haser R, Svensson B. (1993) Site-directed mutagenesis of histidine 93, aspartic acid 180, glutamic acid 205, histidine 290, and aspartic acid 291 at the active site

- and tryptophan 279 at the raw starch binding site in barley alfa-amylase 1. **The Journal of Biological Chemistry**. vol. 268, p. 22480–22484.
- Sonan GK, Receveur-Brechot V, Duez C, Aghajari N, Czjzek M, Gerday C. (2007) The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium. **Biochemical Journal**. vol. 407, p. 293–302.
- Sonenshein AL, Hoch JA, Losick R. *Bacillus subtilis* and Its Closest Relatives: From Genes to Cells. Washington, DC: Amer Society for Microbiology, 2002 - 629 p.
- Sulzenbacher G, Driguez H, Henrissat B, Schülein M, Davies GJ. (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. **Biochemistry**. vol. 35, p. 15280–15287.
- Svergun DI (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **Journal Applied Crystallography**. vol. 25, p. 495–503.
- Svergun DI. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. **Biophysical Journal**. vol. 76, p. 2879–2886.
- Swings JG, Civerolo EL. *Xanthomonas*. London: Chapman & Hall, 1993. 399p.
- Tailford LE, Ducros VM, Flint JE, Roberts SM, Morland C, Zechel DL, Smith N, Bjørnvad ME, Borchert TV, Wilson KS, Davies GJ, Gilbert HJ. (2009) Understanding how diverse beta-mannanases recognize heterogeneous substrates. **Biochemistry**. vol. 48, p. 7009–7018.
- Takahata Y, Nishijima M, Hoaki T, Maruyama T. (2001) *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. **International Journal of Systematic and Evolutionary Microbiology**. vol. 51, p. 1901–1909.
- Tanabe T, Morinaga K, Fukamizo T, Mitsutomi M. (2003) Novel chitosanase from *Streptomyces griseus* HUT 6037 with transglycosylation activity. **Bioscience, Biotechnology, and Biochemistry**. vol. 67, p. 354–364.
- Teeri TT. (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. **Trends in Biotechnology**. vol. 15, p. 160–167.
- Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. (1996) Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. **Nature Structural & Molecular Biology**. vol. 3, p. 638–648.
- Tomme PH, Warren RA, Gilkes NR. (1995) Cellulose hydrolysis by bacteria and fungi. **Advances Microbial Physiology**. vol. 37, p. 1–81.
- Toth IK, Bell KS, Holeva MC, Birch PRJ. (2003) Soft rot erwiniae: from genes to genomes. **Molecular Plant Pathology**. vol. 4, p. 17–30.
- Vagin A, Teplyakov A. (2010) Molecular replacement with MOLREP. **Acta Crystallographica Section D, Biological Crystallography**. vol. 66, p. 22–25.
- Varrot A, Frandsen TP, Driguez H, Davies GJ. (2002) Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate

- analogue, methyl cellobiosyl-4-thio-beta-cellobioside, at 1.9 Å. **Acta Crystallographica Section D, Biological Crystallography**. vol. 58, p. 2201-2204.
- Varrot A, Schulein M, Davies GJ. (2000) Insights into ligand-induced conformational change in Cel5A from *Bacillus agaradhaerens* revealed by catalytically active crystal form. **Journal Molecular Biology**. vol. 297, p. 819-828.
- Vedadi M, Arrowsmith CH, Allali-Hassani A, Senisterra G, Wasney GA. (2010) Biophysical characterization of recombinant proteins: a key to higher structural genomics success. **Journal of Structural Biology**. vol. 172, p. 107-119.
- Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H, Hui R, Marsden BD, Nordlund P, Sundstrom M, Weigelt J, Edwards AM. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. **PNAS - Proceedings of the National Academy of Sciences**. vol. 103, p. 15835-15840.
- Violot S, Aghajari N, Czjzek M, Feller G, Sonan GK, Gouet P, Gerday C, Haser R, Receveur-Bréchet V. (2005) Structure of a full length psychrophilic cellulase from *Pseudoalteromonas haloplanktis* revealed by X-ray diffraction and small angle X-ray scattering. **Journal of Molecular Biology**. vol. 348, p. 1211-1224.
- Voget S, Steele HL, Streit WR. (2006) Characterization of a metagenome-derived halotolerant cellulase. **Journal of Biotechnology**. vol. 126, p. 26-36.
- Vyas NK. (1991) Atomic features of protein carbohydrate interactions. **Current Opinion in Structural Biology**. vol. 1, p. 732-740.
- Wang Q, Graham RW, Trimbur D, Warren RAJ, Withers SG. (1994) Changing enzymatic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. **Journal of the American Chemical Society**. vol. 116, p. 11594-11595.
- Wang Q, Tull D, Meinke A, Gilkes NR, Warren RA, Aebersold R, Withers SG. (1993) Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-β-1,4-glucanase. **The Journal of Biological Chemistry**. vol. 268, p. 14096-14102.
- Wårdell K, Jakobsson A, Nilsson GE. (1993) Laser Doppler perfusion imaging by dynamic light scattering. **Biomedical Engineering**. vol. 40, p. 309-316.
- Watanabe K, Kasai Y. (2008) Emerging technologies to analyze natural attenuation and bioremediation. In: Microbial biodegradation: Genomics and molecular biology. Diaz, E. (ed.). Norfolk, UK Caister academic press, p. 298-318.
- Wilson DB. (2009) Cellulases and biofuels. **Current Opinion in Biotechnology**. vol. 20, p. 295-299.
- b Wilson DB (2009) Three microbial strategies for plant cell wall degradation. **Annals of New York Academy of Sciences**. vol. 1125, p. 289-297.
- Wyman CE, Spindler DD, Grohmann K. (1992) Simultaneous saccharification and fermentation of several lignocellulosic feedstocks to fuel ethanol. **Biomass & Bioenergy**. vol. 3, p. 301-307.
- Yaniv O, Halfon Y, Shimon LJ, Bayer EA, Lamed R, Frolow F. (2012) Structure of CBM3b of the major cellulosomal scaffoldin subunit ScaA from *Acetivibrio cellulolyticus*. **Acta**

Crystallographica Section F: Structural Biology and Crystallization Communications. vol. 68, p. 8-13.

Yoshida S, Park DS, Bae B, Mackie R, Cann IK, Nair SK. (2011) Structural and functional analyses of a glycoside hydrolase family 5 enzyme with an unexpected β -fucosidase activity. **Biochemistry.** vol. 50, p. 3369 –3375.

Zheng Y, Pan Z, Zhang R, Wang D. (2009) Enzymatic saccharization of dilute acid pretreated saline crops for fermentable sugar production. **Applied Energy.** vol. 86, p. 459-2465.

Anexo I

A cristalografia como ferramenta para elucidação de estruturas tridimensionais

As proteínas podem apresentar até quatro níveis estruturais. A estrutura primária que é a sequência de aminoácidos determinada geneticamente e específica para cada proteína. A estrutura secundária que descreve as estruturas regulares formadas por segmentos da cadeia polipeptídica, sendo determinadas principalmente por interações locais. As estruturas secundárias podem ser do tipo hélice α definida como a torção da cadeia ao redor de um eixo (figura A1a). Tal estrutura é mantida por ligações de hidrogênio entre a cadeia principal de um resíduo e do quarto resíduo subsequente, de forma que as interações apresentam-se paralelamente ao eixo da hélice. As cadeias laterais dos aminoácidos ficam projetadas para fora da hélice, pois não participam das ligações de hidrogênio. A fita β , outro tipo de estrutura secundária, também é mantida por ligações de hidrogênio, estabelecidas entre segmentos das cadeias polipeptídicas adjacentes, não necessariamente próximos de uma mesma cadeia. Ela exibe uma conformação mais estendida que a hélice, dispostas lado a lado, o que dá um aspecto de uma folha pregueada (figura A1b). As ligações de hidrogênio neste caso são perpendiculares ao eixo das cadeias e os grupos laterais dos aminoácidos ficam projetados para cima e para baixo do plano da folha β (Nelson e Cox, 2011).

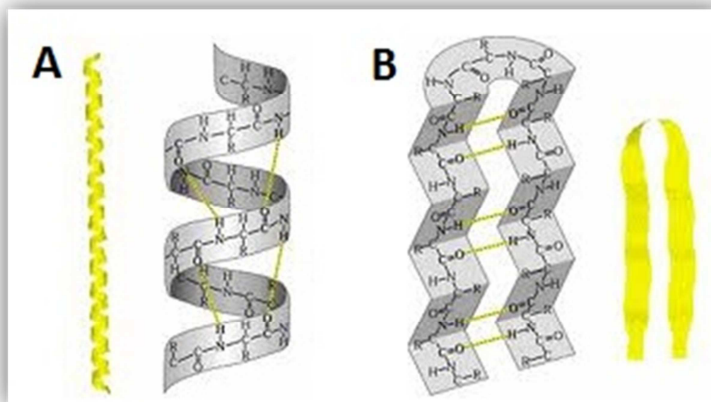


Figura A1: Representação das estruturas secundárias encontradas nas cadeias polipeptídicas. **A** Estrutura secundária do tipo hélice α . **B** Estrutura secundária do tipo folha β .

Como próximo nível estrutural, temos a estrutura terciária, o enovelamento da cadeia polipeptídica por interações de regiões com estrutura regular (hélice ou folha β) ou de regiões sem estrutura definida, permitindo a aproximação e a interação de regiões distantes na

sequência polipeptídica, através de ligações não-covalentes e covalentes, principalmente entre as cadeias laterais dos aminoácidos. Por fim, a estrutura quaternária descreve a associação de duas ou mais cadeias polipeptídicas, também chamadas de subunidades, formando uma proteína funcional oligomérica, mantida por ligações, em geral não-covalentes entre as subunidades, essas ligações não-covalentes são ligações de hidrogênio, interações hidrofóbicas, força de van der Waals e ligações iônicas (Berg et al., 2006; Nelson e Cox, 2011).

Para resolução da estrutura terciária, foco do trabalho, a técnica utilizada foi a cristalografia. A cristalografia de proteínas, por difração de raios X, é uma técnica bem estabelecida para a obtenção de informação de estruturas tridimensionais de macromoléculas biológicas (Kleywegt e Jones, 2002). A determinação do arranjo espacial dos átomos nas moléculas e no cristal é extremamente importante para o entendimento das propriedades químicas, físico-químicas e biológicas das macromoléculas.

A cristalização é o primeiro passo para a obtenção da estrutura por difração de raios X de monocristais. O cristal é um arranjo ordenado das moléculas de forma periódica, caracterizados por operações de simetria e parâmetros de rede que definem exatamente a disposição e periodicidade das unidades fundamentais (cela unitária) das quais são compostos. O cristal de proteína é mantido por interações eletrostáticas (ponte salinas), ligações de hidrogênio, interação de van der Waals e interações de dipolos nas regiões de contato (MacPherson, 1999).

A cristalização de uma macromolécula ocorre pela diminuição lenta de sua solubilidade. Em geral se este processo ocorrer rapidamente tem-se a precipitação desorganizada (MacPherson, 1999). O processo de cristalização pode ser dividido em três eventos: nucleação, crescimento e cessação do crescimento (figura A2). A nucleação ocorre numa região de supersaturação, aglomerados são simultaneamente formados e solubilizados até que consigam transpor uma barreira de energia de ativação e comecem a crescer, incorporando, de maneira ordenada moléculas de proteína do meio. Assim, o sistema antes supersaturado passa a um estado insaturado (mais estável) e, quando a incorporação de moléculas é igual à liberação das moléculas do cristal, o crescimento do mesmo cessa (MacPherson, 1999).

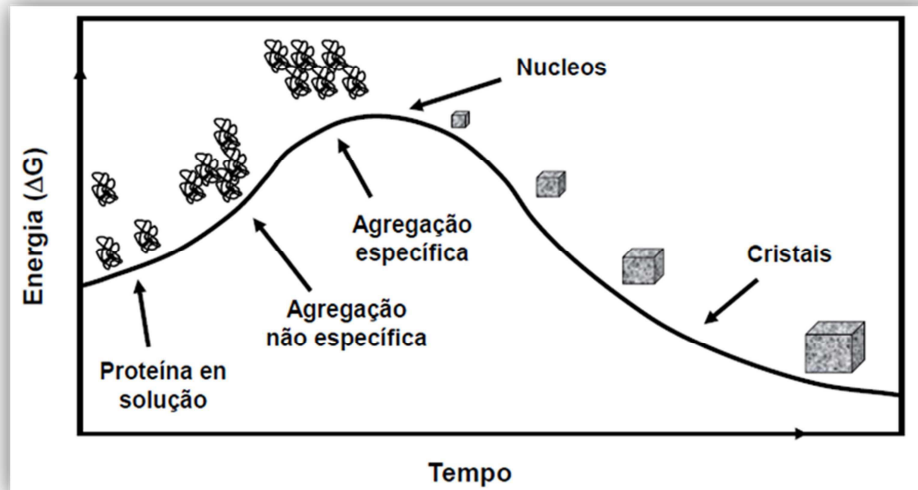


Figura A2: Diagrama da termodinâmica durante o crescimento de um cristal.

Existem diversos métodos para a obtenção de cristais de proteínas, todos eles têm o objetivo de trazer a solução de macromoléculas a um estado de “sólido”. Dentre os métodos, podemos citar: método em *batch*; o método por diálise; e o método por difusão de vapor - este último foi a técnica empregada nos estudos aqui descritos.

Na cristalização por difusão de vapor é necessário elevar a concentração de precipitante na gota, até que o ponto de supersaturação de proteína seja atingido. Na medida em que o volume da solução de cristalização diminui pela perda de água por vapor, as concentrações da proteína e do agente precipitante aumentam. Uma vez que a concentração de precipitante do poço é maior que a concentração da gota, e o volume da solução no poço é muito maior que o volume da gota, esta tende a perder solvente, ou seja, troca de vapor ocorrerá até que o equilíbrio do sistema seja atingido, e as concentrações de precipitante sejam iguais no poço e na gota.

Frequentemente, um número extenso de condições são empiricamente testadas para que se obtenha cristais. Existem inúmeros parâmetros que influenciam a formação dos cristais incluindo supersaturação (concentração da proteína e do precipitante); pureza da amostra; estabilidade conformacional da proteína; nucleação homogênea; temperatura e pH; força iônica e pureza dos reagentes; associação e ligantes (substratos, cofatores, íons metálicos); entre outros (Ducruix e Giegé, 1999). Além disso, o evento é multi-paramétrico acoplado, significando que comumente não é possível alterar um parâmetro sem a variação de outros

correlacionados. Essa propriedade dificulta a racionalização do processo e mantém à cristalização de macromoléculas ainda um experimento essencialmente empírico.

Após obter os cristais de proteína, o próximo passo é expô-los ao feixe de raios X monocromático, pois para distinguir átomos individuais, o comprimento de onda da radiação utilizada deve ser próximo às distancias que se pretende medir. A difração de raios X pelos cristais foi descoberta em 1911 pelo físico alemão Max von Laue (Faria, 1990). Mas foi após os anos 60, que ocorreu a primeira resolução da estrutura cristalográfica de uma proteína, a mioglobina, realizada por Kendrew e Perutz (1963).

Atualmente, de forma geral, há duas formas de gerar raios X, através de uma fonte convencional (de tubo selado ou ânodo rotatório) ou com um acelerador de elétrons síncrotron. A primeira produz raios X de média intensidade e comprimento de onda fixo. Esse comprimento de onda varia com o material de que o ânodo é feito. Já o síncrotron produz uma luz policromática de alto brilho, que ainda pode ser intensificada por dispositivos de inserção, e selecionada por elementos óticos específicos para monocromatização, colimação e focalização do feixe de raios X.

O uso do feixe de raios X de maior brilho possibilita registrar os dados de difração com maior rapidez e resolução. A alta intensidade amplifica os sinais de difração dos cristais de proteínas, que é essencial para solucionar a estrutura, principalmente aqueles que produzem reflexões fracas (cristais pequenos) e pouco espaçadas (com celas unitárias grandes) – uma vez que a distância entre as reflexões no padrão de difração é inversamente proporcional ao tamanho da cela unitária.

A coleta de dados é uma etapa importante na resolução de uma estrutura cristalográfica. A estratégia é vinculada ao tipo de rede cristalina e ao método escolhido para resolução do problema das fases. Quando a estrutura não é solucionada por substituição molecular, a acuidade e a precisão na coleta dos dados são essenciais, já que a obtenção das fases será realizada a partir de pequenas diferenças nos fatores de estrutura.

Durante a coleta dos dados de difração, a posição de cada átomo da estrutura do cristal influencia na intensidade de todas as reflexões, e de um modo oposto, a intensidade de cada reflexão, depende das posições de todos os átomos da célula unitária (Wlodawer et al., 2001). De acordo com a lei Bragg ou sua interpretação geométrica dada pela construção de Ewald, somente as reflexões provenientes de planos na rede cristalina de forma dependente

do ângulo de rotação serão medidas e neste experimento a informação da fase é perdida (Taylor, 2010).

A recuperação das fases da onda espalhada pelo cristal, ou seja, obter informações sobre as fases relativas com as quais os elétrons dos átomos, no interior do cristal espalham a radiação, e assim ter a resolução da estrutura, é o problema central da cristalografia (Taylor, 2003). Para solucioná-lo podem ser utilizados métodos diretos, o método da substituição molecular e métodos com a utilização de átomos pesados (substituição isomorfa e/ou espalhamento anômalo). Aqui, serão discutidos o método de substituição molecular e o de átomos pesados, que foram os utilizados para resolução das estruturas aqui apresentadas.

A substituição molecular, método primeiramente descrito por Michel Rossamann e David Blow (1962), é uma forma de se obter a fase para os fatores de estrutura através de um modelo de busca homólogo. A ideia é construir um modelo e então simular a sua difração e comparar os seus fatores de estrutura com os fatores coletados experimentalmente. Este método exige que haja um conhecimento prévio de uma estrutura tridimensional de uma proteína, que servirá de modelo, tal molde tem que conter pelo menos 30% de sua sequência de aminoácidos idêntica a da proteína alvo, junto com um desvio do RMSD (*Root Mean Square Deviation*) menor que 2.0 Å entre os átomos C α do modelo e da nova estrutura (Taylor, 2010).

O modelo prévio permite atribuir o valor das fases iniciais para os dados de difração coletados, assumindo que as sequências similares tenham enovelamento geral também conservado. Dessa maneira, as posições dos átomos do modelo devem estar relativamente próximas à posição dos átomos da proteína que pretende-se determinar a estrutura. Esse método vem se tornando mais comum à medida que os bancos de dados de estruturas resolvidas crescem, o que aumenta as chances de encontrar um modelo de busca adequado (Giorgetti et al., 2005).

A substituição molecular consiste, basicamente, no processo de orientar (através de um operador de rotação) e posicionar (por de um operador de translação) o modelo de busca no sistema cristalino alvo no espaço recíproco pela função de Patterson (Lebedev et al., 2008; Isupov e Lebedev, 2007). Nos casos em que há um modelo de busca adequado, as operações de rotação e translação serão encontradas com facilidade e as fases calculadas estarão próximas das fases verdadeiras. Em caso de baixa identidade, o mapa obtido pode ser não-interpretável, mesmo que as soluções sejam encontradas, devendo-se então recorrer a métodos experimentais para a solução do problema das fases, como a substituição isomorfa com espalhamento anômalo (Evans e McCoy, 2008).

A resolução do problema das fases pelos métodos de espalhamento anômalo se baseia na absorção dos raios X de elementos específicos incorporados na rede cristalina. Neste caso, deve existir alguma espécie no cristal que absorva raios X em um comprimento de onda apropriado, como alguns átomos de metal pesado, com um elevado número de elétrons. Esses átomos são introduzidos em cada macromolécula seja durante expressão em meio mínimo suplementado com aminoácidos modificados (selênio-metionina) ou após a formação do cristal pela técnica de *soaking*. Os metais pesados mais comumente usados são Pt, Hg, U, Au e I. Esses compostos frequentemente se ligam em sítios específicos na superfície da proteína sem causar mudanças conformacionais ou alterar o retículo cristalino, formando complexos isomorfos. Essa adição de átomos pesados, na estrutura causa mudanças significativas nas intensidades do padrão de difração, as quais são então usadas para obter uma estimativa das fases para cada reflexão, e esses dados são usados como ponto de partida para a construção do modelo.

A introdução de um átomo pesado muda significativamente a intensidade do espalhamento. Os átomos pesados contribuem para a intensidade do espalhamento na proporção do quadrado do número de elétrons que eles contêm. Por exemplo, um átomo de urânio contém 15 vezes mais elétrons do que um átomo de carbono, e então a sua contribuição para a intensidade será equivalente aquela de 225 átomos de carbono. Como resultado, a mudança na intensidade provocada pela adição de um átomo pesado à uma proteína pode ser facilmente medida (Giacovazzo, 2002).

São raros os dados que permitem a recuperação das fases com tal precisão que admitem interpretar todo o mapa da densidade. As fases experimentais são apenas o ponto de partida. Para obter um melhor resultado utilizam-se métodos de modificação da densidade, que são baseados no conhecimento, mesmo que mínimo, da estrutura em questão. *Solvent flattening*, *solvent flipping* e *NCS averaging* são as técnicas que podem ser utilizadas para obter uma melhora da densidade eletrônica, e das fases.

O processo de refinamento geralmente envolve ciclos de otimização intercalados de algoritmos computacionais que utilizam restrições estéreo-químicas e informações estruturais elementares para o refinamento e correções manuais que melhoram a concordância dos mapas da densidade eletrônica com as coordenadas atômicas (Wlodawer et al., 2008). Alguns indicadores estatísticos de qualidade precisam ser analisados para que a estrutura seja validada, como R-factor e R_{free} , RMSDs de distâncias e ângulos de ligações, e a estereoquímica por meio do gráfico de Ramachandran.

O R-factor, normalmente expresso em porcentagem, mede a discrepância global relativa entre o fator de amplitude da estrutura experimental, F_{obs} , e o fator de amplitude da estrutura obtido no modelo, F_{calc} . O R-factor combina os erros intrínsecos dos dados experimentais e o desvio do modelo do real. Um bom refinamento mostra um R-factor em torno de 20% dependendo da resolução dos dados e o nível de interpretação do mapa de densidade eletrônica (Wlodawer et al., 2008).

O R_{free} , é calculado de forma análoga ao R-factor, porém não é influenciado pelo refinamento. Para a determinação do R_{free} usa-se de 5 a 10% reflexões únicas escolhidas de forma randômica, e tais reflexões são excluídas do conjunto de dados que são refinados, desta forma o R_{free} é calculado a partir dessa fração do conjunto de dados, servindo como um indicador que a estrutura modelada e refinada é compatível ou não com os dados coletados a partir do cristal (Kleywegt, 2007). Os valores médios para o R_{free} tende a ser próximos a 5 a 10% maior que o R-factor. Quando a discrepância é maior, pode estar indicando um *over-fitting* dos dados experimentais, o que pode resultar em modelo não compatível com os dados experimentais, ou seja, incorreto. Um caso comum de *over-fitting* é a adição em excesso de moléculas de água, que faz com que o R-factor seja reduzido, mas não o R_{free} (Wlodawer et al., 2008).

O RMSD (*root-mean-square deviation*) dos comprimentos e ângulos de ligações é o índice que verifica os padrões geométricos típicos das ligações químicas presentes na proteína. Os valores do RMSD esperados para o comprimento estão em torno de 0,02 Å, mas algumas variações podem ser encontradas, já que os cálculos são realizados através das médias entre as ligações de estruturas disponíveis em banco de dados (Jaskolski et al., 2007; Wlodawer et al., 2008).

O gráfico de Ramachandran fornece um mapa dos ângulos de torção φ/ψ da cadeia principal polipeptídica (figura A3). Tais ângulos são utilizados na validação da estrutura, porque eles não são restringidos no refinamento (Wlodawer et al., 2008).

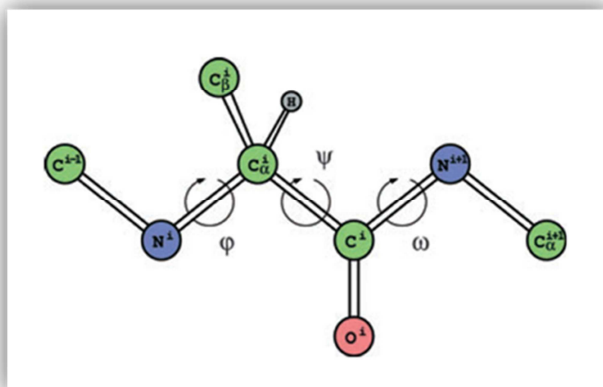


Figura A3: Representação esquemática de um fragmento da cadeia central da proteína, com a definição de ângulos de torção φ , ψ e ω (Wlodawer et al., 2007).

A cristalografia de macromoléculas biológicas pode ser uma importante ferramenta a ser utilizada: para descobertas de novos fármacos, sendo fundamental para observar as interações fármaco-receptor; estudos de otimização de enzimas, analisando quais são as regiões a serem racionalmente editadas molecularmente e as que são essenciais para a interação com o substrato; compreender correlações estrutura-função; resolução da estrutura de proteínas sabidamente relacionadas a distúrbios biológicos pode ser uma ferramenta inicial para entendimento global e pleno, observando as diferenças de uma molécula “sadia” e a que produz a doença. Pode-se afirmar que a cristalografia de macromoléculas sé uma ferramenta que as aplicações ainda estão longe de atingir sua saturação e continuará por muito tempo auxiliando no entendimento de eventos biológicos ao nível atômico de forma complementar a outras técnicas como ressonância magnética nuclear, espalhamento de raios X a baixos ângulos e microscopia eletrônica.

Referências bibliográficas

Ducruix A, Giegé R. Crystallization of Nucleic Acids and Proteins. A practical approach. 2 edição. New York. Oxford University Press, Oxford, 1999 – 464p.

Evans P, McCoy A. (2008) An introduction to molecular replacement. **Acta Crystallographica Section D: Biological Crystallography**. vol. 64, p. 1-10.

Faria JL. Historical Atlas of Crystallography. Ed Lima de Faria e MJ Buerger. Kluwer Academic Publishers, Springer, 1990 – 156p.

Giacovazzo C, Monaco HL, Artioli G, Viterbo D, Milanesio M, Gilli G, Gilli P, Zanotti G, Ferraris G, Fundamentals of Crystallography. International Union of Crystallography, Oxford University Press, Oxford, 2002 – 825p.

Giorgetti A, Raimondo D, Miele A, Tramontano A. (2005) Evaluating the usefulness of protein structure models for molecular replacement. **Bioinformatics**. vol. 21, p. 72-76.

Isupov MN, Lebedev AA. (2007) NCS-constrained exhaustive search using oligomeric models. **Acta Crystallographica Section D: Biological Crystallography**. vol. 64, p. 90-98

Jaskolski M, Gilski M, Dauter Z, Wlodawer A. (2007) Stereochemical restraints revisited: how accurate are refinement targets and how much should protein structures be allowed to deviate from them? **Acta Crystallographica Section D: Biological Crystallography**. vol. 63, p. 611–620.

Kleywegt GJ. (2007) Separating model optimization and model validation in statistical cross-validation as applied to crystallography. **Acta Crystallographica Section D**. vol. 63, p. 939–940.

Kleywegt GJ, Jones TA. (2002) Homo crystallographicus – quo vadis? **Structure**. vol. 10, p. 465–472.

Krendrew J. (1963) Myoglobin and protein structure. **Biofizika**. vol. 8, p. 273-287.

Lebedev AA, Vagin AA, Murshudov GN. (2008) Model preparation in MOLREP and examples of model improvement using X-ray data. **Acta crystallographica Section D, Biological Crystallography**. vol. 64, p. 33-39.

MacPherson A. Crystallization of Biological Molecules. Cold Spring Harbor Laboratory Academic Press, 1999 – 586p.

Nelson DL, Cox MM. Princípios de Bioquímica de Lehninger. 5 edição. New York. W.H. Freeman and Company, 2011 - 1100p.

Rossmann MG, Blow DM. (1962) The detection of sub-units within the crystallographic asymmetric unit. **Acta Crystallographica**. vol. 15, p. 24-31.

Taylor GL. (2003) The phase problem. **Acta Crystallographica Section D: Biological Crystallography**. vol. 59, p. 1881-1890.

Taylor GL. (2010) Introduction to phasing. **Acta Crystallographica Section D: Biological Crystallography**. vol. 66, p. 325-338.

Wlodawer A, Li M, Dauter Z, Gustchina A, Uchida K, Oyama H, Dunn BM, Oda K. (2001) Carboxyl proteinase from *Pseudomonas* defines a novel family of subtilisin-like enzymes. **Nature Structural & Molecular Biology**. vol. 8, p. 442-446.

Wlodawer A, Minor W, Dauter Z, Jaskolski M. (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. **FEBS Journal**. vol. 275, p. 1-21.

Anexo 2 Artigos publicados

Dissecting structure–function–stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168.

Dissecting structure–function–stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168

Camila R. SANTOS^{*1}, Joice H. PAIVA^{*1}, Maurício L. SFORÇA^{*}, Jorge L. NEVES^{*}, Rodrigo Z. NAVARRO^{*}, Júnio COTA[†], Patrícia K. AKAO^{*}, Zaira B. HOFFMAM[‡], Andréia N. MEZA^{*}, Juliana H. SMETANA^{*}, Maria L. NOGUEIRA^{*}, Igor POLIKARPOV[‡], José XAVIER-NETO^{*}, Fábio M. SQUINA[‡], Richard J. WARD[§], Roberto RULLER[‡], Ana C. ZERI^{*} and Mário T. MURAKAMI^{*2}

^{*}Laboratório Nacional de Biotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas SP, Brazil, [†]Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas SP, Brazil, [‡]Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil, and [§]Departamento de Química, Universidade de São Paulo, Ribeirão Preto, Brazil

Cellulases participate in a number of biological events, such as plant cell wall remodelling, nematode parasitism and microbial carbon uptake. Their ability to depolymerize crystalline cellulose is of great biotechnological interest for environmentally compatible production of fuels from lignocellulosic biomass. However, industrial use of cellulases is somewhat limited by both their low catalytic efficiency and stability. In the present study, we conducted a detailed functional and structural characterization of the thermostable BsCel5A (*Bacillus subtilis* cellulase 5A), which consists of a GH5 (glycoside hydrolase 5) catalytic domain fused to a CBM3 (family 3 carbohydrate-binding module). NMR structural analysis revealed that the *Bacillus* CBM3 represents a new subfamily, which lacks the classical calcium-binding

motif, and variations in NMR frequencies in the presence of cellopentaose showed the importance of polar residues in the carbohydrate interaction. Together with the catalytic domain, the CBM3 forms a large planar surface for cellulose recognition, which conducts the substrate in a proper conformation to the active site and increases enzymatic efficiency. Notably, the manganese ion was demonstrated to have a hyper-stabilizing effect on BsCel5A, and by using deletion constructs and X-ray crystallography we determined that this effect maps to a negatively charged motif located at the opposite face of the catalytic site.

Key words: accessory domain, cellulase 5A, carbohydrate-binding module, kinetics, structure, thermal stability.

INTRODUCTION

The production of ethanol from lignocellulosic biomass is well-placed among other possibilities to produce energy, owing to its potential sustainability and agro-economic benefits [1]. Different chemical and enzymatic strategies have been proposed for the saccharification of lignocellulosic biomass [2,3]. The latter has been implemented at different organization levels, from point mutations of cellulolytic enzymes to the engineering of whole metabolic pathways in micro-organisms [4]. Furthermore, these two approaches have been employed synergistically to increase the yields of fermentable sugars from lignocellulosic biomass. Although enzymatic hydrolysis has been successfully implemented as a biomass-to-bioenergy technology, it is widely regarded as an expensive and wasteful link in the whole chain of bioethanol production, mainly because of the large amounts of enzymes required to compensate for their low catalytic efficiency and stability [5,6]. Driven by this need, a number of enzymes with biological and commercial value have been systematically modified through structure-based rational approaches, or by intelligent use of serendipity [7–9]. In spite of these efforts, the molecular basis of protein stability and its intricate correlation with catalysis is still elusive, putting a premium on studies that address structure–function–stability

relationships and the discovery of new alternative pathways to improve thermal stability of the enzymes.

Endo- β -1,4-glucanases (EC 3.2.1.4), also referred to as endoglucanases, are the major enzymes responsible for the breakdown of internal glycosidic bonds of cellulose chains. Several endoglucanases from different *Bacillus subtilis* strains (BsEgls) have been cloned and characterized, aiming at their potential applications in the biofuels industry [10–17]. BsEgls encompass a catalytic domain belonging to family 5A, containing a CBM3 (family 3 carbohydrate-binding module) appended to their C-terminus. BsEgls show maximum activity at approximately pH 6 and 60 °C [10,14] and are capable of hydrolysing CMC (carboxymethylcellulose) and lichenan, but not xylan, chitosan or laminarin [10,18]. Typically, BsEgls are thermostable enzymes, retaining 90 % of activity after incubation for 2 h at 65 °C, 70 % after 30 min at 75 °C, and 12 % after 10 min at 80 °C [11,13,14].

It has been proposed that the non-catalytic domain of BsEgls is important to bind insoluble substrates [13], increasing the catalytic efficiency by a mechanism that involves the disruption of the compact cellulose structure and delivering of the substrate to the CC (catalytic core) in a favoured configuration for catalysis [19–21]. As it is often the case with GHs (glycoside hydrolases), deletion of the accessory domain decreases the thermal stability

Abbreviations used: BsCel5A, *Bacillus subtilis* cellulase 5A; CBM, carbohydrate-binding module; BsEgl, *Bacillus subtilis* endoglucanase; CC, catalytic core; CMC, carboxymethylcellulose; CiCipACBM, *Clostridium thermocellum* CBM3; GH, glycoside hydrolase; HSQC, heteronuclear single-quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; PEG, poly(ethylene glycol); SAXS, small-angle X-ray scattering; T_m , melting temperature; TmCel5A, *Thermotoga maritima* Cel5A; WT, wild-type.

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed (email mario.murakami@lnb.org.br).

Structural factors and atomic co-ordinates of the *Bacillus subtilis* cellulase 5A catalytic core have been deposited in the Protein Data Bank under accession codes 3PZV (Form I), 3PZU (Form II) and 3PZT (Form II'). NMR data of the *Bacillus subtilis* cellulase 5A CBM3 were deposited in the Biological Magnetic Resonance Bank and Protein Data Bank under codes 17399 and 2L8A respectively.

of the catalytic domain [22–30]; however, in BsEgls it has a contrary effect, by increasing enzyme stability [13]. Interestingly, the thermal stability conferred by a CBM to its parental protein is not transferred when it is recombinantly fused to a new CC, suggesting a complex molecular modulation and co-evolution of both modules [25,31–34]. In addition, experimental and *in silico* studies regarding the influence of the linker region [35] and domain–domain interactions [26,29] in the thermal stability of GHs are inadequate. Thus, in spite of extensive functional and structural studies of CBMs, the consequences of the fusion of a given CBM to a specific GH are as yet unpredictable [22,24,36,37].

In the present study we have described in detail the role exerted by the accessory domain in the operation and stability of BsEgls. Furthermore, by studying the specific nature of BsEgl thermal stability, we uncovered a novel stabilization mechanism based on metal ion co-ordination. This in-depth understanding of the modulation of the function and stability of BsEgls at a molecular level is a required step towards its optimization and use in biotechnological processes.

MATERIALS AND METHODS

Cloning, expression and purification

Three constructions of BsCel5A (*B. subtilis* cellulase 5A) were amplified by PCR using *B. subtilis* 168 genomic DNA and specific oligonucleotides [5'-GCTAGCGCAGCAGGGACAAAACG-3' and 5'-GGATCCTTAGGTGCCGAGAATGTTTCTC-3' for CC; 5'-GCTAGCATTTCGTACAGTACAGAGCA-3' and 5'-GGA-TCCCTAATTTGGTCTGTTCCC-3' for CBM3; and 5'-GCT-AGCGCAGCAGGGACAAAACG-3' and 5'-GGATCCCTAA-TTGGTCTGTTCCC-3' for the WT (wild-type) protein]. The amplified sequences were cloned into pGEM-T Easy vector (Promega) and further subcloned into the pET28a expression vector using the restriction sites present in the oligonucleotides (underlined sequences). The proteins were expressed in BL21(DE3)*slyD*⁻ cells at 37°C for 4 h after induction with 0.5 mM IPTG (isopropyl β -D-thiogalactopyranoside) in Luria–Bertani broth. For NMR experiments, the BL21(DE3)*slyD*⁻ cells transformed with pET28a-CBM3 plasmid were grown in M9 minimal medium supplemented with 2 g/l [¹³C]glucose and 1 g/l [¹⁵N]ammonium chloride [38] using the same conditions described above. The harvested cells were resuspended in lysis buffer (50 mM sodium phosphate, pH 7.2, 100 mM NaCl, 1 mM PMSF and 5 mM benzimidazole) and disrupted by lysozyme treatment (80 μ g/ml, 30 min, on ice), followed by sonication (10 pulses of 10 s on ice using a Vibracell VCX 500, Sonics & Materials). The solutions were centrifuged at 10000 g for 30 min and the supernatants were loaded on to nickel-charged 5 ml HiTrap Chelating columns (GE Healthcare) using a flow rate of 1 ml/min. The proteins were eluted using a non-linear gradient of 0–500 mM imidazole. For WT construction, an anionic-exchange chromatography using a 5 ml HiTrap SP HP column (GE Healthcare), a flow rate of 1 ml/min and a non-linear gradient of 0–1 M NaCl was used to improve sample purity. All constructions were concentrated further using Ultrafree-CL (Millipore) and submitted to size-exclusion chromatography using a Superdex 75 16/60 column (GE Healthcare) and a flow rate of 0.5 ml/min. All chromatographic steps were performed using an ÄKTA FPLC system (GE Healthcare). Sample homogeneity was confirmed by SDS/PAGE under denaturing conditions [39]. Protein concentration was estimated by absorbance at 280 nm using molar extinction coefficients derived from the sequence of the constructs.

Enzymatic characterization and optimization of endo- β -1,4-glucanase activity

The standard enzymatic assays for evaluation of CC and WT constructs were performed in triplicate using different substrates at 0.25 % and 100 ng of enzyme in 50 mM citrate-phosphate buffer, pH 6.0. The mixture was incubated at 60°C for 60 min to determine activity in a set of 12 polysaccharides including β -glucan, CMC, larch arabinogalactan, laminarin, galactomannan, xyloglucan, debranched arabinan, 1,4- β -D-mannan, rye arabinoxylan, xylan from beechwood, pectin and lichenan (purchased from Megazyme and Sigma–Aldrich). The enzymatic activity was determined by the amount of reducing sugar liberated from different polysaccharides using the DNS (3,5-dinitrosalicylic acid) method [40]. One unit of endoglucanase activity was defined as the quantity of enzyme to release 1 μ mol of reducing sugars per min under standard conditions. A response surface methodology was performed to optimize the reaction conditions for both CC and WT constructs of BsCel5A in the absence or presence of manganese ion (10 mM). The variables analysed in the present study were the pH together with the temperature. A central composite design ($k = 2$) with four central points and a total of 12 experiments was considered for optimization of these variables. Details concerning the statistical approaches for these experiments have been described by Myers and Montgomery [41]. The regression and graphical analysis of the data were performed using the software Statistica™ version 8.0, considering a P value of 0.05 as significant. The significance of the regression coefficients was given by Student's t test. The second-order model equation was determined by Fisher's test and the multiple coefficient of determination (R^2) gave the variance explained by the model. All glucanase activity assays were carried out in triplicate using 10 μ l of enzyme (100 ng), 40 μ l of 0.1 M citrate-phosphate buffer and 50 μ l of β -glucan (stock 0.5 % in water). The enzymatic reaction was performed at different pH values in the citrate-phosphate buffer system and various temperatures as indicated in Supplementary Table S1 (at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Metal ions, chelating agent (EDTA), surfactants (SDS and Triton X-100) and organic solvents (methanol, ethanol and DMSO) were added to the enzyme assay to test their effects on activity. Control experiments using a citrate-free buffer were carried out, and the results were statistically identical with those obtained in the citrate buffer, indicating no chelating effect of the citrate over the manganese ion under the tested conditions. The kinetic study was performed with β -glucan as substrate at 60°C and pH 6.0 according to the method described above. The kinetic parameters k_{cat} and K_m were determined by a non-linear regression fit of the saturation curves using the Michaelis–Menten equation.

CD spectroscopy and thermal unfolding studies

Far-UV CD spectra (190–260 nm) were recorded on a Jasco J-810 spectropolarimeter (Jasco International Co.) coupled to a Peltier temperature controller using a 1 mm quartz cuvette. Proteins were used at a concentration of 10 μ M. Data collection parameters were set to scan rate of 50 nm/min, response time of 4 s, sensitivity of 100 mdeg, accumulation of 10, heating rate of 1°C/min and delay time for spectrum collection of 60 s. Results were expressed as mean residue ellipticity ($\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1} \cdot \text{residue}^{-1}$). All thermal unfolding experiments were monitored at 222 nm, except for the CBM3 construct, which was monitored at 230 nm.

SAXS (small-angle X-ray scattering)

SAXS data for CC, CBM3 and WT constructs were collected using a 165 mm MarCCD detector at the D02A/SAXS2 beamline (Brazilian Synchrotron Light Laboratory, Campinas, Brazil). The radiation wavelength was set to 1.48 Å (1 Å = 0.1 nm) and the sample-to-detector distance was set to 1559.1 mm to give a scattering vector range from 0.1 nm⁻¹ to 2.3 nm⁻¹ ($q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle). The samples at 2, 4 or 8 mg/ml, in 150 mM NaCl and 20 mM Tris/HCl, pH 7.5, were centrifuged for 10 min at 20000 g (4°C) and then filtered to remove any aggregates. Measurements were carried out at room temperature (~23°C). Buffer scattering was recorded and subtracted from the corresponding sample scattering. The integration of SAXS patterns was performed using Fit2D software [42] and the curves were scaled by protein concentration. Data analysis was performed using the GNOM program [43]. Low-resolution envelopes of each construction were calculated from the experimental SAXS data using an *ab initio* procedure implemented in the DAMMIN program [44]. Averaged models were generated from 10 runs using the DAMAVER package [45]. SAXS envelopes and NMR or XTAL structures were superimposed using the SUPCOMB 2.0 program [46]. BUNCH software [47], which employs rigid body model and simulated annealing routines, was used to find the best relative positions of NMR and XTAL structures in relation to the SAXS envelope. The CRY SOL program [48] was used to calculate the scattering curve from high-resolution structures, including the WT model obtained from BUNCH.

NMR spectroscopy

NMR experiments were performed using a Varian Inova spectrometer at the National Biosciences Laboratory (LNBio/CNPq), operating at a ¹H Larmor frequency of 599.887 MHz and temperature of 20°C. For these experiments, the spectrometer was equipped with a triple resonance cryogenic probe and a Z pulse-field gradient unit. The ¹⁵N-¹³C-labelled CBM sample was dissolved in 20 mM phosphate buffer, pH 7.2, containing 50 mM NaCl and 5% (v/v) ²H₂O, at a final concentration of approximately 0.3 mM. Water suppression was achieved by low-power continuous wave irradiation over the relaxation delay or using the WATERGATE method [49]. All data were processed using NMRPipe and NMRVIEW software [50,51]. Prior to Fourier transformation, the time domain data were zero-filled in all dimensions. When necessary, a fifth-order polynomial baseline correction was applied after transformation and phasing. To obtain distance constraints, cross-peak volumes were estimated from NOESY (nuclear Overhauser enhancement spectroscopy). Protein backbone resonance peaks were assigned using ¹H-¹⁵N-HSQC (heteronuclear single-quantum coherence) and standard triple resonance experiments, including HNCACB, CBCACONH, HNCO and HNCACO [52] of the ¹⁵N-¹³C-labelled CBM3 sample. All spectra for resonance assignment were recorded by setting up a resolution of 2048 × 128 × 96 points and spectral window of 8.00 kHz, 12.07 kHz and 1.95 kHz for ¹H, ¹³C and ¹⁵N respectively. In order to assign the side chain resonances, the ¹⁵N-¹³C-labelled CBM sample was freeze-dried and dissolved in 100% ²H₂O, followed by acquisition of HCCH-TOCSY, hCCH-TOCSY spectra. NOE (nuclear Overhauser effect)-derived distance restraints were obtained from the ¹⁵N-HSQC-NOESY and ¹³C-HSQC-NOESY (separately optimized for aliphatics and aromatics), both collected with 80 ms mixing time. The CBM3 structure was calculated in a semi-automated iterative manner with the CYANA program (version 2.1) [53], using 100

starting conformers. The CYANA 2.1 protocol was applied to calibrate and assign NOE cross-peaks. After the first few rounds of automatic calculations, the NOESY spectra were analysed again to identify additional cross-peaks consistent with the structural model and to correct misidentified NOEs. The structures obtained were further refined by restrained minimization and molecular dynamic studies using the CNS software [54]. The 20 structures with the lowest target function were selected to represent the ensemble of protein structures. The quality of the structures was analysed with PROCHECK-NMR [55]. The summary of NMR data is presented in Supplementary Table S2 (at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>).

Proton/deuterium exchange

Proton/deuterium exchange measurements were carried out with the ¹⁵N-¹³C-labelled CBM3 sample dissolved in 100% ²H₂O. In order to monitor the intensity loss of labile protons, a series of ¹H-¹⁵N-HSQC spectra were collected first at 15 min (time required to dissolve the protein and perform the first experiment) and subsequently at 40 min time intervals over two days. Additionally, a last spectrum was collected 10 days after the protein dissolution.

X-ray crystallography

The highly purified CC sample was concentrated to 75 mg/ml and its buffer was exchanged to water by ultrafiltration using Vivaspin6 centrifugal filters (GE Healthcare). Crystallization experiments were performed by the vapour diffusion method using a HoneyBee 963 robot (Genomic Solutions). Sitting drops were prepared by mixing 0.5 µl of the protein solution with an equal volume of mother liquor and equilibrated against 80 µl of the reservoir solution at 18°C. Formulations based on the following commercial crystallization kits were tested: SaltRX, Crystal Screen, Crystal Screen 2 (Hampton Research), Precipitant Synergy, Wizard I and II (Emerald BioSystems), PACT and JCSG+ (Qiagen/Nextal), with a total of 544 solutions. Large crystals grew within 45 days in the following conditions: 100 mM Bis-Tris-propane pH 6.5, 20% (w/v) PEG [poly(ethylene glycol)] 3350 and 200 mM sodium nitrate (form I); 100 mM SPG (succinic acid, sodium dihydrogen phosphate monohydrate, glycine) pH 6.0 and 25% (w/v) PEG 1500 (form II). Crystals were transferred to new drops containing the mother liquor and 20% (v/v) glycerol for cryoprotection and then flash-cooled in a 100 K nitrogen gas stream. In order to obtain an ion-protein complex, crystals belonging to form II were soaked for 3 min in a cryoprotectant solution containing 50 mM of manganese chloride. X-ray diffraction data were collected on the W01B-MX2 beamline at the Brazilian Synchrotron Light Laboratory (Campinas, Brazil). The wavelength used was 1.4586 Å and the intensities were recorded in a Mar Mosaic 225 mm CCD (charged-coupled-device) detector. Data were indexed, integrated and scaled using the HKL2000 package [56]. The structure of the crystal form II* was solved by the molecular replacement method using the MOLREP program [57] and the atomic co-ordinates of Cel5A from *Bacillus agaradhaerens* (PDB code 7A3H), which shares 70% identity with the catalytic domain of BsCel5A as the search model. Initial cycles of refinement involved a restrained and overall B-factor refinement in the REFMAC5 program [58] using automatic weighting between X-ray and geometry terms. After each cycle of refinement, the model was inspected and manually adjusted into the (2F_o - F_c) and (F_o - F_c) electron density maps using the COOT program [59]. Solvent molecules were manually added at positive peaks above 3.0 σ in the difference Fourier

Table 1 Kinetic parameters of BsCel5A and its catalytic core using β -glucan as substrate

Parameter	CC		WT	
	0 mM Mn ²⁺	10 mM Mn ²⁺	0 mM Mn ²⁺	10 mM Mn ²⁺
$V_{\max}$	143.7 $\pm$ 6.6	92.3 $\pm$ 5.9	46.2 $\pm$ 3.1	35.7 $\pm$ 1.8
K_m	1.4 $\pm$ 0.2	1.6 $\pm$ 0.3	0.32 $\pm$ 0.2	0.5 $\pm$ 0.2
k_{cat}	78.7	50.5	40.4	31.2
k_{cat}/K_m	56.2	31.6	127.8	58.8

maps, taking into consideration hydrogen bonding potential. This structure was extensively refined and then used as the search model for solving other crystalline forms.

Accession numbers

Structure factors and atomic co-ordinates of the BsCel5A CC were deposited in the PDB under accession codes 3PZV (Form I), 3PZU (Form II) and 3PZT (Form II*). NMR data of the BsCel5A CBM3 were deposited in the Biological Magnetic Resonance Bank and PDB under codes 17399 and 2L8A respectively.

RESULTS AND DISCUSSION

Accessory domain enhances catalytic efficiency

The WT protein and its truncated domains were expressed in *Escherichia coli* cells and purified by nickel-affinity and size-exclusion chromatographies (see Supplementary Figure S1 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Biochemical characterization showed that the CBM3 deletion does not alter the substrate specificity (see Supplementary Figure S2 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>) and optimal conditions for catalysis (see Supplementary Figure S3 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Despite the fact that the physiological substrate of BsCel5A is cellulose, we have used soluble substrates in kinetic studies owing to the very slow rate of hydrolysis of insoluble substrates, such as Avicel[®] (results not shown). In addition, it has been shown that BsCel5A is only able to bind Avicel[®] when CBM3 is appended to its C-terminus [13].

The kinetics study showed that CBM3 increased BsCel5A affinity to the substrate owing to a reduced K_m , but the enzyme turnover ($V_{\max}$) is reduced (Table 1). However, the reduction in K_m largely compensates for the lower $V_{\max}$ and, as a result, CBM3 enhances catalytic efficiency (k_{cat}/K_m) (Table 1). We determined that BsCel5A displays an extended substrate binding surface owing to the strategically three-dimensional arrangement of CBM3 in relation to CC (see details below), which favours additional protein-carbohydrate interactions. These additional interactions are reflected in a higher substrate affinity and may also explain the lower turnover of the WT enzyme ($V_{\max}$).

Deletion of the accessory domain reduces enzyme stability

The thermal stability of the BsCel5A constructions was assessed by unfolding studies using CD spectroscopy (Figure 1). As expected, CD spectra of CC and CBM3 displayed characteristic profiles of α/β and all- β proteins respectively (Figure 1A). It is worth noting that BsCel5A and its truncated forms showed a single structural transition in the thermal unfolding experiments (Figure 1B), despite the fact that BsCel5A is a two-domain protein and therefore two transitions corresponding to each domain were

expected. Perhaps the 23-residue-long linker and inter-domain contacts drive the thermal denaturation to a single-transition process. Interestingly, unfolding studies revealed that deletion of the accessory domain increases the thermal stability from 64 to 78 °C (T_m , melting temperature). The reduced thermal stability of the WT protein is probably influenced by the significantly lower stability of CBM3 ($T_m = 59$ °C) in relation to the parental catalytic domain ($T_m = 78$ °C). As shown by NMR structural studies, the lower stability of BsCel5ACBM is explained by the absence of a structural calcium ion. Moreover, deletion of CBM3 promotes exposure of polar residues at the CC surface, which becomes favourably hydrated, thus increasing stability [60].

Both biochemical and biophysical studies clearly indicate the importance of CBM3 in the regulation of overall BsCel5A function, in spite of its apparent negative effect on stability.

Bacillus CBM3 is a member of a novel calcium-independent subfamily

To gain insights into the structural determinants of the stability of *Bacillus* CBM3, we performed a NMR study. The ¹H-¹⁵N-HSQC spectrum of CBM3 showed high-quality signal dispersion (see Supplementary Figure S4 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>), indicating a well-folded protein. The most common classification, based on sequence similarities, defines BsCel5A CBM3 as a member of the CBM3 family (<http://www.cazy.org/Carbohydrate-Binding-Modules.html>). So far, five members of CBM3 have been solved, sharing from 13 to 36% of identity with the BsCel5A CBM3. BsCel5A CBM3 folds into a β -sandwich, a common scaffold for CBMs, encompassing a core formed by two β -sheets of four and five β -strands, and one short α -helix at the C-terminal region (Figure 2A). Structural comparisons showed that the β -sandwich core is conserved among CBMs, but the surface loops are different in both composition and conformation (Figure 2B). Canonically, the CBM3 members with solved structures have a highly conserved calcium-binding site, which is not present in the BsCel5A CBM3 structure (Figures 2B and 2C). A search for sequences sharing similarity with BsCel5A CBM3 in GenBank[®] resulted in a set of CBMs that do not have the structural requirements for calcium binding (see Supplementary Figure S5 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Interestingly, all of these CBMs are from *Bacillus* spp., suggesting that the ancestor of this bacterium genus has lost the CBM3 calcium-binding site during evolution.

Therefore, we propose a new classification of CBM3 into two groups: the calcium-binding proteins, which include the already defined subfamilies CBM3a, CBM3b, CBM3b' and CBM3c; and the calcium-independent proteins, termed in the present study as the CBM3d subfamily, with BsCel5A CBM3 being its first structurally characterized member. The classification of CBM3s in **a**, **b** and **c** refers to the type of substrate that is preferentially recognized by the CBM [61]. The **a** and **b** groups bind crystalline cellulose; **a** is specific for cellulosomal-scaffolding proteins and **b** is found in soluble cellulases. Previously, it was shown that certain CBM3b members bind neither crystalline nor soluble cellulose and hence, they were collectively designated as CBM3b' [62]. The **c** group binds single-chain cellulose and is found in cellulosomal cellulases. The new CBM3d subfamily is closely related to the CBM3a subfamily, owing to its substrate preference; however, it lacks the calcium-binding site, which seems to play a structural role. Indeed, our biophysical studies showed that the *Bacillus* CBM3 is less stable than its parental protein, which can be attributed to the lack of structural calcium.

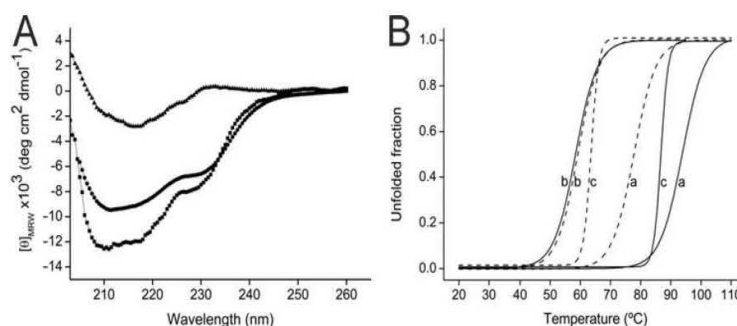


Figure 1 Biophysical characterization of BsCel5A constructions

(A) Far-UV CD spectra of CC (●), CBM (▲) and WT (■). (B) Thermal denaturation in the absence (dashed lines) or presence of manganese (solid lines) of CC (a), CBM (b) and WT (c).

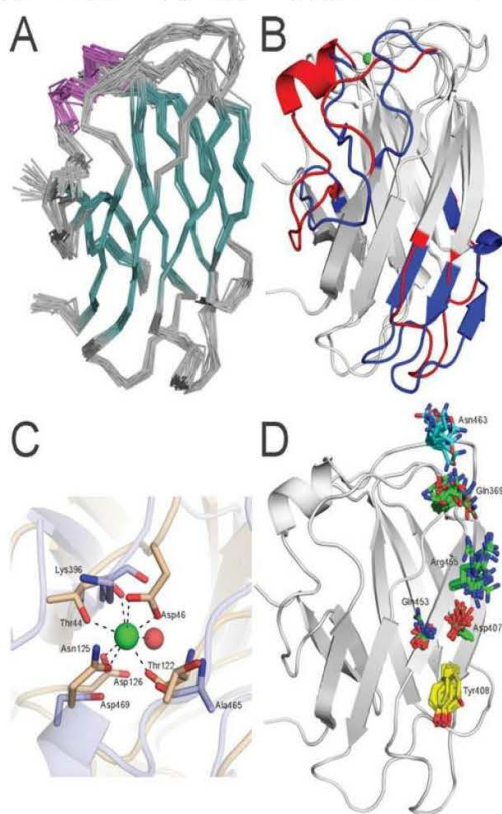


Figure 2 NMR solution structure of BsCel5A CBM3

(A) View of the ensemble of 20 lowest energy structures, with β -strands in cyan, α -helices in magenta and loops in grey. (B) Structural superposition of BsCel5A CBM3 structure on CtCipACBM (PDB code 1NBC) with structural differences coloured in blue and red respectively. The Ca^{2+} ion from the CtCipACBM structure is shown as a green sphere. (C) Calcium-binding site of CtCipACBM (wheat) and the corresponding region in the BsCel5A CBM3 structure (light blue). (D) BsCel5A CBM3 residues involved in carbohydrate recognition: residues conserved in CtCipACBM and identified by NMR CSP (chemical shift perturbation) analysis (carbon atoms in green); residues conserved but not identified by NMR CSP (carbon atoms in yellow); and residues not conserved but identified by NMR CSP (carbon atoms in cyan).

It is worth mentioning that there is another classification of CBMs taking into account both structural and functional similarities, which assorts these modules in types A, B and C

[19]. The type A binds crystalline cellulose and chitin through a flat surface, type B binds a single-chain polysaccharide via an extended groove or cleft, and type C binds mono-, di- or tri-saccharides. In this classification, the BsCel5A CBM3 is considered a type A member, owing to its ability to bind crystalline cellulose [13] and higher structural similarity with type A (PDB code 1NBC) than other type B (PDB code 1J84) or type C (PDB code 1I82) CBMs.

Polar residues play a pivotal role in oligosaccharide recognition by CBM3d

The molecular surface of CBMs has been associated with the type of substrate that they can bind, e.g. CBM3a presents a very flat surface compatible with its target, crystalline cellulose. Despite the fact that BsCel5A binds crystalline cellulose and the binding occurs only when the CBM3 is present [13], the *Bacillus* CBM3 surface is not as flat as those of CBM3a members (Figure 3). This unexpected surface topology of BsCel5A CBM3 can be attributed to the absence of the CBM3a $\beta 4'$ strand (Figure 3) and to the inherent flexibility of its surface loops, as assessed by hydrogen–deuterium exchange (see Supplementary Figure S6 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>) and NMR relaxation experiments (results not shown). On the basis of these dynamics studies, we suggest that the BsCel5A CBM3 can undergo conformational changes, adapting its topology for cellulose recognition and binding. Moreover, the exact protein–carbohydrate interactions are not known, since sugar–complex structures of a CBM3a or even a type A CBM are not available. Thus, in order to map the residues involved in the recognition of cellulose, we analysed variations in NMR frequencies of BsCel5A CBM3 in the presence of cellopentaose. The major perturbations in chemical shifts occurred in Asn⁴⁶³ and Arg⁴⁵⁵, and minor changes were observed in Gln³⁶⁹, Gln⁴⁵³ and Asp⁴⁰⁷ (Figure 2D and Supplementary Figure S7 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). The residues Arg⁴⁵⁵, Gln³⁶⁹, Gln⁴⁵³ and Asp⁴⁰⁷ are conserved in CtCipACBM (*Clostridium thermocellum* CBM3, PDB code 1NBC), a well-characterized type A CBM [63]. Tyr⁴⁰⁸ is also conserved, but it is not affected by the presence of cellopentaose, perhaps owing to the short length of this oligosaccharide.

The molecular recognition and affinity towards carbohydrates is typically achieved by van der Waals interactions arising from stacking of aromatic residues with the furanose/pyranose rings and by hydrogen bonds between polar groups of saccharides and amino acid side chains. The mapped interactions of the BsCel5A CBM3 with cellopentaose revealed that the carbohydrate-binding

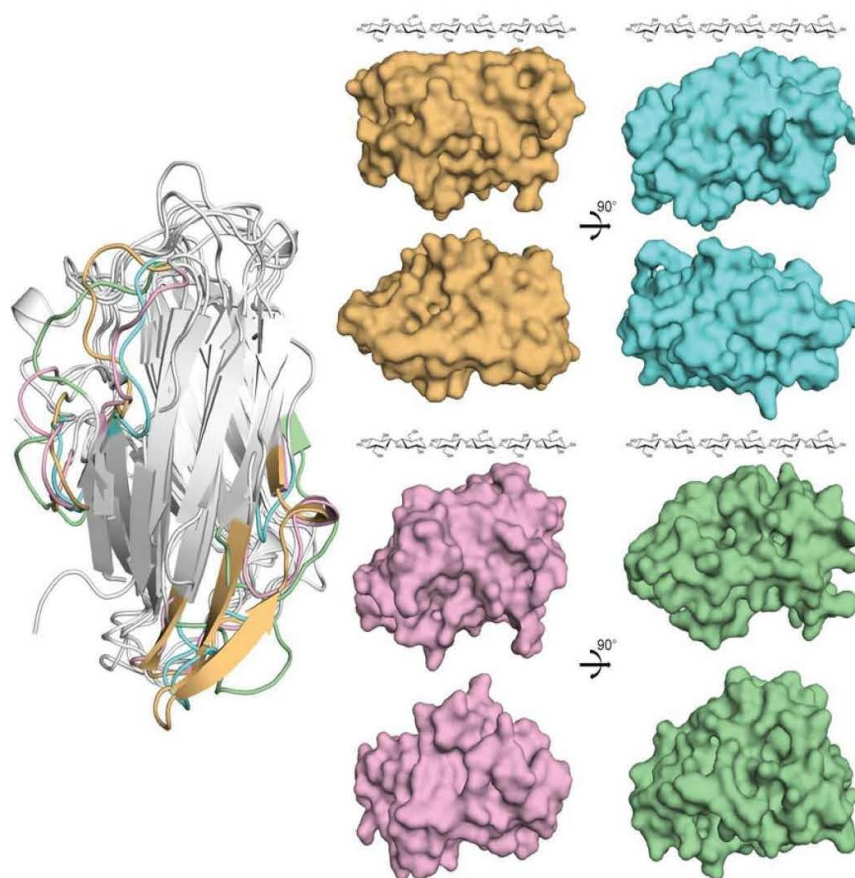


Figure 3 Structural comparison of BsCel5A CBM3 (CBM3d) with CBM3a, b' and c

The CBM3s from CipA (CBM3a, PDB code 1NBC), E4 (CBM3c, PDB code 4TF4) and Cel9V (CBM3b', PDB code 2WOB) were superposed on the CBM3d structure and are shown as cartoon representations (left-hand panel). The main differences at the carbohydrate-binding face were coloured in green (CBM3d), orange (CBM3a), cyan (CBM3c) and pink (CBM3b'). The molecular surfaces of CBM3a, b', c and d are shown as side (middle panels) and frontal (right-hand panels) views following the colours above.

face is rich in charged and polar residues, such as glutamine, arginine, aspartate, lysine and asparagine. The only aromatic residue on this face is Tyr⁴⁰⁸, whereas a set of aromatic residues is buried in the β -sandwich core making hydrophobic contacts. These findings indicate that the interactions of BsCel5A CBM3 with cellulose are mainly by polar side chains. Moreover, we suggest that these polar residues may not only work on the carbohydrate affinity/recognition, but could specifically replace inter-chain hydrogen bonds, acting to disrupt the tight cellulose structure [63]. Therefore it may result in the local separation of cellulose single chains to be delivered to the CC, thus contributing to a higher catalytic efficiency.

Three-dimensional domain architecture of BsCel5A reveals a flat cellulose-binding platform

BsCel5A and its separated domains were analysed by SAXS, showing that all forms are monomeric in solution with R_g values of 1.75, 1.63 and 3.04 nm, and D_{max} values of 4.65, 4.8, and 9.5 nm, for the CC, CBM3 and WT protein respectively (Table 2 and Supplementary Figure S8 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Both CC and CBM3 were nicely fitted into the corresponding envelopes (Figure 4A). The good

quality of SAXS data allowed us to observe the active-site cavity in the CC. The averaged low-resolution model of the WT protein calculated from 20 independent *ab initio* models gives a good notion of the three-dimensional arrangement of the accessory domain with respect to its parental catalytic domain (Figure 4B). The molecular envelope reveals a large continuous planar surface, which has been considered essential for crystalline cellulose binding. The atomic co-ordinates of both domains were fitted into the SAXS model of the WT protein, which showed that this planar surface is formed by the active-site interface of the CC in line with the cellulose recognition face of CBM3. This model is in full agreement with the substrate-recognition face mapped by NMR and with the crystallographic structures of the multidomain endoglucanase from *Clostridium cellulolyticum* (Cel9G) and cellulase E4 from *Thermobifida fusca* (Figure 4C).

Manganese metal ion hyper-stabilizes the CC of BsCel5A

Divalent ions are well-known cofactors of enzymes, either activating their catalytic machinery or having a structural role. In the present study we tested the effect of different divalent ions, including Mn²⁺, Mg²⁺, Ca²⁺, Zn²⁺, Ni²⁺ and Co²⁺. None of them affected the stability, except the Mn²⁺ ion, which drastically

Table 2 Summary of SAXS calculations

	CC			CBM			WT		
	Exper [*]	Envel [†]	XTAL [‡]	Exper [*]	Envel [†]	NMR [‡]	Exper [*]	Envel [†]	Assem [‡]
R_g (nm)	1.75	1.73	1.86	1.63	1.63	1.65	3.04	2.96	2.87
D_{max} (nm)	4.65			4.80			9.50		
χ^2		2.13	2.89		1.89	2.75		2.37	2.01

^{*}Calculated from experimental data.

[†]Calculated from the SAXS-derived envelope.

[‡]Calculated from XTAL, NMR or assembly.

[§]Comparison between experimental and theoretical curves.

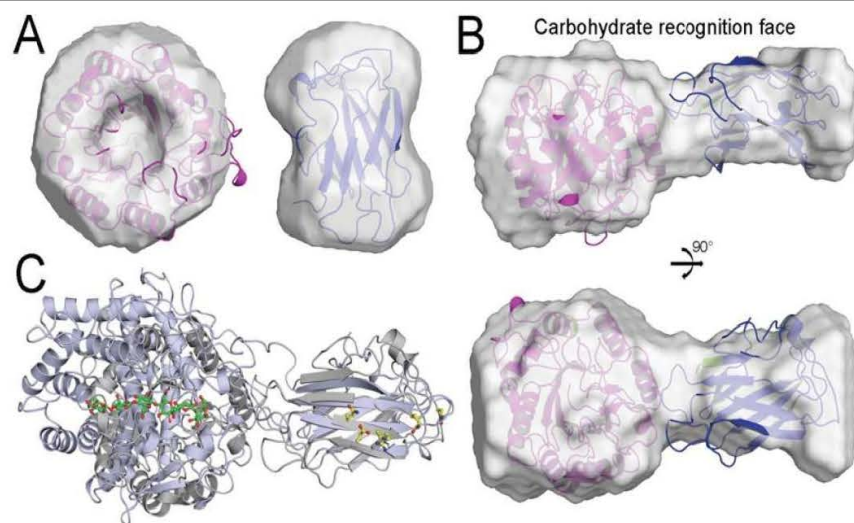


Figure 4 SAXS analysis of BsCel5A constructions

(A) SAXS envelopes from CC (left-hand panel) and CBM3d (right-hand panel) are shown as transparent surfaces and the fitted XTAL or NMR structures are shown as cartoon representations. (B) Different views of the CC and CBM3d arrangement into the SAXS envelope from the WT protein. (C) Overlay of the E4 structure from *T. fusca* (light blue) on the BsCel5A model (grey). The oligosaccharide bound in the E4 catalytic domain is shown in green and the BsCel5A CBM3 residues identified by NMR CSP are shown in yellow.

improved the thermal stability of BsCel5A, by elevation of the T_m by 23 °C, from 64 to 87 °C (Figure 1B). This result indicates that Cel5As (family 5A cellulases) display a very specific motif for recognition of manganese ions. Thus, in order to map the domains involved in the manganese ion interactions, truncated forms of the protein were analysed. As shown in Figure 1(B), the CC had its thermal stability improved by the Mn^{2+} ion, whereas that of CBM remained unaltered, mapping the specific interactions of the Mn^{2+} ion with the CC.

Structural basis for BsCel5A stabilization by manganese ion

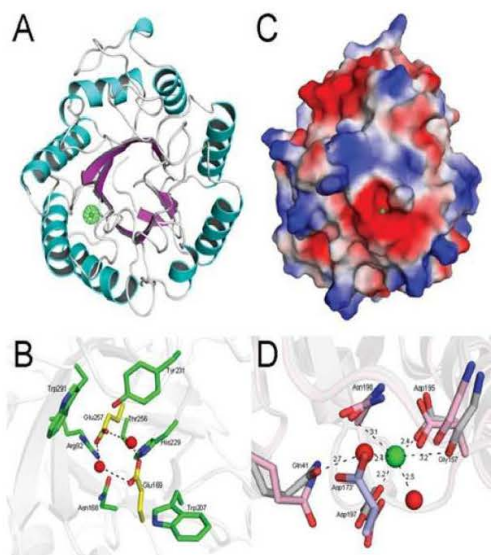
In order to identify the manganese-binding site and to delineate the structural basis for the enhanced thermal stability of BsCel5A by this ion, we undertook an X-ray crystallographic study. BsCel5A CC crystallized in two different crystalline forms (form I and form II), which diffracted up to 2.90 Å and 2.10 Å respectively. However, only the crystal form II was successfully soaked with manganese ions (Table 3). Crystals belonging to form I display a very particular crystallographic symmetry in which the manganese-binding site is concealed by an aspartate residue of the neighbouring molecule (Figure 5D), which prevented ion incorporation by soaking methods.

The BsCel5A CC structure encompasses a classical (β/α)₈-barrel fold (Figure 5A), with the active site located at the crevice formed by the C-terminal ends of β -strands. The catalytic machinery involves two acid residues, Glu²²⁹ and Glu¹⁶⁹, the aromatic gate keepers Trp²⁰⁷ and Trp²⁹¹, and the polarizers His²²⁹, Arg⁹² and Thr²⁵⁶ (Figure 5B). From homology with other GH5 cellulases, Glu¹⁶⁹ is expected to be the proton donor, which protonates the glycosidic bond and catalyses the bond fission, and the Glu²⁵⁷ is the nucleophile, which facilitates the reaction by stabilizing the resulting carbonium ion intermediate. Several studies demonstrated the key role of a histidine residue in the protonation of the acid/base glutamate residue, which is controlled by a catalytic triad, similar to the serine proteases [64]. In this scheme, a third element, which could be a serine/threonine residue, an aspartate residue or a water molecule, would protonate the histidine, permitting a classification of the glycosyl hydrolases according to this histidine proton donor [64]. BsCel5A contains a threonine residue in the equivalent position, which allows classification as type A catalytic triad. Cel5As from *Thermotoga maritima* and *C. thermocellum* have a water molecule in the corresponding position and are classified as type C enzymes (see Supplementary Figure S9 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Water-dependent catalytic machineries are labile at alkaline conditions, since the water molecule

Table 3 Data collection and refinement statistics

Values in parentheses are for the highest-resolution shell. RMSD, root mean square deviation.

Parameter	Form I	Form II native	Form II* 10 mM Mn ²⁺
<i>Data collection</i>			
Space group	C2	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell parameters (Å, °)	a = 155.36, b = 81.08, c = 114.27, β = 104.29	a = 50.24, b = 110.76, c = 121.38	a = 49.53 b = 110.62 c = 122.26
Resolution	50.00–2.90 (3.00–2.90)	50.00–2.10 (2.18–2.10)	50.00–1.97 (2.04–1.97)
Number of unique reflections	30 407 (2607)	39 069 (3393)	47 366 (4163)
Multiplicity	3.5 (2.7)	6.9 (6.2)	4.8 (4.3)
Completeness (%)	97.1 (83.7)	97.2 (85.9)	97.6 (87.8)
R _{merge} (%)	8.3 (43.6)	11.6 (31.9)	8.4 (37.3)
I/σ (I)	11.3 (2.0)	13.5 (5.0)	14.7 (3.8)
<i>Refinement</i>			
R _{work} /R _{free} (%)	23.0/29.6	17.1/22.1	16.4/20.6
Average B-factor	54.5	32.7	29.1
RMSD for bond lengths (Å)	0.006	0.023	0.025
RMSD for bond angles (°)	1.261	1.903	1.938
Protein chains	4	2	2
Water molecules	30	330	452
<i>Ramachandran plot</i>			
Favoured region (%)	90.63	96.44	97.09
Additionally allowed region (%)	8.02	3.39	2.91
Outlier region (%)	1.35	0.17	0.00
PDB code	3PZV	3PZU	3PZT

**Figure 5** Crystal structure of the BsCel5A CC

(A) Cartoon representation of the protein structure, with the manganese ion shown as a green sphere. (B) Zoom of the active site showing the glutamate residues (carbon atoms in yellow), other catalytically relevant residues (carbon atoms in green) and two water molecules (red spheres). (C) Electrostatic surface coloured by charge, from red (–) to blue (+), indicating the presence of the manganese ion in an acidic groove. (D) Interactions of protein residues (grey) and water molecules (red spheres) with the manganese ion (green sphere). The 2.9 Å resolution structure (pink) has an aspartate residue of a neighbouring molecule (blue) occluding the manganese-binding site.

deprotonates at high pH, disrupting the hydrogen network required for catalysis. On the other hand, enzymes with the type A catalytic triad would remain active even in basic conditions, which is advantageous for biotechnological purposes.

In the manganese complex structure, the ion is located at the opposite face of the active site, near to the lid-like β-

hairpin (Figure 5C), and is hexa-co-ordinated by the two negatively charged residues Asp¹⁹⁵ and Asp¹⁹⁷, the polar Asn¹⁹⁸, the carbonyl oxygen from Gly¹⁵⁷ and two water molecules (Figure 5D). The strategic location of the manganese-binding site permits simultaneous water-mediated interactions between the N-terminus and distant regions of the protein, increasing the overall structural rigidity and consequently a notable gain in the protein stability. Comparison between forms II (without manganese ion) and II* (with manganese ion) showed that the ion only substituted a water molecule in the highly negatively charged aperture without affecting either the cleft architecture or the conformational states of catalytically relevant residues. This result also suggests that the region comprising the lid-like β-hairpin, which is subsequent to the N-terminal end, can be a strategic motif that could be engineered with the aim of improving protein stability. Indeed, in the analogous hyperthermostable TmCel5A (*T. maritima* Cel5A) [65], the β-hairpin is replaced by an α-helical element, which enables additional polar contacts, supporting the fact that the stabilization of this region may implicate a higher overall stability.

The effect of the manganese ion on kinetics and specificity of BsCel5A

Our biochemical data showed that the presence of Mn²⁺ ions prolonged the lifetime of the enzyme at high temperatures (90 °C) (see Supplementary Figure S10 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>) without interfering with the optimal conditions (pH and temperature) for activity (see Supplementary Figure S3). Although it was expected that the gain in thermal stability by Mn²⁺ ions could produce higher relative activities at high temperatures, both univariate (Supplementary Figure S11 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>) and response surface (Supplementary Figure S3) experiments indicated no alteration in the temperature-dependent activity profile. This native-like behaviour of BsCel5A could be related to the unaltered configuration of the active site, since the ion binds to a remote site, ~30 Å away from the active-site cleft. In terms of protein dynamics, we suggest that, although the ion stabilizes the structural framework of the enzyme, the catalytic machinery remains susceptible to temperature-induced changes.

As expected, the presence of Mn^{2+} ions did not modify substrate preference, but it slightly reduced the total activity by approximately 30 % (Supplementary Figure S2). This partial inhibition was initially attributed to a possible secondary binding site involving the catalytic glutamate residues, since the manganese-binding site is sterically and electrostatically similar to the active-site pocket, and in the crystal structure of TmCel5A, a divalent (cadmium) ion was found interacting with the catalytic glutamate residues [66]. However, kinetic experiments showed that the Mn^{2+} ion did not interfere with the substrate binding (K_m) (Table 2), thus discarding the possibility of an inhibitory effect by a competitive mechanism. In fact, only V_{max} was decreased by the presence of Mn^{2+} ions, which reflected as a lower enzymatic efficiency (Table 2). It could be a consequence of the higher overall molecular rigidity, which prevents the conformational changes required for optimal catalysis. This hypothesis is supported by previous studies that described the pivotal role of ligand-induced conformational alterations in the catalytic activity of family 5A cellulases [66].

As a complement to the thorough characterization of the binding and effects of the Mn^{2+} ions, other ligands were tested, including several ions (Fe^{3+} , Cu^{2+} , Mn^{2+} , Zn^{2+} , Ni^{2+} , Co^{2+} , Mg^{2+} and Ca^{2+}), a chelating agent (EDTA), surfactants (SDS and Triton X-100) and organic solvents (methanol, ethanol and DMSO) (see Supplementary Figure S12 at <http://www.BiochemJ.org/bj/441/bj4410095add.htm>). Briefly, none of the additives increased the enzymatic activity; Fe^{3+} and Cu^{2+} almost abolished the activity; methanol and ethanol did not reduced the enzymatic activity at concentrations up to 10 %; and SDS caused a decrease of 30 % in enzymatic activity.

Biotechnological implications

Modern biotechnology lies at the interface between basic and applied science, and continuously demands novel enzymes with improved catalytic efficiency and stability under extreme and non-natural conditions. Cellulases have many industrial applications, including food processing, finishing of textiles and generation of cellulosic bioethanol, one of the sustainable energy sources. These industrial processes require cellulolytic activity under a broad range of pH, temperature and ionic strengths. In this context, the understanding of enzymes at a molecular level and the mechanisms to improve their stability and to modify their catalytic efficiency and specificity is essential for rational protein design aiming to ameliorate their use in biotechnological processes.

In the present study, we investigated the role of the accessory domain in the stability and function of family 5 cellulases, and a novel mechanism for thermal stabilization was proposed. Our results showed that *Bacillus* CBM3 represents a novel subfamily, which lacks the calcium-binding site and its binding to carbohydrate is mediated by polar contacts. Moreover, *Bacillus* CBM3 decreases the overall enzyme stability, which can be attributed to the lack of the structural calcium ion, but on the other hand it improves the catalytic efficiency. The discovery of a specific manganese-binding site that is conserved among 5A cellulases suggests a novel approach for thermal stability enhancement without enzyme engineering.

AUTHOR CONTRIBUTION

Camila Santos and Mário Murakami designed the research. Camila Santos, Igor Polikarpov, José Xavier-Neto, Richard Ward, Ana Zeri and Mário Murakami conceived and designed the experiments. Crystal structure determination was carried out by Camila Santos and Joice Paiva. Spectroscopic analyses were performed by Joice Paiva, Patrícia Akao and Andréia Meza. Functional and kinetic studies were performed by Júnio Cota, Zaira Hoffmann, Fábio Squina and Roberto Ruller. NMR studies were carried out by Maurício Storça, Jorge

Neves, Rodrigo Navarro, Juliana Smetana, Maria Nogueira and Ana Zeri. Camila Santos and Mário Murakami analysed the data and wrote the paper. Mário Murakami supervised the project. All authors read and approved the final version of the paper.

ACKNOWLEDGEMENTS

We are thankful for the provision of time on the MX2 beamline (LNLS), Robolab (LNBio), LEC (LNBio) and NMR (LNBio) at the National Center for Research in Energy and Materials (Campinas, Brazil).

FUNDING

This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [grant number 10/51890-8] Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 476043/2011-5] and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

- Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'Hare, M. and Kammen, D. M. (2006) Ethanol can contribute to energy and environmental goals. *Science* **311**, 506–508
- Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., Himmel, M., Keller, M., McMillan, J. D., Sheehan, J. and Wyman, C. E. (2008) How biotech can transform biofuels. *Nat. Biotechnol.* **26**, 169–172
- Binder, J. B. and Raines, R. T. (2010) Fermentable sugars by chemical hydrolysis of biomass. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4516–4521
- Shaw, A. J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., Hogsett, D. A. and Lynd, L. R. (2008) Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13769–13774
- Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R. and Monot, F. (2009) New improvements for lignocellulosic ethanol. *Curr. Opin. Biotechnol.* **20**, 372–380
- Wen, F., Nair, N. U. and Zhao, H. (2009) Protein engineering in designing tailored enzymes and microorganisms for biofuels production. *Curr. Opin. Biotechnol.* **20**, 412–419
- Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshall, J. and Arnold, F. H. (2009) A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5610–5615
- Morais, S., Heyman, A., Barak, Y., Caspi, J., Wilson, D. B., Lamed, R., Shoseyov, O. and Bayer, E. A. (2010) Enhanced cellulose degradation by nano-complexed enzymes: synergism between a scaffold-linked exoglucanase and a free endoglucanase. *J. Biotechnol.* **147**, 205–211
- van der Veen, B. A., Skov, L. K., Potocki-Véronèse, G., Gajhede, M., Monsan, P. and Remaud-Simeon, M. (2006) Increased amylase activity and specificity, and identification of regions important for activity, specificity and stability through molecular evolution. *FEBS J.* **273**, 673–681
- Liu, M., Wang, J., Liu, J., Yao, J. and Yu, Z. (2006) Expression of *Bacillus subtilis* JA18 endo- β -1,4-glucanase gene in *Escherichia coli* and characterization of the recombinant enzyme. *Ann. Microbiol.* **56**, 41–45
- Li, W., Zhang, W. W., Yang, M. M. and Chen, Y. L. (2008) Cloning of the thermostable cellulase gene from newly isolated *Bacillus subtilis* and its expression in *Escherichia coli*. *Mol. Biotechnol.* **40**, 195–201
- Li, W., Huan, X., Zhou, Y., Ma, Q. and Chen, Y. (2009) Simultaneous cloning and expression of two cellulase genes from *Bacillus subtilis* newly isolated from Golden Takin (*Budorcas taxicolor bedfordi*). *Biochem. Biophys. Res. Commun.* **383**, 397–400
- Wang, Y., Yuan, H., Wang, J. and Yu, Z. (2009) Truncation of the cellulose binding domain improved thermal stability of endo- β -1,4-glucanase from *Bacillus subtilis* JA18. *Bioresour. Technol.* **100**, 345–349
- Yang, D., Weng, H., Wang, M., Xu, W., Li, Y. and Yang, H. (2010) Cloning and expression of a novel thermostable cellulase from newly isolated *Bacillus subtilis* strain 115. *Mol. Biol. Rep.* **37**, 1923–1929
- Deka, D., Bhargavi, P., Sharma, A., Goyal, D., Jawed, M. and Goyal, A. (2011) Enhancement of cellulase activity from a new strain of *Bacillus subtilis* by medium optimization and analysis with various cellulosic substrates. *Enzyme Res.* **2011**, 151656
- Liao, H., Zhang, X. Z., Rollin, J. A. and Zhang, Y. H. (2011) A minimal set of bacterial cellulases for consolidated bioprocessing of lignocellulose. *Biotechnol. J.*, doi:10.1002/biot.201100157
- Zhang, X. Z. and Zhang, Y. H. (2011) Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. *Microb. Biotechnol.* **4**, 98–105
- Wolf, M., Geçici, A., Simon, O. and Borriss, R. (1995) Genes encoding xylan and β -glucan hydrolysing enzymes in *Bacillus subtilis*: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase. *Microbiology* **141**, 281–290

- 19 Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781.
- 20 Shoseyov, O., Shani, Z. and Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* **70**, 283–295.
- 21 Hashimoto, H. (2006) Recent structural studies of carbohydrate-binding modules. *Cell. Mol. Life Sci.* **63**, 2954–2967.
- 22 Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H. and Gilbert, H. J. (1995) Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *Biochem. J.* **307**, 151–158.
- 23 Winterhalter, C., Heinrich, P., Candussio, A., Wich, G. and Liebl, W. (1995) Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*. *Mol. Microbiol.* **15**, 431–444.
- 24 Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J. Bacteriol.* **180**, 1709–1714.
- 25 Kataeva, I. A., Blum, D. L., Li, X. L. and Ljungdahl, L. G. (2001) Do domain interactions of glycosyl hydrolases from *Clostridium thermocellum* contribute to protein thermostability? *Protein Eng.* **14**, 167–172.
- 26 Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M. and Bolam, D. N. (2001) *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry* **40**, 9167–9176.
- 27 Shin, E. S., Yang, M. J., Jung, K. H., Kwon, E. J., Jung, J. S., Park, S. K., Kim, J., Yun, H. D. and Kim, H. (2002) Influence of the transposition of the thermostabilizing domain of *Clostridium thermocellum* xylanase (XynX) on xylan binding and thermostabilization. *Appl. Environ. Microbiol.* **68**, 3496–3501.
- 28 Zhao, G., Ali, E., Araki, R., Sakka, M., Kimura, T. and Sakka, K. (2005) Function of the family-9 and family-22 carbohydrate-binding modules in a modular β -1,3-1,4-glucanase/xylanase derived from *Clostridium stercoarium* Xyn10B. *Biosci. Biotechnol. Biochem.* **69**, 1562–1567.
- 29 Zhang, S., Yin, Q. Y., Li, Y. H., Ding, M., Xu, G. J. and Zhao, F. K. (2007) Molecular and biochemical characterization of Ba-EGA, a cellulase secreted by *Bacillus* sp. AC-1 from *Ampullaria crosseana*. *Appl. Microbiol. Biotechnol.* **75**, 1327–1334.
- 30 Sunna, A. (2010) Modular organisation and functional analysis of dissected modular β -mannanase CsMar26 from *Caldicellulosiruptor R18B.4*. *Appl. Microbiol. Biotechnol.* **86**, 189–200.
- 31 Kittur, F. S., Mangala, S. L., Rus'd, A. A., Kitaoka, M., Tsujibo, H. and Hayashi, K. (2003) Fusion of family 2b carbohydrate-binding module increases the catalytic activity of a xylanase from *Thermotoga maritima* to soluble xylan. *FEBS Lett.* **549**, 147–151.
- 32 Mahadevan, S. A., Wi, S. G., Lee, D. S. and Bae, H. J. (2008) Site-directed mutagenesis and CBM engineering of Cel5A (*Thermotoga maritima*). *FEMS Microbiol. Lett.* **287**, 205–211.
- 33 Mammo, G., Hatti-Kaul, R. and Mattiasson, B. (2007) Fusion of carbohydrate binding modules from *Thermotoga neopolitana* with a family 10 xylanase from *Bacillus halodurans* S7. *Extremophiles* **11**, 169–177.
- 34 Zhang, Y., Chen, S., Xu, M., Cavaco-Paulo, A., Wu, J. and Chen, J. (2010) Characterization of *Thermobifida fusca* cutinase-carbohydrate-binding module fusion proteins and their potential application in bioscouring. *Appl. Environ. Microbiol.* **76**, 6870–6876.
- 35 Sonan, G. K., Receveur-Brechot, V., Duez, C., Aghajari, N., Czjzek, M., Haser, R. and Gerday, C. (2007) The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium. *Biochem. J.* **407**, 293–302.
- 36 Morris, D. D., Gibbs, M. D., Chin, C. W., Koh, M. H., Wong, K. K., Allison, R. W., Nelson, P. J. and Bergquist, P. L. (1998) Cloning of the xynB gene from *Dictyoglomus thermophilum* R46B.1 and action of the gene product on kraft pulp. *Appl. Environ. Microbiol.* **64**, 1759–1765.
- 37 Blanco, A., Diaz, P., Zueco, J., Parascandola, P. and Javier Pastor, F. I. (1999) A multidomain xylanase from *aBacillus* sp. with a region homologous to thermostabilizing domains of thermophilic enzymes. *Microbiology* **145**, 2163–2170.
- 38 Li, M. X., Corson, D. C. and Sykes, B. D. (2002) Structure determination by NMR isotope labeling. *Methods Mol. Biol.* **173**, 255–265.
- 39 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- 40 Miller, G. L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–42.
- 41 Myers, R. H. and Montgomery, D. C. (2002) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* (2nd edn), John Wiley & Sons, New York.
- 42 Hammersley, A. P., Brown, K., Burmeister, W., Claustre, L., Gonzalez, A., McSweeney, S., Mitchell, E., Moy, J. P., Svensson, S. O. and Thompson, A. W. (1997) Calibration and application of an X-ray image intensifier/charge-coupled device detector for monochromatic macromolecular crystallography. *J. Synchrotron Radiat.* **4**, 67–77.
- 43 Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503.
- 44 Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886.
- 45 Volkov, V. V. and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864.
- 46 Kozin, M. B. and Svergun, D. I. (2000) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41.
- 47 Petoukhov, M. V. and Svergun, D. I. (2005) Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250.
- 48 Svergun, D. I., Barberato, C. and Koch, M. H. J. (1995) CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
- 49 Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661–665.
- 50 Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293.
- 51 Johnson, B. A. and Blevins, R. A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614.
- 52 Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158.
- 53 Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
- 54 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Panu, N. S. et al. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.
- 55 Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486.
- 56 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- 57 Vagin, A. and Teplov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25.
- 58 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255.
- 59 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
- 60 Vogt, G., Woell, S. and Argos, P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **269**, 631–643.
- 61 Bayer, E. A., Chanzy, H., Lamed, R. and Shoham, Y. (1998) Cellulose, cellulases and cellosomes. *Curr. Opin. Struct. Biol.* **8**, 548–557.
- 62 Jindou, S., Xu, Q., Kenig, R., Shulman, M., Shoham, Y., Bayer, E. A. and Lamed, R. (2006) Novel architecture of family-9 glycoside hydrolases identified in cellosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum*. *FEMS Microbiol. Lett.* **254**, 308–316.
- 63 Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. and Steitz, T. A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751.
- 64 Shaw, A., Bott, R., Vonrhein, C., Bricogne, G., Power, S. and Day, A. G. (2002) A novel combination of two classic catalytic schemes. *J. Mol. Biol.* **320**, 303–309.
- 65 Pereira, J. H., Chen, Z., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A. and Adams, P. D. (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. *J. Struct. Biol.* **172**, 372–329.
- 66 Varrot, A., Frandsen, T. P., Driège, H. and Davies, G. J. (2002) Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio- β -cellobioside, at 1.9 Å. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2201–2204.

Received 16 May 2011/22 August 2011; accepted 1 September 2011
 Published as BJ Immediate Publication 1 September 2011, doi:10.1042/BJ20110869



SUPPLEMENTARY ONLINE DATA

Dissecting structure–function–stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168

Camila R. SANTOS^{*1}, Joice H. PAIVA^{*1}, Maurício L. SFORÇA^{*}, Jorge L. NEVES^{*}, Rodrigo Z. NAVARRO^{*}, Júnio COTA[†], Patrícia K. AKAO^{*}, Zaira B. HOFFMAM[‡], Andréia N. MEZA^{*}, Juliana H. SMETANA^{*}, Maria L. NOGUEIRA^{*}, Igor POLIKARPOV[‡], José XAVIER-NETO^{*}, Fábio M. SQUINA[†], Richard J. WARD[§], Roberto RULLER[†], Ana C. ZERI^{*} and Mário T. MURAKAMI^{*2}

^{*}Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas SP, Brazil, [†]Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas SP, Brazil, [‡]Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil, and [§]Departamento de Química, Universidade de São Paulo, Ribeirão Preto, Brazil

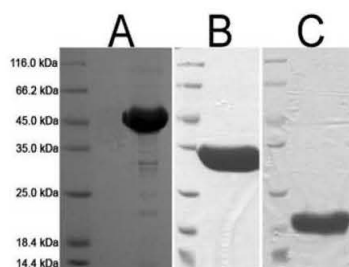


Figure S1 SDS/PAGE analysis of the purified BsCel5A constructions

(A) WT, (B) CC and (C) CBM3.

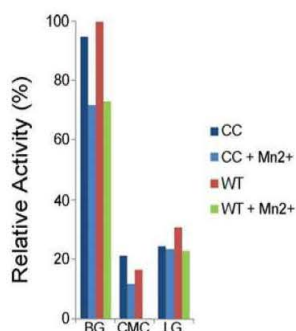


Figure S2 Relative activity of BsCel5A (WT) and the CC on different substrates

Substrates used were β -glucan (BG), carboxymethylcellulose (CMC) and lichenan (LG).

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed (email mario.murakami@lnbio.org.br).

Structural factors and atomic co-ordinates of the *Bacillus subtilis* cellulase 5A catalytic core have been deposited in the Protein Data Bank under accession codes 3PZV (Form I), 3PZU (Form II) and 3PZT (Form II*). NMR data of the *Bacillus subtilis* cellulase 5A CBM3 were deposited in the Biological Magnetic Resonance Bank and Protein Data Bank under codes 17399 and 2L8A respectively.

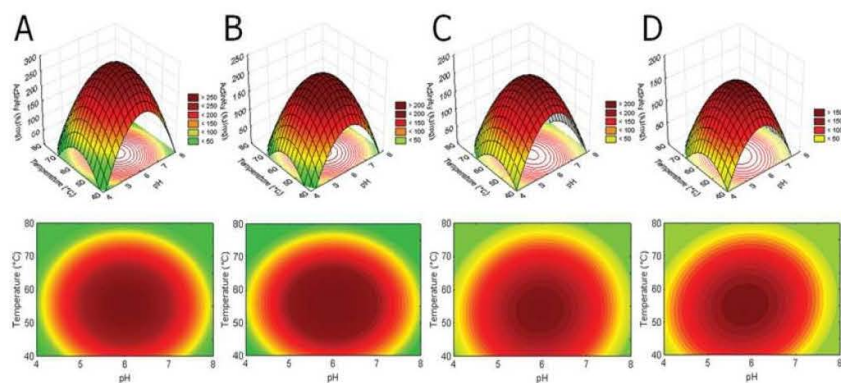


Figure S3 Response surface analysis

Effect of pH and temperature on the glucanase activity of BsCel5A (**A**, without Mn^{2+} and **B**, with 10 mM Mn^{2+}) and on the catalytic core (**C**, without Mn^{2+} and **D**, with 10 mM Mn^{2+}).

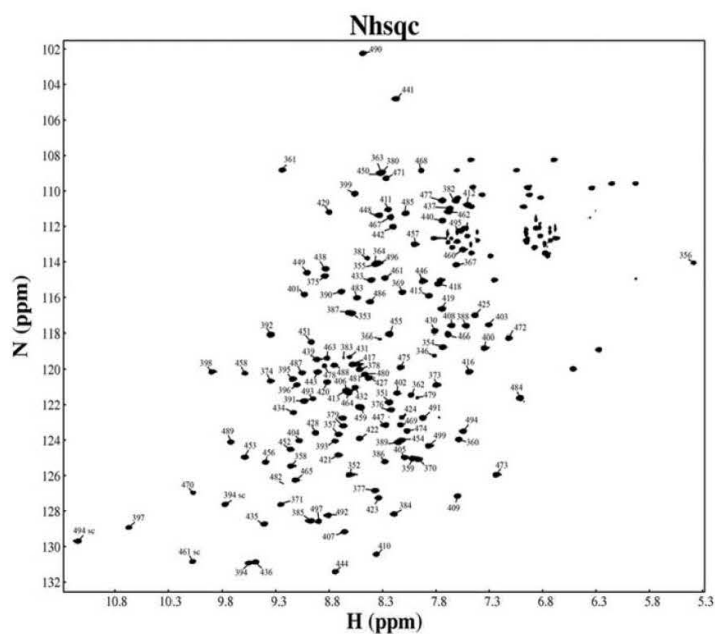


Figure S4 Assigned ^{15}N - ^1H -HSQC spectrum of the CBM3 domain

	392		406	453		473
<i>Bsu</i>	RYWY	ANKGQN-FDC...	QLRLHNDWSN-Y	QSDY	SFF	
<i>Bst</i>	RYWY	ANKGQN-FDC...	QLRLHNDWSN-Y	QSDY	SFF	
<i>Bli</i>	RYWY	ANKGQN-FDC...	QLRLHNDWSN-Y	QSDY	SFL	
<i>Bpu</i>	RYWY	ANKGQN-VDC...	QLRLHNDWSN-Y	QSDY	SFF	
<i>Bam</i>	RYWY	ANKGQN-FDC...	QLRLHNDWSN-Y	QSDY	SFF	
<i>4TF4</i>	RYWFL	DE-----GV...	QFRIAGGPG---	WDPSN	WSFQ	
<i>1G87</i>	KYFMD	LSIVAA--GI...	QFRIAPQGTTYW	NPKN	DFSVD	
<i>2WOB</i>	RYWFL	SGNEQN	NFVC...	PFRIEG--AAE-Y	DTDDYSYN	
<i>1G43</i>	RYYYT	QADKPLTFWC...	QTRFARN	WSN-FDQSN	WSYT	
<i>1NBC</i>	RYYYT	VGQKDQTFWC...	QGRFAKNDWSN-Y	TQSN	YSFK	
	40		55	110		130

Figure S5 Multiple sequence alignment of the region corresponding to the classical calcium-binding site in CBMs belonging to family 3

CBM3a (PDB codes 1NBC and 1G43), CBM3b⁺ (PDB code 2WOB), CBM3c (PDB codes 1G87 and 4TF4) and CBM3d sequences: *Bsu*, *B. subtilis* (GB: NP_389695.2); *Bst*, *Bacillus stearothermophilus* (GB: ADQ21451.1); *Bli*, *Bacillus licheniformis* (GB: AAP51020.1); *Bpu*, *Bacillus pumilus* (GB: ACY72384.1); and *Bam*, *Bacillus amyloliquefaciens* (GB: ABS70711.1). The residues that interact with the calcium ion by side-chain or main-chain atoms are highlighted in green or cyan respectively. The residue conserved in all CBM3 subclasses is in yellow and those conserved only in CBM3d members are in red. The numbers at the top correspond to the BsCel5A numbering and the numbers at the bottom correspond to the 1NBC numbering.

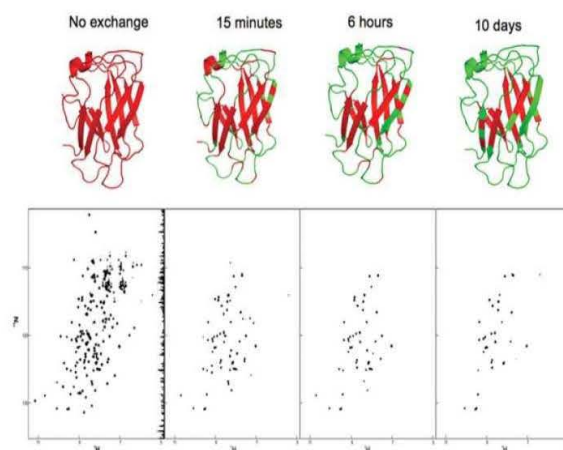


Figure S6 ^{15}N - ^1H -HSQC spectra (bottom panels) of hydrogen-deuterium exchange experiments at different time intervals and schematic representation of the structure (top panels) indicating more (green) or less (red) solvent-exposed residues

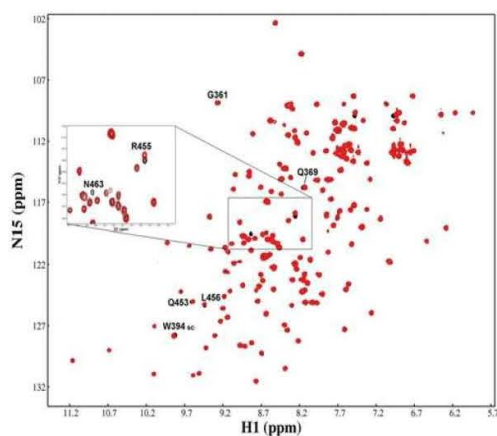


Figure S7 ^{15}N - ^1H -HSQC spectra of CBM3 as isolated (black) and in the presence of cellopentaose (overlaid in red)

The residues perturbed by the ligand are labelled. The inset shows a magnification of a representative spectral region with chemical shift changes.

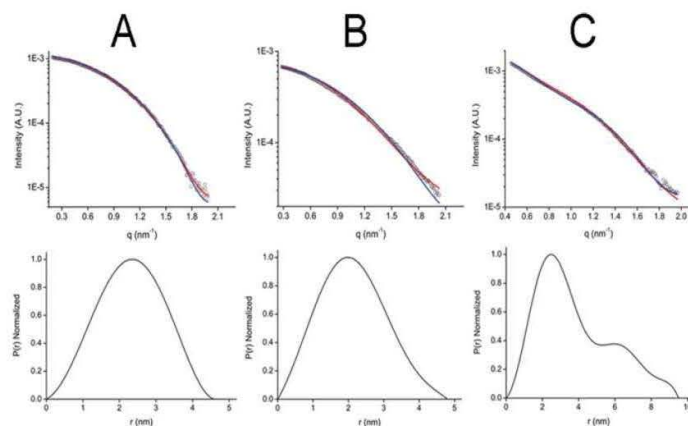


Figure S8 SAXS scattering curves (top panels) and normalized distance distribution $p(r)$ functions (bottom panels)

(A) Catalytic core, (B) CBM3 and (C) full-length protein.

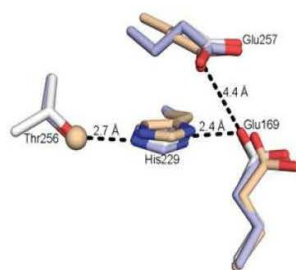


Figure S9 Schematic representation of the three possible catalytic triads in cellulases belonging to family 5A

Carbon atoms are coloured in white, blue and brown for type A, B and C respectively.

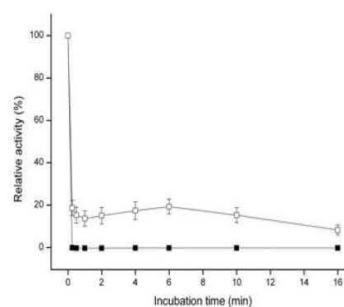


Figure S10 Thermotolerance of BsCel5A at 90 °C in the absence (■) or in the presence (□) of manganese ion

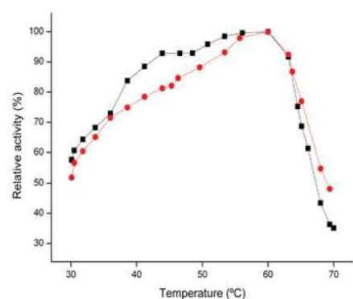


Figure S11 Temperature-dependent activity profile of BsCel5A in the presence (black) and in the absence (red) of manganese ion

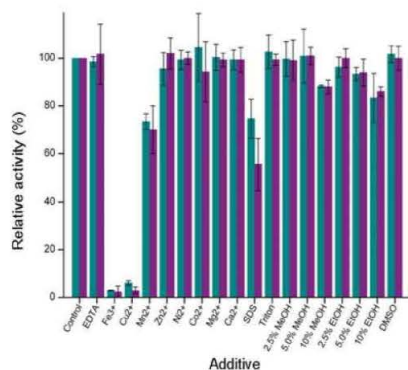


Figure S12 Effect of different additives on the endo-1,4-β-glucanase activity of BsCel5A

Table S2 Summary of the NMR data for *Bacillus* CBM3

RMSD, root mean square deviation.

Factor	Result
NOES	
Total	1373
Sequential	672
Medium-range distance	105
Long-range distance	596
Ramachandran plot	
Favoured region (%)	97.4
Generously allowed region (%)	2.6
Disallowed region (%)	0.0
RMSD (Å)	
Backbone atoms (full-length*)	0.91
Backbone atoms (structured†)	0.66
Non-hydrogen atoms (full-length)	1.66
Non-hydrogen atoms (structured)	1.53

*Full-length CBM3d comprises the amino acid segment from 351 to 499.

†Structured regions are the segments 354–359, 373–378, 388–395, 402–408, 417–422, 431–438, 452–457, 467–473, 484–488 and 491–496.

Table S1 Central composite design matrix (2^2) and glucanase activity after 10 min of incubation

Run number	Coded levels ($X_1 = \text{pH}$; $X_2 = T$)		Actual levels ($X_1 = \text{pH}$; $X_2 = T$)		Glucanase activity (IU/mg protein)
	X_1	X_2	X_1	X_2	
1	−1	−1	4.6	74.2	139.8
2	1	−1	7.4	74.2	170.9
3	−1	1	4.6	94.8	175.2
4	1	1	7.4	94.8	204.2
5	−1.414	0	4.0	84.5	68.6
6	1.414	0	8.0	84.5	154.6
7	0	−1.414	6.0	70.0	159.7
8	0	1.414	6.0	99.0	238.9
9	0	0	6.0	84.5	234.6
10	0	0	6.0	84.5	241.3
11	0	0	6.0	84.5	247.8
12	0	0	6.0	84.5	247.1

Received 16 May 2011/22 August 2011; accepted 1 September 2011

Published as BJ Immediate Publication 1 September 2011, doi:10.1042/BJ20110869

Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- β -D-mannanase



Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

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

Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- β -D-mannanase

Camila Ramos dos Santos^a, Joice Helena Paiva^a, Andreia Navarro Meza^a, Junio Cota^b, Thabata Maria Alvarez^b, Roberto Ruller^b, Rolf Alexander Prade^c, Fabio Marcio Squina^b, Mario Tyago Murakami^{a,*}

^a Laboratório Nacional de Biotecnologia (LNBio), Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil

^b Laboratório Nacional de Ciência e Tecnologia do Biotanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil

^c Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

ARTICLE INFO

Article history:

Received 9 July 2011

Received in revised form 4 November 2011

Accepted 18 November 2011

Available online xxxx

Keywords:

Mannan endo-1,4- β -mannosidase

Glycoside hydrolase family 5

Carbohydrate binding module 27

Thermotoga petrophila RKU-1

Crystal structure

Substrate recognition

ABSTRACT

The breakdown of β -1,4-mannoside linkages in a variety of mannan-containing polysaccharides is of great importance in industrial processes such as kraft pulp delignification, food processing and production of second-generation biofuels, which puts a premium on studies regarding the prospection and engineering of β -mannanases. In this work, a two-domain β -mannanase from *Thermotoga petrophila* that encompasses a GH5 catalytic domain with a C-terminal CBM27 accessory domain, was functionally and structurally characterized. Kinetic and thermal denaturation experiments showed that the CBM27 domain provided thermo-protection to the catalytic domain, while no contribution on enzymatic activity was observed. The structure of the catalytic domain determined by SIRAS revealed a canonical $(\alpha/\beta)_8$ -barrel scaffold surrounded by loops and short helices that form the catalytic interface. Several structurally related ligand molecules interacting with TpMan were solved at high-resolution and resulted in a wide-range representation of the subsites forming the active-site cleft with residues W134, E198, R200, E235, H283 and W284 directly involved in glucose binding.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Mannan endo-1,4- β -D-mannosidase or 1,4- β -D-mannan mannanohydrolase (EC 3.2.1.78), commonly referred as β -mannanase, catalyzes the hydrolysis of β -1,4-mannoside linkages in various mannan-containing polysaccharides, such as glucomannans and galactomannans (Stålbrand et al., 1993; de Vries and Visser, 2001). Degradation of these polysaccharides represents a key step for a number of industrial applications including delignification of kraft pulps (Tenkanen et al., 1997; Montiel et al., 2002), food processing (Sachslehner et al., 2000; Dhavan and Kaur, 2007) and production of second-generation biofuels (Dhavan and Kaur, 2007). In general, these biotechnological processes such as biomass pre-treatments, are performed under extreme environmental

conditions regarding pH, osmolarity and temperature. Thus, β -mannanases being stable and functional at high temperatures offer substantial techno-economical advantages.

In addition to their biotechnological relevance, mannan-degrading enzymes also participate in a number of biological processes such as fruit ripening (Pressey, 1989), seed germination (Black, 1996) and remodeling of plant cell walls (reviewed in Schröder et al., 2009). These enzymes have also been used in structural characterization of polysaccharides having β -mannosidic linkages and sequencing of heteropolysaccharides and carbohydrates attached to glycoproteins (Dhavan and Kaur, 2007).

Thermotoga petrophila strain RKU-1 (T) is a hyperthermophilic bacterium isolated from the Kubiki oil reservoir in Niigata (Japan) that grows optimally at 80 °C (Takahata et al., 2001). Some hyperthermostable enzymes produced by this microorganism have demonstrated great potential for industrial applications and served as models for investigating structure-function-stability relationships in multidomain glycosyl hydrolases (Santos et al., 2010, 2011; Squina et al., 2010; Cota et al., 2011).

The β -mannanase from *T. petrophila* RKU-1, TpMan, consists of a CaZy GH5 catalytic core connected to a CBM27 accessory domain by an 100-residue-long linker. To date, only six structures of GH5 endo- β -1,4-mannanases have been solved: *Thermobifida fusca*

Abbreviations: Mannanases: TpMan, from *Thermotoga petrophila*; TfMan, from *Thermobifida fusca*; TrMan, from *Trichoderma reesei*; LeMan, from *Lycopersicon esculentum*; CmMan, from *Cellvibrio mixtus*; CJMan, from *Cellvibrio japonicus*; BaMan, from *Bacillus agaradhaerens*; BsMan, from *Bacillus subtilis*; VsMan, from *Vibrio* sp. strain MA-138.

* Corresponding author. Address: Laboratório Nacional de Biotecnologia (LNBio), Centro Nacional de Pesquisa em Energia e Materiais, Rua Giuseppe Maximo Solfaro, 10000 Campinas, 13083-970 SP, Brazil. Fax: +55 19 3512 1004.

E-mail address: mario.murakami@lnbio.org.br (M.T. Murakami).

1047-8477/\$ - see front matter © 2011 Elsevier Inc. All rights reserved.

doi:10.1016/j.jsb.2011.11.021

Please cite this article in press as: Santos, C.R.d., et al. Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- β -D-mannanase. J. Struct. Biol. (2011), doi:10.1016/j.jsb.2011.11.021

KW3 (TrMan, PDB codes 1BQC, 2MAN, 3MAN; Hilge et al., 1998), *Trichoderma reesei* (TrMan, PDB codes 1QNO, P, Q, R, S; Sabini et al., 2000), *Bacillus* sp. JAMB-602 (PDB code: 1WKY; Akita et al., 2004), *Lycopersicon esculentum* (LeMan, PDB code 1RH9; Bourgault et al., 2005), *Bacillus agaradhaerens* (BaMan, PDB code 2WHJ; Tailford et al., 2009), and *Bacillus* sp N16-5 (PDB code: 3JUG; not published). From those, only LeMan and TrMan share some sequence identity (~30%) with TpMan. Both enzymes are non-thermophilic proteins and are poorly related to bacterial β -mannanases because of their low sequence similarity and high degree of glycosylation. Several β -mannanases from both eukaryotic and prokaryotic organisms display high affinity for glucomannans (Pressey, 1989; Tenkanen et al., 1997; Hilge et al., 1998; Tailford et al., 2009; Tanaka et al., 2009); however, the identification of glucose binding sites and their respective mode of interaction remain unknown.

Thus, in order to shed light on the molecular basis of thermal stability and substrate specificity of thermostable bacterial β -mannanases, we performed an extensive biochemical and structural characterization of a β -mannanase from the hyperthermophilic bacterium *T. petrophila* RKU-1. Unfolding studies of deletion mutants demonstrated that the CBM27 accessory domain confers thermo-protection to the catalytic core. The TpMan catalytic core structure was solved by the SIRAS method and several ligand-protein structures were obtained at high resolution revealing the details of the substrate-binding channel and the mechanism of glucose binding.

2. Material and methods

2.1. Cloning, protein expression and purification

The catalytic domain of the endo-1,4- β -D-mannanase (residues 32–393) from *T. petrophila* RKU-1 (TpMan Δ CT, GenBank Accession Code: YP_001245126) was cloned, expressed and purified according to Santos et al. (2010). Briefly, *Escherichia coli* BL21(DE3) Δ SlyD cells harboring the TpMan Δ CT/pET-28a vector were grown in selective LB medium to an OD_{600nm} of 0.8 and 0.5 mM IPTG was added to induce heterologous expression for 4 h. The harvested cells were lysed and TpMan isolated from the soluble fraction by nickel-affinity and size-exclusion chromatographies. Sample quality was assessed by polyacrylamide gel electrophoresis under denaturing conditions (Laemmli, 1970) and dynamic light scattering experiments.

Prior to crystallization, the sample was dialyzed against 25 mM Tris-HCl pH 7.5 and concentrated to 12 mg ml⁻¹ with Amicon centrifugal ultrafiltration units (Millipore).

2.2. Enzyme assays

Mannan endo-1,4- β -mannosidase activity was determined using a modified 3,5-dinitrosalicylic acid (DNS) method (Miller, 1959). The reaction was performed by mixing 10 μ L of the diluted enzyme with 50 μ L of mannan or glucomannan at 5 mg/mL for 5 min at 85 °C. The reaction was stopped by the addition of 100 μ L DNS reagent followed by boiling the sample in a ~100 °C water bath for 5 min. A response surface methodology was employed to optimize the reaction conditions for TpMan using glucomannan as substrate. The variables pH and temperature were used to design a central composite ($k = 2$) with four central points, totalizing 12 experiments (Table S1) (Myers and Montgomery, 2001; Cota et al., 2011). One unit of mannan endo-1,4- β -mannosidase activity was defined as the amount of enzyme needed to release 1 μ mole of mannose equivalents per minute. All experiments were done in triplicate, and average values are reported.

2.3. Capillary zone electrophoresis of oligosaccharides

The oligosaccharide 1,4- β -D-mannohexaose (Megazyme) was derivatized with 8-aminopyrene-1,3,6-trisulfonic acid (APTS) by reductive amination (Chen and Evangelista, 1995). Capillary zone electrophoresis (CZE) was performed on a P/ACE MQD instrument (Beckman Coulter) equipped with laser-induced fluorescence detection. A fused-silica capillary (TSP050375, Polymicro Technologies) of internal diameter of 50 μ m and total length of 31 cm was used as separation column for oligosaccharides. Samples were injected by application of 0.5 psi pressure for 0.5 s. Electrophoresis conditions were 15 kV/70–100 μ A at a controlled temperature of 20 °C. Oligomers labeled with APTS were excited at 488 nm and emission was collected through a 520 nm band pass filter.

2.4. Circular dichroism spectroscopy

Far-UV CD spectra of the full-length protein (TpMan) and the truncated catalytic domain (TpMan Δ CT) were measured between 190 and 260 nm in 10 mM phosphate buffer, pH 6.0 at 25 °C with a Jasco J-810 spectropolarimeter (Hachioji City, Tokyo, Japan) using a 2-mm-path-length cuvette and a protein concentration of 0.1 mg/mL. For each experiment, a total of 20 spectra were collected, averaged and corrected by subtraction of a buffer blank and ellipticity was reported as the mean residue molar ellipticity ($[\theta]$; deg cm² dmol⁻¹).

In order to investigate the thermal stability, CD spectra were analyzed at different temperatures ranging from 20 to 100 °C. Thermal unfolding was monitored by far-UV CD at 222 nm. The reversibility of thermal denaturation was verified by cooling the denatured sample and subsequent reheating.

2.5. Crystallization, iodine derivatization and data collection

Since attempts to crystallize the intact TpMan were unsuccessful, due to the inherent flexibility of the long linker, crystallization experiments were carried out with a truncated form (TpMan Δ CT) in which the linker and CBM27 domain were removed.

TpMan Δ CT crystals were obtained by the sitting drop vapor diffusion method at 293 K by mixing 0.5 μ L of protein solution with an equal volume of reservoir solution consisting of 0.1 M citrate pH 5.5, 1 M ammonium phosphate and 0.2 M sodium chloride. The native structure was obtained when a single crystal was directly flash-cooled in a 100 K nitrogen stream without the addition of cryoprotectants. The iodine derivative was prepared according to the quick cryo-soaking method (Dauter et al., 2000) by soaking TpMan Δ CT crystals in the cryosolution (12.5% (v/v) glycerol) with 0.5 M sodium iodide, for 1 min. Glucose was incorporated by soaking crystals in the reservoir solution containing 0.5 M glucose, for 5 min. Likewise the maltose complex was obtained by adding 0.1 M maltose in the drop solution. The J222 crystal form was obtained from a solution containing 0.1 M phosphate pH 4.2, 5% (w/v) PEG-1000, 36% (w/v) ethanol and 10% (v/v) glycerol (Santos et al., 2010).

Diffraction intensities were measured at the tunable-energy MX2 beamline (Brazilian National Synchrotron Light Source, Campinas, Brazil). All datasets were collected with X-ray energy set to 8500 eV except for the derivative crystal whose data were collected at 7797 eV in order to increase the anomalous signal. Data were indexed, integrated and scaled using the HKL2000 package (Otwinowski and Minor, 1997).

2.6. Structure determination and refinement

TpMan Δ CT structure was determined by single isomorphous replacement anomalous scattering (SIRAS) method using the

native and iodide derivative datasets (Table 1). The positions of thirteen iodine atoms with occupancy over 0.4 were obtained by using the SHELXD program (Schneider and Sheldrick, 2002). The iodine substructure was further used to estimate phases using the SHELXE program, which also employs solvent flattening method to improve the quality of phases (Sheldrick, 2002). Automated model building using the amino acid sequence of TpMan (GenBank code YP_001245126) was performed with the ARP/wARP program (Langer et al., 2008). Around 99% of the molecule was automatically traced. Initial cycles of refinement involved a restrained and overall *B*-factor refinement using the REFMAC5 program (Murshudov et al., 1997). After each cycle of refinement, the model was inspected and manually adjusted to correspond to computed σA -weighted ($2F_o - F_c$) and ($F_o - F_c$) electron density maps using the COOT program (Emsley and Cowtan, 2004). Water molecules were manually added at positive peaks above 3.0σ in the difference Fourier maps, taking into consideration hydrogen-bonding potential. Crystal structures with resolution better than $1.85 \text{ \AA}$ resolution were subjected to restrained and anisotropic refinement in later cycles. The I222 form and complex structures were solved by molecular-replacement method using the refined native structure as template with the MOLREP program (Vagin and Teplyakov, 1997). Data collection and refinement statistics are summarized in Table 1.

3. Accession Numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with the Accession Codes 3PZG, 3PZ9, 3PZM, 3PZN, 3PZI, 3PZO and 3PZQ.

4. Results and discussion

4.1. Functional characterization

TpMan was incubated with APTS-labeled mannohexaose and cleavage products resolved by capillary zone electrophoresis (Fig. S1). The major cleavage product was mannotriose with detection of minor amounts of mannobiose and mannotetraose, confirming that TpMan is an endo- β -1,4-mannanase.

A Central Composite Design with two variables (pH and temperature) was used to optimize TpMan reaction conditions (Table S1). Both wild-type (WT) and catalytic-domain proteins established maximal activity at the acidic pH range of 4.5–6.5 (Fig. S2). The truncated catalytic domain had optimal activity at a lower temperature range (79–89 °C) when compared to the WT protein (81–93 °C) (Figs. S2B and D). This difference may be an indication that the full-length WT protein is more thermotolerant than the truncated catalytic domain. A similar thermoprotection effect was also observed for Man26 from the thermophilic bacterium *Caldicellulosiruptor* strain Rt8B.4, which contains two tandem N-terminal CBM27s (Roske et al., 2004). The deletion of both accessory domains reduced the optimal temperature for catalysis from 80 to 60 °C (Roske et al., 2004).

4.2. How useful is the CBM27 accessory domain for catalytic function?

A kinetic study was performed with full-length WT and catalytic-domain proteins using 1,4- β -D-mannan and konjac glucomannan as substrates (Table 2). Both TpMan and TpMan Δ CT had a clear preference for konjac glucomannan, exhibiting a significant higher catalytic efficiency (Table 2). Preference for konjac glucomannan has also been shown for endo-1,4- β -mannosidase from *Aspergillus niger* BK01 (Bien-Cuong et al., 2009), which shares 32% of amino acid sequence identity with TpMan. Moreover, the

A. niger mannosidase contains no accessory domain indicating that the molecular determinants for glucomannan specificity reside exclusively in the catalytic domain. Indeed, deletion of the CBM27 domain from TpMan had no effect on enzyme specificity (Table 2). In contrast, Man5C from *Vibrio* sp., a GH5 mannanase with a CBM27 accessory domain, showed strong positive interaction of the accessory domain with enzyme kinetics improving catalytic efficiency (k_{cat}/K_m) (Tanaka et al., 2009). On the other hand, Man5C was isolated from a mesophilic bacterium and shares only 37% of amino acid sequence identity with TpMan thus suggesting distinctive roles for the accessory domain.

Unconventionally, the deletion of the accessory domain has no statistically relevant effect on the catalytic efficiency upon both mannan and konjac glucomannan (Table 2). For both substrates, the truncated catalytic domain showed higher V_{max} . In addition, substrate-binding assays showed that TpMan was not adsorbed by insoluble mannan corroborating our kinetic results (results not shown). These data suggest that the accessory domain in TpMan contributes neither to increase catalytic efficiency nor for substrate specificity.

4.3. The CBM27 accessory domain is critical for thermal stability

The apparent minor effect of the CBM27 accessory domain on TpMan enzymatic function as well as the lower temperature optimum of the truncated catalytic domain led us to examine a possible role of CBM27 in enzyme stability. Circular dichroism of full-length WT and truncated catalytic-domain proteins showed typical α/β fold spectra (Fig. 1A), which are in full agreement with crystallographic data. The full-length protein displayed a slighter global residual molar ellipticity when compared to the truncated catalytic domain (Fig. 1A). This effect can be attributed to the built-in unstructured 100-amino acid residue-long linker in the full-length protein (Fig. 3S).

To evaluate the role of CBM27 on enzyme stability, thermal denaturation experiments were performed. As predicted, the deletion of the CBM27 accessory domain reduced the melting temperature from 100 to 88 °C (Fig. 1B), confirming the importance of CBM27 for thermal stability. Thus, based on these biochemical and biophysical evidences, the accessory domain of TpMan should be properly referred as a thermostabilizing domain instead of a carbohydrate-binding module.

4.4. TpMan structure

TpMan shares low amino acid sequence identity with other structurally determined mannanases (~30%). The most similar mannanases are LeMan, TrMan and CmMan (from *Cellvibrio mixtus*, PDB codes 1UUQ and 1UZ4), which present 33%, 33% and 26% of identity and structural superposition rmsd values of 1.6, 1.9 and 2.0 Å, respectively (Fig. 2A, left). CmMan is an exo-cleaving enzyme with only one substrate binding subsite (–1) (Dias et al., 2004). BaMan and TfMan, are endo-cleaving mannanases with a clear ability to cleave glucomannan, but are poorly related to TpMan.

The three-dimensional structure of the TpMan catalytic domain folds into a classical (β/α)₈ barrel (Fig. 2A). The C-terminal barrel core of β -strands is surrounded by loops and short helices that form a flat molecular surface around the catalytic groove. The active-site cavity has a volume of $1.507 \text{ \AA}^3$ (Fig. S4A) and is populated by a number of aromatic and acidic amino acids including catalytic residues, E198 and E317 (Figs. S4B and S5). In the TpMan active-site geometry, E198 is hydrogen bonded to H278 and W134 side chains, whereas E317 is coordinated by R71 and Y280 side chains (Fig. S5). Residues Y45, W73, N197, R200, E235, W284 and W350 forming the catalytic interface are absolutely conserved within

ARTICLE IN PRESS

4

Camila Ramos dos Santos et al. / Journal of Structural Biology xxx (2011) xxx–xxx

Table 1
Data and refinement statistics for TpManΔCT structures.

	Crystalline form I	Crystalline form II	Iodine	Glucose	Maltose	Maltose + Glycerol
<i>Data collection</i>						
Space group	I222	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
<i>Cell dimensions</i>						
a, b, c (Å)	91.03, 89.97, 97.90	55.46, 83.35, 92.18	55.55, 83.63, 92.66	55.43, 83.23, 91.92	55.11, 83.19, 91.52	55.29, 83.30, 92.08
Resolution (Å)	30.00–1.40 (1.45–1.40) ^a	30.00–1.42 (1.47–1.42)	30.00–1.82 (1.89–1.82)	30.00–1.55 (1.61–1.55)	30.00–1.92 (1.99–1.92)	30.00–1.55 (1.61–1.55)
Mosaicity (°)	0.9	0.3	0.5	0.4	1.0	0.6
R _{sym} (%)	5.1 (33.7)	7.2 (56.8)	9.0 (31.1)	7.3 (52.2)	9.5 (50.9)	7.1 (56.0)
I/σ (I)	40.3 (4.9)	30.2 (2.2)	24.5 (8.9)	23.2 (2.9)	25.0 (5.0)	24.7 (2.4)
Completeness (%)	96.5 (78.7)	98.4 (86.0)	99.3 (98.2)	99.9 (99.1)	99.6 (99.3)	97.7 (84.5)
Unique reflections	76,427 (6160)	80,074 (6891)	39,216 (3790)	62,462 (6068)	32,825 (3225)	61,007 (5208)
Multiplicity	12.1 (8.1)	9.1 (5.3)	13.6 (12.5)	6.1 (5.3)	10.0 (10.5)	7.1 (5.7)
<i>Structure refinement</i>						
R _{work} /R _{free} (%)	15.4/18.4	14.5/17.8	–	14.0/18.3	16.3/21.0	17.1/20.2
Overall B factor	17.4	17.8	–	17.0	18.5	17.4
No. protein atoms	2950	2943	–	2937	2942	2944
No. water molecules	319	417	–	370	265	291
Ligands	TRS–GOL	–	–	GLC	MAL	MAL–GOL
RMSD bond lengths (Å)	0.030	0.030	–	0.027	0.025	0.033
RMSD bond angles (°)	2.301	2.229	–	2.073	1.997	2.639
<i>Ramachandran plot</i>						
Most favored regions (%)	96.9	96.6	–	96.9	96.1	96.9
Allowed regions (%)	3.1	3.4	–	3.1	3.9	3.1
PDB entry code	3PZG	3PZ9	–	3PZI	3PZO	3PZQ

^a Values in parentheses are for highest-resolution shell.**Table 2**
Apparent kinetic parameters of TpMan and TpManCT.

	β-1,4-Mannan		Glucomannan (Konjac)	
	TpManΔCT	TpMan	TpManΔCT	TpMan
V _{max} (IU/mg)	250 ± 10	100 ± 5	203 ± 6	114 ± 4
K _m (mg/mL)	1.71 ± 0.16	1.52 ± 0.23	0.80 ± 0.11	1.02 ± 0.19
K _{cat} (s ^{−1})	178	126	148	145
K _{cat} /K _m	104	83	165	142

all other β-mannanases with significant amino acid sequence identity (>30%) to TpMan (Fig. 2B).

In order to determine the mechanisms that guide substrate binding interactions of a hyperthermostable β-mannanase, six protein crystals complexed with natural and mimetic oligosaccharides were obtained (Fig. S6). The seven TpMan structures, one native and six carbohydrate complexes shared high overall similarity (RMSD < 0.4 Å); however, local differences were observed in side chain conformations of residues at the protein surface due to distinct crystalline contacts.

4.5. Protein–ligand interactions

Glucose and maltose ligands were incorporated using the soaking method, which yielded high-resolution structures enabling accurate interpretations of binding modes (Fig. S6). The preference for glucomannan as substrate by TpMan (Table 2) motivated the soaking with glucose (GLC) and maltose (MAL) molecules, since structural determinants for glucose binding remain unknown.

In the glucose–TpMan complex crystal, the monosaccharide occupied the +1 subsite and displayed an unexpected orientation in which the reducing O1 atom was pointing to the negative binding subsite (Fig. 3A). This steric positioning was stabilized by several hydrogen bonds with residues E198, R200, E235, H283 and W284.

When TpMan crystals were soaked in the reservoir solution containing 100 mM maltose, three maltose (MAL) molecules were incorporated; one at a remote site participating in crystal packing (Fig. S7A) with no apparent biological relevance and two located at the active-site pocket (Fig. 3B). MAL1 was found at −3 and −2 subsites, and MAL2 was attached to +1 and −1 subsites (Fig. 3B). Both ligands made extensive interactions with the catalytically-relevant residues and aromatic gate-keepers. Analogously to the glucose complex, MAL1 also adopted a pose in which the free O1 atom was turned to the substrate recognition area (Fig. 3B).

Residue N197 is conserved in all GH5 β-mannanases (Sabini et al., 2000) and was found in LeMan structure interacting with the −1 mannosyl group (Bourgault et al., 2005). Mutating this residue to alanine caused loss of activity in *E. chrysanthemi* mannanase (Bortoli-German et al., 1995). In TpMan complexes with glucose and maltose, this residue was not making any contacts with ligands, indicating that N197 is exclusively involved in mannose recognition.

A complex with carbohydrate mimetics including hydroxylated organic compounds, tris and glycerol, were also obtained. They were found at the active-site pocket spanning −3 to +1 subsites and a detailed interaction map is shown in Fig. S7B. The TpMan crystal complex with maltose and glycerol showed a similar configuration observed in the maltose complex with the exception of MAL1, which was replaced by two glycerol molecules (Fig. S7C).

The presence of multiple ligands at the +1 subsite is compatible with the broad substrate specificity exhibited by TpMan, which hydrolyzes glucomannan, mannan and other mannan-containing polysaccharides. The ligand–protein complexes examined in this study revealed a thorough recognition and interaction map of glycosyl groups with residues forming the catalytic interface of TpMan.

4.6. Mapping of substrate-binding subsites

Comparative structural analysis of TpMan complexes with other mannanase complexes permitted us to map the substrate

Please cite this article in press as: Santos, C.R.d., et al. Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4-β-D-mannanase. J. Struct. Biol. (2011), doi:10.1016/j.jsb.2011.11.021

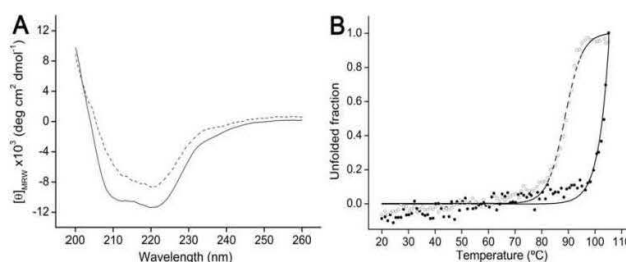


Fig. 1. Biophysical characterization of TpMan constructs. (A) Far-UV CD spectra and (B) thermal denaturation of wild-type (continuous line) and Δ CT (dashed line) proteins. Experimental denaturation data are shown as open and full circles for Δ CT and WT constructs, respectively.

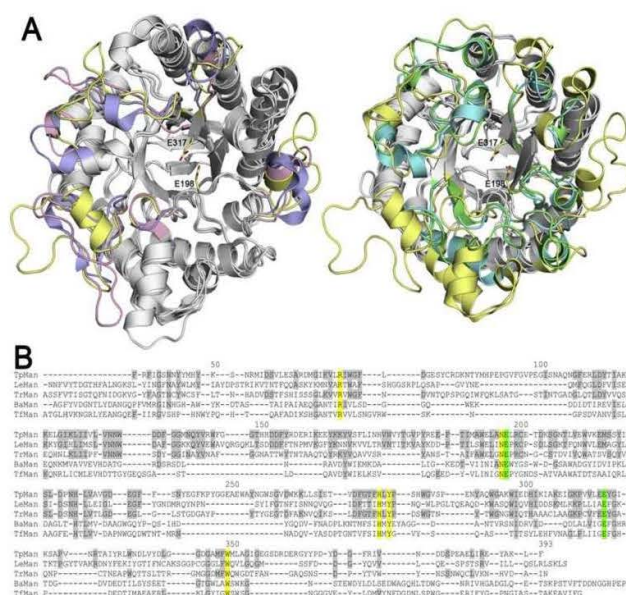


Fig. 2. Sequence and structural analysis of TpMan. (A) Structural superposition of TrMan (pink, left), LeMan (blue, left), BaMan (cyan, right), TfMan (green, right) on TpMan (yellow, both). The catalytic residues, E198 and E317, are drawn as sticks with carbon atoms in yellow. (B) Multiple sequence alignment of TpMan with other β -mannanases with known 3-D structures. Residues identical to those of TpMan are shaded in gray. The residues strictly conserved in all β -mannanases and cellulases belonging to GH5 family are highlighted in blue (catalytic residues) and yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

binding subsites (Fig. 4). The ligands found in TpMan structures covered -3 , -2 , -1 and $+1$ substrate-binding subsites and the -4 and other positive binding subsites were suggested by structural comparisons.

From the eight residues strictly conserved among GH5 mannanases and cellulases (Hilge et al., 1998), seven are present in TpMan. All conserved residues encompass -1 and $+1$ subsites including R71, N197, E198, H278, Y280, E317 and W350 (Fig. 4Bi). The -2 subsite formed by residues Y45, W73 and D371 is also conserved in LeMan and TrMan (Fig. 4Bii). However, in the $+1$ subsite of TpMan, there is a non-conserved histidine residue (H283), which is replaced by a glutamine in LeMan and by a serine in TrMan and with no equivalent residue in both TfMan and BaMan (Fig. 4Bi). In the native structure, H283 displayed a double side-chain conformation (Fig. S8), whereas in the sugar-complex structures, it adopts a defined rotamer conformation in proximity to the aromatic residues W280 and F373 (Fig. S8). In addition, this

residue participates in the coordination of MAL and GLC molecules, suggesting a role in substrate recognition and interaction.

The structure of TfMan with a mannotriose bound to the active site (PDB code 3MAN) revealed the residues forming -2 , -3 and -4 subsites (Hilge et al., 1998). In this complex structure, a flat surface was observed at the -4 subsite (Fig. S9D), whereas in the TpMan there is a narrow cleft at the equivalent position (Fig. S9A). The residues D84, K85, S359 and D360, which made contacts with the glycosyl residue at the -3 subsite in the TpMan-maltose complex, are closing the aperture and may establish the -4 subsite (Fig. 4Biii). This region is also different in LeMan and TrMan, which contain a wider cleft (Figs. S9B and S9C, respectively), indicating that the -4 subsite in TpMan is unique.

The -3 subsite is also different when compared to other mannanases, mainly because of residue Y370 that makes hydrophobic stacking interactions with the saccharide ring (Fig. 4Biv). This residue is absent in all other mannanases and in the TfMan structure a

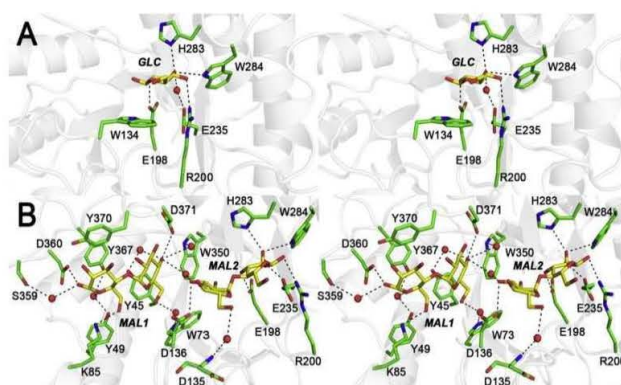


Fig.3. Stereo view of interactions between ligands and active-site residues of TpMan: (A) glucose and (B) maltose complexes.

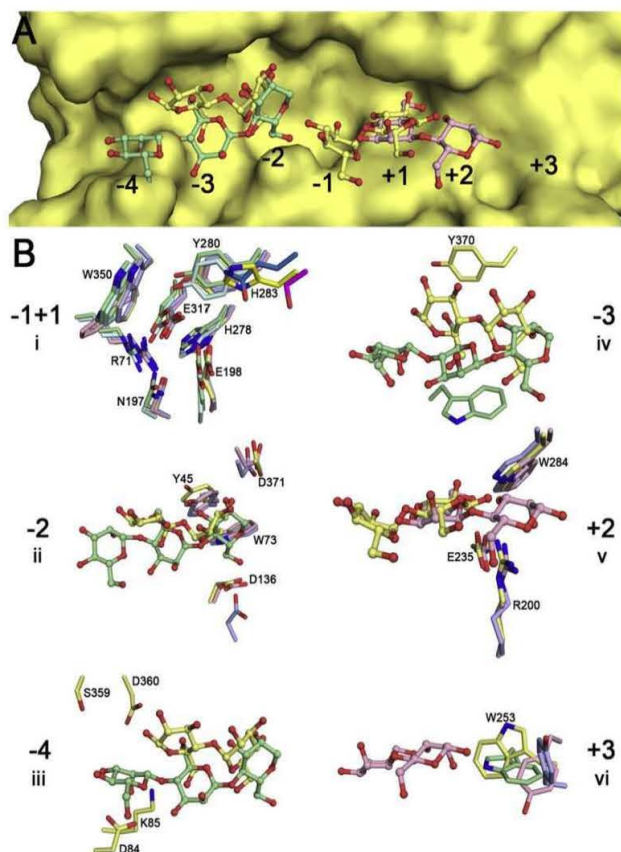


Fig.4. Comparison of TpMan substrate-binding sites with other endo- β -1,4-mannanases. (A) Molecular surface representation of TpMan with substrates bound to the active site. (B) Details of the residues forming the subsites: (i) -1 and +1, (ii) -2, (iii) -4, (iv) -3, (v) +2 and (vi) +3. The carbon atoms of residues and ligands are colored according to each endo- β -1,4-mannanase: TpMan and maltose in yellow, TrMan and mannosiose in pink, TrMan and mannatriose in green, LeMan in blue and BaMan in cyan.

similar role is played by residue W30 (Fig. 4Biv). TpMan has a phenylalanine residue, F137, at the corresponding position, which does not seem to interact with the substrate.

Despite no ligand was observed occupying the +2 subsite in several TpMan structures, residues R200, E235 and W284 located downstream to the +1 subsite are conserved in LeMan and TrMan (Fig. 4Bv). Furthermore, these residues made direct or water-mediated hydrogen bonds with GOL, GLC and MAL molecules that were found at the +1 subsite (Figs. 3 and S7), suggesting that they form the +2 subsite in TpMan.

The +3 subsite was proposed to be formed by residue W171 in TrMan (Hilge et al., 1998) and structural superposition with TpMan indicates that W253 is the equivalent residue (Fig. 4Bvi). In LeMan and TrMan, this residue is substituted by a tyrosine, and in BaMan, there is no corresponding residue.

In summary, multiple ligand-complex structures showed that TpMan can bind glycosyl groups at subsites ranging from –3 to +1. In LeMan and TrMan those residues involved in binding of glycosyl groups, comprising subsites –2, –1, +1 and +2, are conserved indicating a similar molecular recognition pattern of gluco-substituted mannans. Studies on BaMan indicated the –2 subsite in determining specificity however, in TpMan, TrMan and LeMan the –2 subsite as well as additional distal subsites are different, suggesting a distinctive structural mechanism.

5. Concluding remarks

The carbohydrate-binding modules have been associated with targeting the parental catalytic domain to the appropriate substrate enhancing the specific activity of the enzyme (Bolam et al., 1998; Boraston et al., 2003; Carrard et al., 2000). Our biochemical data suggested that CBM27 of TpMan is not implicated in substrate selectivity and has no significant effect on enzyme kinetics. On the other hand, the accessory domain confers thermal stability to the catalytic domain, indicating a structural role. Biophysical characterization of other *T. petrophila* glycosyl hydrolases containing CBMs showed similar thermo-protective effects. However, the mechanism by which CBM27, an independent domain connected to the catalytic domain by a 100-residue-long linker, interferes with enzyme stability remains unknown.

The crystal structure of the catalytic domain, solved by the SIRAS method, unveiled the molecular topology of the catalytic interface of TpMan and several ligand-complex structures allowed us to define each subsite forming the large substrate-binding channel and to identify residues relevant for glucomannan recognition. Glucose and maltose complexes revealed that subsites ranging from –3 to +1 are able to bind glycosyl groups so that the saccharide adopts an unusual orientation with the free O1 atom turned to the substrate-binding subsites.

The biochemical and structural dissection of this hyperthermostable GH5-CBM27 endo-1,4- β -mannanase from *T. petrophila* contributes to a better understanding of the molecular determinants for substrate binding, particularly for gluco-substituted mannans and presents an unconventional facet of some carbohydrate-binding modules more related to structural stability instead of enzyme function.

Acknowledgments

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo to MTM (10/51890-8) and FMS. RAP has funding from the Department of Energy, awards 06103-OKL and ZDJ-7-77608-01. We gratefully acknowledge the provision of time on the facilities MX2 beamline (LNLS), Robolab

(LNBio) and LEC (LNBio) at the National Center for Research in Energy and Materials (Campinas, Brazil).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2011.11.021.

References

- Akita, M., Takeda, N., Hirasawa, K., Sakai, H., Kawamoto, M., et al., 2004. Crystallization and preliminary X-ray study of alkaline mannanase from an alkaliphilic *Bacillus* isolate. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 1490–1492.
- Bien-Cuong, D., Thi-Thu, D., Berrin, J.-G., Haltrich, D., Kim-Anh, T., et al., 2009. Cloning, expression in *Pichia pastoris*, and characterization of a thermostable GH5 mannan endo-1,4- β -mannosidase from *Aspergillus niger* BK01. *Microb. Cell Factor* 8, 59.
- Black, M., 1996. Liberating the radicle: a case for softening up. *Seed Sci. Res.* 6, 39–42.
- Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., et al., 1998. *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* 331, 775–781.
- Boraston, A.B., Revett, T.J., Boraston, C.M., Nurizzo, D., Davies, G.J., 2003. Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate binding module, TmCBM27. *Structure* 11, 665–675.
- Bortoli-German, L., Haiech, J., Chippaux, M., Barras, F., 1995. Informational suppression to investigate structural functional and evolutionary aspects of the *Erwinia chrysanthemi* cellulase EGZ. *J. Mol. Biol.* 246, 82–94.
- Bourgault, R., Oakley, A.J., Bewley, J.D., Wilce, M.C., 2005. Three-dimensional structure of (1,4)- β -D-mannan mannanohydrolase from tomato fruit. *Protein Sci.* 14, 1233–1241.
- Carrard, G., Koivula, A., Söderlund, H., Béguin, P., 2000. Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. USA* 97, 10342–10347.
- Chen, F.T., Evangelista, R.A., 1995. Analysis of mono- and oligosaccharide isomers derivatized with 9-aminopyrene-1,4,6-trisulfonate by capillary electrophoresis with laser-induced fluorescence. *Anal. Biochem.* 230, 273–280.
- Cota, J., Alvarez, T.M., Citadini, A.P., Santos, C.R., Oliveira-Neto, M., et al., 2011. Mode of operation and low resolution structure of a multi-domain and hyperthermophilic endo- β -1,3-glucanase from *Thermotoga petrophila*. *Biochem. Biophys. Res. Commun.* 406, 590–594.
- Dauter, Z., Dauter, M., Rajashankar, K.R., 2000. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 56, 232–237.
- de Vries, R.P., Visser, J., 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol. Mol. Biol. Rev.* 65, 497–522.
- Dhavan, S., Kaur, J., 2007. Microbial mannases: an overview of production and applications. *Crit. Rev. Biotechnol.* 27, 197–216.
- Dias, F.M., Vincent, F., Pell, G., Prates, J.A., Centeno, M.S., et al., 2004. Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of *Cellvibrio mixtus* mannosidase 5A. *J. Biol. Chem.* 279, 25517–25526.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2126–2132.
- Hilge, M., Gloor, S.M., Rypniewski, W., Sauer, O., Heightman, T.D., et al., 1998. High-resolution native and complex structures of thermostable beta-mannanase from *Thermomonospora fusca* – Substrate specificity in glycosyl hydrolase family 5. *Structure* 6, 1433–1444.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- Langer, G.G., Cohen, S.X., Perrakis, A., Lamzin, V.S., 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171–1179.
- Miller, G.L., 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31, 426–428.
- Montiel, M.D., Hernández, M., Rodríguez, J., Arias, M.E., 2002. Evaluation of an endo-beta-mannanase produced by *Streptomyces ipomoea* CECT 3341 for the biobleaching of pine kraft pulps. *Appl. Microbiol. Biotechnol.* 58, 67–72.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 53, 240–255.
- Myers, R.H., Montgomery, D.C., 2001. *Response Surface Methodology*, second ed. John Wiley & Sons, New York.
- Otwowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- Pressey, R., 1989. Endo- β -mannanase in tomato fruit. *Phytochemistry* 28, 3277–3280.
- Roske, Y., Sunna, A., Pfeil, W., Heinemann, U., 2004. High-resolution crystal structures of Caldicellulosiruptor strain R88.4 carbohydrate-binding module CBM27-1 and its complex with mannohexaose. *J. Mol. Biol.* 340, 543–554.
- Sabini, E., Schubert, H., Murshudov, G., Wilson, K.S., Siika-Aho, M., et al., 2000. The three-dimensional structure of a *Trichoderma reesei* beta-mannanase from

ARTICLE IN PRESS

8

Camila Ramos dos Santos et al. / Journal of Structural Biology xxx (2011) xxx–xxx

- glycoside hydrolase family 5. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 56, 3–13.
- Sachtlehner, A., Foidl, G., Foidl, N., Gübitz, G., Haltrich, D., 2000. Hydrolysis of isolated coffee mannan and coffee extract by mannanases of *Sclerotium rolfsii*. *J. Biotechnol.* 80, 127–134.
- Santos, C.R., Squina, F.M., Navarro, A.M., Ruller, R., Prade, R., et al., 2010. Cloning, expression, purification, crystallization and preliminary X-ray diffraction studies of the catalytic domain of a hyperthermostable endo-1,4-beta-D-mannanase from *Thermotoga petrophila* RKU-1. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 66, 1078–1081.
- Santos, C.R., Squina, F.M., Navarro, A.M., Oldiges, D.P., Leme, A.F., et al., 2011. Functional and biophysical characterization of a hyperthermostable GH51 α -L-arabinofuranosidase from *Thermotoga petrophila*. *Biotechnol. Lett.* 33, 131–137.
- Schneider, T.R., Sheldrick, G.M., 2002. Substructure Solution with SHELXD. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 58, 1772–1779.
- Schröder, R., Atkinson, R.G., Redgwell, R.J., 2009. Re-interpreting the role of endo-beta-mannanases as mannan endotransglycosylase/hydrolases in the plant cell wall. *Ann. Bot.* 104, 197–204.
- Sheldrick, G.M., 2002. Macromolecular phasing with SHELXE. *Z. Kristallogr.* 217, 644–650.
- Squina, F.M., Santos, C.R., Ribeiro, D.A., Cota, J., Oliveira, R.R., et al., 2010. Substrate cleavage pattern, biophysical characterization and low-resolution structure of a novel hyperthermostable arabinanase from *Thermotoga petrophila*. *Biochem. Biophys. Res. Commun.* 399, 505–511.
- Stålbrand, H., Siika-aho, M., Viikari, L., 1993. Purification and characterization of two β -mannanases from *Trichoderma reesei*. *J. Biotechnol.* 29, 229–242.
- Tailford, L.E., Ducros, V.M., Flint, J.E., Roberts, S.M., Morland, C., et al., 2009. Understanding how diverse beta-mannanases recognize heterogeneous substrates. *Biochemistry* 48, 7009–7018.
- Tanaka, M., Umemoto, Y., Okamura, H., Nakano, D., Tamaru, Y., et al., 2009. Cloning and characterization of a beta-1,4-mannanase 5C possessing a family 27 carbohydrate-binding module from a marine bacterium, *Vibrio* sp. strain MA-138. *Biosci. Biotechnol. Biochem.* 73, 109–116.
- Takahata, Y., Nishijima, M., Hoaki, T., Maruyama, T., 2001. *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. *Int. J. Syst. Evol. Microbiol.* 51, 1901–1909.
- Tenkanen, M., Makkonen, M., Perttula, M., Viikari, L., Teleman, A., 1997. Action of *Trichoderma reesei* mannanase on galactoglucomannan in pine kraft pulp. *J. Biotechnol.* 57, 191–204.
- Vagin, A., Teplyakov, A., 1997. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* 30, 1022–1025.

SUPPLEMENTARY MATERIAL

Molecular insights into substrate specificity and thermal stability of a bacterial GH5-CBM27 endo-1,4- β -D-mannanase

Camila Ramos dos Santos^a, Joice Helena Paiva^a, Andreia Navarro Meza^a, Junio Cota^b,
Thabata Maria Alvarez^b, Roberto Ruller^b, Rolf Alexander Prade^d, Fabio Marcio
Squina^b, Mario Tyago Murakami^{a*}

Table S1: Central composite design matrix (2^2) and the response of TpMan activity after 10 minutes of incubation.

Run no.	Coded levels ($X_1 = \text{pH}$; $X_2 = T$)		Actual levels ($X_1 = \text{pH}$; $X_2 = T$)		Laminarinase activity (IU/mg protein)
	X_1	X_2	X_1	X_2	
1	-1	-1	4.6	74.2	139,8
2	1	-1	7.4	74.2	170,9
3	-1	1	4.6	94.8	175,2
4	1	1	7.4	94.8	204,2
5	-1.414	0	4.0	84.5	68,6
6	1.414	0	8.0	84.5	154,6
7	0	-1.414	6.0	70.0	159,7
8	0	1.414	6.0	99.0	238,9
9	0	0	6.0	84.5	234,6
10	0	0	6.0	84.5	241,3
11	0	0	6.0	84.5	247,8
12	0	0	6.0	84.5	247,1

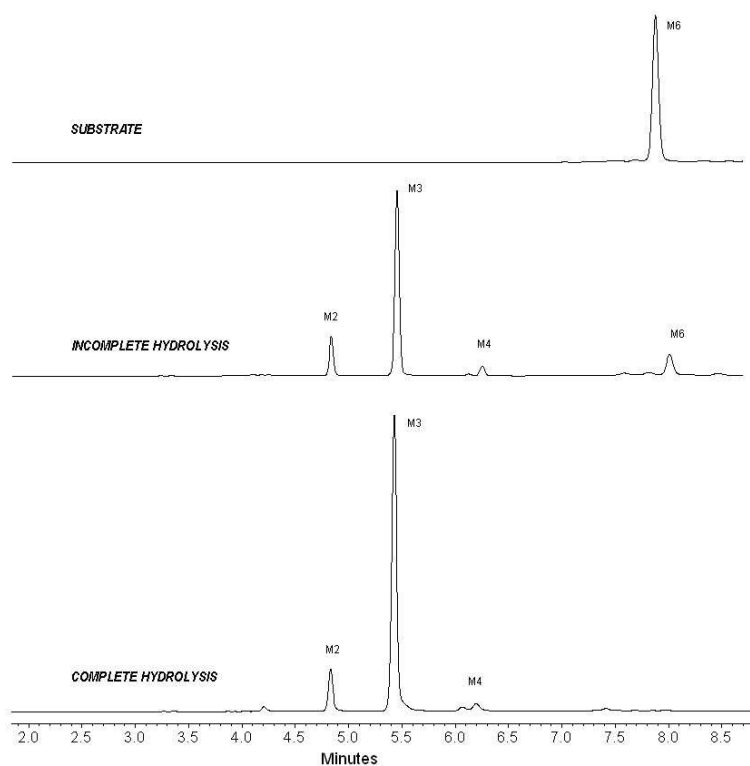


Figure S1: Capillary zone electrophoresis of APTS-labeled oligosaccharides released by TpMan during enzymatic hydrolysis of mannohexaose.

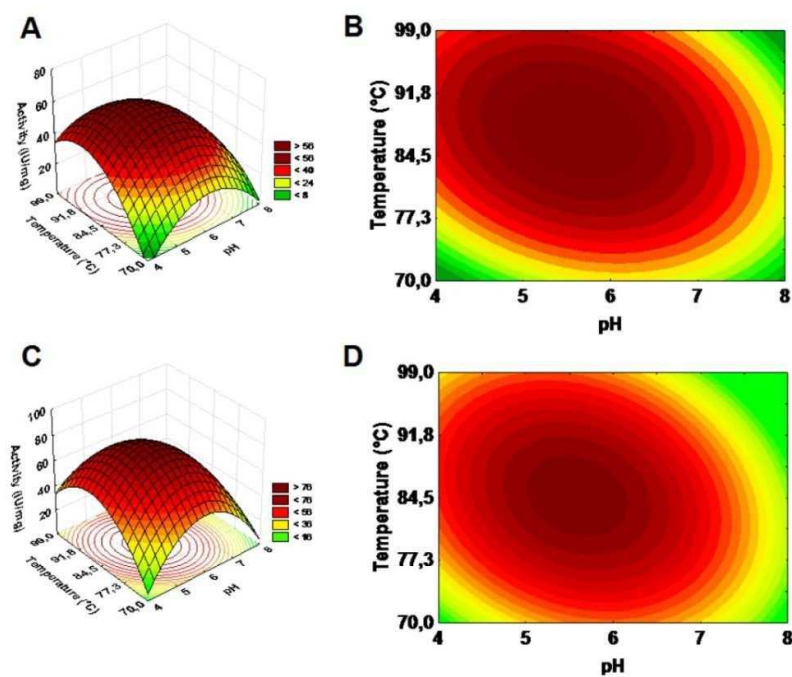


Figure S2: Functional characterization of TpMan constructions. 3D and 2D profiles of response surface results of TpMan (A, B) and TpMan Δ CT (C, D).

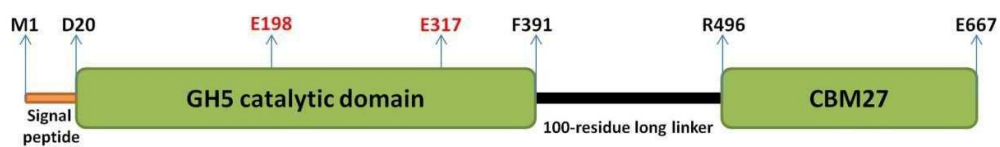


Figure S3: Schematic representation of TpMan domain architecture.

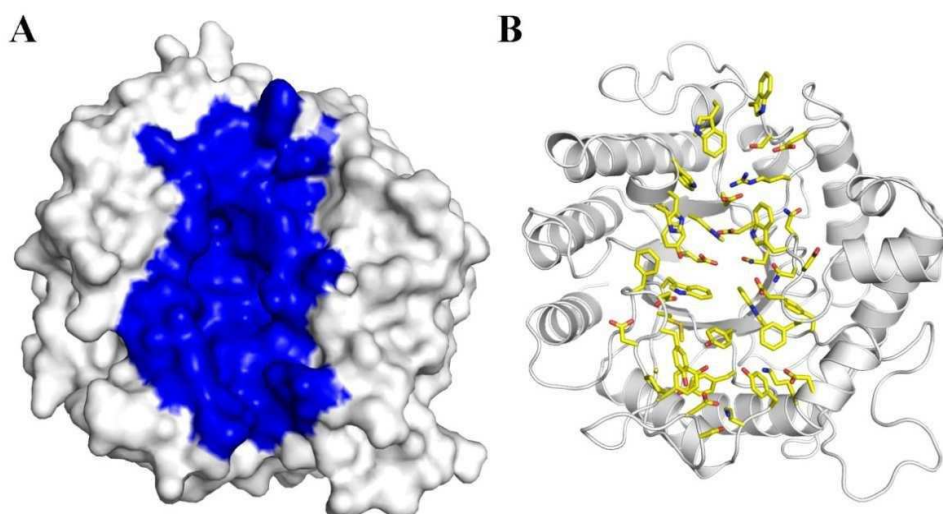


Figure S4: Details of the TpMan catalytic groove. (A) Molecular surface representation with the active-site pocket colored in blue. (B) Stick representation of the residues forming the pocket.

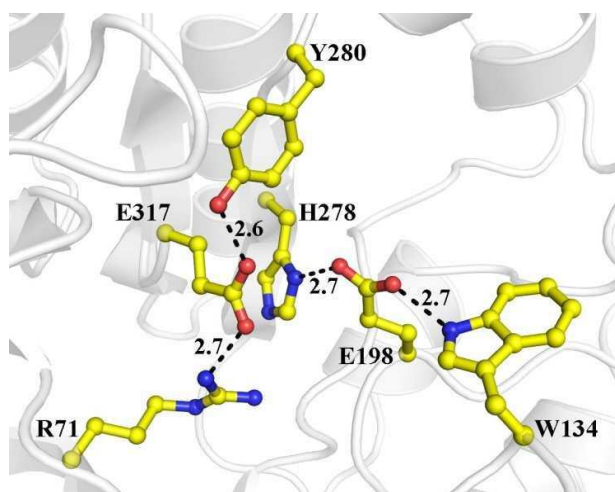


Figure S5: Hydrogen bond network supporting the catalytic residues E198 and E317.

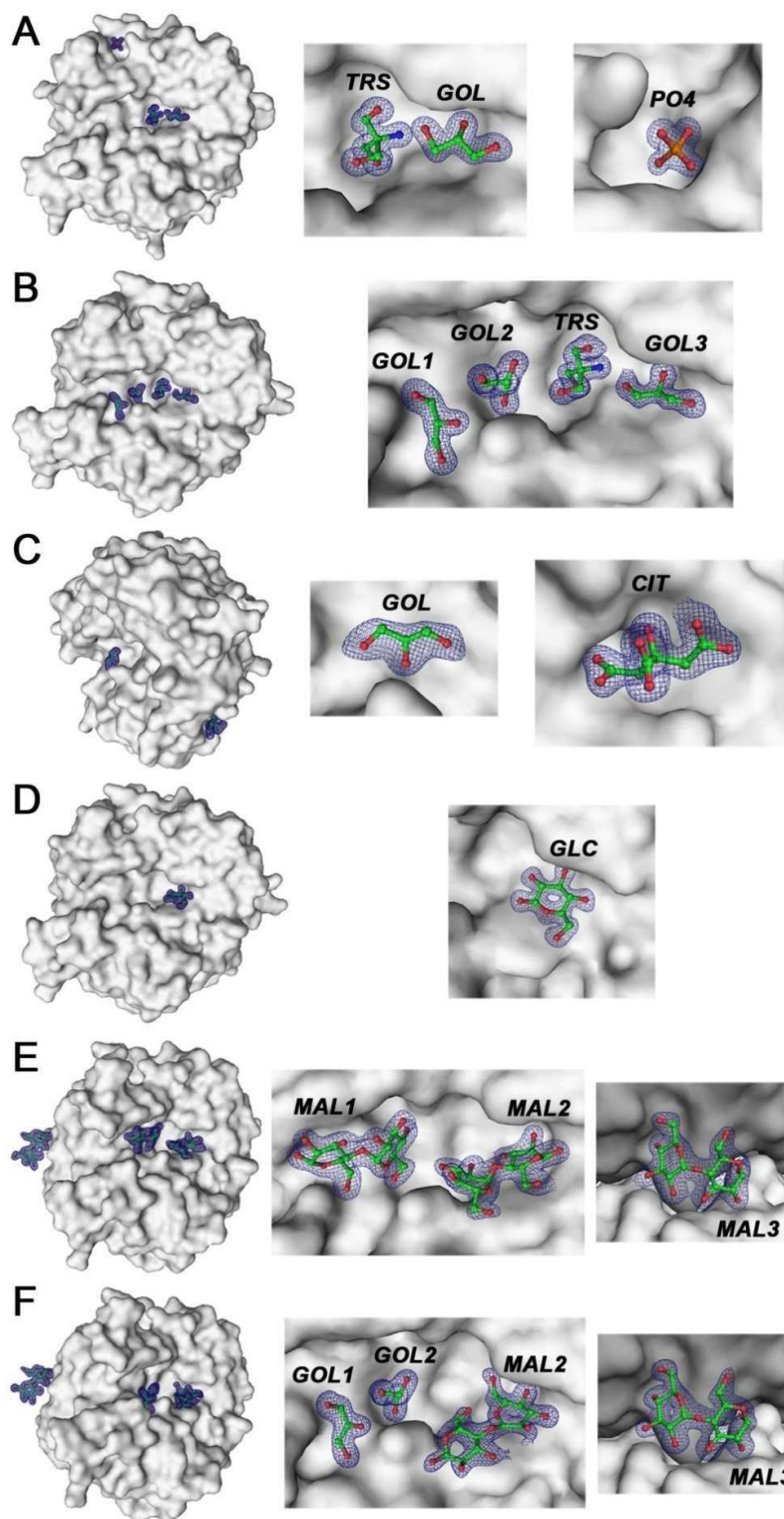


Figure S6: Electron density map (1.0 sigma) for the ligands in TpMan complexes. (A) *I*222 space group structure, (B) TRS+GOL complex, (C) CIT+GOL complex, (D) GLC complex, (E) MAL complex and (F) MAL+GOL complex.

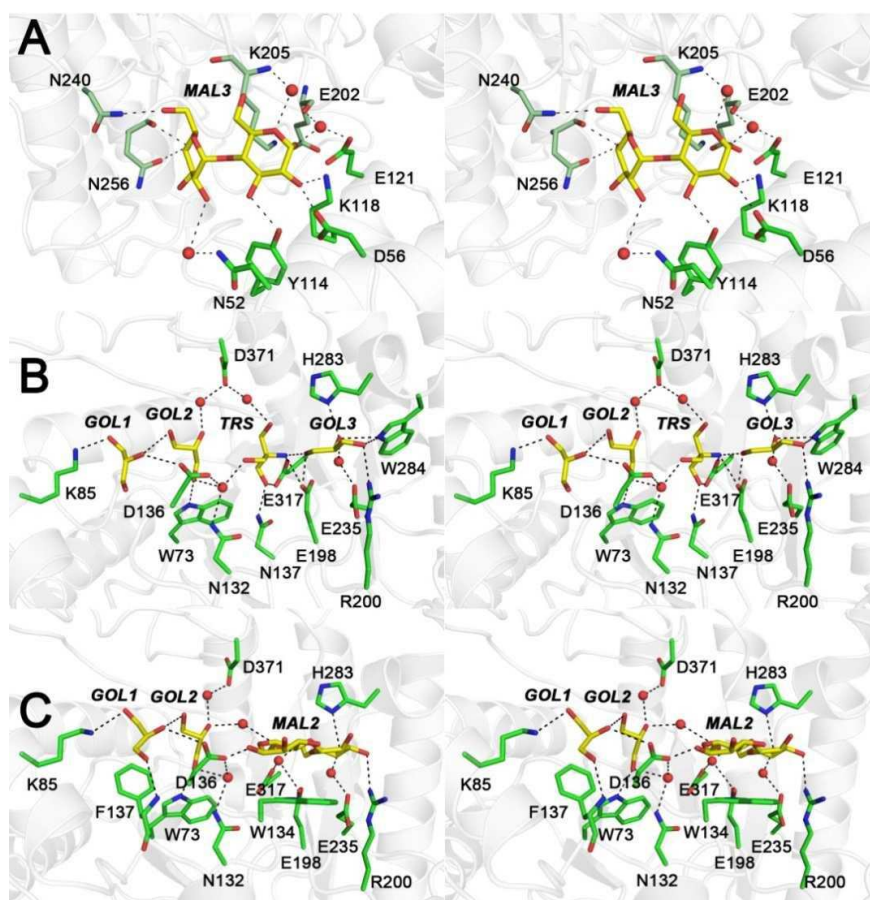


Figure S7: Interactions between ligands and protein residues in TpMan complexes: (A) maltose at a remote site, (B) glycerol and tris molecules at the active-site pocket, and (C) maltose and glycerol also at the active site.

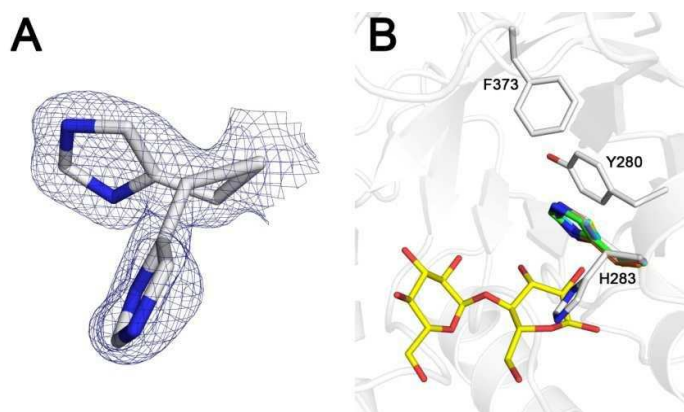


Figure S8: (A) Double side-chain conformation of the residue H283 with electron density map (1.0 sigma). (B) Representation of the +1 subsite in the presence of a maltose molecule.

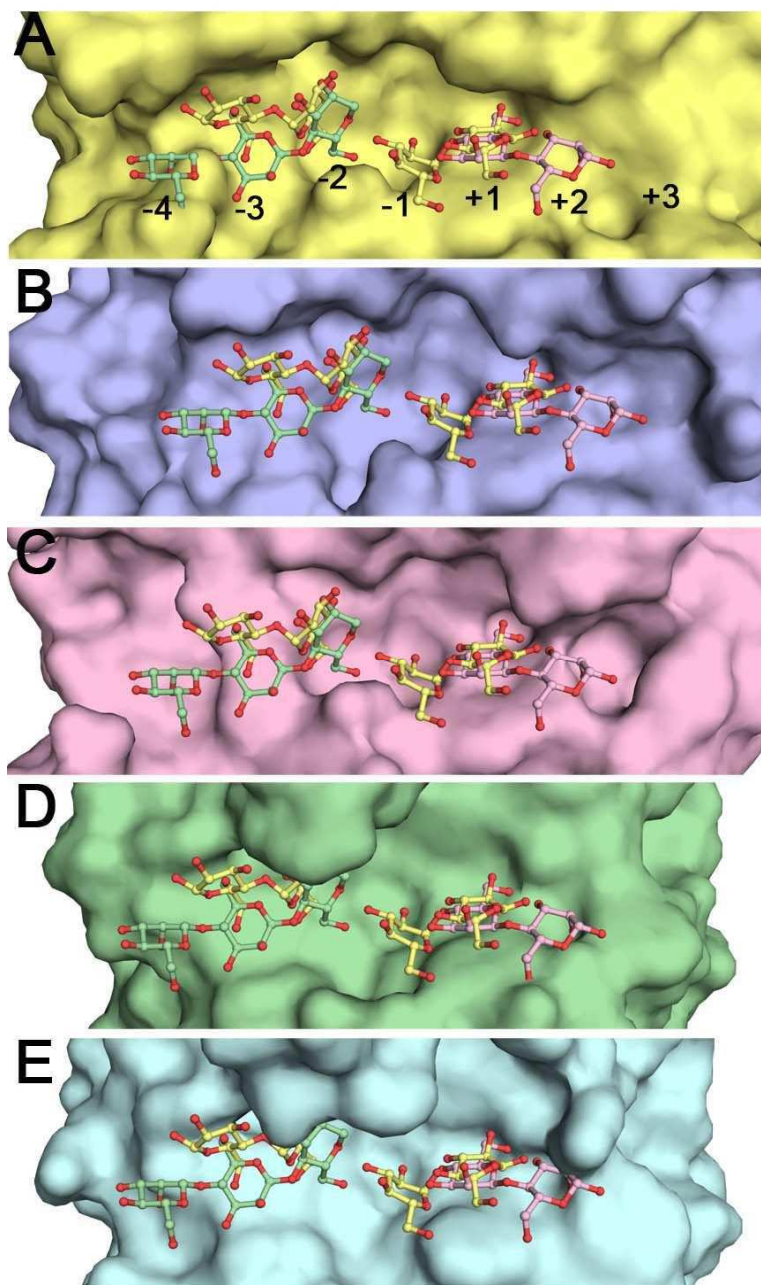


Figure S9: Molecular surface representation of the active-site cleft of TpMan (A), LeMan (B), TrMan (C), TfMan (D) and BaMan (E). The surface and carbon atoms of ligands are colored according to each endo- β -1,4-mannanase: TpMan and maltose in yellow, LeMan in blue, TrMan and mannobiose in pink, TfMan and mannotriose in green and BaMan in cyan.

The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction



The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction

Mário Tyago Murakami,* Mauricio Luis Sforça, Jorge Luiz Neves, Joice Helena Paiva, Mariane Noronha Domingues, André Luiz Araujo Pereira, Ana Carolina de Mattos Zeri, and Celso Eduardo Benedetti*

Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, 13083-970, Campinas, SP, Brazil

ABSTRACT

Many plant pathogenic bacteria rely on effector proteins to suppress defense and manipulate host cell mechanisms to cause disease. The effector protein PthA modulates the host transcriptome to promote citrus canker. PthA possesses unusual protein architecture with an internal region encompassing variable numbers of near-identical tandem repeats of 34 amino acids termed the repeat domain. This domain mediates protein–protein and protein–DNA interactions, and two polymorphic residues in each repeat unit determine DNA specificity. To gain insights into how the repeat domain promotes protein–protein and protein–DNA contacts, we have solved the structure of a peptide corresponding to 1.5 units of the PthA repeat domain by nuclear magnetic resonance (NMR) and carried out small-angle X-ray scattering (SAXS) and spectroscopic studies on the entire 15.5-repeat domain of PthA2 (RD2). Consistent with secondary structure predictions and circular dichroism data, the NMR structure of the 1.5-repeat peptide reveals three α -helices connected by two turns that fold into a tetratricopeptide repeat (TPR)-like domain. The NMR structure corroborates the theoretical TPR superhelix predicted for RD2, which is also in agreement with the elongated shape of RD2 determined by SAXS. Furthermore, RD2 undergoes conformational changes in a pH-dependent manner and upon DNA interaction, and shows sequence similarities to pentatricopeptide repeat (PPR), a nucleic acid-binding motif structurally related to TPR. The results point to a model in which the RD2 structure changes its compactness as it embraces the DNA with the polymorphic residues facing the interior of the superhelix oriented toward the nucleotide bases.

Proteins 2010; 78:3386–3395.
© 2010 Wiley-Liss, Inc.

Key words: TAL effectors; tetratricopeptide repeat; pentatricopeptide repeat (PPR); *Xanthomonas citri*.

INTRODUCTION

Xanthomonas axonopodis pv. *citri* (Xac), the causal agent of citrus canker, induces the formation of raised pustules on the surface of the host plant. The pustules typically develop into larger corky and water-soaked lesions, which break the epidermis favoring bacterial dissemination.^{1,2} Although it is known that canker lesions result from the intense division and expansion of the host cells at the site of infection, how exactly Xac induces cell division and growth is not yet clear. We have shown that Xac modulates the synthesis of cell growth regulators during the onset of infection³ and that the common action of auxin and gibberellin is required for initial canker development.⁴ Interestingly, these hormones alone or in combination had no apparent effect on citrus cell growth or division, suggesting that another factor is required for canker formation.⁴ This additional factor is thought to be the Xac effector protein PthA, which is sufficient to promote cell hypertrophy when transiently expressed in citrus leaves.⁵ Accordingly, expression of PthA proteins in citrus cells provokes transcriptional changes that overlap with those triggered by Xac infection associated with auxin and gibberellin action (Pereira and Benedetti, unpublished data).

PthA proteins differ from each other primarily by the number of near-identical repeats of 33–34 amino acids that are tandemly located in the central region of the protein. This domain confers host selectivity and is critical to determine pathogenicity.^{6,7} PthA proteins are 95–97% identical to AvrBs3, the best known *Xanthomonas* type III effector protein. AvrBs3 is targeted to the nucleus of host cells where it modu-

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Fundação de Amparo à Pesquisa do Estado de São (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

*Correspondence to: M.T. Murakami, Laboratório Nacional de Biociências, Centro Nacional de Energia e Materiais, R. Giuseppe Máximo Solfaro, 10000, Campinas, SP, 13083-970, Brazil. E-mail: mario.murakami@lnbio.org.br or C.E. Benedetti, Laboratório Nacional de Biociências, Centro Nacional de Energia e Materiais, R. Giuseppe Máximo Solfaro, 10000, Campinas, SP, 13083-970, Brazil. E-mail: celso.benedetti@lnbio.org.br

Received 26 May 2010; Revised 16 July 2010; Accepted 24 July 2010

Published online 17 August 2010 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/prot.22846

lates transcription.⁸ Notably, AvrBs3 recognizes different plant promoters and activates transcription of particular genes in both susceptible and resistant plants.^{9,10} The interaction of AvrBs3 with its target DNA is mediated by its repeat domain,^{9–12} which is also essential for AvrBs3 dimerization before its nuclear import.¹³ Likewise, we found that the repeat domain of PthA proteins is critical for protein–protein interactions.¹⁴ Therefore, it is becoming clear that the repeat domain of such effectors has a dual character as a protein scaffold that allows protein–protein and protein–DNA contacts with host targets.

An intriguing aspect of the repeat domain of such effectors is the variability found in every repeat unit at certain amino acid positions. In the case of the PthA variants, the polymorphism occurs preferentially at Positions 4 (D, E, Q, or A), 12 (N or H), 13 (I, D, G, or S), and 24 (R or A) of the repeat units.¹⁴ Since most members of the AvrBs3/PthA protein family share high-sequence identity, it has been postulated that, besides the variation in the number of repeat units, the polymorphism within the repeat units would play a critical role in conferring the specificity required for the interactions with particular protein or DNA targets. Indeed, recent studies have demonstrated that the pair of residues at Positions 12–13 determines the DNA sequence specificity^{15,16}; however, the structural basis for these interactions is presently unknown. Thus, structural information on the repeat region becomes highly relevant for understanding the molecular function of this unique protein domain. Although numerous homologues of AvrBs3/PthA proteins have been identified, no three-dimensional (3D) structure is yet available for any transcription activator-like (TAL) effector. Despite considerable efforts, we have not yet been able to crystallize any of the PthA variants or their repeat domains. Thus, we used a combination of circular dichroism (CD), dynamic light scattering (DLS), nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS) and molecular modeling approaches to gain insights into the 3D structure of the PthA repeat domain. In this work, we present the first 3D structure of a peptide corresponding to 1.5-repeat units of PthA2 (PDB code: 2KQ5) and a low-resolution envelope for the entire repeat domain of PthA2 (RD2) that is in agreement with its theoretical tetratricopeptide repeat (TPR)-like structure. Spectroscopic studies also revealed that RD2 undergoes conformational changes in a pH-dependent manner and in the presence of DNA. Furthermore, we show that the repeat units of RD2 are similar to pentatricopeptide repeat (PPR), a 35 amino acid motif that is structurally related to TPR and have nucleic acid binding activity.^{17,18}

MATERIALS AND METHODS

Protein expression and purification

The 15.5-RD2 was amplified from *X. citri* plasmids and cloned into the *Bam*HI/*Sac*I sites of pET28a

(Novagen). The construct was sequenced and used to transform *Escherichia coli* strain BL21 (DE3). Protein expression was induced at OD₆₀₀ = 0.6 with 0.4 mM IPTG at 25°C. The cells were harvested after 3 h, centrifuged, and suspended in binding buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 15 mM imidazole, 5% (v/v) glycerol, 1 mM PMSF, 0.5% Nonidet P-40) and incubated on ice with lysozyme (1 mg mL^{−1}) and RNase (10 µg mL^{−1}) for 30 min. Bacterial cells were disrupted by sonication and the soluble fraction was incubated with DNase I (0.5 µg mL^{−1}) and MgCl₂ (2 mM). The suspension containing the 6xHis-RD2 was loaded on a 5-mL HiTrap chelating HP (GE Healthcare) column pre-equilibrated with binding buffer. The column was washed with 20 column volumes of the same buffer and the protein was eluted with imidazole gradient. After purification, dithiothreitol (DTT) was added to a final concentration of 1 mM and fractions displaying the highest purity were concentrated and loaded on a G75 16/60 Superdex (GE Healthcare) column, pre-equilibrated with 20 mM Tris–HCl pH 8.0, 200 mM NaCl, 5% (v/v) glycerol, and 1 mM DTT. RD2 was eluted as a single peak with a molecular mass corresponding to the monomer.

Dynamic light scattering

DLS experiments were carried out using a DynaPro 810 (Protein Solutions) apparatus equipped with a Peltier module for temperature control. The wavelength of the laser light and the output power were set to 830 nm and 30 mW, respectively. About 100 measurements were made at intervals of 20 s for each run. The DLS experiments were repeated several times with intervals of 30 min to check stability. Protein solutions of 1 mg mL^{−1} were prepared in 50 mM sodium acetate pH 5.0, 50 mM sodium phosphate pH 6.0 and 7.0, and 20 mM Tris–HCl pH 8.0. The complex of RD2:DNA was prepared in a molar ratio of 1.0:1.2, incubated for 12 h at 20°C, and centrifuged at 10,000g for 15 min at 4°C before analysis. Standard curves of bovine serum albumin were used for calibration and the experiments were conducted at 18°C. Hydrodynamic parameters were determined using the software DYNAMICS v.6.10.1.2.

CD measurements

CD spectroscopy experiments were conducted on a JASCO J-810 CD spectrophotometer equipped with a Peltier temperature control using 1-mm path quartz cuvettes. Spectra were acquired with a final protein concentration of 5.0–10 µM in 20 mM Tris–HCl pH 8.0, containing 1.0 mM DTT and 100 mM NaCl. The RD2:DNA mixes were incubated for 3 h at 20°C and centrifuged (10,000g) for 15 min at 4°C before analyses. CD measurements were collected between 185 and 260 nm using a scanning rate of 50 nm min^{−1} with an average response time of 4 s.

Secondary structure variations were monitored as a function of changes in the initial CD spectrum upon addition of sodium dodecyl sulfate (SDS) for the 1.5-repeat unit. For all the spectra, an average of 10 scans was accumulated and the background spectrum of the buffer was subtracted.

NMR studies

NMR experiments were performed at 20°C using a Varian Inova 600 MHz spectrometer equipped with a cryogenic probe. Synthetic peptides (Proteimax, São Paulo, Brazil) corresponding to 1.0- or 1.5-repeat units of the repeat domain of PthA were dissolved in 20 mM phosphate buffer, pH 5.0, containing 70 mM SDS, 1.0 mM DTT, and 5% (v/v) D₂O, at a final concentration of ~1.0 mM. Peptide resonance peaks were assigned using standard methods including correlation spectroscopy,¹⁹ total correlation spectroscopy (TOCSY),²⁰ and nuclear Overhauser enhancement spectroscopy (NOESY).²¹ The TOCSY spectra were acquired using a DIPSI spin-lock sequence at a field strength of 10 kHz and a mixing time of 70 ms. NOESY spectra were recorded with mixing times of 150 and 300 ms. All 2D experiments were acquired in the phase-sensitive mode using the method of states.²² Water suppression was achieved by low-power continuous wave irradiation during the relaxation delay or using the WATERGATE method.²³ Data were processed and analyzed using the NMRPipe/NMRVIEW software.²⁴ Before Fourier transformation, the time domain data were zero filled in both dimensions to yield a (4096 by 4096) data matrix. When necessary, a fifth-order polynomial baseline correction was applied after transformation and phasing. To obtain distance constraints, crosspeak volumes were estimated from the NOESY spectra.

The structure of the peptide corresponding to 1.5-repeat units of PthA was calculated in a semiautomated iterative manner with the program CYANA version 2.1,²⁵ using 100 starting conformers. CYANA 2.1 protocol was applied to calibrate and assign NOE crosspeaks. After the first few rounds of automatic calculations, the NOESY spectra were analyzed again to identify additional crosspeaks consistent with the structural model and to correct misidentified NOEs. The structures obtained were further refined by restrained minimization and molecular dynamic (MD) studies using the CNS software.²⁶ The 20 structures with the lowest target function were selected to represent the ensemble of peptide structures. The quality of the structures was analyzed with PROCHECK-NMR.²⁷ The NMR data were deposited in the biological magnetic resonance bank (BMRB) and protein data bank (PDB) under the entry codes 16589 and 2KQ5, respectively.

SAXS data collection and analysis

SAXS data were collected at the D11A-SAXS beamline at the Brazilian Synchrotron Light Laboratory (LNLS). The

radiation wavelength was set to 1.488 Å and a charge-coupled device area detector (MARCCD 165 mm) was used to record the scattering patterns. The sample-to-detector distance was set to 1415.95 mm to give a scattering vector ranging from 0.10 to 3.5 nm⁻¹. The measurements were carried out with a sample concentration of 2 mg mL⁻¹ at 18°C. Each protein sample was previously analyzed by DLS and only monodisperse solutions (polydispersity <20%) were used. Protein samples were centrifuged for 15 min at 10,000g to eliminate any existing aggregates immediately before each measurement. The scattering curves of the protein solutions and buffers were collected in frames of 300 s each to avoid radiation-induced protein damage. Each frame was carefully checked for possible bubbles or radiation-induced aggregation of the protein before calculating the average intensity and the associated experimental error. The experimental intensities were corrected for background, buffer contributions, detector inhomogeneities, and sample transmission.

The radius of gyration (R_g) was evaluated using Guinier approximation²⁸ as implemented in the program PRIMUS.²⁹ The indirect Fourier transform package GNOM³⁰ was used to evaluate the pair-distance distribution function $p(r)$. The low-resolution envelope of the RD2 protein was determined using *ab initio* modeling implemented in DAMMIN.³¹ An averaged model was generated from several runs using the DAMAVER suite of programs.³² The SAXS model and the NMR structures were superimposed with SUPCOMB.³³

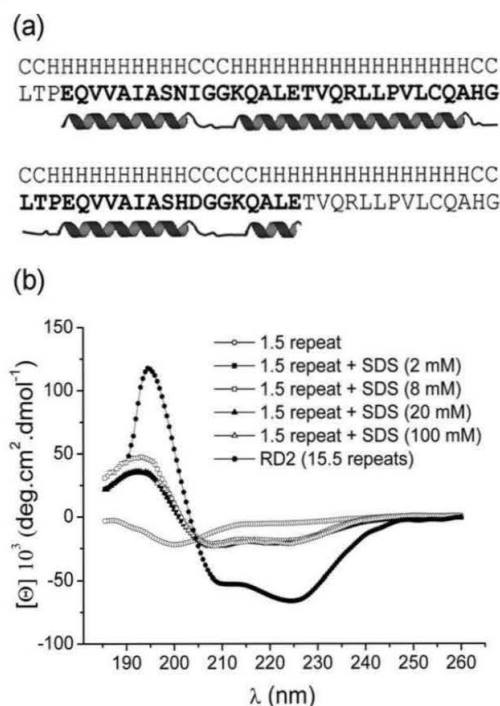
Homology molecular modeling of the RD2

The crystal structures of importin β -1 from *Saccharomyces cerevisiae* (2BPT)³⁴ and TIP120 from *Homo sapiens* (1U6G)³⁵ were used as 3D templates for restraint-based modeling as implemented in the MODELLER program.³⁶ The overall model was improved enforcing the proper stereochemistry using spatial restraints and CHARMM energy terms, followed by conjugate gradient simulation based on the variable target function method.³⁶ Ten models were built for the RD2 sequence based on the (m)GenThreader alignment. All models were evaluated with the DOPE potential and the one with the lower global score was selected for explicit solvent MD simulation using GROMACS³⁷ to check its stability and consistency. The overall and local quality analyses of the final model were assessed by VERIFY3D,³⁸ PROSA,³⁹ and VADAR.⁴⁰ Three-dimensional structures were displayed, analyzed, and compared using the program COOT.⁴¹

RESULTS AND DISCUSSION

The 1.5-repeat peptide of PthA shows a TPR-like fold

Secondary structure algorithms predict only α -helices and turns for the repeat domains of PthA proteins. Not

**Figure 1**

Structural features of the repetitive units of RD2. (a) PSIPRED secondary structure prediction of two consecutive repeat units of RD2 showing coils (C) and helices (H). The 1.5-repeat unit peptide used for structural resolution is shown in bold and the secondary structure elements determined by NMR are represented underneath the sequence. (b) CD curves of the purified 15.5-repeat domain (5 μ M) of PthA2 (RD2) in comparison with that of the 1.5-repeat unit peptide (10 μ M) in the presence and absence of different SDS concentrations.

surprisingly, the predictions are almost identical within each of the repeat units, as shown for two consecutive repeats found in the PthA variants [Fig. 1(a)]. The predictions are consistent with the far-UV CD spectrum of the entire 15.5-RD2 showing two minimum peaks at 208 and 222 nm and a positive peak at 193 nm, which indicate that RD2 has high contents of α -helices [Fig. 1(b)].

CD data analysis has indicated that the 34-amino acid peptide corresponding to one repeat unit of PthA is not structured in solution (not shown). Thus, we have used *in silico* 3D modeling algorithms to find a minimal peptide size that possesses the stereochemical requirements for a structured motif, and we observed that the 1.5-repeat peptide is the minimal structural unit of the repeat region of PthA. CD measurements indicated that the 1.5-repeat peptide is partially unfolded in solution; however, addition of 2 mM SDS was sufficient to increase the contents of secondary structural elements, particularly α -helices [Fig. 1(b)]. Although SDS has been suc-

cessfully used to investigate the structure of membrane peptides, it has also been shown to stabilize conformations in peptides with propensity to form α -helices via electrostatic interactions.⁴² We therefore used NMR techniques to solve the structure of the 1.5-repeat peptide in the presence of SDS since no experimental 3D model of the repeat unit was yet available.

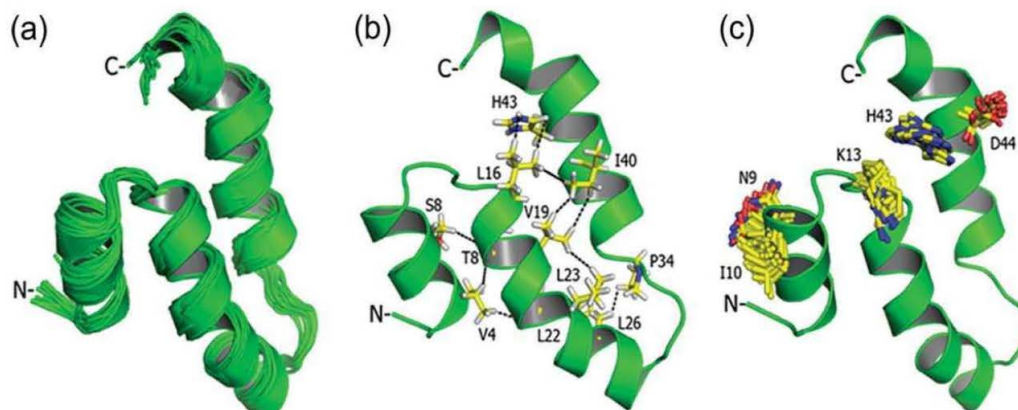
The 1.5-repeat structure was solved using 982 distance constraints derived from the NOESY spectra. The inter-residue NOEs correlate with the chemical shift index and show a dense pattern of $d\alpha N(i, i + 3)$, $d\alpha\beta(i, i + 3)$, and $d\alpha N(i, i + 4)$ NOEs involving residues V3-I10, Q14-Q20, P24-A29, and V37-G45, indicating that these regions possess an helical fold (Supporting Information Figure 1). Statistics of an ensemble of 20 lower energy structures [Fig. 2(a)] show no significant violations of the molecular geometry parameters (Table I). In addition, the Ramachandran plot analysis shows that all ϕ - ψ angles are in allowed regions and the root mean square deviations (RMSD) for the backbone and side chains are within expected values, indicating a good stereochemical quality of the ensemble.

The peptide folds into a helical-bundle structure that is very similar to the TPR topology [Fig. 2(a)], thus corroborating the secondary structure predictions and CD data. In fact, the CD curve of RD2 is quite similar to those of TPR domains with a 222/208 nm ratio of ~ 1.3 , which indicates coiled-coil-like conformations.^{43–46} The specific interactions responsible for these conformational states are the presence of a series of hydrogen bonds between the backbone carbonyl (C=O) of residues “*i*” and the backbone amide (NH) of residues “*i* + 4” spanning the helical regions. The secondary structure is also stabilized by hydrophobic interactions that keep the helical regions close together. Preferential spatial arrangement is confirmed by NOE interactions between side chains of the hydrophobic residues V4-S8 with T18-L22 and L16-C27 with L32-H43 [Fig. 2(b)].

Interestingly, the two hypervariable diresidues NI and HD, known to specify preferential interactions with adenine and cytosine in DNA target sites, respectively,^{15,16} are structurally close in the NMR structures [Fig. 2(c)]. This would be expected considering that each pair of the variable residues recognizes one nucleotide in a linear fashion, that is, one variable diresidue to one nucleotide in the DNA target sequence.^{15,16} We also noticed that the K residues, which typically play a central role in protein–DNA interactions,⁴⁷ are located adjacently to the polymorphic diresidues [Fig. 2(c)].

The RD2 has an elongated shape

The *ab initio* molecular shape of RD2 was determined by SAXS analysis and the corresponding profiles of the scattering curve and distance distribution function $p(r)$ are shown in Figure 3(a). SAXS calculations revealed a maximum dimension of 129 Å from the $p(r)$ curve and a gyra-

**Figure 2**

The NMR structure of the 1.5-repeat peptide of PthA. (a) Ensemble of the 20 lowest energy structures showing three helices connected by two turns. (b) The lowest energy structure showing the hydrophobic interactions between the side chains of residues VVAIA and LLPVL, which stabilize the structure. (c) Orientation of the hypervariable diresidues NI and HD relative to the K residue in the peptide structure.

tion radius of 40.82 ± 0.01 Å for the RD2 molecule. These values are in agreement with those obtained from the *in silico* modeling (see below) and indicate that RD2 is a monomer in solution. Analytical size-exclusion chromatography and DLS data also suggest that RD2 is a monomer in solution with a hydrodynamic radius of around 4.5 nm for a sample with low polydispersity index (15%).

The low resolution envelope of RD2 was derived without imposing any constraints [Fig. 3(b)]. Structural alignments of each individual molecular model, obtained by *ab initio* shape reconstruction, showed that they are very similar as indicated by the normalized spatial discrepancy (NSD) values, which are around 0.8. The RD2 model reveals an elongated or extended shape consistent with the notion that the repeat units are structurally arranged in tandem along the repeat region [Fig. 3(b)]. The elongated shape of the RD2 molecule can also be noted from the $p(r)$ curve profile [Fig. 3(a)].

Molecular modeling predicts a TPR-like superhelical structure for RD2

Fold assignment searches for RD2 using PSIPRED⁴⁸ returned 10 protein structures with significant scores, all displaying TPR domains. The crystal structures of the yeast importin β 1 (PDB code: 2BPT) and human TIP120 (PDB code: 1U6G) were used as templates to generate a restraint-based 3D model of RD2 [Fig. 4(a)]. The model shows good local and global stereochemical properties with a Z-score of 6.8. Analyses of the Ramachandran plot indicate that 93% of the RD2 residues are in most favorable regions, 5% are in additional allowed regions and only 2% are in disallowed regions. In addition, local quality analysis assessed by plotting the energies as a function of

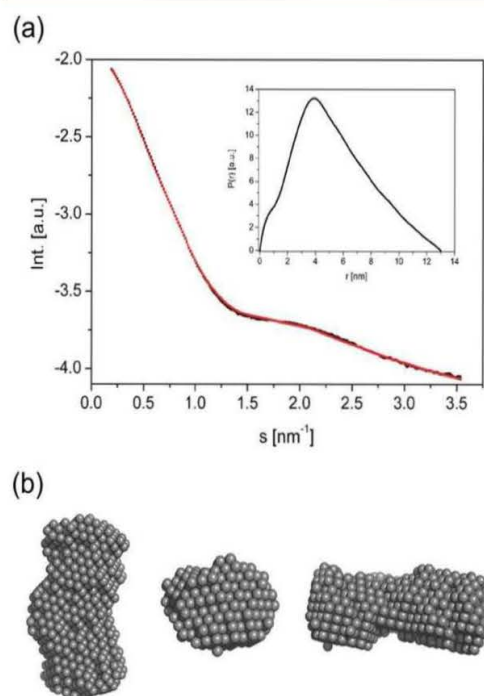
the amino acid positions shows no positive values thus highlighting the good stereochemical quality of the model and its suitability for structural analysis and comparisons. The NMR structure of the 1.5-repeat peptide superimposes adequately with the RD2 model with an RMSD of 2.7 Å [Fig. 4(b)], thus corroborating the *in silico* model.

The RD2 3D model conserves all the structural features found in TPR-containing proteins encompassing 31 anti-parallel α -helices that fold into a superhelical structure [Fig. 4(a)] and is in agreement with previous molecular modeling studies.⁴⁹ Furthermore, the superhelix model is consistent with the RD2 CD data showing an average 222/208 nm ratio >1.1, which is typical of coiled-coil regions. The repeat units of RD2 have the same amino acid segment length of TPRs and their superposition indicates that

Table 1

Structural Statistics of the 1.5-Repeat Unit Peptide

NOEs	
Total number	982
Short range $ i-j \leq 1$	650
Medium range $1 < i-j < 5$	156
Long range $ i-j \geq 5$	176
CYANA	
Target function	1.01 ± 0.15
Distance violation >0.20	0
Dihedral angles $>5^\circ$	0
RMSD	
Backbone (total)	0.72 ± 0.13
Side chain (total)	1.54 ± 0.17
PROCHECK	
Most favorable	84.1%
Additional allowed	15.3%
Generously allowed	0.6%
Not allowed	0.0%

**Figure 3**

Ab initio shape determination of RD2. (a) Experimental SAXS curve and the theoretical fitting of the data (black and red, respectively) using the program GNOM. The inset shows the pair distance distribution function $p(r)$. (b) Different views of the *ab initio* low resolution DAMMIN envelope model of RD2 in solution. The NSD values for a set of 15 DAMMIN and GASBOR models ranged from 0.85 to 0.95 in both calculations.

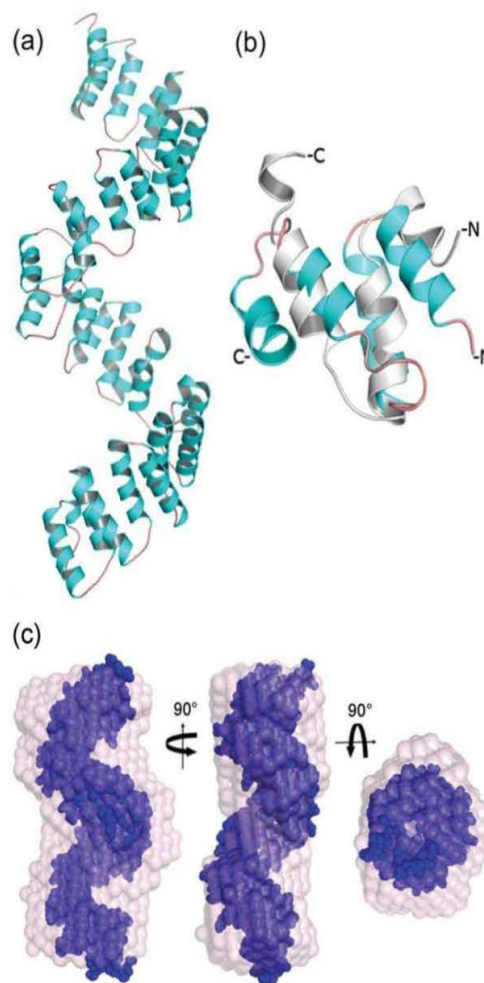
they are quite similar to one another ($\text{RMSD} \leq 1.5 \text{ \AA}$). As observed in the structure of the 1.5-repeat peptide [Fig. 2(b)], the superhelix is stabilized by contacts of hydrophobic residues across the α -helical faces resulting in a continuous hydrophobic core. Additional polar interactions between close side chains contribute to the protein packing and stabilization. Notably, the *in silico* model fits remarkably well on the SAXS envelope and displays a NSD of 1.25, indicating high-shape complementarity between the *ab initio* envelope and the surface of the theoretical superhelix model [Fig. 4(c)].

TPR domains are a well-known protein scaffolds that mediate interprotein associations and assembly of multi-protein complexes in numerous organisms. The domain was first identified in yeast proteins required for mitosis and RNA synthesis as a degenerated 34-amino acid sequence arranged in tandem repeats.^{50,51} Since then, studies have highlighted its diversity of arrangements and functions.^{52,53} More recently, it was found that TPR domains can self-interact and promote protein oligomerization.^{43,54,55} Thus, the TPR-superhelical structure of

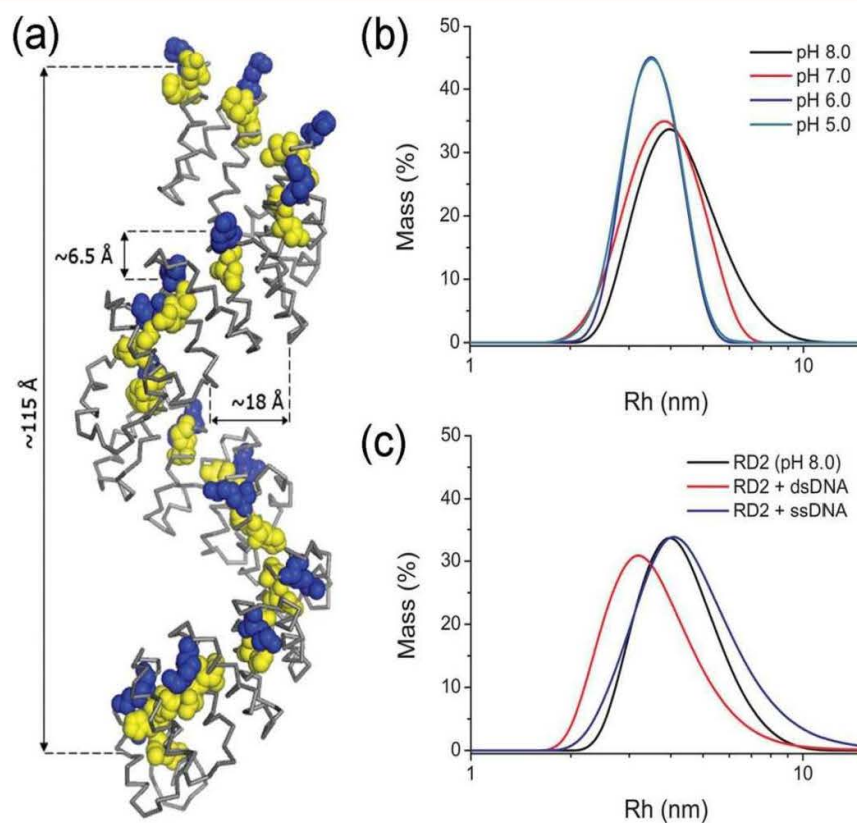
RD2 is consistent with the fact that RD2 also self-interacts in two-hybrid assays and associates with a number of citrus proteins, including the TPR domain of TDX.¹⁴

Conformational changes of RD2 upon interaction with DNA

Although TPR domains have not been reported to bind nucleic acids directly, they are found in numerous DNA and RNA-binding proteins.^{51,53,56,57} To gain insights into how the RD2 superhelix could bind DNA with

**Figure 4**

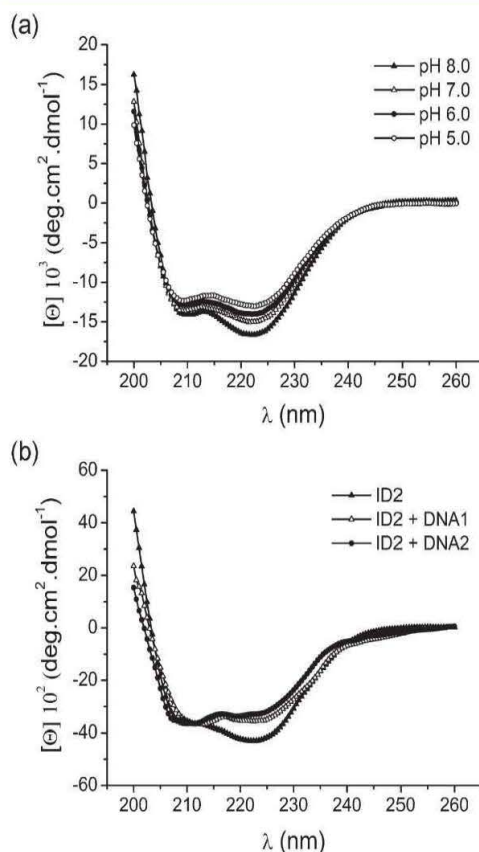
The structural model of RD2. (a) A cartoon representation of the TPR-like superhelical structure of RD2 depicting helices in cyan and loops in magenta. (b) Superposition of the NMR (gray) and modeled structures (cyan) of the PthA 1.5-repeat unit. (c) Orthogonal views illustrating the superposition of the TPR-like superhelical structure of RD2 onto its SAXS envelope.

**Figure 5**

Conformational changes of RD2. (a) A ribbon representation of the RD2 TPR-like structure highlighting the relative location of the hypervariable diresidues (yellow spheres) and the lysines (blue spheres). The approximate dimensions of the height, cleft and spacing between two consecutive hypervariable diresidues are shown. DLS curves of RD2 at different pHs (b) and after incubation with the single-strand (ss) or double-strand (ds) DNA sequence ACACATTCTAAATTTATATAAACCCCTCATCCATTTC derived from a citrus PthA-induced promoter (c).

sequence specificity, we looked at the spatial distribution of the polymorphic residues at Positions 12 (N or H) and 13 (I, D, or G). As observed in the 1.5-repeat peptide structure [Fig. 2(c)], the NI, HD, and NG pairs are located at the same side of the molecule at the tip of helix1, in close proximity to the K residues [Fig. 5(a)]. Furthermore, it is evident from the RD2 model that the polymorphic diresidues are mostly oriented to the inner face of the superhelix [Fig. 5(a)]. Thus, the only way the polymorphic diresidues could contact DNA would be the protein embracing the double strand, which is consistent with the fact that ligand binding in TPR proteins predominantly involves the concave surface of the TPR domain.⁵³ The atomic distances of the superhelix clefts (18 Å wide on average) supports this notion [Fig. 5(a)]. However, both the modeled structure and the SAXS envelop of RD2 have a molecular length of ~115 Å [Fig. 5(a)], which would cover ~34 nucleotides, that is, twice as many nucleotides as predicted to bind RD2.¹⁵ Therefore, to be able to bind a 17-nucleotide target

sequence, as predicted by the linear recognition model of one nucleotide per hypervariable diresidues,^{15,16} the RD2 protein would have to undergo some degree of compaction, like a coil spring being compressed. Surprisingly, when the hydrodynamic behavior of RD2 was investigated, substantial changes in its hydrodynamic radius (R_h) due to pH changes were found [Fig. 5(b)]. At pH 8.0, the RD2 molecule displays an R_h of 4.5 nm that is consistent with the SAXS measurements done at similar pH. However, at more acidic pH, which might mimic the DNA surface, the R_h values are around 3.5 nm, indicating that RD2 can assume a more compact structure [Fig. 5(b)]. Similarly, RD2 significantly increased its compactness (R_h of 3.6 nm) in the presence of a DNA fragment derived from a citrus promoter upregulated by PthA2 [Fig. 5(c)]. Changes in the hydrodynamic radius was not clearly observed when the corresponding single-strand sense DNA was used [Fig. 5(c)], indicating that compaction preferentially occurs upon double-strand DNA interaction.

**Figure 6**

CD spectra of RD2 at different pHs and in the presence of DNA. CD curves of RD2 protein taken at different pHs (a) and after incubation with two DNA fragments (c), showing major changes in the 222/208 nm ratios at acidic pH and in the presence of DNA. The DNA1 sequence (ACACATTCTAAATTTAT ATAAACCTCATCCATTTC) is derived from a citrus promoter upregulated by PthA2 whereas DNA2 contains the predicted PthA2 binding site¹⁵ (ACAC ATTCTAAATTTACACACCTCTTTAATAATTTC).

To further investigate these conformational changes, the secondary structure of RD2 at different pHs and in the presence of DNA was analyzed by CD spectroscopy. At acidic pH, RD2 apparently loses some of its α -helical contents, as judged by the lower intensities of the 208 and 222 nm negative signals at pHs 7.0, 6.0, and 5.0, relative to pH 8.0 [Fig. 6(a)]. However, the 222/208 nm ratio shifted from 1.3 at pH 8.0 to nearly 1.0 at pH 5.0, indicating a decrease in coiled-coil-like conformations at acidic pHs. The 222/208 nm ratio of RD2 also diminished in the presence of DNA. This was observed not only with the citrus promoter fragment but also with a DNA fragment containing the predicted PthA2 binding site¹⁵ [Fig. 6(b)]. Interestingly, it has been shown that a disulfide bond

within the TPR region of the barley Sgt1 protein is thought to promote a compaction of the helical bundles.⁴³ The reduced Sgt1 has a CD 222/208 nm ratio of ~ 1.2 , whereas oxidized Sgt1 has a ratio of ~ 0.9 ,⁴³ indicating that compaction of the TPR helical bundles is associated with a decrease in its coiled-coil-like conformation.

The repeat units of RD2 are similar to PPR motifs

As TPR domains show no obvious sequence similarities to the repeat units of RD2 and so far have not been reported to bind DNA directly, we searched for TPR-related motifs with known nucleic acid binding properties. Surprisingly, we found that the consensus sequence of PPR motifs^{58,59} is very similar to that of the repeat units of RD2 (see Fig. 7). In addition, BLAST searches using the RD2 sequence as query also identified a number of Arabidopsis (Q9MA95.2, NP_187175.1) and rice (Os02g0555100, OsJ_07125, OsJ_12287) proteins containing PPR motifs.

PPRs are formed by 2–26 tandem arrays of a degenerate 35 amino acid motif with an average of 9–12 motifs per protein.^{17,58} PPR proteins have been associated with the transcription and translational machineries, playing roles in mRNA stabilization and RNA editing.^{17,18,60} Interestingly, some of the citrus proteins that were isolated in a two-hybrid screening as interactors of PthAs are associated with RNA stabilization.¹⁴ Although no PPR structures are known, tandem PPR motifs are predicted to fold into a superhelical structure just like the TPR superhelix.^{18,58} Accordingly, the CD spectrum of maize PPR5 resembles that of RD2¹⁷ and a structural model of PPRs superposes well to the RD2 superhelix (results not shown). Thus, the structure of RD2 proposed here is consistent with the notion that nucleic acid-binding PPR motifs display a TPR-like fold. Although PPR proteins are almost absent in prokaryotes, the fact that they are predominantly found in plant mitochondria and chloroplasts⁶¹ raises the possibility that PthAs and related type III effectors might have evolved from a common PPR ancestor protein.

```
ETVQRLLPVLCQAHGLTPEQVVAIASHDG-GKQAL RD2 rep2
ETVQRLLPVLCQAHGLTPDQVVAIASHDG-GKQAL RD2 rep6
ETVQRLLPVLCQAHGLTPQVVAIASNGG-GKQAL RD2 rep8
EEA..LY..M....G..EN..TYNALINAYAK.G. PPR cons.
* . * : * *: . * .. *
```

Figure 7

The repeat units of RD2 are similar to PPR motifs. Protein sequence alignment between three polymorphic repeat units of RD2 and the consensus sequence from 1303 pentatricopeptide (PPR) motifs⁵⁷ performed by ClustalW2 (EMBL-EBI). Identical residues are shaded and indicated by asterisks, whereas similar residues are indicated by single or double dots. Dots in the PPR consensus sequence represent the most variable residues in PPR motifs.

In conclusion, this work provides the first experimental structural data of the PthA repeat domain. Both the NMR structure of the 1.5-repeat peptide and the SAXS envelope of RD2 corroborate the theoretical model that predicts a TPR-superhelical structure for RD2. Moreover, DLS and CD studies show that RD2 undergoes conformational rearrangements upon DNA interaction thus supporting the idea that the protein embraces the DNA with differential compactness as to cover a variable number of nucleotides.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Laboratório Nacional de Biociências (LNBio) and the Laboratório Nacional de Luz Sincrotron (LNLS) for providing the NMR facilities and SAXS beamline time, respectively. They also thank Renata Rocha de Oliveira for technical assistance on the CD measurements.

REFERENCES

- Brunings AM, Gabriel DW. *Xanthomonas citri*: breaking the surface. *Mol Plant Pathol* 2003;4:141–157.
- Wichmann G, Bergelson J. Effector genes of *Xanthomonas axonopodis* pv. *vesicatoria* promote transmission and enhance other fitness traits in the field. *Genetics* 2004;166:693–706.
- Cernadas RA, Camillo LR, Benedetti CE. Transcriptional analysis of the sweet orange interaction with the citrus canker pathogens *Xanthomonas axonopodis* pv. *citri* and *Xanthomonas axonopodis* pv. *aurantifolii*. *Mol Plant Pathol* 2008;9:609–631.
- Cernadas RA, Benedetti CE. Role of Auxin and Gibberellin in citrus canker development and in the transcriptional control of cell-wall remodeling genes modulated by *Xanthomonas axonopodis* pv. *citri*. *Plant Sci* 2009;177:190–195.
- Duan YP, Castañeda GZ, Erdos G, Gabriel DW. Expression of a single, host-specific, bacterial pathogenicity gene in plant cells elicits division, enlargement, and cell death. *Mol Plant Microbe Interact* 1999;12:556–560.
- Swarup S, Yang Y, Kingsley MT, Gabriel DW. An *Xanthomonas citri* pathogenicity gene, *pth A*, pleiotropically encodes gratuitous avirulence on nonhosts. *Mol Plant Microbe Interact* 1992;5:204–213.
- Al-Saadi A, Reddy JD, Duan YP, Brunings AM, Yuan Q, Gabriel DW. All five host-range variants of *Xanthomonas citri* carry one PthA homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. *Mol Plant Microbe Interact* 2007;20:934–943.
- Kay S, Bonas U. How *Xanthomonas* type III effectors manipulate the host plant. *Curr Opin Microbiol* 2009;12:37–43.
- Kay S, Hahn S, Marois E, Hause G, Bonas U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 2007;318:648–651.
- Römer P, Hahn S, Jordan T, Strauss T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* 2007;318:645–648.
- Kay S, Hahn S, Marois E, Wieduwild R, Bonas U. Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3Delta_{rep16}. *Plant J* 2009;59:859–871.
- Römer P, Strauss T, Hahn S, Scholze H, Morbitzer R, Grau J, Bonas U, Lahaye T. Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. *Plant Physiol* 2009;150:1697–1712.
- Gürlebeck D, Szurek B, Bonas U. Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. *Plant J* 2005;42:175–187.
- Domingues MN, Souza TA, Cernadas RA, de Oliveira MLP, Docena C, Farah CS, Benedetti CE. The *Xanthomonas citri* effector protein PthA interacts with citrus proteins involved in nuclear transport, protein folding and ubiquitination associated with DNA repair. *Mol Plant Pathol* 2010;11:663–675.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA-binding specificity of TAL-Type III effectors. *Science* 2009;326:1509–1512.
- Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL Effectors. *Science* 2009;326:1501.
- Williams-Carrier R, Kroeger T, Barkan A. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 2008;14:1930–1941.
- Pfalz J, Bayraktar OA, Prikryl J, Barkan A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J* 2009;28:2042–2052.
- Aue WP, Bartholdi E. Ernst RR2-dimensional spectroscopy—application to nuclear magnetic resonance. *J Chem Phys* 1976;64:2229–2246.
- Braunschweiler L, Erns RR. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J Magn Reson* 1983;53:521–528.
- Kumar A, Ernst RR, Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton crossrelaxation networks in biological macromolecules. *Biochem Biophys Res Commun* 1980;95:1–6.
- States DJ, Haberkorn RA, Ruben DJ. A two-dimensional nuclear Overhauser experiment with pure absorption phase in 4 quadrants. *J Magn Reson* 1982;48:286–292.
- Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 1992;2:661–665.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–293.
- Güntert P. Automated nmr structure calculation with cyana. *Methods Mol Biol* 2004;278:353–378.
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Panuu NS, Read RJ, Rice LM, Simonson T, Warren LG. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905–921.
- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8:477–486.
- Guinier A, Fournet G. Small angle scattering of X-rays. New York: Wiley; 1955. pp17–19.
- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 2003;36:1277–1282.
- Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 1992;25:495–503.
- Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 1999;76:2879–2886.
- Volkov VV, Svergun DI. Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr* 2003;36:860–864.
- Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 2001;34:33–41.
- Liu SM, Stewart M. Structural basis for the high-affinity binding of nucleoporin Nup1p to the *Saccharomyces cerevisiae* importin-beta homologue, Kap95p. *J Mol Biol* 2005;349:515–525.

35. Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N. Structure of the Cnd1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell* 2004;119:517–528.
36. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* 2003;374:461–491.
37. Lindahl E, Hess B, van der Spoel DL. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 2001;7:306–317.
38. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 1997;277:396–404.
39. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007;35:407–410.
40. Willard L, Ranjan A, Zhang H, Monzavi H, Robert FB, Sykes BD, Wishart DS. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* 2003;31:3316–3319.
41. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–2132.
42. Montserret R, McLeish MJ, Böckmann A, Geourjon C, Penin F. Involvement of electrostatic interactions in the mechanism of peptide folding induced by sodium dodecyl sulfate binding. *Biochemistry* 2000;39:8362–8373.
43. Nyarko A, Mosbahi K, Rowe AJ, Leech A, Boter M, Shirasu K, Kleanthous C. TPR-mediated self-association of plant SGT1. *Biochemistry* 2007;46:11331–11341.
44. Cortajarena AL, Regan L. Ligand binding by TPR domains. *Protein Sci* 2006;15:1193–1198.
45. Carrigan PE, Sikkink LA, Smith DE, Ramirez-Alvarado M. Domain:domain interactions within Hop, the Hsp70/Hsp90 organizing protein, are required for protein stability and structure. *Protein Sci* 2006;15:522–532.
46. Mirus O, Bionda T, von Haeseler A, Schleiff E. Evolutionary evolved discriminators in the 3-TPR domain of the Toc64 family involved in protein translocation at the outer membrane of chloroplasts and mitochondria. *J Mol Model* 2009;15:971–982.
47. Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 2001;29:2860–2874.
48. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000;16:404–405.
49. Schornack S, Meyer A, Römer P, Jordan T, Lahaye T. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J Plant Physiol* 2006;163:256–272.
50. Sikorski RS, Boguski MS, Goebel M, Hieter P. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* 1990;60:307–317.
51. Hirano T, Kinoshita N, Morikawa K, Yanagida M. Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc2+. *Cell* 1990;60:319–328.
52. Lamb JR, Tugendreich S, Hieter P. Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* 1995;20:257–259.
53. D'Andrea LD, Regan L. TPR proteins: the versatile helix. *Trends Biochem Sci* 2003;28:655–662.
54. Bansal PK, Nourse A, Abdulle R, Kitagawa K. Sgt1 dimerization is required for yeast kinetochore assembly. *J Biol Chem* 2009;284:3586–3592.
55. Krachler AM, Sharma A, Kleanthous C. Self-association of TPR domains: lessons learned from a designed, consensus-based TPR oligomer. *Proteins* 2010;78:2131–2143.
56. Weber P, Fulgosi H, Piven I, Müller L, Krupinska K, Duong VH, Herrmann RG, Sokolenko A. TCP34, a nuclear-encoded response regulator-like TPR protein of higher plant chloroplasts. *J Mol Biol* 2006;357:535–549.
57. Tanaka A, Takano Y, Ohnishi Y, Horinouchi S. AfsR recruits RNA polymerase to the afsS promoter: a model for transcriptional activation by SARPs. *J Mol Biol* 2007;369:322–333.
58. Small ID, Peeters N. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 2000;25:46–47.
59. Bentolila S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA* 2002;99:10887–10892.
60. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmüller R. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 2006;18:176–197.
61. O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Lineweber C, Sugita M, Small I. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* 2008;25:1120–1128.