**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS**

Evandro Ares de Araújo

**Explorando as relações entre estrutura e função das
hidrolases de glicosídeos das famílias 9, 48 e 74**

São Carlos

2018

Evandro Ares de Araújo

**Explorando as relações entre estrutura e função das
hidrolases de glicosídeos das famílias 9, 48 e 74**

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular

Orientador: Prof. Dr. Igor Polikarpov

Versão corrigida
(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2018

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica revisada pelo Serviço de Biblioteca e Informação Prof. Bernhard Gross,
com os dados fornecidos pelo(a) autor(a)

S856m	<p>Araújo, Evandro Ares de Explorando as relações entre estrutura e função das hidrolases de glicosídeos das famílias 9, 48 e 74 / Evandro Ares de Araújo ; Igor Polikarpov – versão corrigida – São Carlos, 2018. 146 p. : il. (algumas color.) ; 30 cm.</p> <p>Tese (Doutorado - Programa de Pós-Graduação em Física Aplicada) – Instituto de Física de São Carlos, Universidade de São Paulo, 2018.</p> <p>1. Raios X. 2. Cristalografia. 3. Espalhamento de raios X a baixo ângulo. 4. Enzimas. 5. Carboidrases. 6. I. Polikarpov, Igor, orientador. II. Título.</p>
-------	--

À minha mãe, Maria, por ser uma mulher forte e sem igual. Ao meu pai, José, por sua simplicidade e dedicação à família.

AGRADECIMENTOS

Agradeço ao Prof. Dr. Igor Polikarpov por ter aceitado me orientar neste projeto de pesquisa. Sem ele nada disso seria possível.

Agradeço ao CNPq pelo apoio financiamento e bolsa (IFSC-DAI/USP, processo número: 158752/2015-5), especialmente ao professor Prof. Dr. Tito José Bonagamba (IFSC-USP). Agradeço ainda à FAPESP pelo apoio financeiro, sem o qual o desenvolvimento deste projeto não seria possível.

Agradeço também aos colaboradores com quem pude colaborar em seus trabalhos e que também puderam colaborar no meu trabalho. Obrigado Lívia Manzine, Marco Kadowaki, Atílio Tomazini, Vasily Piyadov, Hévila Brognaro Mateus Urio, Lorenzo e Danilo.

Aos técnicos do Grupo de Biotecnologia Molecular Maria, Lívia, Josimar e João. Agradeço à Lívia por sua atenção e constante ajuda nas dúvidas de biologia.

Aos colegas de laboratório: Lorenzo, Danilo, Vassíly, Darlan, Renata, Atílio, Marco Antônio, Aline, Vanessa, Ana, Paula, Caio e outros que por ventura esqueci.

Aos amigos de São Carlos: Antônio, Samuel, Vassíly, Carol, Marília, Lorenzo, Danilo e Tiago. Aos amigos de fora de São Carlos: Akira, Lilian, Júlio, Graciana, Cláudio, Natalino, Carlos e jhonatan. Obrigado pelo apoio. Agradeço ainda à Jacqueline Mattos, por me fazer ver que inglês prende-se falando.

Aos professores Richard e Eduardo, pelas aulas de cristalografia e estrutura de proteínas.

Às bibliotecárias, aos técnicos do ScInfor, aos técnicos da gráfica do IFSC-USP e do serviço de Pós-Graduação do IFSC-USP por sempre estarem dispostos a ajudar.

Ao Laboratório Nacional de Luz Síncrotron - LNLS e Centro Nacional de Pesquisa em Energia e Materiais-CNPEM pela infraestrutura e apoio para realização dos experimentos.

Agradeço à Universidade de São Paulo e Instituto de Física de São Carlos, por me possibilitarem toda a estrutura e apoio necessários à minha pesquisa.

Enfim, a todos que contribuíram de forma direta ou indireta para a realização deste trabalho.

Agradeço ainda aos meus familiares, especialmente à minha mãe Maria. Agradeço ainda ao seu Alcides e toda a sua família por estes anos em São Carlos.

“Um galo sozinho não tece uma manhã: ele precisará sempre de outros galos. De um que apanhe esse grito que ele e o lance a outro; de um outro galo que apanhe o grito de um galo antes e o lance a outro; e de outros galos que com muitos outros galos se cruzem os fios de sol de seus gritos de galo, para que a manhã, desde uma teia tênue, se vá tecendo, entre todos os galos (...).”

João Cabral de Melo Neto

RESUMO

ARAÚJO, E. A. **Explorando as relações entre estrutura e função das hidrolases de glicosídeos das famílias 9, 48 e 74.** 2018. 146p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

A parede das plantas é formada por uma matriz composta principalmente por celulose, hemicelulose e lignina. Celulose é o principal polissacarídeo das paredes das plantas, apresentando alta cristalinidade e recalcitrância. Xiloglucano (XyG) é um polissacarídeo complexo envolvido no controle da expansão celular e na biossíntese de componentes da parede celular vegetal. Uma complexa rede entre XyG e celulose é mediada por ligações de hidrogênio. A eficiente hidrólise de XyG e celulose é uma estratégia promissora na desconstrução da biomassa lignocelulósica pela ação orquestrada de CAZymes. Aqui são descritas as caracterizações funcional e estrutural de três novas enzimas das hidrolase de glicosídeos das famílias 9 (*BtCel9*), 48 (*BtCel48*) e 74 (*XcGH74*). *XcGH74* é uma xiloglucanase de *Xanthomonas campestris* altamente específica para XyG. Durante a hidrólise do XyG por *XcGH74* XX e XG são os produtos finais. A estrutura cristalográfica dessa enzima foi resolvida e as razões moleculares para sua alta permissibilidade no reconhecimento de XyG analisadas. Os resultados sugerem que *XcGH74* cliva XyG preferencialmente entre motivos **X–X**; no entanto, não hidrolisa entre os motivos **L–L**, onde uma substituição da cadeia lateral é um pré-requisito para o reconhecimento do substrato. *BtCel9* e *BtCel48* são celulasas de *Bacillus licheniformis*. *BtCel48* é cataliticamente estável em uma ampla gama de temperaturas e pHs exibindo atividade em celulose inchada com ácido fosfórico (PASC) e celulose bacteriana (BC). *BtCel48* libera predominantemente celobiose, e também pequenas quantidades de celotriose, celotetraose como produtos do PASC e tem processividade aparente 4,6 vezes maior em BC do que em PASC. Análises de espalhamento de raios X a baixo ângulo (SAXS) mostraram que essa enzima é globular e monomérica em solução. A estrutura cristalográfica de *BtCel48* foi resolvida na presença de ligantes nas posições -5 a -2 e +1 a +2 no sítio catalítico. A especificidade no reconhecimento de celo-oligossacarídeo foi investigada por cromatografia de troca aniônica de alto desempenho (HPLC-PAD), cristalografia de proteínas e análise de acoplamento estatístico, mostrando que esta enzima possui endo iniciação durante a hidrólise de PASC. *BtCel9* é uma endoglucanase que exibe eficiência catalítica máxima em pH 7,0 e 60 °C. Tem maior atividade em PASC, seguida por BC e, em menor grau, carboximetilcelulose (CMC). A análise por HPAEC-PAD dos produtos hidrolíticos demonstrou que o produto final da hidrólise é principalmente celobiose. Análises de dados cristalografia de raios X e SAXS mostraram que essa enzima é monomérica em solução, conforme estimado a partir dos dados do SAXS. Tem uma forma alongada composta por um módulo de ligação à

carboidratos (CBM3c) N-terminal ligado ao domínio catalítico (GH9) C-terminal por um *linker* de 20 aminoácidos. Os domínios estão intimamente justapostos em uma conformação estendida e formam uma estrutura relativamente rígida em solução, indicando que as interações entre os domínios catalíticos e CBM3c desta enzima têm um papel cooperativo no reconhecimento da celulose. Juntos, esses resultados lançam alguma luz sobre a relação entre estrutura e função das hidrolases glicosídicas das famílias 9, 48 e 74.

Palavras-chave: Raios X. Cristalografia. Espalhamento de raios X a baixo ângulo. Enzimas. Carboidrases.

ABSTRACT

ARAÚJO, E. A. **Exploring structure-function relationships of glycoside hydrolases from families 9, 48 and 74.** 2018. 146p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

Plant cell walls form a matrix composed mainly of cellulose, hemicellulose, and lignin. Cellulose is the main polysaccharide of the plant cell walls with high crystallinity and recalcitrance. Xyloglucan (XyG) is a complex polysaccharide involved in the control of cell expansion and biosynthesis of cell walls components. A complex crosslink between XyG and cellulose is mediated by H-bonds. An efficient hydrolysis of XyG and cellulose is a promising strategy to achieve an effective lignocellulosic biomass deconstruction by orchestrated action of CAZymes. Here are described the functional and structural characterization of three novel enzymes belonging to glycoside hydrolase families 9 (*BlCel9*), 48 (*BlCel48*) and 74 (*XcGH74*). *XcGH74* is a highly specific xyloglucanase from *Xanthomonas campestris*. During the XyG hydrolysis, XX and XG are its end products. We also solved the structure of this enzyme and analyzed molecular reasons for its high permissibility in XyG recognition. These results suggest that the *XcGH74* is able to cleave XyG preferentially between **X-X** motifs; however, it is unable to hydrolyze the polysaccharide between **L-L** motifs where a side-chain substitution is a prerequisite to improved substrate recognition. The *BlCel9* and *BlCel48* are cellulases from *Bacillus licheniformis*. *BlCel48* is catalytically stable in a broad range of temperatures and pH conditions, exhibiting hydrolytic activity against phosphoric acid swollen cellulose (PASC) and bacterial cellulose (BC). *BlCel48* releases predominantly cellobiose, and also small amounts of celotriose and cellotetraose as products from PASC with apparent processivity 4.6-times greater performance on BC than on PASC. Small-angle X-ray scattering (SAXS) data analysis revealed a globular molecular shape and monomeric state in solution. The crystal structure of *BlCel48* was solved and used in presence of ligands spanning -5 to -2 and +1 to +2 subsites into the catalytic site. The specificity of the recognition of the cello-oligosaccharide was investigated by high-performance anion exchange chromatography (HPLC-PAD), protein crystallography, and statistical coupling analysis, which showed that this enzyme has an endo-like initiation on PASC. *BlCel9* is a processive endoglucanase exhibiting maximum catalytic efficiency at pH 7.0 and 60 °C, exhibiting highest hydrolytic activity against PASC, followed by BC, and to a lesser extent carboxymethyl-cellulose (CMC). The HPAEC-PAD analysis of the hydrolytic products demonstrated that the end product of the enzymatic hydrolysis is primarily cellobiose. SAXS and X-ray crystallographic data analyses revealed that this enzyme adopts a monomeric state in solution, as estimated from SAXS data; has an elongated shape composed of an N-terminal

family 3 carbohydrate-binding module (CBM3c) and a C-terminal GH9 catalytic domain joined together by 20 amino acid residue long linker peptides. The domains are closely juxtaposed in an extended conformation and form a relatively rigid structure in solution, indicating that the interactions between the CBM3c and GH9 catalytic domains might play a key role in cooperative cellulose recognition. Together, these results shed some light on the structure-function relationship of glycoside hydrolases from families 9, 48 and 74.

Keywords: X-ray. Crystallography. Small-angle X-ray scattering. Enzymes. Carbohydrate-active enzymes.

LISTA DE TABELAS

Tabela 1	– Estatísticas. Coleta, processamento e análises da estrutura cristalo- gráfica de XcGH74 obtida por difração de raios.	51
Tabela 2	– Comparação bioquímica. Comparação de propriedades bioquímicas e atividade enzimática de XcGH74 em XyG com outras enzimas da família GH74. Nr = não reportado.	58
Tabela 3	– Estatísticas de processamento. Estatística da coleta, processamento dos dados de difração, determinação estrutural e refinamento.	73
Tabela 4	– Parâmetros de coleta, processamento e análise. São apresenta- dos os parâmetros da coleta, da análise dos dados, bem como parâme- tros estruturais do processamento e análise.	86
Tabela 5	– Estatísticas de coleta e processamento. Estatística da coleta, pro- cessamento dos dados de difração, determinação estrutural e refinamento.	92
Tabela 6	– Parâmetros gerais. Os números entre parênteses são a discrepân- cia obtida a partir dos valores esperados para os pesos moleculares esperados de proteína calculados usando o software SAXSMoW 2.0 (http://saxs.ifsc.usp.br/). # Resolução calculada pelo SASRES.	104

LISTA DE ABREVIATURAS E SIGLAS

SAXS	Espalhamento de raios X a baixo ângulo
$\vec{q}$	Vetor de espalhamento
$I(q)$	Intensidade de espalhamento
$I(0)$	Intensidade de espalhamento a ângulo zero
R_g	Raio de giração
D_{max}	Dimensão máxima
MW	Massa molecular
V_m	Volume molecular
$P(r)$	Função de distribuição de distância (radial) entre pares eletrônicos
Q	Invariante de Porod
1D	Unidimensional
2D	Bidimensional
3D	Tridimensional
SEC	Cromatografia de exclusão por tamanho
UV-vis	Espectroscopia no ultravioleta visível
PDB	Repositório de estruturas biomoléculas (no inglês <i>Protein Data Bank</i>)
MR	Substituição molecular
GHs	Hidrolases de Glicosídeos
CAZy	Repositório de sequências de enzimas ativas em carboidratos (no inglês <i>Carbohydrate-Active enZymes Database</i>)
CD	Domínio catalítico
CBMs	Módulos de ligação a carboidratos
HPLC	Cromatografia líquida de alta eficiência (em inglês: <i>High performance liquid chromatography</i>)

Tris	Hidroximetil)aminometano
PEG	Polietilenoglicol
LNLS	Laboratório Nacional de Luz Síncrotron
rmsd	Valor médio para o desvio médio dos átomos (no inglês: <i>Root Mean Square deviation</i>)

LISTA DE SÍMBOLOS

Da	Dalton
λ	Comprimento de onda
Å	Angstrom
nm	nanômetro
mM	Milimolar
μ M	Micromolar
nM	Manomolar

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Panorama	23
1.2	Hidrolases de Glicosídeos e degradação da parede celular de plantas	23
1.3	Estrutura, função e arquitetura das Hidrolases de Glicosídeos	25
1.3.1	Famílias GH74, GH9 e GH48	27
1.4	Breve introdução à cristalografia de raios X e espalhamento de raios X a baixo ângulo	29
1.4.1	Breve introdução ao espalhamento de raios X	30
1.4.2	Cristais, difração de raios X e lei de Bragg	31
1.4.2.1	Difração em mono cristal, reconstrução da densidade eletrônica e problema das fases	32
1.4.3	Espalhamento de raios-X a baixo ângulo (SAXS)	33
1.4.3.1	Parâmetros estruturais	33
1.4.3.2	SAXS e cromatografia de exclusão por tamanho: SEC-SAXS	36
1.4.4	OBJETIVOS	37
2	MATERIAIS E MÉTODOS	39
2.0.1	Clonagem, expressão e purificação	39
2.0.2	Estudos bioquímicos: pH, temperatura, estabilidade térmica e especificidade por substrato	40
2.0.3	Padrão de clivagem	41
2.0.4	Estudos estruturais por espalhamento de raios X a baixo ângulo	42
2.0.5	Estudos estruturais por cristalografia de raios X	43
2.0.6	Alinhamento e análises de bioinformática	44
3	RESULTADOS E DISCUSSÕES	47
3.1	Estudos estruturais e bioquímicos de uma nova Xiloglucanase XcGH74 de <i>Xanthomonas campestris</i>: base molecular para a preferência da ramificação de xilose no sub-sítio -1	47
3.1.1	Análise filogenética baseada na sequência das GH74 caracterizadas	47
3.1.2	Caracterização bioquímica e análise da hidrólise de XyG	48
3.1.3	Estudos estruturais por cristalografia de raios X	50
3.1.4	Explorando as interações substrato-proteína no sítio ativo de XcGH74	55
3.1.5	Interações substrato-enzima nos subsítios -1/+1	57
3.1.6	Relações entre estrutura e função	58
3.1.6.1	Caracterização enzimática, especificidade pelo substrato e perfil de hidrólise	58

3.1.6.2	Base estrutural para o reconhecimento da ramificação Xyl no sub-sítio -1	60
4	RESULTADOS E DISCUSSÕES	63
4.1	Caracterização bioquímica e estrutural da exo-glucanase de <i>B/Cel48</i> de <i>Bacillus licheniformis</i>: processividade e taxa hidrolítica são limitados pelo tamanho e morfologia da celulose	63
4.1.1	Produção e caracterização bioquímica	63
4.1.1.1	Produção, caracterização bioquímica e análise dos produtos liberados da hidrólise de PASC	63
4.1.2	Perfil do produto liberado em função do tempo da hidrólise e processividade da enzima <i>B/Cel48</i> em PASC e BC	66
4.1.3	Superfície do substrato - morfologia e tamanho da fibra dirigem a processividade aparente	69
4.1.4	Estudos estruturais por cristalografia de raios X	73
4.1.4.1	sub-sítios: mapeando as interações com celo-oligosacarídeos	75
4.1.4.2	Explorando as interações substrato-proteína para a família GH48	77
4.1.5	Percepções sobre afinidades e interações de substrato/produtos no sítio catalítico usando análise de acoplamento de estatística (SCA)	80
4.1.5.1	Estudos estruturais por espalhamento de raios X a baixo ângulo	85
5	RESULTADOS E DISCUSSÕES	89
5.1	Caracterização bioquímica e estrutural de uma celulase <i>B/Cel9</i> de <i>Bacillus licheniformis</i>: correlações entre modularidade, flexibilidade e modo de ação	89
5.1.1	Produção, caracterização bioquímica, estabilidade estrutural e padrão de clivagem	89
5.1.2	Estudos estruturais por cristalografia de raios X	92
5.1.3	Modelo estrutural, domínios catalítico e acessório	93
5.1.3.1	Domínio de ligação à celulose	96
5.1.4	Estrutura cristalográfica de <i>B/Cel9</i> em complexo com celotriose e celoexaose revela o modo de ligação nos subsítios -2 a -7	97
5.1.5	Subsítios envolvidos nas interações substrato-proteína na família GH9: relacionando estrutura e função para <i>B/Cel9</i>	100
5.1.6	Estudos estruturais por espalhamento de raios X a baixo ângulo	104
5.1.6.1	<i>B/Cel9</i> é flexível em solução	106
5.1.6.2	Estudos de flexibilidade induzida por pH e temperatura	108
6	CONCLUSÕES E PERSPECTIVAS	111
	REFERÊNCIAS	115

ANEXOS	131
ANEXO A – ARTIGOS PUBLICADOS	133
A.1 Biochemical characterization and low-resolution SAXS structure of two-domain endoglucanase <i>BICel9</i> from <i>Bacillus licheniformis</i> . . .	133
A.2 SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale	135
A.3 Biochemical characterization, low-resolution SAXS structure and an enzymatic cleavage pattern of <i>BICel48</i> from <i>Bacillus licheniformis</i>	137
A.4 Biochemical characterization and low-resolution SAXS structure of an exo-polygalacturonase from <i>Bacillus licheniformis</i>	139
A.5 Biochemical Characterization and Low-Resolution SAXS Molecular Envelope of GH1 β -Glycosidase from <i>Saccharophagus degradans</i> .	141
A.6 Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism	143
A.7 Expression, purification, crystallization and preliminary X-ray diffraction analysis of <i>Aspergillus terreus</i> endo- β -1,4-glucanase from glycoside hydrolase family 12	145

1 INTRODUÇÃO

1.1 Panorama

As recentes descobertas e estudos detalhados de enzimas envolvidas em processos biotecnológicos têm tido grandes avanços, em grande parte, devido ao acesso em larga escala de genomas de diversos organismos. Isso tem contribuído para o desenvolvimento de áreas estratégicas, tais como agricultura, bioenergia, saúde e indústria de alimentos. Particularmente, o entendimento das bases físico-químicas e moleculares de enzimas ativas em carboidratos (CAZymes) tem servido de alicerce para a consolidação das indústrias de enzimas e de biocombustíveis derivados da biomassa lignocelulósica.

A matriz lignocelulósica é formada um arranjo estrutural altamente recalcitrante para a conversão à polissacarídeos fermentáveis, tais como pentoses e hexoses. Fortuitamente, alguns organismos são capazes de secretar uma mistura de CAZymes eficientes na biodegradação desses polímeros. *Bacillus licheniformis* e *Xanthomonas campestris* constituem potentes microrganismos lignocelulolíticos capazes de secretar enzimas capazes de hidrolisar os distintos componentes da biomassa. A caracterização de enzimas presentes no genoma destes organismos tem permitido a descoberta de CAZymes que podem ser aplicadas nas indústrias de biocombustíveis e produtos de química verde. Nesses organismos foram identificadas as enzimas das famílias GH9, 48 (de *Bacillus licheniformis*) e GH74 de *Xanthomonas campestris*.

Neste trabalho, foram realizados os estudos estruturais e bioquímicos das enzimas BlGH9, BlGH48 e XcGH74 afim de entender a base molecular e o modo de ação para cada uma delas.

1.2 Hidrolases de Glicosídeos e degradação da parede celular de plantas

Biomassa vegetal, composição e disponibilidade. A biomassa vegetal representa a maior fonte de carbono na biosfera terrestre - 80% da biomassa total do planeta¹ - com potencial para a produção de biocombustíveis.^{2,3} As principais fontes renováveis de biomassa provêm de produtos derivados do bagaço da cana-de-açúcar, resíduos gerados pela prática da agricultura, da indústrias madeira e de reflorestamento.^{4,5}

No Brasil, o bagaço da cana-de-açúcar é uma das principais fontes renováveis de biomassa vegetal (75% da massa total da cana-de-açúcar é bagaço) obtida como resíduo nas indústrias sucroalcooleiras e depois usado principalmente para gerar energia térmica e elétrica.^{6,7} Em média, a biomassa lignocelulósica é composta por vários polissacarídeos e polímeros fenólicos, incluindo celulose (30–50%), hemicelulose (20–40%), e lignina (5–25%).^{8,9} No caso específico da cana-de-açúcar, o bagaço é composto por cerca 40% de

celulose, 27% de hemicelulose e 28% de lignina.¹⁰

Biomassa e hidrólise enzimática. A hidrólise enzimática representa a mais importante etapa para a transformação da biomassa lignocelulósica em combustíveis e/ou produtos biotecnológicos.¹¹ No entanto, devido à sua recalcitrância e heterogeneidade, a hidrólise enzimática da biomassa ainda é um grande desafio técnico-científico para as indústrias.¹¹ Somando a isso, a hidrólise enzimática desse tipo de material ainda é um processo oneroso, em parte devido ao elevado custo e às quantidades de enzimas usadas na escala industrial.² Adicionalmente, os rejeitos dos pré tratamentos químicos/termoquímicos da biomassa são fortes inibidores da ação de CAZymes.¹² Apesar destes problemas, os estudos envolvendo a descoberta, caracterização e melhoramento de CAZymes para fins de pesquisa básica e aplicação industrial têm sido o foco de diversos grupos de pesquisa ao longo das últimas décadas.

Inúmeros microrganismos são capazes de utilizar celulose, hemicelulose e lignina como fonte de carbono por meio da hidrólise enzimática.^{13,14} Para isso, eles utilizam uma sofisticada maquinaria para produzir um arsenal de enzimas líticas e oxidativas.¹⁵⁻¹⁷ As bactérias e fungos, por exemplo, utilizam dois mecanismos de produção. *i*) Secretam um conjunto de CAZymes independentes entre si;¹⁵ *ii*) secretam complexos multi enzimáticos formando uma estrutura supramolecular chamada celulosoma;¹⁷ e por fim, *iii*) secretam uma mistura de CAZymes constituídas por diversos módulos catalíticos em uma mesma cadeia polipeptídica.¹⁶ Essas enzimas estão envolvidas na bioconversão de celulose, hemicelulose e compostos aromáticos,^{13,14} assim como polissacarídeos da microbiota intestinal de ruminantes¹⁸ e de humanos.¹⁹

CAZymes: distribuição, especificidade por substratos e modo de ação.

Atualmente mais de 12500 genomas de distintos organismos lignocelulolíticos estão disponibilizados na base de dados do CAZy (<http://www.cazy.org/>, acessado em 03-01-2019), dos quais aproximadamente 92% são bacterianos e 2% fúngicos. Nesses genomas são encontradas CAZymes envolvidas na quebra, biossíntese e modificação de carboidratos ou glicoconjugados. De acordo com o modo de ação e especificidade por substratos, essas enzimas são classificadas nas superfamílias das Hidrolases de Glicosídeos (GHs), Transferases de Glicosídeos (GTs), Liasas de Polissacarídeos (PLs), Esterases de Carboidratos (ECs) e oxidativas (AAs) (<http://www.cazy.org/>, acessado em 03-01-2019).

A superfamília das hidrolases de glicosídeos (EC 3.2.1.-) abrange enzimas com notável diversidade funcional e estrutural. São caracterizadas pela catálise da ligação glicosídica de substratos celulósicos, hemicelulósicos e pectínicos.^{20,21} Considerando o mecanismo de ação, essas enzimas são agrupadas em 156 famílias (www.cazy.org/, acessado em 09-01-2019). Celulases (famílias 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, e 124) e hemicelulases (famílias GH9, GH12, GH26, GH43, GH44, GH51 e GH74, entre outras)²² têm chamado a atenção pela enorme diversidade funcional e suas variadas aplicações na

pesquisa básica, aplicada e biotecnológica. As celulasas são divididas em três grandes grupos: endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.91) e beta-glicosídeses (EC 3.2.1.21).^{2,15} Além das celulasas, as hemicelulasas são enzimas que promovem a degradação da hemicelulose, tais como xilanas e xiloglucanas.²³ Hemicelulasas apresentam maior diversidade em termos no número de genes quanto à especificidade por substratos.

Conjuntamente com as enzimas líticas, as enzimas oxidativas - (no inglês *Lytic Polysaccharide MonoOxygenases* LPMOs) - são classificadas em quinze famílias.²² Elas pertencem a uma classe de enzimas dependentes de cobre capazes de oxidar polissacarídeos celulósicos e/ou hemicelulósicos.²⁴⁻²⁶ Para isso, utilizam o oxigênio molecular ou do seu parceiro celobiose desidrogenase para oxidar os polissacarídeos.^{24,26}

No modelo atual de ação das CAZymes, durante a hidrólise as endoglucanases clivam internamente as cadeias da fibra de celulose em regiões preferencialmente amorfas enquanto que as celobioidrolases hidrolisam as partes cristalinas da celulose. E, por fim, a celobiose é convertida em glicose pela ação das betaglicosídeses.¹⁵ Entretanto, nem a ação conjunta destas enzimas é capaz de hidrolisar completamente a celulose. Para isso, as enzimas oxidativas são capazes de contribuir fortemente no aumento da taxa hidrolítica. A ação das LPMOs torna a biomassa mais acessível para a hidrólise das enzimas lítica. São capazes produzir aumentos na atividade de 46-54% (mistura com celobioidrolase) e de 79-87% (mistura com celobioidrolase, endoglucanase e beta-glicosídeses).²⁷

Acredita-se que as hemicelulasas promovam o afrouxamento da parede celular vegetal, tendo como consequência o aumento da acessibilidade para a ação das celulasas²⁸ e enzimas auxiliares. Essa ação reflete-se no aumento da capacidade de extensão e nos efeitos de alongamento da parede.²⁹

1.3 Estrutura, função e arquitetura das Hidrolases de Glicosídeos

Algumas famílias de GHs têm mostrado uma grande diversidade na ação sobre polissacarídeos extremamente distintos entre si em estrutura e composição. Um exemplo disso é refletido na especificidade de enzimas da família 5 por substratos altamente ramificados/heterogêneos³⁰ e ao mesmo tempo por substratos não ramificados.³¹ A diversidade na função, na especificidade por substratos e no modo de ação destas enzimas é regulada não só pela topologia estrutural mas também por detalhes particulares da estrutura tridimensional dessas enzimas.²⁰ Nesse cenário inclui-se os efeitos introduzidos pela organização estrutural, orientação entre domínios e alterações conformacionais induzidas por pH e/ou temperatura.

Em termos de organização e modularidade, tanto celulasas quanto hemicelulasas podem ser compostas por um ou mais domínios (domínio(s) catalítico(s) e domínio(s) de ligação a celulose).³² O domínio catalítico está relacionado à quebra da ligação glicosídica

e o domínio(s) acessório(s) à adsorção ou ancoragem sobre a superfície do substrato com diferentes graus de especificidade.³³ Desde de que o primeiro modelo estrutural em alta resolução de uma GH foi determinado na década de 60,²⁰ a riqueza de informações derivadas de estruturas das mais distintas famílias de GHs vem pavimentando o caminho da correlação entre estrutura e função destas enzimas.

No geral, a maioria dos modelos estruturais de GHs foram obtidos a partir do domínio catalítico e pouquíssimos foram determinados na forma dessas enzimas intactas. Provavelmente, a dificuldade em cristalizar e resolver estruturas dessas enzimas com múltiplos domínios é determinada pela flexibilidade entre domínio catalítico e acessório que os *linkers*, por sua flexibilidade intrínseca, adicionam ao sistema. A natureza desses conectores (tipos de amino ácidos e tamanho) facilita a ação do domínio catalítico durante a catálise enzimática.³⁴ Essa flexibilidade pode fazer com a cristalização se torne uma tarefa ainda mais difícil do que habitualmente pode ser. Nesse sentido, os *linkers* apresentam uma dinâmica própria durante a associação enzimática do CBM com o substrato,³⁵ o que por sua vez em muitos casos pode melhorar a atividade enzimática.³⁶ Já os domínios acessórios, por sua vez, podem aumentar a atividade enzimática pelo aumento populacional nas imediações do substrato.³⁷ Por outro lado, esse aumento não implica em uma hidrólise mais eficiente.^{38,39}

Entretanto, a determinação dos modelos estruturais a baixa resolução para GHs com múltiplos domínios tem permitido avaliar o comportamento destas enzimas intactas em solução sob diversas condições. A determinação dos modelos para CBHI e CBHII de *T. reesei* permitiu observar que essas enzimas têm uma forma semelhante a um girino com o domínio catalítico formando uma cabeça isotrópica e a cauda formada pelo *linker* mais o CBM.^{40,41} Já para a Cel6A-Cel6B de *Humicola insolens* importantes informações sobre a estrutura terciária e efeito da flexibilidade foram observados.⁴² A investigação estrutural da Cel7A de *T. harzianum* forneceu também informações sobre a conformação e sobre a mobilidade em solução do módulo de ligação à celulose.⁴³ Alinhado com isso, as informações derivadas do modelo a alta resolução do domínio catalítico e do modelo a baixa resolução da enzima intacta em solução para a Cel7A de *T. harzianum* IOC 3844 foi possível observar a influência da temperatura na mobilidade, enovelamento e compacidade desta enzima em solução.⁴⁴

Por outro lado, em alguns poucos casos o modelo estrutural em alta resolução foi determinado para GHs multi modulares.^{31,45} Uma particularidade que chama atenção é o fato de elas terem um *linker* razoavelmente curto, variando de 10-20 resíduos de amino ácidos, que provavelmente reduz o grau de flexibilidade entre os domínios. Além disso, a orientação espacial entre os domínios catalítico e acessório forma uma superfície, que a priori, pode funcionar como uma extensão do sítio catalítico para a acomodação e reconhecimento do substrato.^{31,45}

1.3.1 Famílias GH74, GH9 e GH48

A maioria dos estudos mencionados anteriormente têm permitido obter informações fundamentais sobre a ação, mecanismo e a importância de CAZymes para a hidrólise da biomassa lignocelulósica, especialmente para a produção de biocombustível e bioprodutos voltados à química verde. Neste sentido, algumas classes de CAZymes vêm sendo estudadas extensivamente afim de entender as relações entre estrutura e função destas enzimas, como por exemplo celulasas e xiloglucanase.

Família GH74. Xiloglucano (XyG) é um heteropolissacarídeo hemicelulósico altamente ramificado presente em plantas, frutas e vegetais^{28,46,47} envolvido na reserva de carboidratos, biossíntese e no controle da expansão celular de vegetais.^{28,47} Estruturalmente, esse polissacarídeo tem uma cadeia central de glicose com ligações glicosídicas do tipo β -1,4 e alto grau de variação no padrão de ramificação,^{28,47} fazendo com que XyG tenha grande variação estrutural. XyG pode ser representado com base em um sistema de nomenclatura de uma letra de acordo com o padrão e espécies da ramificação: **G** para glicose (Glc) sem ramificação, **X** para a ramificação de xilose (Xyl: α -D-Xyl-Glc), **L** para a ramificação de xilose acrescida do resíduo de galactose (Gal: β -Gal- α -D-Xyl-Glc) e assim por diante.^{48,49}

XyG pode ser hidrolisado por GHs das famílias 5, 9, 12, 16, 26, 44 e 74. As hidrolases de glicosídeos da família 74 (GH74) são hemicelulasas que hidrolisam principalmente xiloglucano. O modo de ação das GH74 tem sido motivo de discussões acerca da especificidade por XyG e reconhecimento das ramificações. Tipicamente, GH74 que agem como endo-enzimas clivam internamente a ligação glicosídica do XyG no resíduo de glicose não ramificado, originando como produtos clássicos os oligossacarídeos XXXG, XLXG/XXLG e XLLG; consequentemente, as GH74 que clivam o xiloglucano de forma endo acabam liberando quatro resíduos de glicose a cada vez que atacam o carboidrato.⁵⁰

Por outro lado, algumas GH74 têm preferência por hidrolisar oligossacarídeos de XyG (XyGOs) e, neste caso, são classificadas como oligoxiloglucanases por reconhecerem especificamente as extremidades redutoras dos XyGOs, liberando dois segmentos de resíduos de glicose a partir destas extremidades.⁵¹ Martinez-Fleites et al.⁵⁰ demonstraram que o sítio catalítico da endoxiloglucanase Xgh74A (PDB id: 2CN2 e 2CN3) de *Ruminiclostridium thermocellum* tem uma grande superfície em forma de fenda, formada, na maior parte, por *loops*. Neste mesmo trabalho foi possível identificar dois resíduos como tendo função de ácido e base (Asp480 e Asp70, respectivamente), localizados na fenda catalítica.

Sobre o modo de ação das GH74 semelhantes à celobiohidrolases desempenhado pela exo-xiloglucanase OXG-RCBH (PDB id: 1SQJ e 2EBS) de *Geotrichum sp. M128*, Yaoi et al.⁵² demonstraram que a preferência por XyGOs está relacionada diretamente com a pre-

sença de um *exo-loop* Gly375–His385. Neste sentido, esse *loop* causa impedimento estérico para a acomodação do substrato no sítio ativo da enzima. A retirada deste *loop* produziu hepta, nona e octassacarídeo, tal qual a xiloglucanase de *Ruminiclostridium thermocellum* F7/YS. Em termos topológicos, essas enzimas são similares entre si mas com modos de ação muito distintos.

Família GH9. Hidrolases de glicosídeos da família 9 são encontradas principalmente em classes de organismos como arqueias, bactérias, fungos e cupins. Geralmente possuem um ou mais módulos acessórios. A Cel9A de *Thermomonospora fusca* contém um domínio catalítico e um de ligação à celulose da família III, separados por um *linker* rico em resíduos Ser/Pro/Thr, de tal modo que a orientação e a interação entre estes dois domínios com o substrato são responsáveis pelo modo de ação desta enzima.⁴⁵

Essa ideia foi melhor explorada na estrutura cristalográfica da endoglucanase Cel9G de *Clostridium cellulolyticum* complexada com celo-oligossacarídeos.⁵³ Foi possível observar que o módulo acessório (CBM) tem uma superfície plana e estendida que pode orientar as cadeias de celulose no sítio catalítico. Quando o domínio catalítico atua isoladamente na hidrólise, a atividade dessa enzima é significativamente reduzida, sugerindo a estreita relação de dependência catalítica entre o domínio catalítico e o CBM de Cel9G e Cel9A.

Os detalhes estruturais de GHs da família 9 proporcionaram informações importantes sobre interação dos resíduos no sítio catalítico com o substrato. Assim como nas celobioidrolases da família 7, *loops* localizados no sítio ativo de GHs da família 9 são essências para acomodar o substrato. A exocelulase GH9-CbhA de *Ruminiclostridium thermocellum* tem uma estrutura que adota uma conformação mais estendida que outras GH9, tais como as Cel9A de *T. fusca*, Cel9M de *C. Cellulolyticum* e Cel9A de *C. Thermocellum*.⁵⁴ No sítio ativo desta enzima há um *loop* (composto pelos resíduos 590 à 620) que bloqueia o sub-sítio -2, de modo que há o impedimento da clivagem de substratos com cadeia longa.⁵⁴ O movimento deste *loop*, provavelmente, pode permitir o acesso do substrato ao sítio ativo.⁵⁵

Essas enzimas são descritas com baixa ou nenhuma atividade sobre celulose cristalina.⁵⁶ Por outro lado, são muito eficientes na hidrólise de CMC⁵⁷ e outros polissacarídeos como beta-glucomanano, xilano⁵⁸ e xiloglucano.⁵⁹ Outra característica importante destas enzimas é sua ótima performance na hidrólise a altas temperaturas (geralmente entre 80 a 90 °C).⁶⁰

Família GH48. GH48 são hidrolases de glicosídeos capaz de despolimerizar a matéria lignocelulósica. Elas têm enovelamento estrutural similar ao das GH9 e na maioria dos casos são específicas para substratos celulósicos. A Cel48S de *C. thermocellum* é um exemplo de uma típica celulase.⁶¹ Essas enzimas têm o sítio ativo constituído por um túnel, seguido por uma fenda aberta. A endocelulase processiva CelF de *Clostridium celluloly-*

ticum tem o sítio ativo constituído por um túnel (aproximadamente 25 Å), seguido por uma fenda aberta, onde, possivelmente, os resíduos Glu55, doador de prótons, e Asp225,⁶² acceptor, promovem a catálise. Comparando CelF com outras enzimas, percebe-se que a arquitetura da CBHI de *T. reesei*⁶³ da família GH7 reserva muitas semelhanças estruturais, principalmente quanto aos aspectos do túnel catalítico. Na CelF foi possível observar que apenas subsídios +1 a +4 estão localizados na região de fenda aberta, enquanto os outros estão no túnel.

Comparando as famílias de GH9 e GH48, há muitas similaridades estruturais. Apesar de suas propriedades distintas, a ação combinada entre GHs das famílias 9 e 48 tem alto grau de sinergismo na hidrólise de celulose.⁶⁴ Em certos organismos, essas duas famílias evoluíram mantendo sua funcionalidade como proteínas multi modulares.⁶⁵ Assim, Su, Mackie e Cann⁶⁶ descreveram a bioquímica de uma enzima modular (Cel9B/Man5A) de *C. bescii* com atividades em celulose e manana.⁶⁶ Essa enzima é composta por três módulos de ligação a carboidratos (CBM3), flanqueados pelo módulo de uma GH9 (no N-terminal) e um módulo de GH5 (no C-terminal). No caso da CbCel9A/Cel48A (CelA) de *C. bescii* (com um módulo catalítico GH9 no N-terminal e um módulo catalítico GH48 no C-terminal), a configuração é muito parecida com Cel9B/Man5A. Entretanto, dos três CBMs, dois são CBM3b e um é CBM3c. Inicialmente, essa enzima foi caracterizada como multifuncional.⁶⁷ Quando se compara a hidrólise em Avicel de CelA com uma de suas truncagens (truncagem T1: CelA sem o módulo catalítico GH48), CelA é duas vezes mais eficiente na liberação de glicose.⁶⁸ Porém, curiosamente, quando se mistura o módulo GH9 (truncagem T2 de CelA) e a truncagem T1 a hidrólise de celulose cristalina ocorre de forma sinérgica entre estas duas construções, levando a pressupor a hipótese de que a truncagem T1 (módulo GH48 e os CBM3bs) de CelA desestabiliza determinadas regiões da celulose cristalina para o ataque do módulo GH9 (T2).⁶⁸

Em outro trabalho mais recente com CelA, com base nos resultados estruturais (módulo GH48 PDB id: 4EL8 e módulo GH48 PDB id: 4DOD), dinâmica molecular, microscopia eletrônica de transmissão e ensaios funcionais, observou-se que, ao incubar Avicel PH101 com CelA, essa enzima foi capaz de escavar extensas áreas na superfície do substrato, demonstrando que essa enzima é eficiente na hidrólise de biomassa não pré-tratada.⁶⁹ Além disso, a comparação de CelA com a combinação da exocelobioidrolase I de *Trichoderma reesei*⁷⁰ (55) e E1 (GH5) de *Acidothermus cellulolyticus*⁷¹ mostrou que CelA é sete vezes mais eficiente em Avicel.⁶⁹

1.4 Breve introdução à cristalografia de raios X e espalhamento de raios X a baixo ângulo

Muito do entendimento da estrutura de proteínas se deve aos trabalhos pioneiros em cristalografia de proteínas, pelo entendimento das propriedades dos raios X para a

difração⁷² e espalhamento de raios X a baixo ângulo. Somando-se a isso, o uso de fontes de radiação síncrotron voltadas à determinação estrutural de macromoléculas vem impactando fortemente a área de biologia estrutural.⁷³ Essas duas técnicas são complementares entre si têm respondido pela maior parte de estruturas depositas nos repositórios públicos PDB (no inglês *Protein Data Bank*) e SASBDB (no inglês *Small Angle Scattering Biological Data Bank*).

1.4.1 Breve introdução ao espalhamento de raios X

O entendimento da função e estrutura de biomoléculas em grande parte se deve ao desenvolvimento de técnicas que permitam a reconstrução em escala atômica. No caso de raios X, a reconstrução estrutural destas moléculas tem sido obtida com sucesso ao longos dos anos por cristalografia de proteínas e espalhamento de raios X a baixo ângulo (SAXS). Essas duas técnicas são complementares e são duas poderosas ferramentas para a biologia estrutural. Entretanto, existem outras técnicas que também permitem a determinação estrutural destas moléculas mas não serão abordada neste trabalho.

Espalhamento de raios X é implementante usado em Física e ciências da vida. No eletromagnetismo clássico pode-se considerar que um feixe de raios X ao interagir com a matéria seu campo elétrico $\vec{E}$ produz uma força ($\vec{F} = q_e \vec{E}$) nos elétrons livres dos espalhadores. Esses elétrons ressonam na mesma frequência dos raios X incidentes e emitem ondas secundárias com intensidade dada pela equação de Thomson:^{74,75}

$$I_s(\theta) = \beta I_p \frac{1}{r^2} \left(\frac{1 + \cos 2\theta}{2} \right) \quad (1.1)$$

, onde 2θ é o ângulo de espalhamento, I_p é a intensidade das ondas primária, I_s a intensidade das ondas secundárias, $\beta = \frac{e^2}{mc^2}$ o raio do elétron clássico, r é a distancia do objeto ao detector e $\frac{1+\cos 2\theta}{2}$ o fator de polarização. Assumindo que as ondas secundárias têm amplitude e intensidade definidas. Cada onda secundária tem uma fase ϕ que permite a distinção destas ondas entre si.⁷⁴ A fase da onda espalhada depende da posição dos elétrons no espaço de espalhamento.⁷⁵ A notação vetorial permite uma análise simples e intuitiva dos princípios do espalhamento. Para um elemento de volume com densidade $\rho(\vec{r})$ centrado na origem com espalhamento até um ponto arbitrário com distância $\vec{e}$, o feixe incidente pode ser denotado pelo vetor unitário $\vec{S}_0$ e o espalhado por $\vec{S}_1$.^{74,75} A diferença de caminho entre as ondas incidente e espalhada será $\vec{S}_1 - \vec{S}_0 = \vec{S}$. A fase da onda secundaria será $2\pi/\lambda$ vezes a diferença de caminho $\vec{S}$. A diferença de fase poderá ser calculada como $\phi = -\frac{2\pi}{\lambda} \vec{r} \cdot (\vec{S}_1 - \vec{S}_0) = -\frac{2\pi}{\lambda} \vec{r} \cdot \vec{S}$. O termo $\vec{S}_1 - \vec{S}_0$ é simétrico em relação aos feixes incidente e espalhado com magnitude $2 \sin \theta$, onde θ representando metade do ângulo de espalhamento. Algumas vezes o vetor $\vec{S}$ é substituído por $\vec{q}$ normalizado por um fator de 2π ($\vec{q} = 2\pi \vec{S}$),^{74,75} como por exemplo em SAXS. Nesse último caso o módulo do vetor $\vec{q}$ pode ser dado por:

$$|q| = \frac{4\pi \sin \theta}{\lambda} \quad (1.2)$$

, onde 2θ é o ângulo de espalhamento.

Cada onda secundária pode ser descrita vetorialmente pela exponencial complexa $e^{i\phi}$ definida por uma fase $\phi = 2\pi\vec{S}\cdot\vec{r}$ e magnitude.^{74,75} A onda resultante será obtida pela integração da contribuição de cada espalhador representada pelo fator de espalhamento atômico ou fator de forma:

$$F(\vec{S}) = \int_{volume} \rho(\vec{r})e^{(2\pi i\vec{r}\cdot\vec{S})} dv. \quad (1.3)$$

Essa expressão tem simetria esférica e depende somente da magnitude de $\vec{S}$. Isso significa que $\vec{S} = 0$, $\vec{S}$ ($F(S \rightarrow 0) = \int_{volume} \rho(\vec{r})dv = Z$) será proporcional ao número de elétrons Z do elemento espalhador. Em outras palavras $F(\vec{S})$ é a transformada de Fourier da densidade eletrônica $\rho(\vec{r})$ ($\mathcal{L}\{\rho(\vec{r})\} = F(\vec{S})$).

1.4.2 Cristais, difração de raios X e lei de Bragg

Os estudos estruturais por cristalografia de macromoléculas têm ajudado no avanço de diversas áreas das ciências da vida, incluindo medicina e biotecnologia, por exemplo. A interação entre fármaco e moléculas-alvos, o mecanismo molecular do reconhecimento por antibióticos, e até mesmo a elucidação da estrutura do ribossomo, são alguns destes avanços.⁷⁶

A cristalização de proteínas é um passo essencial para obtenção de uma estrutura molecular e tem sido feita por diversas técnicas, dentre elas difusão a vapor é a mais comumente usada por sua simplicidade e taxa de sucesso no crescimento de cristais.^{76,77} A base disso está na supersaturação da solução de proteínas:⁷⁸ na fase de supersaturação as proteínas em solução excedem seu limite de solubilidade, há início da nucleação e, por fim, crescimento do cristal. Entretanto, nem sempre esse é um processo rápido e fácil. Na maior parte do tempo é um processo demorado e requer melhorias na qualidade dos cristais. O processo de semeadura⁷⁹ é uma das muitas intervenções que se podem fazer para melhorar a qualidade de um cristal. A qualidade do cristal é um dos fatores que definem o sucesso no processamento e determinação estrutural de uma proteína.

De posse de um cristal, os experimentos de difração de raios X podem ser então conduzidos. A incidência do feixe de raios X sobre o cristal tem como resultado a difração das ondas nos múltiplos planos cristalográfico que são iluminados pelos raios X. Essa ideia geométrica para a difração de raios X é matematicamente expressa pela lei de Bragg:^{75,80}

$$2d \sin \theta = n\lambda, \quad (1.4)$$

onde λ é o comprimento de onda, n um número natural, d é a distância entre planos atômicos e θ é o ângulo de espalhamento. A interpretação do espalhamento como reflexões nos planos da cela permite estabelecer a relação entre o espalhamento no espaço real e no espaço recíproco.^{75,80} Em outras palavras, permite quantificar d para um conjunto de

planos de reflexões hkl (d_{hkl}) e extrair as relações entre o vetor de espalhamento d_{hkl} e seu recíproco d_{hkl}^* na cela unitária com índices de Miller hkl , assim como as condições de difração nos planos cristalográficos:^{75,80}

$$2d_{hkl} \sin \theta = n\lambda \iff \frac{1}{d_{hkl}} = \frac{2 \sin \theta}{n\lambda} = d_{hkl}^*. \quad (1.5)$$

1.4.2.1 Difração em mono cristal, reconstrução da densidade eletrônica e problema das fases

Dado que um cristal tem um arranjo 3D periódico de átomos em uma molécula, a função da onda total espalhada nos planos do cristal na direção $\vec{S}$ é dada por $F(\vec{S}) = \sum_{j=1}^N f_j e^{(2\pi i \vec{r}_j \cdot \vec{S})}$ e o fator de estrutura será então:

$$F_{hkl} = \sum_{j=1}^N f_j e^{[2\pi i (hx_j + ky_j + lz_j)]}, \quad (1.6)$$

onde f_j é o fator de espalhamento atômico para cada átomo, $\vec{r}_j = \vec{a}x_j + \vec{b}y_j + \vec{c}z_j$ e $\vec{r}_j \cdot \vec{S} = hx_j + ky_j + lz_j$. Essa representação permite calcular o padrão de difração numa cela unitária dado que as posições de todos os átomos sejam conhecidas.⁷⁵ Entretanto, no experimento de difração é possível somente quantificar as intensidades ($I_{hkl} = F_{hkl} \cdot F_{hkl}^* = |F_{hkl}|$) das ondas difratadas durante o espalhamento.

Uma vez que o padrão de difração de um cristal pode ser calculado a partir da equação para o fator de estrutura F_{hkl} , é possível reconstruir o mapa da densidade eletrônica $\rho(\vec{r})$. Para fazer isso é preciso reescrever $F(\vec{S}) = \sum_{j=1}^N f_j e^{(2\pi i \vec{r}_j \cdot \vec{S})} = \int_{\text{volume}} \rho(\vec{r}) e^{(2\pi i \vec{r} \cdot \vec{S})} dv$ multiplicada por $e^{(-2\pi i \vec{r}_j \cdot \vec{S})}$, o que resulta na função $\rho(\vec{r}) = \int_{\text{volume}} F(\vec{S}) e^{(-2\pi i \vec{r} \cdot \vec{S})} dv$. Entretanto, $\rho(\vec{r})$ é uma função não contínua que pode ser escrita como:

$$\rho_{xyz} = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl}^* e^{[-2\pi i (hx_j + ky_j + lz_j)]}. \quad (1.7)$$

A partir dessa equação é possível o cálculo da densidade eletrônica, que por sua vez representa a estrutura eletrônica do cristal na unidade da cela.⁷⁵ Ou seja, o fator de estrutura é a transformada de Fourier da densidade eletrônica.

A transformada de Fourier estabelece uma relação entre fator de estrutura e densidade eletrônica. Isso significa que para calcular a densidade eletrônica de uma proteína, por exemplo, é preciso conhecer as amplitudes e as fases do fator de estrutura.⁷⁵ Entretanto, para cada intensidade I_{hkl} medida no experimento de difração são observados fatores de estrutura F_{obs} equivalentes mas a fase da onda associada é perdida. Isso se aplica a terminação da estrutura cristalográfica de qualquer molécula. Esse problema é chamado de problema das fases.⁷⁵

Para a determinação de estrutura cristalográfica de proteínas, em particular, existem diversas metodologias que podem ser usadas para resolver o problema das fases,^{75,80}

incluindo o método de substituição molecular⁸¹ (no inglês *Molecular replacement* - MR). Os pré-requisitos para a utilização do método de MR são:^{81–84} um padrão de difração observado para a estrutura desconhecida e as coordenadas atômicas de uma estrutura de uma proteína homóloga (sonda), de tal modo que a tarefa do MR envolve o posicionamento da sonda no interior da cela unitária do cristal da proteína-alvo, de maneira que o padrão de difração teórico que resultaria deste modelo se aproxima do experimental, com orientação correta da sonda e o posicionamento da molécula corretamente orientada dentro da cela unitária via resolução das funções de rotação e translação.

O método de substituição molecular é largamente utilizado para resolver o problema das fases de estruturas cristalográficas.⁸⁵ Entretanto é estritamente dependente da existência da estrutura de um homólogo com identidade sequencial de pelo menos 25%.^{83,85,86} Mais recentemente tem sido possível a determinação do modelo estrutural por MR sem a restrição da similaridade sequencial de um homólogo.⁸⁷ Por outro lado, caso as condições para o MR não seja satisfeitas, é preciso investir na solução do problema das fases buscando determiná-las experimentalmente.^{88–90}

1.4.3 Espalhamento de raios-X a baixo ângulo (SAXS)

Espalhamento de raios-X a baixo ângulo (SAXS) é um método de caracterização estrutural experimental para análise rápida e robusta de macromoléculas em solução.^{74,91–93} Uma das grandes vantagens para o uso de SAXS é que as amostras não precisam ser cristalizadas, podendo ser estudadas em diferentes ambientes de pH.^{94,95} Isso permite entender a estrutura biológica de macromoléculas o mais próximo de suas condições fisiológicas.

1.4.3.1 Parâmetros estruturais

Intensidade de espalhamento. SAXS é uma técnica que ao longo dos anos vem sendo cada vez mais usada para a análise de biomoléculas em solução. Os fundamentos de SAXS foram inicialmente estabelecidos nos trabalhos pioneiros de Guinier⁹⁶ e Debye⁹⁷ e posteriormente, à luz destas ideias, diversos outros trabalhos contribuíram para o estabelecimento desta técnica.^{74,91} Considerando o espalhamento de uma proteína solução, a função de espalhamento $F(\vec{q})$ depende do número de espalhadores N e do vetor de espalhamento $\vec{q}$:

$$F(\vec{q}) = \sum_{i=1}^N f_j(\vec{q}) e^{i\vec{q} \cdot \vec{r}} \quad (1.8)$$

, onde $\vec{r}$ é a distância do centro de espalhamento até um espalhador k arbitrário. $\vec{q}$ é o vetor de espalhamento definido no espaço recíproco ($|q| = \frac{4\pi \sin \theta}{\lambda}$, com λ sendo o comprimento de onda e 2θ o ângulo de espalhamento) e $f_j(\vec{q})$ é o fator de forma atômico.

Entretanto, $F(\vec{q})$ não pode ser medido diretamente durante o experimento. No sistema de detecção é possível medir as intensidades de espalhamento dos objetos espalhadores ($I(\vec{q}) = |F(\vec{q})|^2 = F(\vec{q}) * F(-\vec{q})$). Uma das consequências imediatas disso é que a

intensidade I não carrega as informações associadas às fases das ondas. Adicionalmente, ao se assumir que a) o espalhamento seja isotrópico, b) o meio homogêneo, c) o contraste $\Delta\rho = \rho_{amostra} - \rho_{solvente}$ não seja variável e que d) as partículas tenham movimento independente entre si, a média sobre todas as orientações é dada por $\langle e^{-i\vec{q}\cdot\vec{r}_{ij}} \rangle = \frac{\sin qr_{ij}}{qr_{ij}}$. Isso permite reescrever a intensidade como:

$$I(\vec{q}) = \sum_{i=1}^N \sum_{j=1}^N f_i(\vec{q}) f_j(\vec{q}) \frac{\sin qr_{ij}}{qr_{ij}} \quad (1.9)$$

, onde $r_{ij} = |\vec{r}_i - \vec{r}_j|$.

Tanto em laboratórios quanto em modernas linhas de luz baseadas em radiação síncrotron dedicados à SAXS, os dados de SAXS são obtidos por detectores bidimensionais (2D), geralmente usando sistema de detecção de um detector CCD ou Pilatus. A média angular de padrões de detectores 2D produz uma intensidade $I(q)$ de espalhamento unidimensional (1D) que é função do módulo do vetor de espalhamento $|q|$. É a partir deste dado experimental que todos os parâmetros estruturais (incluindo raio de giro R_g , dimensão máxima D_{max} da molécula, volume molecular V_m e massa molecular MW , por exemplo), envelope molecular e flexibilidade ou mudanças conformacionais serão extraídos.

Função distribuição de pares de distâncias $P(r)$ e lei de Guinier. A função $P(r)$ é o histograma das distâncias na partícula e representa a distribuição de distâncias entre pares de elétrons limitados pelo volume da partícula.^{74,91} Essa função está diretamente conectada com a intensidade de espalhamento por uma transformada de Fourier:^{74,91,92}

$$P = \frac{r^2}{2\pi^2} \int_0^\infty I(q) q^2 \frac{\sin qr}{qr} dq. \quad (1.10)$$

Da mesma forma, se a $P(r)$ é conhecida, é possível calcular a curva de intensidade de espalhamento I por uma transformada de Fourier inversa:^{74,91,92}

$$I(\vec{q}) = (4\pi) \int_0^\infty P(\vec{r}) \frac{\sin qr}{qr} dr. \quad (1.11)$$

Na prática, a $P(r)$ está limitada entre zero e a dimensão máxima D_{max} da partícula ($0 \leq r \leq D_{max}$). Essa função é a representação dos dados de SAXS, como também contém informação sobre o tamanho e forma da molécula no espaço real. A obtenção da $P(r)$ a partir das intensidades é possível através de uma transformada de Fourier indireta das intensidade de espalhamento.⁹⁸ Essa metodologia foi inicialmente proposta por Glatter⁹⁸ e posterior implementada no programa GNOM⁹⁹ por Svergun.¹⁰⁰

A partir da função $P(r)$ é possível obter dois importantes parâmetros da molécula em solução: o raio de giro e a intensidade a ângulo zero $I(q \rightarrow 0)$. O raio de giro ($R_g^2 = \frac{1}{2} \frac{\int r^2 P(\vec{r}) dr}{\int P(\vec{r}) dr}$) descreve a distribuição de massa da partícula em torno do centro de massa.

Já a $I(q \rightarrow 0)$ ($I(q \rightarrow 0) = \int_0^{D_{max}} P(\vec{r})dr$) traz informação sobre o número de partículas em solução. Esses dois parâmetros também podem ser obtidos no espaço recíproco através da aproximação de [Guinier](#).⁹⁶

Para sistemas monodispersos e com q tendendo a zero, a intensidade $I(\vec{q})$ pode ser aproximada por uma expansão de Taylor na forma:

$$I(\vec{q}) = 4\pi \int_0^{D_{max}} P(\vec{r}) \frac{1}{qr} [qr - \frac{1}{3!}(qr)^3 - \frac{1}{5!}(qr)^5 + \dots] dr. \quad (1.12)$$

Nessa expansão somente os termos de ordem $\leq q^3$ vão contribuir para $I(\vec{q})$. Neste caso $I(\vec{q}) = 4\pi \int_0^{D_{max}} P(\vec{r})dr [1 - \frac{q^2}{3!} \frac{\int_0^{D_{max}} r^2 P(\vec{r})dr}{\int_0^{D_{max}} P(\vec{r})dr}]$. Lembrando da forma das equações do raio de giro e da intensidade a ângulo zero, é fácil ver que elas estão contidas nessa última equação de $I(\vec{q})$. Levando isso em consideração, nós finalmente podemos expressar $I(\vec{q})$ como:

$$I(q) = I(q \rightarrow 0) e^{-\frac{q^2 R_g^2}{3}}. \quad (1.13)$$

Essa aproximação é conhecida como lei Guinier.⁹⁶

Essa aproximação é uma maneira de obter $I(q \rightarrow 0)$ e R_g no espaço recíproco. Para isso, pode-se fazer $\ln I(q) = \ln I(q \rightarrow 0) - q^2 R_g^2/3$. Os valores do R_g e $I(q \rightarrow 0)$ podem ser extraídos simplesmente fazendo o ajuste linear dessa equação, fazendo $\ln I(q)$ versus q^2 . A região a baixo q utilizada no ajuste linear depende do limite $q \cdot R_g$. Para proteínas globulares esse limite é $q \cdot R_g < 1,3$ e, à medida que as proteínas vão tornando-se alongadas, este limite tende a ser maior que 1,3.^{91,96}

Invariante de Porod, volume e massa molecular. Outros importante parâmetros derivados da curva de intensidades de SAXS são o volume e o invariante de Porod.^{91,101,102} Assumindo a situação em que q tende a zero, Porod propôs^{74,91} que o termo $\int_0^\infty I(q)q^2 dq$ da equação 1.10 é invariante sobre quaisquer modificações particular do sistema e/ou da sua função de espalhamento. O invariante pode ser obtido pelo cálculo:

$$Q = \int_0^\infty I(q)q^2 dq. \quad (1.14)$$

Uma consequência de extrapolar as intensidades a ângulo zero é que $I(q \rightarrow 0)$ é proporcional à área da $P(r)$ ($I(q \rightarrow 0) = (\Delta\rho \cdot V)^2 = 4\pi \int_0^{D_{max}} P(\vec{r})dr$). Isso permite observar que o invariante Q é proporcional à densidade e ao volume da partícula:

$$\frac{1}{2\pi^2} \int_0^\infty I(q)q^2 dq = (\Delta\rho)^2 V, \quad (1.15)$$

e por seguinte do volume:

$$V = 2\pi^2 \frac{I(q \rightarrow 0)}{Q}. \quad (1.16)$$

Uma consequência imediata do cálculo de V é a estimativa de MW . Uma metodologia que usa esse tipo de informação para o cálculo de massa molecular foi implementada

no programa SAXSMoW.^{101,102} O diferencial neste trabalho é que os autores calculam o Q' truncado ($Q' = \int_0^{q_{max}} I(q)q^2 dq$), e com isso é possível então calcular volume V' ($V' = 2\pi^2 \frac{I(q \rightarrow 0)}{Q'}$) aparente da proteína, e por fim a massa molecular da proteína em solução.^{101,102}

$$MW[\text{kDa}] = \frac{\rho_m[\text{g/cm}^3]V[\text{cm}^3]}{1,662 \cdot 10^{-21}[\text{g/kDa}]} \quad (1.17)$$

, onde $\rho_m = 1,37\text{g/cm}^3$.

Cálculo de flexibilidade. Por fim, os dados de SAXS podem trazer informações sobre mudanças conformacionais e flexibilidade de proteínas induzidas pelo microambiente da solução. Informações sobre mudanças conformacionais e flexibilidade de proteínas em solução podem ser obtidas pela análise do gráfico de Kratky.^{74,103,104} Proteínas globulares tipicamente apresentam no gráfico de Kratky ($I(q)q^2$ versus q) um pico (pico quase Gaussiano) proeminente a baixo ângulo, enquanto que proteínas não estruturadas mostram um aumento contínuo em $I(q)q^2$ à medida que q aumenta.^{104,105} Já para proteínas com múltiplos domínios, e flexíveis entre si, exibe-se um perfil no gráfico de Kratky que pode ser a mistura dos perfis para proteínas globulares junto com perfil das proteínas desestruturadas.^{104,105}

Entretanto, há casos em que proteínas mesmo sendo flexíveis podem não apresentar indícios de flexibilidade pela análise de Kratky.¹⁰⁶ Além disso, esse método é altamente sensível à subtração do tampão.¹⁰⁴ Alternativamente, o gráfico de Kratky adimensional ($I(q)(qR_g)^2$ versus (qR_g))¹⁰⁷ permite a análise mais confiável da flexibilidade e/ou desordem para biomoléculas em solução.^{104,108} Para proteínas globulares e que respeitam a aproximação de Guinier, o gráfico de Kratly adimensional terá um pico a $\sqrt{3}$ (em y) com magnitude 1,104 (em x), independente do tamanho, massa e concentração da amostra.^{104,108} Desvios deste perfil sugerem que a molécula apresenta flexibilidade ou assimetria na forma do envelope.^{104,107,108} Informações similares e/ou complementares podem ser obtidas fazendo a análise de Porod-Debye ($I(q)q^4$ versus q^4).^{104,108}

1.4.3.2 SAXS e cromatografia de exclusão por tamanho: SEC-SAXS

Bio SAXS tem-se tornado cada vez mais uma poderosa técnica na caracterização, obtenção de parâmetros estruturais e observação de mudanças estruturais em proteínas e complexos macromoleculares em solução.^{109,110} Entretanto, para se obter informação estrutural ou até mesmo a reconstrução da estrutura da molecular em baixa resolução, o sistema precisa ser monodisperso e diluído.¹⁰⁹

Preparar amostras para SAXS ainda é um desafio porque, de modo geral, moléculas biológicas existem, particularmente, em múltiplas conformações ou estados conformacionais ou fazendo complexos com outras proteínas. Associado a isso, problemas com contaminantes provenientes de uma purificação incompleta podem levar a interações inespecíficas, agregação, degradação,¹⁰⁹ como também á formação de complexos com baixa

constante de afinidade.¹¹¹ Esses são fatores que dificultam ou até mesmo impossibilitam as análises posteriores à coleta de dados já que cada espécie na mistura vai contribuir para a intensidade total dos objetos espalhadores.¹⁰⁹ Na maioria dos casos o efeito de contaminantes e polidispersividade é detrimental para a correta interpretação estrutural dos dados de SAXS.¹⁰⁹

Uma técnica complementar à medida de SAXS é o uso de cromatografia de exclusão por tamanho acoplada imediatamente antes de amostras serem submetidas aos raios X (SEC-SAXS).¹¹² Diversos trabalhos usando essa metodologia foram conduzidos em alguns laboratórios Síncrotron.^{113–116} Neste tipo de cromatografia uma solução qualquer (amostra e tampão, por exemplo) - fase móvel - passa através de uma matriz porosa - fase estacionária - na qual as partículas são retidas nesta matriz de acordo com o tamanho e forma - partículas menores são retidas por mais tempo enquanto partículas maiores são eluídas em menor tempo. A resolução da coluna (definida como a separação entre dois máximos no espectro de absorbância no UV-vis a 280 nm, por exemplo) depende do tipo de coluna (material, tamanho do poro e volume) e das condições escolhidas durante a cromatografia (fluxo, tipo de tampão usado e volume de injeção). SEC apresenta duas grandes vantagens para a purificação de biomoléculas: i) os experimentos podem ser conduzidos em diferentes tampões, o que permite a otimização da amostra para condições de maior estabilidade; ii) o uso combinado de leitura espectroscópica UV-vis, o que possibilita a avaliação qualitativa do estado oligomérico e presença de agregados presentes na amostra de biomolécula.¹⁰⁹

SEC-SAXS permite não somente superar as dificuldades com contaminantes e polidispersividade,¹¹⁵ mas também a análise de sistemas intrinsecamente desordenados e flexibilidade macromolecular. Permite ainda a avaliação individual dos componentes da amostra independente da medida de SAXS, o que permite a determinação do estado oligomérico em solução seja utilizando métodos baseados na intensidade de espalhamento quando tende a ângulo zero⁹⁹ ou método que calcula a massa molecular a partir da intensidade da curva de espalhamento.^{101,102} Isso facilita a obtenção de dados e a correta interpretação estrutural obtidos por SAXS.

1.4.4 OBJETIVOS

Nesta tese realizamos os estudos estruturais utilizando cristalografia de raios X, espalhamento de raios-X a baixo ângulo (SAXS) para elucidar a estrutura e bioquímicos das hidrolases de glicosídeos das famílias 9, 48 (*B. licheniformes*) e 74 (*X. campestris*). Mais especificamente, das enzimas *BlGH9*, *BlGH48* e *XcGH74*.

- Especificamente, nossos objetivos foram: clonagem e expressão heteróloga; definir condições ótimas para pHs, temperaturas e atividade ótima de cada enzima; caracterização de seus perfis de liberação de produtos gerados pela hidrólise destas enzimas;

conduzir estudos estruturais por SAXS e cristalografia de proteínas; resolver as estruturas destas proteínas na presença e ausência de ligantes; avaliar os efeitos do pH e temperatura para a flexibilidade de proteínas multimodulares e por identificar os modos de ação das enzimas na depolimerização da biomassa lignocelulósica.

2 MATERIAIS E MÉTODOS

2.0.1 Clonagem, expressão e purificação

A partir do DNA genômico de *Bacillus licheniformis* DSM-13 (ATCC 14580) e *Xanthomonas campestris* pv. *campestris* (ATCC33913), os genes para *BlCel9* (GenBank: AAU40775.1), *BlCel48* (GenBank: AAU234116.1) e *XcGH74* (GenBank: NP_637119.1), respectivamente, foram clonados usando o sistema independente de ligase (LIC)¹¹⁷ no vetor de expressão pETTRXA-1a/LIC e transformados individualmente em *Escherichia coli* rosetta(DE3)pLysS (Novagem). *XcGH74* foi clonada como descrito por [de Araujo, E. A.](#)¹¹⁸

O processo de expressão se deu por crescimento celular de 5 h a 37 °C e expressão por 20, 24 e 25 h, respectivamente *BlCel9*, *BlCel48* e *XcGH74*, a 20 °C, usando meio de cultura auto indutor¹¹⁹ suplementado com canamicina e cloranfenicol em concentrações de 50 e 25 µg/mL, respectivamente. Ao final, as células foram concentradas por centrifugação a 10000g e suspendidas em tampão Tris-HCl 20 mM (pH 8,0), complementado de 300 mM de NaCl e 1mM de fluoreto de fenilmetilsulfonila (PMSF do inglês *phenylmethylsulfonyl fluoride*). As células foram submetidas a um processo de ruptura celular por ultrassom e separadas por centrifugação a 20000g.

Três etapas de purificação foram realizadas: *i*) cromatografia de afinidade Ni^{2+} (Ni-NTA, Quiagen); *ii*) segunda passo cromatográfico com Ni^{2+} e; por fim, *iii*) cromatografia por exclusão molecular em colunas Superdex 75 HiLoad 16/60 utilizando o sistema AKTA purifier 10 (GE Healthcare Life Sciences). Na etapa *i*, as etapas de lavagens foram realizadas com tampão Tris-HCl 20 mM (pH 8,0) complementado com 300 mM de NaCl e gradiente de 10-30 mM de imidazol; posteriormente, realizou-se a eluição das proteínas de interesse, aumentando a concentração de imidazol do tampão de eluição para 50-300 mM.

Posteriormente as proteínas foram então dialisadas (em tampão Tris-HCl 20 mM (pH 8,0) complementando-se com 300 mM de NaCl para remoção do imidazol e submetidas a clivagem da cauda de histidina com TEV protease (do inglês *Tobacco Etch Virus*) com incubação por 24 h. Por fim, a solução proteica foi submetida novamente à coluna de Ni^{2+} , etapa *ii*, no tampão Tris-HCl 20 mM (pH 8,0) 150 mM de NaCl). As proteínas então clivadas foram submetidas ao último passo de purificação, etapa *iii*, por cromatografia de exclusão por tamanho (SEC) em coluna Superdex 75 HiLoad 16/60 previamente equilibrada com tampão Tris-HCl 20 mM (pH 8,0) e NaCl 200 mM.

Finalmente, as frações referentes à proteína de interesse foram coletadas e a pureza final avaliada por gel SDS-PAGE 15% corado com Azul Brilhante de Coomassie¹²⁰ e a

concentração determinada utilizando o espectrofotômetro (NanoDrop, ThermoScientific).

2.0.2 Estudos bioquímicos: pH, temperatura, estabilidade térmica e especificidade por substrato

Experimentos de varredura diferencial da fluorescência (ThermoFluor) foram realizados para investigar os efeitos de temperatura e pH sobre a estabilidade estrutural de cada enzima. As medidas foram conduzidas usando o sistema CFX96 Real-Time (Bio-Rad) em placas de 96 poços (Bio-Rad) com 0,1 $\mu\text{g/mL}$ de proteína em diferentes tampões (variando pH de 1,2 a 10,0 com 50 mM de cada solução final) e sonda 2000 X SYPRO Orange dye (Molecular Probes). Em cada reação variou-se a temperatura de 25 °C a 90 °C com incrementos de 1 °C e tempo de estabilização de 30 segundos. A variação de fluorescência da sonda foi medida usando excitação a 490 nm e emissão a 530 nm.

Os experimentos para determinar a atividade enzimática foram realizados incubando-se as proteínas em diferentes substratos celulósicos e hemicelulósicos (Sigma-Aldrich ou Megazyme). A atividade de *BtCel48* foi avaliada em celulose inchada com ácido fosfórico (PASC),¹²¹ Avicel PH-101, Sigmacell 20, celulose bacteriana (BC), carboximetilcelulose (CMC), xilano, pectina, quitina, pNPC (*p-nitrophenyl-beta-D-cellobioside*), pNPG (*p-Nitrophenyl-beta-D-glucopyranoside*) (Sigma, USA), papel filtro (N.1), 1,4-beta-manano, xiloglucano (XyG), liquenano, arabinogalactano, arabinano, arabinana, arabinoxilano e beta-glucano; as misturas reacionais contendo 1,0 mg/mL de *BtCel48* e 10 mg/mL (w/v) de cada substrato foram incubadas a 50 °C durante 60 min em 50 mM do tampão McIlvaine a pH 5,5.¹²²

A atividade de *BtCel9* foi avaliada para CMC, Avicel PH-101, papel filtro, xilano, PASC e BC. As misturas reacionais contendo 3,0 μM de *BtCel9* e 10 mg/mL (w/v) de cada substrato foram incubadas a 50 °C durante 60 min em tampão *McIlvaine* a pH 7,0. A atividade de *XcGH74* foi testada para beta-glucano, CMC, Avicel, 1,4-beta-manano, XyG, liquenano, laminarina, pectina, arabinogalactano, arabinano, arabinano linear, arabinoxilano e xilano. Nas reações foram usados 0,2 mg/mL de *XcGH74* em tampão *McIlvaine* a pH 6,5 a 50°C durante 60 minutos.

Em seguida, fez-se a determinação do pH ótimo usando tampão *McIlvaine*¹²² com valores de pH de 3,0 a 10,0. Ainda, fez-se o gradiente térmico a fim de obter o perfil de temperatura num intervalo de 30 a 80 °C. A hidrólise de vários polissacarídeos foi quantificada utilizando o método colorimétrico do DNS¹²³ usando glicose como padrão e absorbância monitorada colorimetricamente a 540 nm. Cada reação foi realizada em triplicata¹²³ utilizando leitor de microplacas *Infinite 200 PRO* (TECAN).

Tanto em *BtCel48* quanto em *BtCel9* o efeito de CaCl_2 , NiSO_4 , KCl, MgCl_2 , CdSO_4 , CuSO_4 , SrCl_2 , DTT, etanol, Triton X-100 e Tween-20 foi avaliado na atividade destas enzimas. Todos estes reagentes estavam na concentração final de 5 mM.

Amostras de *BICel9* (1,2 μM e *BICel48* (10,0 μM) foram submetidas à incubação nas temperaturas de 40, 50 e 60 °C por 1, 3, 6, 12, 24 e 48 h. Posteriormente, estas amostras foram conduzidas à hidrólise de celulose inchada em ácido fosfórico (PASC) (1,0%) em tampão citrato/fosfato 150 mM (pH 5,5) por 1 h em volume reacional de 200 μL nas respectivas temperaturas de incubação acima. Ao fim do tempo reacional, as enzimas foram desnaturadas por fervura (5 min). Os açúcares foram quantificados segundo o método de DNS¹²³ e as atividades foram determinadas novamente e relacionadas à atividade, em cada temperatura das celulasas não submetidas à incubação.

A atividade específica para *BICel48* foi medida para os seguintes substratos: PASC, Avicel, BC, papel de filtro, CMC e beta-glucano; o ensaio foi realizado utilizando uma concentração final de enzima de 1,1 mg/mL durante 6 horas em condições pré-estabelecidas.

Para a determinação dos parâmetros cinéticos de *XcGH74*, a atividade específica foi determinada em diferentes concentrações de XyG, variando de 0,01 mg/mL a 20 mg/mL. As reações foram iniciadas adicionando *XcGH74* a uma concentração final de 0,01 mg/mL e incubando durante 10 minutos a 50 °C. DNS¹²³ foi adicionado para parar as reações e para a quantificação da quantidade de açúcar reduzido. Constantes cinéticas aparentes foram calculadas usando taxas iniciais com gráficos de Lineweaver-Burk ou usando regressão não linear para se ajustar diretamente à equação de Michaelis-Menten.

2.0.3 Padrão de clivagem

Os produtos gerados pela clivagem de PASC pelas enzimas *BICel48* e *BICel9* foram avaliados por HPAEC usando o método adaptado de Borne et al 2013.¹²⁴ 20 μM de *BICel9* e 50 μM *BICel48* foram incubados a 50 °C em tubos de 2 mL (1 mL de reacional) com PASC 0,5%. Alíquotas de 900 μL foram coletadas nos tempos 0, 1 e 48 horas, sendo então centrifugado (17000g) a 4 °C e a fração solúvel dos açúcares liberados analisada no sistema ICS-5000 HPAEC (Dionex, EUA) equipado com um detector amperométrico pulsado.

Os açúcares foram eluídos em tampão 100 mM NaOH e 500 mM de acetato de sódio + 100 mM NaOH como eluentes A e B, respectivamente, com os seguintes procedimento: separação isocrática (5 min, 95% A + 5% B), gradiente de separação (8 min, 1037% B), lavagem da coluna (2 min, 99% B), e subsequente equilíbrio da coluna (2,5 min, 95% A + 5% B). O fluxo da coluna foi de 1 mL/min. Uma mistura de glicose, celobiose, celotriose, celotetraose, celopentaose e celoexaose (Sigma-Aldrich) em concentrações conhecidas foi injetada como amostra padrão para quantificar os açúcares liberados.

Nas análises dos produtos liberados no decurso temporal pela hidrólise de PASC, BC e celo-oligossacarídeos por *BICel48** as hidrólises foram realizadas em tampão 50 mM citrato de sódio (pH5,5) com 150 mM de NaCl e 5 mM cloreto de cálcio a 40°C usando 10 μM de *BICel48* e 1 g/L de PASC ou BC com tempo de reação de 0-2880 minutos.

* Estes experimentos foram realizados em colaboração com o Dr. Marco A. S. Kadowaki.

Celotriose, celotetraose, celopentaose e celoexaose foram incubados por 10 minutos a 40 °C. Em seguida *BlCel48* na concentração de 0,1 μ M foi adicionada à reação durante 180 minutos. Além disso, celoexaose foi incubado com *BlCel48* em intervalos distintos de 0-60 minutos a 40 °C.

Os produtos gerados na hidrólise foram separados da parte insolúvel do substrato por centrifugação durante 10 min a 17000g. Os produtos da reação foram quantificados usando o sistema ICS-5000 HPAEC (Dionex, EUA) conectado a uma coluna CarboPac PA1 4X250 mm (DIONEX). A coluna foi pré-equilibrada com NaOH 100 mM durante 5 minutos a 1 mL / min. Os produtos liberados da hidrólise foram separados por tempo de retenção usando o gradiente de 100 mM NaOH/0 mM 100 mM NaOH/150 mM NaAc durante 2 minutos e equilíbrio com NaOH 100 mM durante 5 minutos. Um padrão de oligos com DP=1 a DP=6 foi utilizado para quantificar os celo-oligossacarídeos liberados na hidrólise.

Para *XcGH74* as análises dos produtos liberados de XyG da semente de foram realizadas no sistema HPAEC-PAD acoplado a um massas tipo MALDI-TOF[†]. Para o HPAEC-PAD, o perfil de hidrólise foi realizado utilizando 1 mL de 0,5% do XyG das sementes de *Copaifera langsdorffii*. XyG foi obtido de acordo com o procedimento descrito por Buckeridge et al.¹²⁵ O substrato foi incubado com XEG GH12¹²⁶ ou *XcGH74* em dosagem de 2,5U por 24 h a 30 °C. Os oligossacarídeos produzidos durante a reação enzimática foram analisados por HPAEC-PAD em coluna CarboPac PA-100 (Dionex, CA, EUA) utilizando 88 mM como linha de base em gradiente 0–200 mM de acetato de sódio em 45 min com o fluxo de 0,9 mL/min. Nos experimentos MALDI-TOF, as amostras foram suspensas em 50 mL de água desionizada misturadas 1:1 com a matriz (ácido 2,5-Diidroxibenzóico, 20 mg/mL em 30 % acetonitrila, contendo 0,1% de TFA (v/v)). Estas amostras foram então medidas no espectrômetro de massa Autoflex MALDI-TOF no modo refletivo positivo (Bruker Daltonics, Bremen, Germany).

2.0.4 Estudos estruturais por espalhamento de raios X a baixo ângulo

BlCel48. Dados de espalhamento de raios X a baixo ângulo (SAXS) foram coletados na linha de luz D02A-SAXS2 no Laboratório Nacional de Luz Síncrotron (Campinas, Brasil) usando detector bidimensional CCD (MarResearch, EUA) cobrindo uma faixa de transferência de momento q variando de 0,01 – 0,35 Å⁻¹.

As medidas de SAXS foram realizadas com temperatura controlada a 20 °C utilizando duas concentrações proteicas (2 e 6 mg/mL). Dez imagens foram coletados em intervalos de exposição de 30 segundos cada. A imagem média foi gerada e depois reduzida usando o programa FIT2D¹²⁷ na pipeline da linha D02A-SAXS2. Em seguida, a

[†] Esses experimentos foram realizados em colaboração com o professor Marcos Buckeridge, Universidade de São Paulo (USP), Instituto de Biociências.

curva de espalhamento obtida para solução de tampão foi subtraída da curva de espalhamento obtida para cada concentração de proteína. O raio de giro (R_g) foi estimado pela equação de Guinier⁹⁶ e a função de distribuição de distâncias $P(r)$ e dimensão máxima D_{max} foram avaliadas usando o programa GNOM.¹⁰⁰ A modelagem *ab initio* dos modelos foi realizada usando o programa DAMMIF¹²⁸ e o modelo médio calculado pelo programa DAMAVER.¹²⁹ A partir do modelo DAMSTART,⁹⁹ o modelo final foi melhorado usando o programa DAMMIN.¹³⁰ Para a seleção do modelo mais típico, a mínima discrepância entre as curvas experimentais e calculadas foi quantificada pelo parâmetro χ -minimizado. Para cada modelo, a média de 15 modelos gerados pelo DAMMIN¹³⁰ foi realizada usando o programa DAMAVER.¹²⁹ O modelo médio foi sobreposto ao modelo da estrutura cristalográfica de *BlCel48* e *BlCel9* usando o programa SUPCOMB.¹³¹ Em seguida, a resolução do modelo médio calculada pelo programa SASRES.¹³²

Com base na estrutura cristalográfica, a curva de espalhamento de raios X foi calculada usando o programa FoXS¹³³ e depois comparada com a experimental. Por fim, a massa molecular foi calculada usando o programa SAXSMoW2.0 (<http://saxs.ifsc.usp.br>)¹⁰².

***BlCel9*.** Parte dos dados para *BlCel9* foram coletados na linha de luz D02A-SAXS2 no Laboratório Nacional de Luz Síncrotron (Campinas, Brasil), como já mencionado para *BlCel48*. Além disso fizemos medidas de SEC-SAXS para *BlCel9* no Síncrotron Diamond.

As medidas de SEC-SAXS foram conduzidas na linha de luz B21 no Síncrotron Diamond (Didcot, Oxfordshire, UK). A proteína foi centrifugada a 17000*g* por 2 horas a 4 °C e depois injetada na uma coluna KW-403 de 4,8 mL (Shodex, New York, EUA) equilibrada com solução tampão (50 mM Tris-HCl pH 7,0, 150 mM NaCl, 5 % de glicerol) acoplado ao sistema HPLC (Agilent 1200) com fluxo de 0,16 mL/min a temperatura de 25 °C.

Os dados foram medidos utilizando o detector Pilatus 2M (Baden-Daettwil, Suíça) ajustado para cobrir q no intervalo 0,004-0,47 Å⁻¹. As imagens de espalhamento foram coletadas com tempo de exposição de três segundos e posteriormente reduzidas usando o programa Scatter (<http://www.bioisis.net/tutorial/9>). As análises para gerar os modelos e cálculo de massa molecular foram realizadas da mesma forma que *BlCel48*.

2.0.5 Estudos estruturais por cristalografia de raios X

Cristalização. Os experimentos de cristalização inicialmente foram realizados por meio de soluções de cristalização comerciais usando o robô de pipetagem automática *Gryphon LCP* (Art Robbins Instruments, LLC), monitorando a dinâmica de cristalização no sistema automático de magiamento *Imager 1000* (Formulatrix), equipado com leitor na região do UV-vis. Após a condição de cristalização ser obtida, utilizou-se a técnica manual para cristalização do tipo gota sentada por difusão a vapor.¹³⁴

Para *BICel9*, dois cristais foram submetidos a experimentos de soaking: 1) com celotriose e 2) com celoxaose. Para *BICel48* os cristais foram obtidos em 500 mM de LiCl, 28% (p/v) de PEG 6000 e 100 mM de Tris-HCl a pH 8,5; estes cristais foram otimizados usando a *seeding* e *microseeding*⁷⁹ na condição 500 mM de LiCl, 24% de PEG 6000, 100 mM de Tris-HCl na pH 8,5, 5mM de $CaCl_2$ e 20 mM de celobiose. Para *XcGH74* os cristais foram obtidos na condição 200 mM de iodeto de sódio, 100 mM de Bis-Tris propano (pH 7,5) e 20% (w/v) de PEG 3350. Os cristais foram congelados à temperatura criogênica após adição de 20% (v/v) de etilenoglicol.

Coleta, processamento, determinação estrutural, refinamento e validação. Os conjuntos de dados de difração de raios X foram obtidos na linha para difração de macromoléculas MX2 c¹³⁵ no Laboratório Nacional de Luz Síncrotron -LNLS- usando o detetor PILATUS 2M (Dectris).

Os dados de difração coletados foram processados (integração, indexação e escalonamento) usando o programa XDS.¹³⁶ O programa POINTLESS¹³⁷ foi utilizado para examinar as simetrias cristalográficas. O conteúdo de moléculas na unidade assimétrica foi estimado utilizando o coeficiente de Mathews e possíveis operações de geminação foram estimadas com o programa phenix.xtriage do pacote PHENIX.¹³⁸

Para a obtenção do modelo com as fases iniciais, usou-se o programa Phaser¹³⁹ aplicando o método de substituição molecular.⁸¹ Para *BICel9A* foi usado o modelo de busca da estrutura da enzima *TfCel9A* de *Thermobifida fusca* (PDB id: 1JS4)⁴⁵ com identidade sequencial de 49% a *BICel9A*. Para *BICel48* foram usadas as coordenadas atômicas do modelo da enzima *BpCel48* de *Bacillus pumilus* (PDB id 5BV9; 70% de identidade¹⁴⁰). Para a enzima *XcGH74*, usou-se as coordenadas do modelo da enzima *CtGH74A* de *Clostridium thermoCellum* (PDB id 2CN2⁵⁰). A solução inicial para cada enzima acima mencionada foi então submetida ao programa Autobuild do pacote PHENIX¹³⁸ para a construção automática da cadeia polipeptídica e ajuste na densidade eletrônica. A solução obtida foi modificada manualmente com o programa Coot¹⁴¹ e as posições atômicas corrigidas por ciclos de refinamento no programa Refine do pacote PHENIX.¹³⁸

Os modelos cristalográficos obtidos neste trabalho estão em processamento final com os aspectos estereoquímicos em validação por meio do MolProbity¹⁴² para posterior deposição no PDB. O alinhamento estrutural entre os modelos cristalográficos e as imagens das estruturas tridimensionais foram obtidos com o programa PyMOL (PyMOL, versão 1.5.0.4, Schrödinger, LLC).

2.0.6 Alinhamento e análises de bioinformática

Alinhamento de sequências. Os alinhamentos das sequências referente às estruturas depositadas para as enzimas das famílias GH9, GH48 e GH74 foram realizados usando o programa MUSCLE¹⁴³ e análises pelo programa ESPript3.¹⁴⁴

Análises de bioinformática por acoplamento estatístico - SCA[‡]. Para realizar as análises usando o método de SCA as sequências de enzimas da família GH48 foram obtidas fazendo uma busca por sequências homologas com programa BLAST¹⁴⁵ (e-value de 0,0001 com 5 interações) na base de dados do UNIREF-90.¹⁴⁶ O alinhamento das sequências de homólogos com identidade maior que 90% ou menor que 20% foram realizado no programa MUSCLE.¹⁴³ AS sequências redundantes foram removidas pelo método de clusterização do programa D-HIT,¹⁴⁷ resultando em 2000 sequências únicas.

Os cálculos de SCA foram realizados usando a *web-pipeline* pySCA, como descrito por Rivoire, Reynolds e Ranganathan.¹⁴⁸ Após a determinação e clusterização dos componentes independentes (IC), esses IC foram normalizados usando os valores da matriz de correlação maior que 30%. Cada IC construído foi ordenado com base nos valores de κ calculados, como sugerido por Rivoire, Reynolds e Ranganathan.¹⁴⁸

[‡] Estas análises foram realizadas em colaboração com o Dr. Vassili Piiadov.

3 RESULTADOS E DISCUSSÕES

3.1 Estudos estruturais e bioquímicos de uma nova Xiloglucanase XcGH74 de *Xanthomonas campestris*: base molecular para a preferência da ramificação de xilose no sub-sítio -1

3.1.1 Análise filogenética baseada na sequência das GH74 caracterizadas

De acordo com a base de dados CAZy existem atualmente 490 sequências classificadas como membros da família GH74 (<http://www.cazy.org/GH74>, acessado em 10-11-2018). Destas, 3,7% foram caracterizadas bioquimicamente e 1,3% foram caracterizadas estruturalmente. Elas têm alta especificidade para XyG, exceto *TmCel74* de *Thermotoga maritima* com maior atividade em β -glucano. Dentre estas enzimas, incluem-se, por exemplo, *TrCel74A* de *Trichoderma reesei*,¹⁴⁹ *AnGH74C* de *Aspergillus niger*,¹⁵⁰ *GeoGH74 XEG*⁵¹ e *GeoGH74 CBH*^{52, 151, 152} de *Geotrichum sp.*, *PcGH74B* de *Phanerochaete chrysosporium*,¹⁵³ *AnGH74* de *Aspergillus nidulans*,¹⁵⁴ *CtGH74A* de *Clostridium thermocellum*,⁵⁰ *CjGH74* de *Cellvibrio japonicus*,⁴⁶ *TfGH74* de *Thermobifida fusca*,¹⁵⁵ *PaGH74* de *Paenibacillus sp.*,¹⁵⁶ *ScGH74* de *Streptomyces coelicolor*,¹⁵⁷ e as duas xiloglucanases *SaGH74A* e *SaGH74B* de *Streptomyces avermitilis*.¹⁵⁸ Com base na classificação do CAZy²² e no alinhamento de sequências, as enzimas caracterizadas como pertencentes à família GH74 podem ser separadas em dois ramos distintos (Figura 1).

Figura 1 – Análise de sequências. Dendrograma do alinhamento de sequências das enzimas caracterizadas (acessado em 10-11-2016). Neste grupo, são incluídas as enzimas que tiveram sua bioquímica (cor preta) e estrutura (cor azul). Por fim, em vermelho é mostrado a localização no dendrograma da enzima XcGH74. O dendrograma foi construído no o programa MEGA 7¹⁵⁹ usando o modelo *Maximum Likelihood*.

Fonte: Elaborada pelo autor.

As endo-xiloglucanases *CtGH74A*, *TfGH74* e *CjGH74* estão agrupadas no ramo 1. O ramo 2 contém a enzima XcGH74 de *Xanthomonas campestris pv. campestris*, localizada dentro do mesmo cluster que as endoxiloglucanases *GeoGH74 XEG* e *SaGH74B*. Nesse

mesmo ramo estão as duas exo-xiloglucanases *GeoGH74* CBH e *AnGH74*. Para estas últimas duas enzimas, a atividade exo-glucanásica está relacionada com o *loop* adicional localizado na saída do sítio catalítico, contribuindo para a permissividade de XyGOs curtos no sítio ativo destas enzimas.^{151,160} No entanto, a região equivalente ao *exo-loop* está ausente em *XcGH74*.

3.1.2 Caracterização bioquímica e análise da hidrólise de XyG

Para avaliar as propriedades da enzima *XcGH74*, foram realizadas as caracterizações bioquímica e de especificidade pelo substrato. A análise das curvas de transição térmica mostra que a enzima é mais estável na faixa de pH entre 4,7 a 7,3 (Figura 2A), com pH ótimo de 6,2 e temperatura de 61 °C.

Figura 2 – Caracterização bioquímica. Os estudos de caracterização de *XcGH74* consistiram de análise de estabilidade estrutural em função do pH (A), da especificidade por substratos (B), da dependência da atividade em função do pH (C) e temperatura (D).

Fonte: Elaborada pelo autor.

A especificidade por substrato de *XcGH74* foi testada usando um painel de polissacarídeos incluindo XyG da semente de tamarindo e outros polissacarídeos hemicelulósicos (Figura 2B). A análise da atividade em diferentes pHs mostrou que *XcGH74* é mais ativa

na faixa de pH entre 5,0 e 7,0, com atividade máxima no pH 6,5 (Figura 2C). A atividade em função da temperatura mostra ainda que a temperatura ótima para a atividade de XcGH74 é atingida em torno de 50 °C (Figura 2D). Esta enzima mostrou uma especificidade acentuada para XyG e atividade basal em carboximetilcelulose (3%) e Avicel (2,5%), comparado com XyG. XcGH74 possui atividade específica de (70 ± 8) U/mg em XyG. Os parâmetros cinéticos foram determinados usando XyG como substrato e mostraram um K_m de (5 ± 1) mg/mL e k_{cat} de $89,8 \text{ s}^{-1}$, definindo a eficiência catalítica de XcGH74 (k_{cat}/K_m) como $17,9 \text{ (mg/mL/s)}^{-1}$.

Quando examinados os produtos da hidrólise de XyG, XcGH74 é capaz de liberar como principais produtos XX, LG, XGX, LGX, LLGXX e, em maior quantidade, LLGX (Figura 3). Os fragmentos de XyG mostrados foram gerados pela endoxiloglucanase AnGH12 (A-B) de *Aspergillus niveus*,¹²⁶ assim como os gerados pela XcGH74 (C-D).

Figura 3 – Produtos da hidrólise de XyG gerados por XcGH74. Os fragmentos do XyG da sementes de *Copaifera langsdorffii* foram quantificados por cromatografia líquida em HPLC-PAD e espectrometria de massas do tipo MALDI-TOF. São mostrados os produtos gerados por uma típica endoxiloglucanase AnGH12 (A-B) de *Aspergillus niveus*,¹²⁶ assim como os produtos gerados por XcGH74 após a hidrólise de XyG (C-D).

Fonte: Elaborada pelo autor.

3.1.3 Estudos estruturais por cristalografia de raios X

Para entender a base estrutural da catálise e o potencial papel de *XcGH74* na hidrólise de *XyG*, foi determinada a estrutura tridimensional desta enzima ([Figura 4](#)).

Figura 4 – Modelo estrutural de *XcGH74*. *XcGH74* é representada em cartoon (A) e em superfície (B). Nela são observados elementos da estruturas terciária formada por folhas betas antiparalelas e voltas (A-B). Os dois domínios possuem enovelamento típico para enzimas da família GH74 do tipo *7-β-propeller*. Entre esses domínios, o sítio catalítico é formado pelo bolsão em forma de fenda (B).

Fonte: Elaborada pelo autor.

Cristais de *XcGH74* em complexo com *XyGO XXXG* foram obtidos por cocrystalização e depois submetidos à experimentos de difração. Não obtivemos sucesso em obter cristais na forma apo para *XcGH74*. *XcGH74* foi cristalizada no grupo espacial $P2_12_12_1$ com duas moléculas na unidade assimétrica. A estrutura foi resolvida pela metodologia de substituição molecular usando as coordenadas atômicas da estrutura *CtGH74A* (PDB id: 2CN2). A estrutura 3D de *XcGH74* (resíduos Pro45-Pro750) foi refinada com R_{work}/R_{free} 0,19/0,23. As estatísticas da coleta, processamento e refinamento estão sumarizadas na [Tabela 1](#).

O modelo estrutural consiste de dois domínios formados por sete hélices-beta ([Figura 4](#)) com enovelamento *7-beta-propeller*. Possui estruturas secundárias semelhantes a *CtGH74A* (PDB id: 2CN2,⁵⁰ com 40% de identidade de sequência de aminoácidos), *CjGH74* (PDB id: 5FKS,⁴⁶ com 38% de identidade) e *GeoGH74 CBH* (PDB id: 1SQJ/2EBS,⁵² exibindo 37% de identidade). Pelo alinhamento estrutural, as estruturas

Tabela 1 – Estatísticas. Coleta, processamento e análises da estrutura cristalográfica de XcGH74 obtida por difração de raios.

	XcGH74 +XG
Coleta e Processamento	
Comprimento de onda (Å)	1,54
Grupo espacial	P2 ₁ 2 ₁ 2 ₁
a, b, c (Å)	81,79 100,43 168,27
α, β, γ (°)	90 90 90
Resolução (Å)	3,8 - 1,95 (2,02 - 1,95)
R _{merge}	0,1955 (1,09)
R _{meas}	0,2042 (1,15)
CC _{1/2}	0,997 (0,864)
Fator B de Wilson	19,91
Reflexões únicas	99729 (9225)
$\langle I/\sigma(I) \rangle$	11,01 (2,28)
Completeza (%)	0,98 (0,92)
Multiplicidade	12,2 (10,0)
No. Mol. na cela	2
Refinamento	
R _{work}	0,1988
R _{free}	0,2309
Número de resíduos de aminoácidos	1404
<i>desvios (R.m.s.)</i>	
Comprimentos de ligação (Å)	0,004
Ângulos de ligação (°)	0,78
<i>Ramachandran</i>	
Favorável (%)	97
Permitido (%)	3,1
outliers (%)	0,072
<i>Fator B</i>	
Médio	20,69
macromoléculas	20,27
solvente	23,93
<i>Grupo TLS</i>	
Número de grupos	2

Fonte – Elaborada pelo autor.

mais próximas são 2CN2 (rmsd=1,8 Å), 5FKS (rmsd=1,8 Å) e 2EBS/1SQJ (rmsd=1,8 Å), respectivamente.

As diferenças nas estruturas terciárias mais marcantes entre XcGH74 e outras GH74 estão localizadas nos *loops* que conectam os dois domínios. Dois *loops* em XcGH74 destacam-se particularmente em comparação com CjGH74 (PDB id: 5FKS): o *loop* Thr508-Ile517 na posição do sub-sítio +2 em XcGH74 é significativamente maior que em CjGH74; em contraste, o *loop* Ala392-Gly398 na posição +4 em XcGH74 é mais curto que em CjGH74. No entanto, o Thr508-Ile517 em XcGH74 é estruturalmente muito semelhante ao *loop* equivalente em CtGH74A (Ala523-Tyr526).

Na comparação entre XcGH74 e GeoGH74 CBH, duas diferenças notáveis podem

ser mapeadas nos sub-sítios +1 e +2. Há uma inserção em *XcGH74* (Thr214-Ala224) na vizinhança do sub-sítio +1. Esta inserção não está presente *GeoGH74* CBH, mas existe tanto em *CjGH74* como em *CtGH74A*. Enquanto isso, *GeoGH74* CBH possui um único *loop* (Gly375His385, o *exo-loop*) próximo ao sub-sítio +2 na extremidade do bolsão catalítico bloqueando o acesso de XyG à posição +3. A presença do *exo-loop* em *GeoGH74* CBH está relacionada à preferência desta enzima por XyG curtos.^{52,151} A sua localização no sub-sítio +2 causa impedimento estérico, o que desempenha um papel importante na especificidade de XyGOs e padrão de clivagem.^{51,151} Contudo, o *exo-loop* correspondente está faltando na estrutura de *XcGH74*.

De acordo com a classificação para estas enzimas,¹⁶¹ os membros das GH74 realizam a hidrólise enzimática via inversão estereoquímica do carbono anomérico.¹⁶¹ Em *XcGH74*, os resíduos catalíticos, Asp41 (base catalítica) e Asp430 (ácido catalítico), estão situados em lados opostos na fenda catalítica entre os domínios. Estes resíduos correspondem ao Asp70 (base catalítica) e Asp480 (ácido catalítico) em *CtGH74A*; Asp70 (base catalítica) e Asp483 (ácido catalítico) em *CjGH74* e Asp35 (base catalítica) e Asp465 (ácido catalítico) em estruturas *GeoGH74* CBH. Todos eles são preservados na mesma posição e orientação.

Atualmente existem apenas três estruturas de GH74 complexadas com XyGOs (PDB ids: 2CN3, 2EBS e 5FKS). A comparação entre *XcGH74*, *CjGH74*, *CtGH74A* e *GeoGH74* CBH (Figura 5) mostra algumas similaridades e também algumas diferenças no sítio ativo. A *CtGH74A* (Figura 5A) exibe uma topologia em que o sítio ativo é delimitado por quatro sub-sítios negativos (-4 a -1) e quatro sub-sítios positivos (+1 a +4) envolvidos na ligação com o XyGOs. Esses subsítios cobrem uma área de 2756 Å² dentro de dezessete anéis de açúcar correspondentes à molécula de XLLG nos sub-sítios -4 a -1 e XXLG nos sub-sítios +1 a +4.

CjGH74 tem sítios catalíticos e sub-sítios similares a *CtGH74A*. *CjGH74* também apresenta um sulco de ligação do substrato delimitado por oito sub-sítios (quatro positivos e quatro negativos) com dois oligossacarídeos derivados de xiloglucano correspondentes a um fragmento GXLG nos sub-sítios negativos e GLLX nos sub-sítios positivos (PDB id: 5FKS). O sulco do sítio ativo cobre uma área de aproximadamente 3189 Å² (Figura 5B). Todavia, no caso de *GeoGH74* CBH, o bolsão que acomoda o substrato é delimitado por apenas quatro sub-sítios (-2 a +2, PDB id: 2EBS), cobrindo uma área de aproximadamente 1020 Å², onde XXXG está localizado (Figura 5C). Para todas essas enzimas, os sítios catalítico expostos ao solvente são compostos de resíduos carregados e polares altamente conservados entre *CjGH74*, *GeoGH74* CBH e *CtGH74A*. Por exemplo, o motivo GGF/Y exposto ao solvente é conservado em *CtGH74A*, *CjGH74* e *GeoGH74* CBH.

A estrutura tridimensional de *XcGH74* ligada ao XyGO XG (Figura 5D) neste estudo é o único modelo estrutural 3D relatado para as bactérias *Xanthomonas*. O seu

Figura 5 – Subsítios e interação com XyGOs. Comparando as estruturas cristalográficas das enzimas *CtGH74A*, *CjGH74*, *GeoGH74* e *XcGH74*, vê-se que o sítio catalítico destas enzimas tem a forma de uma fenda onde o XyG pode ser alocado. Os subsítios foram marcados tomando como referencia a posição dos aminoácidos catalíticos. Subsítios negativos representam posições ocupadas por substrato e os negativos os produtos.

Fonte: Elaborada pelo autor.

sítio catalítico está localizado na superfície plana formada principalmente por *loops* envolvendo uma área de aproximadamente 2880 Å². As análises dos ensaios enzimáticos indicam que o sítio ativo de *XcGH74* possui pelo menos quatro sub-sítios (sítios 2 a +2). De fato, *XcGH74* foi co-cristalizada com o XyGO XXXG. O mapa de densidade eletrônica mostra claramente a presença da molécula XG ligada nos sub-sítios positivos de ligação do substrato (+1 e +2). Isto indica que *XcGH74* é cataliticamente ativa mesmo no empacotamento cristalográfico clivando XXXG em XX e XG. O ligante XG está localizado nos sub-sítios positivos de ligação +1 (Glc⁺¹ e Xyl⁺¹) e +2 (Glc⁺²) em uma conformação estendida (Figura 6).

As principais interações são compreendidas por uma rede de ligações de hidrogênio e também por ligações salinas. Glc⁺¹ e Xyl⁺¹ estão posicionados próximos aos resíduos catalíticos Asp41 e Asp430. Asp430 faz ligação de hidrogênio com Glc⁺¹ (O4) e também com Xyl⁺¹ (O2), enquanto que Asp41 faz ligação com Glc⁺¹ (O4) mediada por água. Glc⁺¹ também faz ligação de hidrogênio com His359 (O3), Gly379 (O2 e O3), Arg474 (O5 e O6). As interações mediadas por água são observadas para Glc⁺¹ com Arg128, Gly20 e uma molécula de etilenoglicol. O aromático Trp360 faz interação de empilhamento com a

Figura 6 – Resíduos de XcGH74 envolvidos na interação com o ligante nos subsítios +1 e +2. O ligante XG presente na estrutura apresenta principalmente ligações de hidrogênio com resíduos aromáticos e carregados. Outras ligações presentes com distância maior que 3,5 Å não foram mostradas.

Fonte: Elaborada pelo autor.

Glc⁺¹ na posição +1.

Os resíduos de Glc⁺¹ (via O2) e Glc⁺² (via O6) formam consideráveis contatos mediados por água com Gln445, His359, Thr357. Adicionalmente, no sub-sítio +1, a decoração de Xyl⁺¹ faz ligação de hidrogênio com a cadeia principal de Arg474 e interação de empilhamento com o resíduo His182. O mesmo grupamento de açúcar faz as interações mediadas por água com resíduos de Arg128, Leu266, Arg318, Trp360 e Asp475. No sub-sítio +2, Glc⁺² forma ligações de hidrogênio somente com Arg320. No entanto, as interações mediadas por água também são formadas com Asp475, Arg318 e Arg474.

Com base na sobreposição estrutural entre *GeoGH74* CBH (PDB id: 2EBS) e *XcGH74*, a reconstrução da molécula XXXG foi feita posicionando XX nos sub-sítios -1 e -2. Para isso, foi tomado como base a conformação de XXXG observada em *GeoGH74* CBH e também a conformação de XG observada em *XcGH74*. Com XXXG ocupando os sub-sítios -2 a +2 pôde-se mapear os resíduos que podem interagir nas posições -2 e -1 no sítio de *XcGH74*. Glc⁻¹ faz interações com Asp41 (O2), Arg128 (O2 e O3) e Asn124 (O3). Localizada na fenda do sub-sítio 1, a decoração Xyl⁻¹ forma pelo menos três interações com os resíduos Asp430 (O2), Leu413 (O3) e Asn453 (O3), respectivamente. Enquanto isso, no sub-sítio 2, Glc⁻² interage com os resíduos Tyr92 (O3) e Asn124 (O1 e O5); enquanto a cadeia lateral de Xyl⁻² faz interações com Asp265 (O3 e O4), His185 (O5) e Val183 (O5).

3.1.4 Explorando as interações substrato-proteína no sítio ativo de *XcGH74*

A sobreposição estrutural das interfaces dos sítios de ligação das endo-xiloglucanases *CtGH74A* (PDB id: 2CN3), *CjGH74* (PDB id: 5FKS) e exo-xiloglucanase *GeoGH74* CBH (PDB id: 2EBS) permitiu a visualização de diferentes modos de ligação exibidos entre essas três enzimas e *XcGH74* (Figura 7). Um grande número de resíduos que interagem com XyGOs são equivalentes entre estas enzimas. A comparação dos sítios mostra que *XcGH74* preserva os sub-sítios -3 a +4. A ausência do sub-sítio -4 está relacionada com os impedimentos estéricos do *loop* Gly90-Ala100 em *XcGH74*.

Figura 7 – Comparação dos subsítios -4 (A) e +4 (B) e suas interações com XyGOs para *CtGH74A*, *CjGH74*, *GeoGH74* e *XcGH74*. *XcGH74* parece não possuir o subsítio -4 por conta da presença da Arg96. Esse resíduo não tem equivalente estrutural nas outras estruturas analisadas.

Fonte: Elaborada pelo autor.

Em termos de comparação conformacional, os *loops* equivalentes em *CtGH74A* e *CjGH74* são um pouco diferentes que *XcGH74*, exceto para a cadeia lateral de Arg96 que bloqueia a fenda no sub-sítio -4. Esse resíduo cria um obstáculo estérico para a ligação ou acomodação de Glc^{-4} . O *loop* Gly90-Ala100 em *XcGH74* é mais longo que os seus equivalentes nas estruturas de *CtGH74A* e *CjGH74*; o resíduo Arg96 não faz ligação de hidrogênio com Glc^{-4} . Além disso, o Glc^{+4} posicionado na extremidade mais distante da fenda do sítio ativo mostra diferenças conformacionais nos modelos *CtGH74A* e *CjGH74*, sugerindo um certo grau de flexibilidade para Glc^{+4} .

De fato, os resíduos de Glc posicionados nos sub-sítios +2, +3 e +4 em *CtGH74A* e *CjGH74* não fazem ligações de hidrogênio com os resíduos de aminoácidos, embora haja muitos resíduos que fazem interações mediadas por água com estes açúcares. No entanto, para *GeoGH74* CBH a Glc no sub-sítio +2 faz ligações de hidrogênio somente com Asp318, Gly382 e Trp395. Além disso, quase todas as ligações de hidrogênio dos resíduos de aminoácidos são formadas com as ramificações dos sub-sítios +2 e +3. No sub-sítio +3, *XcGH74* pode formar interações com Xyl^{+3} e Gal^{+3} , como também observado nas estruturas *CtGH74A* e *CjGH74*. Em *XcGH74* esses sub-sítios são formados pelo *loop*

Gly471-Arg482, entretanto os correspondentes em *CjGH74* e *CtGH74A* são mais curtos; como consequência, os volumes das fendas de ligação do substrato nas estruturas *CjGH74* e *CtGH74A* são menores nessa região.

O *loop* Gly471-Arg482 no modelo *XcGH74* possui comprimento similar ao *loop* correspondente na estrutura *GeoGH74* CBH (Ser507-Arg518). No entanto, ele não preserva uma conformação similar, sugerindo que pode ser um elemento estrutural para o reconhecimento das ramificações de XyG. O *loop* Ala356-Gly362 em *XcGH74* define a forma do sub-sítio +2; ele é mais curto do que em *CjGH74* e *CtGH74A*. Em *CtGH74A*, Xyl⁺² não faz ligações de hidrogênio, mas em vez disso, faz uma interação de empilhamento com Trp395; na estrutura *CjGH74* este resíduo é substituído por Phe397. Adicionalmente, Xyl⁺² faz ligações de hidrogênio com os resíduos Asp483 e Ser506 na estrutura *CjGH74*, enquanto o resíduo Glu406 faz ligações de hidrogênio com Gal⁺². O resíduo Glu correspondente também está conservado em *CtGH74A*, mas não em *XcGH74*.

Os sub-sítios -3, -2 e -1 estão conservados em *XcGH74*. No sub-sítio -3, o resíduo Glc⁻³ não faz interações com a proteína, mas as ramificações Xyl⁻³ e Gal⁻³ podem ser acomodadas neste sub-sítio. Na estrutura *CtGH74A*, a decoração Xyl⁻³ interage com a cadeia lateral do Asp731, enquanto no modelo *CjGH74* esta decoração está envolvida na interação com Asn735. No entanto, a decoração Gal⁻³ em *CtGH74A* apresenta somente ligações mediadas por água. Glc⁻³ na estrutura *CjGH74* faz ligação de hidrogênio com o resíduo Asn755, que, novamente, não possui equivalente estrutural em *XcGH74*.

O tamanho e a forma do sub-sítio -2 são similares entre *XcGH74*, *CjGH74* e *CtGH74A*, indicando que *XcGH74* pode acomodar as decorações Glc⁻² e Xyl⁻². Nas estruturas *CjGH74*, *CtGH74A* e *GeoGH74* CBH, Glc⁻² interage com o resíduo Asn conservado (Asn755 em *CjGH74*, Asn749 em *CtGH74A* e Asn753 em *GeoGH74* CBH), enquanto que em *XcGH74* este resíduo é substituído por Gly705, o qual não faz ligação de hidrogênio com o Glc⁻². *XcGH74* pode interagir também com Asn124, como observado para o seu equivalente em *CjGH74* (Asn154), *CtGH74A* (Asn154) e *GeoGH74* CBH (Asn116). No entanto, para a posição Xyl⁻², tanto *CtGH74* como *CjGH74* fazem ligação de hidrogênio com Xyl⁻² através dos aromáticos Tyr (Tyr216 e Tyr298 em *CjGH74* e Tyr214 e Tyr295 em *CtXgh74*); em vez disso, em *GeoGH74* CBH e *XcGH74* estes resíduos não são conservados.

Por outro lado, tanto nas outras estruturas como em *XcGH74* a ligação da Arg128 com Xyl⁻² é observada. Em *XcGH74* Asp265 faz ligação de hidrogênio com Xyl⁻² no sub-sítio -2. Esta interação parece ser específica para *XcGH74*, uma vez que não há resíduos equivalentes nas estruturas *CtGH74A* ou *CjGH74* e *GeoGH74* CBH onde esse resíduo é substituído por Trp255. Além disso, a inspeção de *XcGH74* sugere que o *loop* Asn181-Gly184 pode ser um impedimento estérico para a acomodação da decoração Gal⁻² no sub-sítio -2. Em *CtGH74A* e *CjGH74*, o equivalente da decoração Gal⁻² está exposto ao

solvente e parece não interagir com os resíduos destas enzimas.

3.1.5 Interações substrato-enzima nos subsítios -1/+1

A sobreposição estrutural de *XcGH74* com as enzimas homólogas mostra que os sub-sítios -1 e +1 são similares entre elas. Em *XcGH74* o bolsão do sub-sítio +1 pode acomodar Glc^{+1} decorado com resíduo Xyl^{+1} ; dentro deste, Glc^{+1} interage com o aromático His359, Gly379, e Asp430 via ligações de hidrogênio. Em *GeoGH74* CBH estes resíduos possuem as mesmas posições e interações. Além disso, nas estruturas *CtGH74A* e *CjGH74* os resíduos equivalentes Gly379 e Asp430 também fazem ligações de hidrogênio com Glc no sub-sítio +1. No entanto, o resíduo His359 em *XcGH74* é substituído por Trp410 e Trp413 em *CtGH74A* e *CjGH74*, respectivamente.

Semelhante a *GeoGH74* CBH, a cadeia lateral do Trp360 em *XcGH74* está envolvida nas interações de empilhamento com a Glc^{+1} . Este resíduo é substituído por Met414 e Met411, respectivamente, em *CtGH74A* e *CjGH74*. Além disso, nas estruturas *CtGH74A*, *CjGH74*, e *GeoGH74* CBH a decoração Xyl^{+1} faz somente interações mediadas por água e não está envolvida com interações direta via ligação de hidrogênio.

As principais diferenças no sub-sítio +1 entre as últimas enzimas e *XcGH74* incluem o resíduo de aminoácido Arg474 e His182 presente em *XcGH74*. Arg474 faz ligações de hidrogênio com $\text{Glc}^{+1}/\text{Xyl}^{+1}/\text{Glc}^{+2}$ e His182 está empilhada com Xyl^{+1} . Isso sugere que os resíduos Arg474 e His182 poderiam estar fortemente correlacionados com a capacidade de *XcGH74* de clivar XyG , bem como com a especificidade no reconhecimento da ramificação na posição +1 (Figura 8).

No sub-sítio -1, *XcGH74* mostra um volume semelhante a outras GH74 estruturalmente caracterizadas. Em *XcGH74*, os três resíduos de aminoácidos Gly21, Asn124 e Arg128 estão envolvidos em ligações de hidrogênio com Glc^{-1} . Estes resíduos também estão conservados nas enzimas *CtGH74A* (Gly50, Asn154 e Arg158), *CjGH74* (Gly50, Asn154 e Arg158) e *GeoGH74* CBH (Gly15, Asn116 e Arg120), sugerindo que eles definem o reconhecimento de Glc^{-1} no sub-sítio -1 (Figura 8). A sobreposição entre *XcGH74* e *GeoGH74* CBH mostrou ambientes semelhantes em torno das posições Glc^{-1} e Xyl^{-1} . Similarmente à *GeoGH74* CBH, *XcGH74* pode acomodar a decoração de xilose no sub-sítio -1. A permissividade da decoração de Xyl^{-1} no sub-sítio -1 é suportada pela ligação de hidrogênio entre Asn473 e Xyl^{-1} . Em *GeoGH74* CBH, o homólogo Asn453 está localizado em uma posição semelhante a *XcGH74*. Por outro lado, em *CtGH74A* e *CjGH74* os resíduos correspondentes são substituídos por Ser507 e Ser503, respectivamente (Figura 19B). Esta substituição de aminoácido no sub-sítio -1 altera a interação formada com Glc^{-1} e Xy^{-1} . Em *CtGH74A* e *CjGH74*, a Glc^{-1} não interage com o resíduo serina correspondente (Ser507 e Ser503, respectivamente), sugerindo que Asn473 no sub-sítio -1 é essencial para as interações com o substrato e a preferência pela decoração Xyl^{-1} no sub-sítio -1 em

Figura 8 – Comparação do sub-sítio -1 para *CtGH74A*, *CjGH74*, *GeoGH74* e *XcGH74*. São mostrados os resíduos conservados na família GH74 e que estão presentes *CtGH74A*, *CjGH74*, *GeoGH74* e *XcGH74*. O resíduo Asn453 que interage com a Xyl¹ em *XcGH74* está presente também em *GeoGH74* enquanto que em *CtGH74A*, *CjGH74* é substituído por uma Ser.

Fonte: Elaborada pelo autor.

XcGH74.

3.1.6 Relações entre estrutura e função

3.1.6.1 Caracterização enzimática, especificidade pelo substrato e perfil de hidrólise

XcGH74 é uma xiloglucanase com alta especificidade para XyG, quando comparada com seus homólogos (Tabela 2). A Tabela 2 indica que a maioria das xiloglucanases da família GH74, especialmente as de origem bacteriana, é estável ao longo de uma faixa estreita de pH e temperatura, com condições ótimas próximas a pH 6,0 - 7,0 e temperatura em torno de 40-50 °C. As condições catalíticas ótimas para *XcGH74* estão próximas do pH 6,5 e 50 °C.

Tabela 2 – Comparação bioquímica. Comparação de propriedades bioquímicas e atividade enzimática de *XcGH74* em XyG com outras enzimas da família GH74. Nr = não reportado.

pH	Temp., °C	Atividade específica (U/mg)	K _m (mg/mL)	K _{cat} (s ⁻¹)	K _{cat} /K _m ((mg/mL) · s) ⁻¹	Organismo	Ref
6,5	50	70 ± 8	5,0 ± 1,1	89,82	17,96	<i>X. campestris</i>	
7,0	40	282	Nr	Nr	1,62	<i>X. citri</i>	¹⁶²
7,0	50	295	0,76	0,79	1,03	<i>C. thermocellum</i>	⁵⁰
6,0	50	35,2	0,57	196	343,85	<i>S. coelicolor A3(2)</i>	¹⁵⁷
5,5/6,0	60 / 55	1,6/2,4	Nr	Nr	Nr	<i>S. avermitilis</i>	¹⁵⁸
7,5	60	12,2	0,032	16,1	503,12	<i>T. fusca</i>	¹⁵⁵
6,0	70	44	1,2	77,6	64,66	<i>Geotrichum sp. M128</i>	¹⁵⁶
3	42	7	Nr	Nr	Nr	<i>A. nidulans</i>	¹⁵⁴
3,5	50	15,3	Nr	Nr	Nr	<i>Geotrichum sp. M128</i>	¹⁵²

Fonte – Elaborada pelo autor.

XcGH74 é capaz de liberar XX, LG, XGX, LGX, LLGXX e, em maior quantidade LLGX, como principais produtos da hidrólise de *XyG*. Estes resultados mostram claramente que *AnGH12*¹²⁶ e *XcGH74* exibem especificidade muito diferente por *XyG* (Figura 3). Além disso, estes resultados sugerem fortemente que *XcGH74* reconhece e cliva posições em *XyG* entre os motivos XX, ramificados (ou não) pelo L no terminal não-redutor. Tanto *GeoGH74* CBH como *AnGH74* reconhecem a extremidade redutora de vários oligossacarídeos derivados de *XyG* e atuam somente se a segunda xilose da extremidade redutora não for ramificada.^{152,154,163} No entanto, algumas enzimas desta família apresentam diferentes padrões de clivagem, clivando entre os motivos **X-X**.¹⁶² A diversidade nos padrões de clivagem para estas enzimas foi previamente observada para *SaGH74A/SaGH74B* e *PaGH74*, que clivam não somente após G, mas também após X.^{153,156,158,164} Contudo, a endoxiloglucanase de *Xanthomonas citri* (*XcXGHA*) tem uma preferência pela clivagem de *XyG* e *XyGOs* entre o motivo X-X.

Baseado no padrão de clivagem de *XyG*, é proposto aqui que *XcGH74*, similarmente a *XcXGHA*, é capaz de clivar os oligossacarídeos XXXG, convertendo-os em XX+XG, enquanto que XXXXG é degradado em X+XXXXG e XX+XXG ou produtos XXX+XG. Quando os *XyGOs* com ramificação de galactose são analisados, XXXLG pode ser convertido em X+XXLG e XXX+LG, e XXLLG hidrolisado em XX+LLG. O *XyG* estudado neste trabalho tem arquitetura típica com as repetições XXXG, XXLG, XLXG e XLLG, como observado por Buckeridge et al..¹²⁵ Neste caso, quando este *XyG* é incubado com *XcGH74*, os produtos liberados são XX, LG, XGX, LGX, LLGXX e principalmente LLGX. Juntos, estes resultados sugerem que *XcGH74* é capaz de clivar *XyG* preferencialmente entre os motivos **X-X**, todavia, é incapaz de hidrolisar o polissacarídeo entre os motivos **L-L**.

Como previamente também foi demonstrado para *XcXGHA*, que possui 84% de identidade de sequência de aminoácidos com *XcGH74*, *XcGH74* exibe um padrão de clivagem único com preferencia por *XyGs* ou *XyGOs*. Desta forma essa enzima é capaz de degradar desde substratos mais simples, como os *XyGOs*, quanto substratos mais complexos, como os *XyGs*. Curiosamente, *XcGH74* demonstra algumas características muito similares a *AnGH74* e *GeoGH74* CBH. Embora *GeoGH74* CBH seja um exo-oligoxiloglucano devido a presença do *exo-loop* em sua estrutura, ela pode converter XXXG em XX e XG, assim como *XcGH74* faz. De fato, com base nos resultados funcionais e no perfil da hidrólise de *XyG*, *XcGH74* parece clivar preferencialmente os *XyGs* através dos resíduos ramificados, onde a substituição da cadeia lateral é um pré-requisito para aumentar o reconhecimento do substrato. Isso até então foi relatado apenas para *GeoGH74* CBH e *XcXGHA*.

3.1.6.2 Base estrutural para o reconhecimento da ramificação Xyl no sub-sítio -1

Para determinar a base molecular deste perfil incomum de XyG da família GH 74, especialmente *XcGH74*, a análise de suas estruturas em complexo com XyGOs fornece informação sobre a especificidade destas enzimas por XyG. Em *XcGH74*, Glc⁺¹ pode ser acomodado dentro do sub-sítio +1. Este sub-sítio é semelhante a *GeoGH74* CBH, *CtGH74A* e *CjGH74*. No entanto, os resíduos His182 e Arg474 são diferenças estruturais marcantes entre estas enzimas. His182 faz interação de empilhamento com Xyl⁺¹, enquanto que Arg474 faz interação de hidrogênio com Glc⁺¹/Xyl⁺¹/Glc⁺². Estes dois resíduos são restritos a *XcGH74* e sem equivalente estrutural em outras GH74s, sugerindo que eles estão diretamente correlacionados com a especificidade por XyG de *XcGH74*. Essa enzima é capaz de acomodar Xyl⁻³/Gal⁻³, Gal⁻²/Xyl⁻², Glc⁻¹/Xyl⁻¹ e também Xyl⁺¹, Xyl⁺²/Gal⁺², Gal⁺³, Glc⁺⁴. No entanto, devido aos impedimentos estéricos em *XcGH74*, um substituinte galactosil não pode ser acomodado nos sub-sítios -2 e -1.

Nossos dados funcionais e estruturais mostram que *XcGH74* é uma endoxiloglucanase, ao passo que também compartilha características de uma exo-xiloglucanase. Durante a hidrólise de XyG, XX e XG são seus produtos finais principais, sugerindo que *XcGH74* pode aceitar preferencialmente resíduos glicopiranosil ramificados em ambos os sub-sítios -1 e +1. Isto indica que a ligação das cadeias laterais do xilosil nos sub-sítios -1 e +1 é essencial para a hidrólise de XyG.

Além disso, a perfilação de oligossacarídeos (Dionex) e inspeção de espectrografia de massas (MALDI-TOF) de fragmentos enzimaticamente liberados de XyG por *XcGH74* mostram a presença de oligossacarídeos LG, LLG, LGX, LLGX e LLGXX como um dos principais produtos liberados durante a catálise de XyG. Este padrão indica que a clivagem de resíduos ramificados com galactose sempre irá ocorrer. Os resíduos de galactose irão aparecer no sub-sítio +1 enquanto que no sub-sítio -1 não há espaço suficiente para acomodar uma decoração de galactose, mas somente para uma ramificação de xilose.

A [Figura 9](#) mostra uma ramificação com o resíduo de galactose modelado ligado à cadeia lateral de Xyl⁺¹, onde se observa choques estéricos para acomodar tal decoração. Do ponto de vista estrutural, *XcGH74* pode aceitar tanto resíduos de glicopiranosil ramificados e não ramificados no sub-sítio -1. Embora possa aceitar preferencialmente XyG ramificado com decoração Xyl na posição -1. Esta suposição sugere que os resíduos de glicopiranosil ramificados, que ocorrem no sub-sítio -1, dependem do ambiente de ligação neste sub-sítio, bem como do tipo de interação da proteína com o resíduo Glc ramificado.

Como observado anteriormente, as endoxiloglucanases clássicas, como *CtGH74A* e *CjGH74*, são capazes de produzir oligossacarídeos XXXG, XLXG/XXLG e XLLG como produtos finais liberados a partir da hidrólise de XyG. Consequentemente, estas enzimas têm preferência por clivar XyG na ligação glicosídica dos resíduos de glicosil não ramifica-

Figura 9 – Permissividade para *XcGH74* e mapa das posições nos subsítios. Nessa figura uma ramificação de Gal foi posicionada na posição -1 para mostrar a permissividade do motivo **L-L** nas posições -1 e +1. Abaixo são mostrados as posições das ramificações de XyGOs observadas nas estruturas de *CtGH74A*, *CjGH74*, *GeoGH74* e *XcGH74* e as posições das ramificações permitidas em *XcGH74*.

Fonte: Elaborada pelo autor.

dos. Por outro lado, a exo-xiloglucanase *GeoGH74* CBH é capaz de clivar somente XyGOs curtos, para os quais a clivagem ocorre preferencialmente entre as Glc ramificadas, como na posição **X-X**. Entre estas xiloglucanases, o sub-sítio -1 em *XcGH74* pode está correlacionado com a sua habilidade de clivar XyGs de uma maneira distinta (Figura 9). Além disso, os sub-sítios nas posições -2 e -1 parecem serem os únicos elementos estruturais restritivos para corrigir a ligação do substrato na fenda catalítica de *XcGH74*.

4 RESULTADOS E DISCUSSÕES

4.1 Caracterização bioquímica e estrutural da exo-glucanase de *BlCel48* de *Bacillus licheniformis*: processividade e taxa hidrolítica são limitados pelo tamanho e morfologia da celulose

A família 48 das hidrolases de glicosídeos compreende um grupo diversificado e versátil de enzimas envolvidas na quebra da parede celular vegetal. Vários microrganismos têm um único gene codificante para GH48 enquanto outros têm dois ou mais genes pertencentes a esta família. Por exemplo, para *C. thermocellum* são observados dois genes e *Caldicellulosiruptor bescii* três. *Bacillus licheniformis* é um destes microrganismos bacterianos que codifica apenas um gene para enzimas da família GH48.

Estudos estruturais e funcionais foram conduzidos afim de compreender a relação entre estrutura e função desta enzima, que chamaremos daqui por diante de *BlCel48*. Nossos resultados envolvem estudos bioquímicos, do perfil de produtos liberados em celulose, cristalografia de raios X e análises de bioinformática por acoplamento estatístico (SCA).

4.1.1 Produção e caracterização bioquímica

4.1.1.1 Produção, caracterização bioquímica e análise dos produtos liberados da hidrólise de PASC

BlCel48 foi expressa com rendimento estimado de aproximadamente 24 mg/L e pureza maior que 95%. Quando calculada tomando como base a sequência de aminoácidos, essa enzima tem 79 kDa. Sua massa molecular aparente após purificação foi determinada sendo aproximadamente 80 kDa (ver gel SDS-PAGE, [Figura 10](#)).

Tomando como base as atividades até então descritas para enzimas caracterizadas da família GH48, a hidrólise de diferentes substratos foi realizada para testar a especificidade por substrato de *BlCel48*. Essa enzima apresentou maior atividade em PASC e BC. E atividade basal para papel filtro, CMC e beta-glucana ([Figura 11A](#)). Nenhuma atividade mensurável foi observada para Avicel, Sigmacell 20, quitina, xiloglucana, liquenan, arabinana, xilana, ou pNPC e pNPG (dados não mostrados).

BlCel48 reteve cerca de 65 a 90% da atividade máxima em pHs na faixa de 4,0-6,0, tendo pH 5,0 como ótimo ([Figura 11B](#)). Em pHs inferiores a 3,0 ou superiores a 7,0, a atividade diminuiu para cerca de 50%, em comparação com a condição ótima de atividade. A atividade máxima de *BlCel48* foi observada na temperatura de 60 °C ([Figura 11C](#)). Os valores ótimos de pH e temperatura obtidos para o *BlCel48* são semelhantes aos determinados para outras enzimas da família GH48. Para essas enzimas, a atividade ótima varia para pHs entre 5,0–6,5 e temperaturas entre 50–70 °C.

Figura 10 – Expressão e purificação da enzima *BICel48*. As etapas de produção e purificação por afinidade em Ni^{2+} para *BICel48* são mostradas do gel desnaturante - (M) marcador de tamanho molecular, (1) proteína total, (2) fração insolúvel, (3) fração solúvel, (4) *BICel48* clivada. O último passo de purificação foi realizado por SEC, como mostrado na banda 5 do gel SDS-PAGE e no cromatograma.

Fonte: Elaborada pelo autor.

As atividades específicas mais altas encontradas em PASC e BC foram descritas para outras enzimas da família caracterizadas como celobioidrolases, tais como *CtCelS*¹⁶⁵ e *CtCel48Y*.¹⁶⁶ Essas duas enzimas têm atividade em Avicel. Entretanto não sabemos porquê *BICel48* não ter atividade neste substrato. Uma possível explicação para isso seria a ausência de um módulo de ligação a carboidratos, como reportado por Sukharnikov et al..⁶⁵ Sukharnikov et al..⁶⁵ mostraram que a enzima GH48 de *H. chejuensis* não é capaz de hidrolisar Avicel na ausência do CBM.⁶⁵

Em seguida, a influência na atividade de vários íons e agentes químicos foi investigada (Figura 11D). O íon estrôncio inibiu substancialmente a atividade de *BICel48* (96%) enquanto que o íon cálcio aumentou em 29%. O efeito de cálcio na atividade enzimática de outras GH48 está relacionado à estabilidade e à melhora na afinidade proteína-substrato, como reportado anteriormente.¹⁶⁷

A análise das curvas de transição térmica mostrou valores mais altos de T_m na faixa de pHs de 5,0 a 9,0 (Figura 12A). Neste caso, o maior T_m foi de 73 °C e melhor pH de 6,5. Além disso, na presença de cálcio o T_m foi de 75 °C, sugerindo que esse íon impacta tanto a atividade enzimática quanto a estabilidade térmica de *BICel48*.

Posteriormente, a atividade residual de *BICel48* foi testada pré-incubando-a nas temperaturas de 40, 50 e 60 °C por 48 h (Figura 12B) e em seguida medindo a atividade após reação de hidrólise em PASC por 20 min a 50 °C. Os resultados indicam que *BICel48*

Figura 11 – Caracterização bioquímica. Os estudos de caracterização consistiram da determinação da atividade específica de *BICel48* em diferentes substratos (A) e da avaliação do efeito do pH (B), da temperatura (C) e de íons/agentes químicos na atividade específica de *BICel48* (D).

Fonte: Elaborada pelo autor.

é significativamente estável a temperaturas entre 40 a 50 °C. Em 40 °C, 90% da atividade foi mantida mesmo após 12 h e aproximadamente 80% da atividade foi observada após 48 h. Uma tendência semelhante foi observada para 50 °C, onde *BICel48* reteve 80% da atividade hidrolítica após 12 e 50% após 24 h. No entanto, sua atividade a 60 °C foi marcadamente reduzida para menos de 40% após incubação por 1 h. Algumas enzimas desta família têm atividade ótima na faixa de temperatura de 40–50 °C, mas também foram rapidamente inativadas a 60 °C.⁶⁵

Na análise do padrão de clivagem gerado por *BICel48* em PASC, os produtos da hidrólise foram medidos no equipamento HPAEC-PAD após incubação da enzima com PASC a 50 °C por 1 e 48 horas. Nestas condições de reação enzimática, *BICel48* produziu celotetraose (Glc4), celotriose (Glc3) e celobiose (Glc2), **Figura 12C**. Após 1 h de hidrólise, *BICel48* produz principalmente Glc2 (64%), bem como uma pequena quantidade de Glc3 e Glc4 (**Figura 12D**).

Após 48 horas de hidrólise, o principal produto foi celobiose, atingindo aproximadamente 83% dos oligossacarídeos solúveis liberados (**Figura 12D**). A maior parte da celotetraose presente nas fases iniciais da hidrólise (após 1 hora de reação) foi reduzida em aproximadamente 4 vezes após 48 horas. Isto indica uma clivagem lenta, mas considerável,

Figura 12 – Caracterização da estabilidade térmica e determinação do padrão de hidrólise em PASC. A estabilidade estrutural foi avaliada em diferentes pHs monitorando fluorescência dependente da temperatura pela técnica de varredura diferencial da calorimetria (A). Já estabilidade térmica foi avaliada monitorando o efeito temperatura na atividade enzimática residual ao longo do tempo (B). Por fim, os produtos da hidrólise sobre PASC foram avaliados nos tempos de 1 h e 48 h (C-D).

Fonte: Elaborada pelo autor.

de Glc4 para celobiose. Essas características de liberar alta quantidade de celobiose e usar como produto intermediário celotetraose por *BiCel48* foram observadas para *TfCel48A*.⁶⁴

4.1.2 Perfil do produto liberado em função do tempo da hidrólise e processividade da enzima *BiCel48* em PASC e BC

A processividade é um importante fator para entender os mecanismos enzimáticos de muitas enzimas que atuam na hidrólise de substratos insolúveis, tais como celulose. A morfologia do substrato, o grau de polimerização DP, a taxa com que a cadeia de celulose desliza frente ao sítio catalítico e ligação inicial enzima-substrato, são fatores que influenciam na medição e cálculo da processividade.¹⁶⁸ Mesmo que, nos estudos anteriores, o perfil do produto de *BiCel48* já tenha sido avaliado em PASC, o decurso temporal da hidrólise utilizando PASC, BC e celoexaose foram incluídos neste estudo a fim de calcular a processividade aparente.

As hidrólises de PASC e BC foram conduzidas em pH 5,5 a 40 °C variando o tempo de 5 min a 48 h. Em seguida, a capacidade de sacarificação de *BiCel48* nesses

substratos foi avaliada através de medidas dos produtos solúveis de celo-oligossacarídeos por HPAEC-PAD. Como pode ser visto na [Figura 13](#), durante o decurso de tempo da hidrólise, *BiCel48* exibiu uma conversão enzimática maior em PASC do que em BC.

Figura 13 – Taxa de hidrólise no tempo. A quantidade de açúcar liberado ao longo do tempo é mostrada para a hidrólise para PASC (●) e BC (△).

Fonte: Elaborada pelo autor.

No tempo inicial (5 e 60 min) a razão da distribuição do produto para PASC aumentou acentuadamente em 16 vezes comparado com BC. Essa razão diminuiu significativamente em até 4 vezes (de 60 a 2880 min). Isso pode ocorrer porque o PASC é mais acessível do que BC para o reconhecimento da cadeia de celulose no seu túnel catalítico de *BiCel48*. Outra possibilidade seria o fato de esta enzima poder ter um maior grau de processividade em BC do que em PASC, dada a baixa taxa de conversão de BC. Com base na última hipótese, o perfil do produto produzido por *BiCel48* foi revisitado com o objetivo de acessar a sua processividade aparente ([Figura 14](#) e [Figura 15](#)).

A análise do padrão de clivagem enzimática em PASC baseia-se nos resultados anteriores, em que *BiCel48* gerou principalmente celobiose e uma pequena quantidade de celotetraose e celotriose. No entanto, de forma inesperada, essa enzima produz uma pequena quantidade de glicose e celoexaose em reações com tempo maior que 24 h ([Figura 14](#)).

A hidrólise de PASC ao longo do tempo mostra que inicialmente *BiCel48*, em tempos de reação curtos, produz seus produtos clássicos (como a celobiose, celotetraose e

Figura 14 – Padrão de hidrólise. Quantidade de oligossacarídeos liberados ao longo do tempo por *BICel48* da hidrólise de PASC.

Fonte: Elaborada pelo autor.

celotriose) e então produz celobiose e celoexaose (muito lentamente). Isso sugere glicose e celoexaose correspondem a cortes inespecíficos produzidos por *BICel48* em PASC. Embora

Figura 15 – Padrão de hidrólise. Quantidade de oligossacarídeos liberados ao longo do tempo por *BICel48* da hidrólise de celulose bacteriana.

Fonte: Elaborada pelo autor.

BICel48 possa produzir uma pequena quantidade de glicose, celotetraose e celoexaose a partir da hidrólise de PASC, na hidrólise de BC é produzido somente celobiose e celotriose (Figura 15). Surpreendentemente, a glicose e celoexaose observadas aqui não são relatadas

em nenhum trabalho anterior para outras enzimas da família GH48.

4.1.3 Superfície do substrato - morfologia e tamanho da fibra dirigem a processividade aparente

Comparando as taxas das medidas dos perfis de celo-oligossacarídeos de BC e PASC, os períodos de hidrólise indicam que *Bl*Cel48 possui maior processividade em BC do que em PASC.

A processividade é pensada como sendo altamente influenciada pelo padrão dos oligossacarídeos gerados na hidrólise, por exemplo, gerando celobiose, celotriose e celotetraose.¹⁶⁹ Celo-oligossacarídeos solúveis gerados pela ligação inicial ($DP \leq 3$) ou mesmo como produtos intermediários ($4 \leq DP \leq 6$) podem introduzir artefatos na medida da processividade.¹⁶⁸ Para superar essa limitação, os produtos da hidrólise de celotriose, celotetraose, celopentaose e celoeixaose foram medidos. Adicionalmente, o mesmo experimento foi realizado com celoeixaose para obter algumas ideias sobre seu perfil dos produtos no tempo.

Uma vez que celotetraose (Glc4) e celoeixaose (Glc6) foram gerados lentamente pela *Bl*Cel48 em PASC, após 180 min, estes celo-oligossacarídeos foram quase totalmente hidrolisados quando usados como substratos (Figura 16). O menor substrato de *Bl*Cel48 foi celotetraose (Glc4), que foi lentamente clivado em duas unidades de celobiose. No entanto, esta enzima foi incapaz de clivar celotriose. Celopentaose (Glc5) foi convertida principalmente em Glc2, Glc3, e Glc4, e levemente em Glc1. Deste modo, celoeixaose (Glc6) foi clivada em Glc2, Glc3, Glc4, e Glc5 e Glc1 em pequenas quantidades.

Figura 16 – Hidrólise de celo-oligossacarídeos. Padrão de hidrólise de diferente celo-oligossacarídeos com graus de polimerização de 3 a 6.

Fonte: Elaborada pelo autor.

Além disso, a hidrólise de celoexaose, variando o tempo da hidrólise, foi realizada para avaliar a distribuição de produto da quebra inicial (Figura 17). A quantidade de Glc2 foi quase equivalente a Glc4 e as quantidades de Glc1 e Glc3 ou Glc5 foram bastante semelhantes entre si. Juntos, esses resultados sugerem Glc3 é gerado na ligação inicial na extremidade da cadeia de celulose em BC, enquanto que Glc3 e Glc4 do modo similar para PASC.

Figura 17 – Hidrólise de Glc6. Padrão de hidrólise de celoexaose ao longo do tempo.

Fonte: Elaborada pelo autor.

Adicionalmente, a primeira derivada foi aplicada na quantidade total de açúcar e celobiose liberados durante o decurso da hidrólise para avaliar o momento com a menor influência da inibição gerada por Glc2 (Figura 18). Como mostrado na Figura 18, as taxas lineares do período de hidrólise foram observadas entre 167 a 231 min.

A partir destas análises, a processividade aparente foi calculada com os dados coletados a 180 min usando a razão $P^{app} = \frac{(Glc2 - (Glc3 + Glc4))}{Glc1 + Glc3}$. A processividade aparente foi 8,6 e 39,6 para PASC e BC. Estes valores indicam que *BICel48* é altamente processiva em PASC ou em BC. Comparada com os valores de processividade recentemente medidos usando o mesmo método, a processividade derivada da hidrólise de PASC por *BICel48* é bastante similar a *MtCel7A* ($P^{app} = 7$)¹⁷⁰ e *TrCel7A* ($P^{app} = 10$)¹⁷⁰ enquanto que é 2,4-3 vezes maior que a processividade calculada para *TfCel48A* em BC ($P^{app} = 16$)¹⁶⁹.

Entretanto, a medição da processividade aparente parece ser altamente dependente do método e do tipo de substrato, bem como da carga enzimática usada na hidrólise, como demonstrado por Kurašin e Våljamäe.¹⁷¹ Além disso, a velocidade desempenhada pela enzima na superfície do substrato, as baixas taxas de associação à cadeia de celulose e a disponibilidade da superfície do substrato para a adsorção da enzima estão fortemente

Figura 18 – Velocidade hidrolítica em PASC e BC. $S(x)$ é definida como a quantidade de açúcar liberado ao longo do tempo da hidrólise e x representa o tempo em que a reação aconteceu. A primeira derivada de $S(x)$ permite saber em que momento mudanças causadas por saturação acontecem. A taxa de hidrólise da quantidade de açúcar total e de celobiose (para PASC e BC) estão relacionadas às taxas de inibição enzimática, o que pode se refletir na processividade. Para isso, a primeira derivada permite saber o ponto provável da mínima inibição.

Fonte: Elaborada pelo autor.

correlacionadas com a processividade de celobioidrolases das famílias GH6 e GH7, e quitinases da família GH18.^{172–175}

A partir da distribuição do produto da quebra inicial e da sua taxa em PASC e BC, a taxa de produção total foi observada como sendo mais significativa nas primeiras 3 h durante o período da hidrólise. Posteriormente, os dados indicam que a taxa de geração de produtos por *BICel48* muda significativamente, sugerindo mudanças na superfície do substrato. Estas taxas indicam que, no tempo inicial, a enzima foi mais rápida do que em períodos longos, quando a hidrólise foi realizada em celulose insolúvel com uma associação mais lenta ao substrato em BC do que em PASC.

BICel48 mostrou atividade reduzida em BC comparado com PASC, na qual PASC foi degradado mais rapidamente do que BC. A taxa de geração dos produtos a partir de substratos insolúveis foi pelo menos 4 vezes menor [Figura 13](#), indicando que a complexação, formação do produto, e consequentemente, a processividade de *BICel48* foram altamente limitados pelo substrato, como relatado anteriormente para a *TfCel48A* de *Thermobifida fusca*.¹⁶⁹ Estas disparidades na degradabilidade de PASC e BC foram propostas por [Kosty-](#)

lev et al.¹⁶⁹ como tendo relação com a alta acessibilidade e o número de extremidades da cadeia disponíveis em PASC comparada com a BC, o que pode explicar as taxas mais lentas de digestão em BC em comparação com PASC observadas para *B*lCel48.

PASC é um tipo de celulose altamente amorfa obtida após o tratamento da celulose Avicel com ácido fosfórico¹²¹ sem impacto no grau médio de polimerização tendo aproximadamente 202-214 unidades de glicose¹⁷⁶. Além disso, BC é uma celulose pura com grande área de superfície, alta porosidade.^{177,178} Alinhado com isso, o DP medido aqui para Avicel, PASC, e celulose bacteriana foram 219, 237, e 2259 unidades de glicose,¹⁷⁹ respectivamente, os quais são similares aos valores reportados em outros estudos (2000–2600 unidades de glicose.¹⁸⁰ Neste caso, considerando o valor da processividade de *B*lCel48 em PASC ($P^{app} = 8,6$) e BC ($P^{app} = 39,6$), assim como o seu respectivo DP (219 e 2259, respectivamente), a razão da processividade entre PASC e BC foi aproximadamente 5, enquanto que a razão entre os DPs de PASC/BC foi aproximadamente 5,2 unidades de celobiose (Figura 19). Este resultado sugere que a afinidade de ligação de *B*lCel48 à cadeia de celulose dentro do túnel do sítio ativo é altamente dependente da morfologia do substrato, acessibilidade à extremidade da cadeia e tamanho da fibra, o que causa um grande efeito no grau de processividade e na taxa de hidrólise em geral.

Figura 19 – Processividade ao longo tempo. Processividade aparente calculada ao longo do tempo para PASC (○) a celulose bacteriana (△).

Fonte: Elaborada pelo autor.

Como *B*lCel48 gera principalmente Glc2 a Glc4 em PASC e apenas Glc2 a Glc3 em BC, as distribuições destes produtos sugerem que os padrões de produtos gerados a partir de PASC e BC ou celo-oligossacarídeos são resultantes da susceptibilidade dos sub-sítios +1, +2, +3 e +4 em acomodar lá o terminal da unidade glicosil. A estrutura cristalográfica de diversas GH48s^{61,181} com ou sem oligossacarídeos mostrou que no sítio

ativo (-7 a -1 e +1 a +3) destas enzimas a ligação com o substrato é intermediada por uma extensa rede de ligações de hidrogênio e diversos resíduos aromáticos propostos como os responsáveis pela acessibilidade e manutenção das cadeias de celulose no túnel catalítico.

4.1.4 Estudos estruturais por cristalografia de raios X

Para entender o mecanismo molecular envolvido na hidrólise da celulose por *BICel48*, foi determinada a estrutura tridimensional desta enzima. A proteína *BICel48* foi submetida a vários ensaios de cristalização. A presença de celobiose foi crítica para o crescimento de bons cristais para experimentos de difração. *BICel48* foi cristalizada na presença de celobiose no grupo espacial $P12_11$. A estrutura foi determinada com resolução de 1,75 Å com uma molécula por unidade assimétrica. Após extensivos ciclos de refinamento, obteve-se R_{free} final de 0,18. A [Tabela 3](#) resume os parâmetros e as estatísticas recuperadas da coleta de dados de difração de raios X, processamento e do refinamento.

Tabela 3 – Estatísticas de processamento. Estatística da coleta, processamento dos dados de difração, determinação estrutural e refinamento.

	<i>BICel48</i> -G2+G4	<i>BICel48</i> -G2
Coleta e processamento		
Comprimento de onda (Å)	1,458	1,458
Grupo espacial	$P12_11$	$P12_11$
a, b, c (Å)	59,537 98,98 59,91	59,651 98,79 60,19
α, β, γ (°)	90 103,13 90	90,00 103,34 90,00
Resolução (Å)	29,7 - 1,68 (1,74 - 1,68)	29,7 - 1,95 (2,02 - 1,95)
R_{meas}	0,07793 (0,7242)	0,09532 (0,2838)
$CC_{1/2}$	0,997 (0,481)	0,992 (0,84)
CC^*	0,999 (0,806)	0,998 (0,956)
Reflexões totais	420867 (26232)	95600 (9108)
Reflexões únicas	75730 (7148)	48205 (4573)
Multiplicidade	5,6 (3,8)	5,3 (4,1)
Completeza (%)	0,99 (0,91)	0,97 (0,93)
$\langle I/\sigma(I) \rangle$	18,85 (2,40)	21,74 (2,62)
Fator B de Wilson	12,88	12,34
No. Mol. na cela	1	1
Refinamento		
R_{work}	0,1504 (0,2384)	0,2246 (0,2389)
R_{free}	0,1833 (0,3072)	0,2608 (0,2774)
macromoléculas	1	1
Número de resíduos de aminoácidos	698	698
<i>Desvios (R.m.s.)</i>		
Comprimentos de ligação (Å)	0,011	0,003
Ângulos de ligação (°)	1,07	0,59
<i>Ramachandran</i>		
Favorável (%)	98	97
Permitido (%)	2	2,9
Outliers (%)	0,29	0,14
<i>Fator B</i>		
Médio	14,99	12,15
macromoléculas	13,86	11,58
solvente	25,45	18,00
$R_{meas} = \frac{\sum_{hkl} \sum_{i=1}^n (\sqrt{\frac{n}{n-1}})(I_i(hkl) - I(hkl))}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$		

Fonte – Elaborada pelo autor.

Esta estrutura foi obtida por cocrystalização com celobiose, das quais duas moléculas

las de celobiose foram modeladas no túnel do sítio ativo, abrangendo os sítios de ligação -5 a -2. Uma terceira molécula de celobiose foi modelada na parte da fenda catalítica aberta nos sítios +1 e +2 (Figura 20). Um segundo conjunto de dado foi obtido com a celobiose nas posições +1 e +2 (Tabela 3). No entanto, não foi possível obter a estrutura desta enzima na forma apo. *BlCel48* tem uma topologia geral que compreende um barril (α/α)₆

Figura 20 – Estrutura cristalográfica. Modelo cristalográfico da estrutura de *BlCel48*. Os elementos de estrutura terciária são principalmente formados por alfa-hélices com topologia tipo barril com as repetições (α/α)₆. Na estrutura de *BlCel48* os subsítios -5 a -2 e +1 a +2 interagem com a celo-oligossacarídeos.

Fonte: Elaborada pelo autor.

tradicionalmente observada em outras enzimas da família GH48.¹⁸² O alinhamento estrutural e o rmsd calculado de carbonos-alfa realizados no programa DALI Server¹⁸³ mostra similaridades estruturais com outras enzimas da família GH48. A estrutura de *BlCel48* é conservada entre as outras enzimas GH48, incluindo *CtCel48S* de *Clostridium thermocellum* (PDB id 1L1Y; rmsd= 1,3 Å⁶¹), *CcCel48F* de *Clostridium cellulolyticum* (PDB id 1G9G; rmsd= 1,2 Å¹⁸²), *TfCel48A* de *Thermobifida fusca* (PDB id 4JJJ; rmsd= 1,4 Å¹⁶⁹) e *BpCel48* de *Bacillus pumilus* (PDB id 5BV9/5CVY; rmsd= 0,7 Å¹⁴⁰).

O modelo estrutural de *BlCel48* compreende os resíduos Asp2–Ala704. O sítio catalítico observado pode ser subdividido em duas regiões com pelo menos nove subsítios de ligação para celo-oligossacarídeos formados por uma entrada em forma de túnel e uma saída em forma de fenda aberta. *BlCel48* exhibe três *loops* mais longos (Gly462–Gln492, *loop* 1; Val626–Ser640, *loop* 2; Asn666–Gly675, *loop* 3), que representam a maior diferença estrutural quando comparada com outras estruturas previamente determinadas (Figura 21). Os *loops* 1 e 3 estão localizadas na saída do túnel; comparados com as estruturas cristalográficas de outras enzimas GH48, estes dois *loops* estão localizados no final do sítio catalítico e provavelmente podem contribuir para o seu modo de ação.

Figura 21 – Comparação estrutural entre *BICel48* e outras GHs da família 48. No alinhamento estrutural de *BICel48* em relação às outras GHs da família 48 são observados três *loops* presente na estrutura *BICel48*: *loop1*, *loop2* e *loop3*. Os triângulos em cor vermelha mostram a região para cada um destes *loops*. Os *loops* 1 (colorido em azul claro) e 3 (colorido em violeta) localizados na saída do túnel catalítico.

Fonte: Elaborada pelo autor.

O caminho de ligação do substrato é significativamente maior e mais fechado do que em outras estruturas de enzimas GH48 devido aos *loops* 1 e 3. As interações dos ligantes no túnel e os sub-sítios da abertura da fenda delimitam as posições -5 a -1 e +1 a +2, respectivamente. Em relação às características estruturais de *BICel48*, a sua estrutura preserva a forma globular, enquanto que o seu sítio ativo é composto por pelo menos nove sub-sítios (-7 a -1 e +1 a +2).

Próximo à saída do túnel de *BICel48*, um íon de cálcio octacoordenado faz ligações de hidrogênio com os resíduos Gln182, Glu187, Asp406, e com várias moléculas de água, bem como interações mediadas por água com Arg598 e Glu603, onde as distâncias de ligação variam de 2,41–2,63 Å (Figura 22). Um ambiente similar com resíduos de Glu/Asp/Gln para o íon cálcio foi observado na estrutura de *CcCel48F* e *TfCel48A*, o que indica que estes resíduos estão intimamente relacionados com a coordenação do íon cálcio na família GH48. Adicionalmente, este íon está envolvido na interação com dois *loops* formadores do túnel (*loop* Glu400-Trp412 e Asn175-Gly205) e uma hélice-alfa (Glu594-Tyr599) localizada na saída do túnel catalítico, sugerindo que este íon pode conduzir algumas mudanças conformacionais durante a ligação com resíduos proteicos. De fato, a atividade enzimática determinada na presença de 5 mM de $CaCl_2$ foi aumentada em até 29%, como observado anteriormente. No entanto, todas estas interações estão suficientemente longe do sítio catalítico e nenhum contato com o ligante nos sub-sítios +1 e +2 foi observado.

4.1.4.1 sub-sítios: mapeando as interações com celo-oligossacarídeos

Nos sub-sítios (-5 a -2), parte do túnel, uma clara densidade eletrônica foi observada correspondendo a uma molécula de celotetraose. Uma representação esquemática

Figura 22 – Ligantes envolvidos na interação com *BICel48*. Na parte superior é mostrado o íon de cálcio octacoordenado por ligações de hidrogênio. Abaixo são mostrados as principais ligações de hidrogênio dos celo-oligossacarídeos posicionados nos subsítios -5 a -2 (I) e +1 a +2 (II).

Fonte: Elaborada pelo autor.

das subunidades do açúcar nos sub-sítios -5 a -2 na parte do túnel do complexo de açúcares, que faz a interação com as cadeias laterais dos aminoácidos (Figura 22). Olhando para os sub-sítios de clivagem no túnel, os celo-oligossacarídeos têm uma conformação geométrica que se assemelha a uma hélice torcida. Ambas as subunidades de celobiose foram estabilizadas pelas ligações de hidrogênio com resíduos de proteína e água, e pelo empilhamento das interações com os resíduos aromáticos.

No sub-sítio -5 vários resíduos de aminoácidos interagem com a glicose. Glc^{-5} faz ligação de hidrogênio com Thr107 (O6) e Tyr269 (O6); e interações mediadas por água com Gln220 (O3), Arg222 (O2) e Thr107 (O4). Além disso, Glc^{-5} faz interações de empilhamento com Trp312. Considerando a interação de empilhamento, interações direta e medida por água, a orientação do anel Glc^{-5} é altamente estável, sugerindo que a liberdade do anel de Glc é reduzida. De forma similar, o subsítio -4 do anel Glc^{-4} interage com Lys268 (O2 e O3), Gln178 (O6) e Tyr297 (O2) fazendo ligação de hidrogênio. Além disso, várias moléculas de água fazem ligações de hidrogênio com Glc^{-4} . Ademais, no sub-sítio -3 apenas uma molécula de água (O6) faz interação direta com Glc^{-3} enquanto que

pelo menos quatro resíduos de proteína fazem ligações de hidrogênio com Glc^{-3} . O anel de glicopirranose Glc^{-3} interage com resíduos de Tyr404 (O6), Asn175 (O2 e O3), Gln178 (O3) e Thr224 (O2 e O3). No entanto, para Glc^{-2} o anel glicopirranose tem ligações de hidrogênio com várias moléculas de água. Estas interações são muito similares àquela feita no anel Glc^{-5} . Glc^{-2} interage com as cadeias laterais de Trp296 (O3), Tyr404 (O3), Tyr323 (O1) e Asn225 (O6), enquanto o resíduo de triptofano Trp296 está envolvido na interação com Glc^{-2} via interação de empilhamento. No sub-sítio -1 não foi observada a presença de ligantes. O volume da cavidade na posição -1 é ocupado por algumas moléculas de água e duas moléculas de etilenoglicol.

A análise dos sítios de ligação dos produtos nos permite observar uma molécula de celobiose nos sub-sítios +1 e +2. A análise inicial do ambiente destas posições mostrou que este ligante faz várias ligações de hidrogênio com moléculas de água e resíduos de proteína. Em relação a esse complexo, nenhuma densidade eletrônica adicional foi encontrada para ligantes nos sub-sítios +3 e +4. A unidade glicosil Glc^{+1} interage com Glu38 (O2), Glu49 (O3 e O4) e ligação mediada por água com Asp228 (O4); Glc^{+1} também faz interações com os aromáticos triptofano Trp685 (O2) e Trp418 (O4). Glc^{+2} faz ligação de hidrogênio apenas com o resíduo Asp535 (O6) e com Trp412 (O1), esta última é mediada por uma molécula de água. A orientação do Trp412 permite uma interação de empilhamento com Glc^{+2} , mesmo o Trp412 estando levemente desviado da sua posição relativa para realizar uma interação de empilhamento clássica. Embora a unidade de Glc^{+1} faça diversas interações com a cadeia lateral de proteína, Glc^{+2} faz uma rede de ligações de hidrogênio maior na região da fenda aberta do sub-sítio +2. Esta rede de ligações de hidrogênio pode induzir a estabilização da celobiose nestes sub-sítios e, provavelmente, aumentar a ligação dos celo-oligossacarídeos lá. Adicionalmente, as interações eletrostáticas carregadas bem como as interações hidrofóbicas dos resíduos aromáticos com os anéis de celobiose nos sub-sítios +1 e +2, poderiam desempenhar um papel importante na melhoria da estabilização da celobiose. Além disso, os sub-sítios nas posições +1 e +2 parecem ser o único elemento estrutural restritivo para a ligação do substrato na fenda catalítica de *BlCel48*, podendo desempenhar um papel fundamental na baixa atividade de *BlCel48*, especialmente na inibição por produtos finais.

4.1.4.2 Explorando as interações substrato-proteína para a família GH48

A análise da estrutura terciária seguida da sobreposição das três enzimas da família GH48 mais estudadas (*CcCel48*, *CtCel48* e *TfCel48A*) foi realizada para avaliar as diferenças estruturais entre elas e *BlCel48*. Estas enzimas possuem estruturas tridimensionais comparáveis. As duas maiores diferenças estruturais observadas foram os *loops* 1 2, e 3 de *BlCel48*; além dos *loops*, a estrutura de *BlCel48* é altamente semelhante a outras três enzimas da família. Uma vez que os *loops* estão localizados logo na saída do túnel catalítico, eles poderiam contribuir para a interação do produto da celobiose nos sub-sítios

+1 e +2 por fechar a saída do produto da fenda ativa.

As interfaces dos sítios de ligação permitiram a visualização de diferentes ligantes exibidos entre estas enzimas e *BlCel48*, dos quais os sítios de ligação +1 e +2 são muito semelhantes. Um grande número de resíduos de aminoácidos que interagem com a celobiose nos sub-sítios +1 e +2 é equivalente entre estas enzimas, indicando que os sub-sítios positivos possuem modo de ligação similar. A molécula Glc^{+1} forma múltiplas ligações de hidrogênio com Glu39 e Glu49, bem como interações aromáticas com His30, Trp418 e Trp685. Da mesma forma, o anel de glicopirranose Glc^{+2} forma ligações de hidrogênio com Asp535 e interação de empilhamento, levemente deslocada, com Trp412. As interações de empilhamento e ligações de hidrogênio do triptofano com celobiose foram feitas através do lado plano da superfície plana da molécula de celobiose, enquanto que Glu38, Glu591, Arg683 e His30 interagem com o lado oposto desta superfície.

Esta topologia nos sub-sítios +1 e +2 sugere que os membros de GH48 possuem um caminho particular de preferência e ajuste para os sítios de ligação dos produtos. Adicionalmente, este tipo de arranjos e as propriedades químicas dos resíduos de aminoácidos podem aumentar a estabilidade da celobiose nos sub-sítios +1 e +2, e com isso aumentando a energia necessária para a expulsão do produto. Isso sugere que estes sub-sítios estão relacionados com a ligação improdutiva e inibição de produtos, o que pode ser uma das causas para a baixa atividade específica observada para *BlCel48*. [Chen et al.¹⁸⁴](#) sugeriram que o efeito da alta inibição de muitas enzimas desta família é devido a uma aquisição ineficiente da cadeia de celulose na entrada do túnel e ao efeito inibidor do produto final.

Em comparação, *BlCel48* teve um ambiente de ligação similar aos observados por [Chen et al.¹⁸⁴](#) em torno dos sub-sítios +1 e +2, preservando principalmente as ligações de hidrogênio e os contatos aromáticos com o complexo. A modelagem molecular de uma molécula de celobiose na fenda aberta indica que Glc^{+3} e Glc^{+4} podem ser acomodados ao longo dos sub-sítios +3 e +4. No entanto, estes sítios têm muito mais contatos com moléculas de água do que com resíduos de *BlCel48*. De fato, as análises dos complexos de *CcCel48* com a celotriose nos sub-sítios +1, +2 e +3, bem como a sua estrutura complexa com o inibidor de tio-oligossacarídeos nos sub-sítios +1, +2, +3 e +4, mostram que os sub-sítios +3 e +4 desta enzima não fazem muitas interações com os ligantes, sugerindo que o comportamento das interações do ligante nestas posições não requer alta especificidade. Estes achados indicam que nas posições + 3 e +4 o oligossacarídeo possui requisitos mínimos no seu sítio para a ligação, e não requer interações fortes. Além disso, a sobreposição da estrutura de *BlCel48* e todas as outras enzimas da família 48 mostraram que os *loops* 1 e 3 parecem ser uma extensão dos sub-sítios, ao mesmo tempo que um estreitamento do bolsão catalítico ([Figura 21](#)).

Curiosamente, os *loops* localizados na saída do sítio catalítico têm mostrado importância fundamental na liberação do produto das famílias de hidrolases glicosídicas 74¹⁶³

e 6¹⁸⁵ onde os sítios ativo estão cercados por *loops* localizados na saída do bolsão catalítico, indicando terem um papel relevante na ligação do produto e no reconhecimento de substratos, especialmente na expulsão do produto.

Embora os sub-sítios dos produtos sejam uma fenda aberta que contém pelo menos os sub-sítios +1 a +2, a estrutura de *BlCel48* compreende os sub-sítios -1 a -7 dentro do túnel catalítico. Juntamente com o túnel, várias ligações de hidrogênio, assim como interações aromáticas e de empilhamento, foram formadas entre os resíduos de proteína e as unidades glicosil dos substratos de celotetraose e celobiose. Estes tipos de interações ao redor dos sub-sítios do túnel, podem fornecer o caminho para o correto alinhamento do substrato necessário para a clivagem processiva da cadeia de celulose, similarmente ao observado anteriormente para a celobiodrolase *TrCel7A* da família GH7.^{63,186}

Comparativamente, as conformações dos ligantes destas estruturas sobrepostas foram similares aos sub-sítios -2 a -5, exceto para os sub-sítios -6 a -7. O sub-sítio -2 não apresenta diferenças com relação ao local de ligação do substrato, exceto para o resíduo Tyr404 em *BlCel48*, que é substituído por Trp404 em *TfCel48A* (PDB ID 4jjj). Este sub-sítio foi caracterizado por ter ligações de hidrogênio com resíduos de Tyr404, Tyr323 e Trp296, bem como interação de empilhamento envolvendo o resíduo Trp151. Os resíduos Tyr404 e Trp151 também estão envolvidos nas interações de ligação de hidrogênio com Glc⁻³; adicionalmente, o Glc⁻³ possuía ligações de hidrogênio com Gln178, Gln175 e Thr224.

No sub-sítio -4, o Glc⁻⁴ forma interações de ligação de hidrogênio com três resíduos proteicos (Lys268, Gln178 e Tyr297), enquanto que Trp312 e Tyr269 fazem interações de empilhamento e ligação de hidrogênio com Glc⁻⁵, respectivamente. Em *BpCel48* (PDB id:5cvy) Glc⁻⁶ estabelece ligações de hidrogênio com a cadeia lateral dos resíduos Thr106 e Gln219 enquanto que Glc⁻⁷ faz empilhamento com Trp309, bem como ligação de hidrogênio com Gln214 e Glu213. Em *BlCel48*, no sub-sítio -6, os resíduos equivalentes foram Thr107 e Gln220, enquanto que na posição -7 os resíduos Trp310, Gln215 e Glu214 interagem com Glc⁻⁷.

Comparando *BlCel48* e *BpCel48*, os elementos estruturais equivalentes dessas enzimas são altamente conservados, especialmente os resíduos que estão envolvidos na interação com o ligante. No sub-sítio -5, três moléculas de água (HOH161, HOH247 e HOH267) foram altamente coordenadas próximas à face plana de Glc⁻⁵, sugerindo que estas águas podem estabilizar o anel glicopiranosídeo em uma conformação de cadeira localizada neste sub-sítio. A celobiose posicionada na entrada do túnel (sub-sítios -6 e -7) mostra algumas diferenças conformacionais. Isto sugere certo grau de flexibilidade dos anéis Glc⁻⁶ e Glc⁻⁷ nestas posições, provavelmente como consequência da camada de solvente perto da entrada do túnel, mesmo o Trp310 aromático fazendo interação de empilhamento com o anel Glc⁻⁷.

4.1.5 Percepções sobre afinidades e interações de substrato/produtos no sítio catalítico usando análise de acoplamento de estatística (SCA)

Como mencionado anteriormente, *BiCel48* tem um túnel de 45 Å (-7 a -1), e sítios de ligação na fenda aberta (+1 a +3), onde os substratos e os produtos transientes podem ser localizados imediatamente após o ciclo de clivagem. Estes sub-sítios ([Figura 23](#)) foram mapeados com base na análise estrutural de *BiCel48* e na comparação das estruturas de outras enzimas GH48 ligadas ou não. Muitas destas estruturas possuem celobiose firmemente ligada aos sub-sítios +1 a +2; e em alguns casos, a celotriose estava ligada aos sub-sítios +1 a +3, mesmo se os sub-sítios -7 a -1 não estivessem ocupados. De maneira similar, celo-oligossacarídeos longos, como por exemplo a celoexaose, estão ligados a parte do túnel (sub-sítios -7 a -2); em poucas estruturas a molécula de celopentaose está ligada à parte catalítica do túnel (-6 a -2), sugerindo que estes sítios tendem a ser mais flexíveis para acomodar celo-oligossacarídeos com diferentes graus de polimerização.

Figura 23 – Mapa de conservação da sequência e sub-sítios envolvidos na interação com o ligante. As representações em cartoon e em superfície mostram o posicionamento de celo-oligossacarídeos no túnel catalítico *BiCel48*. Os resíduos foram coloridos de acordo com o sistema de cor do programa ConSurf.¹⁴⁶

Fonte: Elaborada pelo autor.

No entanto, o sub-sítio -1 parece ser energeticamente desfavorável para acomodar celo-oligossacarídeos, como evidenciado pela análise estrutural de *BiCel48* e outras enzimas GH48, nas quais os celo-oligossacarídeos tendem a ocupar a parte do túnel enquanto o sítio -1 está desocupado. Isso foi observado anteriormente^{187,188} para a celobioidrolase *TrCel7A*.

A [Figura 24](#) mostra a enzima apo (PDB id: 4EL8; 1G9G; 4XWL; 5YJ6), o complexo improdutivo (7b: PDB id: 5BV9; 4L0G; 1FAE; 4FUS; 1L1Y), ambos os substratos

no túnel (sub-sítios -7 a -2) e o produto nas partes complexas dos sub-sítios de expulsão (sub-sítios -1 a +2) (7c: PDB id: 5CVY; 1FBW; 4JJJ), e, por último, o complexo inibidor preenchendo todo o sítio catalítico (7d: PDB id: 2QNO). Estas comparações estruturais sugerem que, na ausência da cadeia de celulose no sítio catalítico destas enzimas, incluindo *BiCel48*, a celobiose se liga preferencialmente nas posições +1 a +2.

Figura 24 – Sub-sítios conservados. Mapeamento dos sub-sítios conservados na família GH48 e observados em *BiCel48*. Via comparação estrutural, são mostradas as permissividades de celo-oligossacarídeos nos sub-sítios de *BiCel48*.

Fonte: Elaborada pelo autor.

O túnel é formado principalmente por *loops* estabilizados por um grande número de ligações de hidrogênio e algumas interações de empilhamento estabelecidas entre os celo-oligossacarídeos. Contém no mínimo quatro resíduos Trp altamente conservados (Figura 25) que estão envolvidos nas interações de empilhamento com o ligante das unidades de celobiose (Figura 24).

Três destes resíduos estão localizados no sítio de ligação produtivo (Trp151 no sub-sítio -2; Trp312 no sub-sítio -5; e Trp310 no sub-sítio -7), enquanto que o Trp412 está localizado no sítio de ligação improdutivo (isto é, sub-sítio +2). Este tipo de arranjo envolvido nas interações de empilhamento é muito similar ao observado para a *TrCel7A*,¹⁸⁶ que também possui quatro resíduos de triptofano conservados, três dos quais estão localizados na parte do túnel (Trp40 na entrada do túnel no sub-sítio -7; Trp38 no centro do túnel do sub-sítio -4; e Trp367 no sub-sítio -2) e o Trp376 na parte da fenda aberta no sub-sítio +1. Estas observações sugerem que estas duas famílias de enzimas possuem um mecanismo semelhante de reconhecimento da superfície de substratos insolúveis. De fato, a assimetria do sítio catalítico, bem como as interações de empilhamento, em particular os resíduos de

Figura 25 – Resíduos aromáticos ao longo do túnel catalítico de *BICel48*. São mostrados os resíduos aromáticos altamente conservados ao longo do túnel catalítico e que estão envolvidos na interação com o substrato.

Fonte: Elaborada pelo autor.

triptofano, devem favorecer a processividade para as enzimas GH48, especialmente para *BICel48*.

Para investigar a importância destes Trp nas clivagens de celulose por *TfCel48A*, Kostylev e colaboradores¹⁶⁹ realizaram várias mutações pontuais de resíduos aromáticos (Tyr97Ala, Fhe195Ala, Tyr213Ala, Trp313Ala, Trp315Ala e Tyr211Ala), não aromáticos (Asn212Ala, Ser311Ala) e mutações duplas (Tyr213Ala-Ser311Ala e Tyr213Ala-Trp313Ala) em resíduos localizados fora e dentro da entrada do túnel. Como resultado destas mutações, a processividade tanto em PASC quanto em BC foi grandemente afetada quando Trp313 e Trp315 foram mutados para Ala; isto também foi acompanhado pela mudança na atividade específica, que foi aumentada em PASC enquanto que em BC foi diminuída. Em *BICel48*, os resíduos Tyr100, Trp310 e Trp312 estão conservados na mesma posição dos seus equivalentes estruturais Tyr97, Trp313 e Trp315 em *TfCel48A*, enquanto que os equivalentes estruturais de Fhe195, Tyr213 e Asn212A em *TfCel48A* são os resíduos de Gln197, Pro217 e Gln215, respectivamente, em *BICel48*. Comparativamente à *TfCel48A*, os resíduos Trp310 (sub-sítio -7) e Trp312 (sub-sítio -5) na entrada do túnel de *BICel48* podem estar intimamente relacionados com a adsorção da celulose e o seu encaixe no sítio catalítico.

A análise SCA foi usada para identificar os resíduos coevoluídos associados à atividade enzimática e processividade na família GH48. O alinhamento múltiplo de 2000

sequências foi realizado usando o programa PySCA.¹⁴⁸ Através da análise da matriz de correlação, três subgrupos de posições de aminoácidos coevoluídos (IC1-3) foram determinados (Figura 26).

Figura 26 – Grupos de posições de aminoácidos co-evoluídos. É mostrado o resultado dos cálculos de SCA e seus respectivos ICs projetados na estrutura de *BICel48* em termos de resíduos mapeados.

Fonte: Elaborada pelo autor.

Uma vez que os setores podem ser considerados como unidades funcionais na proteína formados por uma rede fisicamente contígua de aminoácidos estruturalmente conectados,¹⁴⁸ apenas os três primeiros grupos de IC (Figura 27) estão mapeados na estrutura 3D de *BICel48* como resíduos proteicos coevoluídos.

Baseado na matriz SCA, o padrão das correlações inter-ICs indica que o IC1 é um setor independente, enquanto que ICs 2-3 estão inter-correlacionados, o que sugere terem uma função comum; neste caso, estes dois IC foram agrupados em um único setor (setor 2). Na rede de resíduos identificados pela matriz SCA em *BICel48*, foram mapeados principalmente os sítios de ligação ao substrato (coloridos em destaque). O setor 1 revela um único setor que abrange os sítios de ligação do túnel (sub-sítios -7 a -1) até a parte dos sub-sítios relacionados aos produtos (+1 a +2). O segundo setor mapeia a superfície da fenda aberta localizada no sub-sítio +3, que inclui alguns resíduos de aminoácidos do *loop*

Figura 27 – Grupos de posições de aminoácidos co-evoluídos. Resultados do SCA mostrando os ICs e seus amino ácidos envolvidos na clusterização das posições co-evoluídas.

Fonte: Elaborada pelo autor.

Trp665-Glu673. Este achado indica uma relação direta entre as posições de aminoácidos coevoluídos nos setores 1–2 e as unidades funcionais da família GH48.

Consistente com a conservação dos resíduos e os sub-sítios substrato/produto nas estruturas 3D das GH48, os resíduos de aminoácidos de *BICel48* interagem extensivamente com os ligantes celobiose (sub-sítios +1 a +2) e celoexaose (sub-sítios -7 a -2); na forma ligada destas enzimas (Figura 22), estão fortemente coevoluídos ao longo dos setores posicionais. As posições coevoluídas do setor 1 capturam os resíduos que são críticos para a interação proteína-ligante, enquanto que as posições no setor 2 capturam os resíduos relacionados com a ligação improdutivo do ligante. Analisando o setor 1, os resíduos coevoluídos com maior peso no cálculo pelo SCA são: Trp130 (sub-sítio -7), Gln220 (sub-sítio -6), Tyr269 (sub-sítio -5), Lys268/Tyr297/Gln178 (-4), Thr224/Gln178/Asn175 (sub-sítio -3) e Trp296/Tyr323 (-2). Da mesma forma, nos sub-sítios da fenda aberta, o Trp418/Glu38 (sub-sítio +1), bem como Trp412/Asp535 (sub-sítio +2) são altamente co-

evoluídos. Curiosamente, o resíduo Trp130 (sub-sítio -7), que é altamente co-evoluído, é essencial para o acesso e processividade da fibra de celulose, como relatado anteriormente para a família das enzimas GH48.¹⁶⁹

Além disso, os resíduos no setor 2 foram localizados principalmente no lado aberto do sítio catalítico na saída do túnel; no entanto, a partir do exame deste setor, não é clara a ligação entre os resíduos coevoluídos e sua função à primeira vista. Todavia, os resíduos coevoluídos com maior peso (Asp595, His593 e Glu591) estão localizados no sub-sítio +3 em *CcCel48F* (PDB id. 1FBW); His593 (Arg544) e Glu591 (Glu542) formam uma rede de ligações de hidrogênio com a celotriose, sugerindo que as posições destes resíduos são importantes para modular a conformação destes resíduos, bem como a relação destes com a expulsão do produto durante a atividade catalítica das enzimas da família GH48.

4.1.5.1 Estudos estruturais por espalhamento de raios X a baixo ângulo

As medidas e estudos de SAXS foram realizados afim de determinar o envelope estado oligomérico de *BiCel48* em solução. Os resultados obtidos para *BiCel48*, assim como os parâmetros estruturais coleta e processamento são mostrados na [Figura 28](#) e [Tabela 4](#).

Figura 28 – Análise e determinação de parâmetros estruturais por SAXS. (A) Intensidade experimental de *BiCel48* (círculos pretos abertos) é sobreposta com as curvas teóricas de espalhamento calculadas a partir do modelo restaurado. (B) Gráfico de Kratky, (C) $P(r)$ determinada a partir dos dados experimentais de espalhamento. (D) Intensidade experimental SAXS de *BiCel48* (círculos pretos abertos) ajustada contra as curvas de espalhamento calculadas das macromoléculas PDB id 4L0G (mostradas como uma linha preta) e PDB id 5BV9 (mostrada como uma linha pontilhada preta).

Fonte: Elaborada pelo autor.

Usando a aproximação de Guinier vê-se que a região inicia da curva de espalhamento exibe um perfil linear na região de ajuste (dados não mostrados), mostrando a ausência de interferência inter-partículas ou contribuição de danos por radiação. O raio de giro determinado partir do gráfico de Guinier foi de 25 Å (Tabela 4).

Tabela 4 – Parâmetros de coleta, processamento e análise. São apresentados os parâmetros da coleta, da análise dos dados, bem como parâmetros estruturais do processamento e análise.

Parâmetros de coleta de dados				
Instrumento		LNLS D02A-SAXS2		
Comprimento de onda $\lambda(\text{\AA})$:		1,54		
q-intervalo ($\text{\AA}^{-1}$),usado no modelo		0,02 a 0,31		
Tempo de exposição (s)		300		
Concentração de proteína		2 e 6 mg/mL		
Temperatura		20 (°C)		
Parâmetros estruturais				
	BICel48	Cristalográfico	PDB id 5BV9	PDB id 4L0G
R_g (Å) (Por Guinier)	25,1 ± 0,5	25,3	24,8	24,7
R_g (Å) (por $P(r)$)	25,0 ± 0,2	25,2	24,8	24,6
D_{max} (Å)	67,0	66	69	67
$\chi_{Exp/DAMMIN}$	1,3			
$\chi_{Exp/GASBOR}$	1,3			
$\chi_{Exp/PDBid}$	NA	1,9	3,4	3,7
Determinação da massa molecular				
MW (kDa) estimado usando SAXSMoW2.0			77,2	
MW (kDa) com base na sequência de aminoácidos			79,5	
Programas				
Processamento			Fit2D, Scatter, e ATSAS	
<i>ab initio</i>			DAMMIF	
Superposição 3D			SUPCOMB	
NA - Não avaliado				

Fonte – Elaborada pelo autor.

Com base nos dados de espalhamento (Figura 28A), a massa molecular molecular de 77,2 kDa foi calculada a partir da intensidade de espalhamento, como descrito por Piiadov et al..¹⁰²

A função de distribuição pares de distâncias, $P(r)$, tem um perfil característico de partículas globulares com distância máxima (D_{max}) de 68 Å (Figura 28B). O R_g calculado do espaço real foi de 24,6 Å, semelhante ao R_g estimado a partir da aproximação de Guinier. O gráfico de Kratky tem um pico bem definido, sugerindo que essa enzima é muito compacta (Figura 28C).

Os modelos *ab initio* foram reconstruídos usando o programa DAMMIN¹³⁰ e cada modelo gerado foi avaliado usando o critério da menor discrepância entre o a curva experimental e a curva representativa para cada modelo. O valor médio de NSD¹³¹ obtido foi de 0,470 com desvio padrão de 0,004. O envelope de BICel48 reconstruído a partir dos dados do SAXS teve ótima superposição com o modelo cristalográfico ($\chi = 0,88$) desta

enzima (Figura 29). O perfil da curva de espalhamento calculada para o modelo cristalográfico de *B*lCel48 reproduziu bem a curva experimental de SAXS. Além disso, as curvas calculadas a partir das coordenadas atômicas dos modelos cristalográficos (PDB ids: id 4L0G e 5BV9) dos homólogos de *B*lCel48 são muito similares aos dados experimentais de SAXS obtidos para *B*lCel48.

Figura 29 – Modelo *ab initio*. Modelo de baixa resolução de *B*lCel48. Vista frontal e lateral de *B*lCel48. Envelope molecular DAM *ab initio* monomérico (superfícies semitransparentes) obtido por SAXS sobreposto ao cristalográfico.

Fonte: Elaborada pelo autor.

5 RESULTADOS E DISCUSSÕES

5.1 Caracterização bioquímica e estrutural de uma celulase *BlCel9* de *Bacillus licheniformis*: correlações entre modularidade, flexibilidade e modo de ação

5.1.1 Produção, caracterização bioquímica, estabilidade estrutural e padrão de clivagem

O gene AAU40775.1 de *Bacillus licheniformis* codifica uma proteína com 654 resíduos de aminoácidos, incluindo um peptídeo sinal de secreção constituído pelos primeiros 33 resíduos de aminoácidos. Essa enzima *BlCel9* é constituída por um domínio catalítico - CD - da família GH9 e um módulo de ligação à carboidratos da família 3c - CBM3c - ligados através de um *linker* rico em Thr/Ser. A enzima foi clonada correspondendo aos resíduos 34 a 654 e massa molecular de 69,9 kDa. A Figura 30 mostra o gel SDS-PAGE da proteína obtida após purificação, mostrando uma única banda com massa molecular aparente de aproximadamente 70 kDa.

Figura 30 – Expressão e purificação da enzima *BlCel9*. Etapas de produção e purificação por afinidade em Ni^{2+} para *BlCel9* são mostradas No gel desnaturante - (M) marcador de tamanho molecular, (1) proteína total, (2) fração insolúvel, (3) fração solúvel, (4) *BlCel9* clivada. O último passo de purificação foi realizado por SEC, como mostrado na banda 5 do gel SDS-PAGE e no cromatograma.

Fonte: Elaborada pelo autor.

As sequências de aminoácidos das enzimas da família GH9 anteriormente caracterizadas foram comparadas com *BlCel9*. Estas enzimas estão envolvidas na hidrólise de uma ampla gama de polissacarídeos, atuando como celulases, liquenases, laminarinases, beta-glicosidases e xiloglucanases. A atividade de *BlCel9* foi testada contra substratos celulósicos, incluindo celulose inchada com ácido fosfórico (PASC), papel filtro (FP), Avicel PH-101 e celulose bacteriana (BC). Sua atividade catalítica foi maior em PASC, seguida por BC, FP e bastante baixo em CMC (Figura 31A).

Figura 31 – Caracterização bioquímica. São mostradas as caracterização de especificidade por substrato (a), influências do pH (b) e da temperatura (c) na atividade enzimática. Efeitos de íons metálicos e agentes químicos na atividade enzimática (d).

Fonte: Elaborada pelo autor.

Além disso, o efeito de diferentes pHs e temperaturas na atividade de *BICel9* sobre PASC foi avaliado. Os pHs usados foram de 3,0 a 10,0 e as temperaturas medidas foram entre 30 a 70 °C. *BICel9* exibiu níveis de atividade maior que 80% dentro de uma ampla faixa de pHs (de 6,0 a 8,5), tendo 7,0 como pH ótimo (Figura 31B). No entanto, a atividade diminuiu para menos de 60% em pHs inferiores a 4,0 ou superiores a 8,5. *BICel9* exibiu a atividade mais alta a 60 °C (Figura 31C). Entre 40 e 50 °C, *BICel9* reteve mais de 80% de atividade, enquanto que a 70 ou 30 °C a atividade diminuiu para menos de 60% (Figura 31C).

Uma vez que muitas CAZymes mostraram mudanças em sua atividade catalítica em resposta a vários produtos químicos, a influência de íons metálicos e agentes químicos na atividade enzimática de *BICel9* foi investigada (Figura 31D). Para a maioria dos íons metálicos e agentes químicos testados, o aumento na atividade enzimática de *BICel9* não foi representativa, incluindo Ca^{2+} . Entretanto, na presença de DMSO, $CuCl_2$, e $FeCl_2$, a atividade de *BICel9* foi fortemente inibida. De acordo com os estudos anteriores realizados para outras enzimas desta família, a presença de Ca^{2+} tem impactado grandemente a estabilidade térmica e afinidade pelo ligante destas enzimas.^{189,190}

Para investigar a estabilidade estrutural de *BICel9* em diferentes pHs, estudos de estabilidade térmica foi conduzida usando a técnica *Thermoflour*. Para pHs próximo ao neutro entre 6,6 e 7,0, o Tm máximo foi de 77 °C (Figura 32A).

No entanto, em pHs entre 5,0 a 6,5, bem como 7,5 a 9,0, as Tm variam entre 68-76

Figura 32 – Estabilidade estrutural e padrão de clivagem. Estabilidade térmica avaliada em diferentes pHs (a). Efeito da temperatura na atividade (b). Perfil dos liberados (c), onde as quantidades de açúcar liberado foram normalizadas para valores relativos (%), (d).

Fonte: Elaborada pelo autor.

°C. Finalmente, para valores de pH acima de 9,0 e menores que 4,5, a T_m foi reduzida para aproximadamente 55 °C. Adicionalmente, a atividade residual de *BICel9* foi avaliada após pré-incubações nas temperaturas de 40, 50 e 60 °C por até 48 horas (Figura 32B). Os resultados mostram que *BICel9* é estável em temperaturas entre 40 a 50 °C por pelo menos 48 horas. A 50 °C, esta enzima foi capaz de reter mais de 80% de sua atividade hidrolítica após 24 horas de incubação. No entanto, a 60 °C sua atividade diminuiu significativamente para menos de 50% após 3 horas de incubação. Os valores ótimos de pH e temperatura obtidos para *BICel9* são similares àqueles determinados para outras enzimas da mesma família caracterizadas anteriormente. A atividade ótima de outros membros da família foi registrada principalmente dentro de uma faixa de pH de 6,0–7,5 e temperaturas entre 50–70 °C. ^{191–195}

Os produtos formados por *BICel9* após 1 hora e 48 horas de hidrólise em PASC foram analisados por HPAEC-PAD; *BICel9* (a 20 nM) foi incubada com PASC a 50 °C (Figura 32C). A identificação dos produtos gerados mostrou que essa enzima produz celotetraose (Glc₄), celotriose (Glc₃), celobiose (Glc₂) e glicose (Glc₁). Após 60 minutos de hidrólise, *BICel9* produz principalmente Glc₂ (46%), bem como uma pequena quantidade de Glc₃ (23%) e Glc₄ (17%), Figura 32D. Após 48 horas de hidrólise, o produto principal foi novamente Glc₂, que atingiu aproximadamente 59% de todos os oligossacáridos liberados. A presença de Glc₄ nos estágios iniciais da hidrólise de PASC (após 1 hora de reação) foi reduzido quatro vezes após 48 horas de hidrólise. Isto indica que Glc₄ é convertido em

Glc2.

Similarmente, *TfCel9A* foi capaz de produzir pequenas quantidades de Glc1 e Glc4).¹⁶⁹ *CcCel9A* de *C. cellulosi* também produz pequenas quantidades de Glc1 e Glc4 como produtos final.¹⁹⁶ Tomando como base a análise dos produtos liberados, foi proposto que a *TfCel9A* inicialmente produz principalmente a celotetraose (Glc4), que posteriormente é convertida em celotriose (Glc3), celobiose (Glc2) e glicose (Glc1).¹⁶⁹ No entanto, *BtCel9* produz mais Glc2 (que é indicativo da sua processividade) do que Glc3 enquanto que *TfCel9A* produz maiores quantidades de Glc3 do que Glc2. Um padrão semelhante foi observado para *CcCel9A*¹⁹⁶ e *CbCel9A*⁶⁸ de *Caldicellulosiruptor*. Em geral, Glc2 foi o produto hidrolítico dominante quando *BtCel9* atua em PASC. Isto demonstra que *BtCel9* é uma celulase endo-processiva típica.

5.1.2 Estudos estruturais por cristalografia de raios X

BtCel9 foi cristalizada na condição 100 mM de Tris no pH 8,5, 15% de PEG 20000 (w/v) a 6 mg/mL. Três cristais foram submetidos a experimentos de difração: 1) cristal apo, 2) cristal submetido a *soaking* com celotetraose e 3) cristal submetido a *soaking* com celotriose (Tabela 5). O cristal apo foi crio-protetido com 15% de etilenoglicol.

Tabela 5 – Estatísticas de coleta e processamento. Estatística da coleta, processamento dos dados de difração, determinação estrutural e refinamento.

	<i>BtCel9-apo</i>	<i>BtCel9-G3</i>	<i>BtCel9-G6</i>
Coleta e processamento			
Comprimento de onda (Å)			
Grupo espacial	C2	C2	C2
a, b, c (Å)	133,325 66,799 100,687	133,019 67,1297 100,76	133,513 67,6 100,674
α, β, γ (°)	90 124,72 90	90 124,98 90	90 124,838 90
Resolução (Å)	29,4 - 2,059 (2,133 - 2,059)	82,56 - 1,963 (2,033 - 1,963)	36,96 - 2,198 (2,277 - 2,198)
R _{meas}	0,1079 (0,7248)	0,09177 (0,4183)	0,1922 (0,6796)
CC _{1/2}	0,997 (0,82)	0,992 (0,763)	0,987 (0,75)
Reflexões totais	253633 (17443)	96222 (7934)	169628 (11471)
Reflexões totais únicas	44849 (4328)	48837 (4121)	37200 (3397)
Multiplicidade	5,7 (4,0)	2,0 (1,9)	4,6 (3,4)
Completeza (%)	0,99 (0,96)	0,93 (0,79)	0,99 (0,91)
$\langle I/\sigma(I) \rangle$	13,10 (1,83)	9,15 (2,82)	8,64 (2,32)
Fator B de Wilson	25,63	15,65	18,67
No.Mol. na cela	1	1	1
Refinamento			
R _{work}	0,1675	0,1666 (0,2158)	0,1682 (0,2989)
R _{free}	0,2019	0,1937	0,2119
Número de resíduos de aminoácidos	615	615	615
Desvios (R.m.s.)			
Comprimentos de ligação (Å)	0,003	0,005	0,008
Ângulos de ligação (°)	0,61	0,78	0,96
Ramachandran			
Favorável (%)	96	96	96
Permitido (%)	4,2	3,9	3,4
Outliers (%)	0,16	0,33	0,49
Fator B			
Médio	31,14	17,90	22,88
macromoléculas	30,49	16,63	21,65
solvente	38,75	28,82	30,56
$R_{meas} = \frac{\sum_{hkl} \sum_{i=1}^n (\sqrt{\frac{n}{n-1}}) (I_i(hkl) - \bar{I}(hkl))}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$			

Fonte – Elaborada pelo autor.

Um total de 720 imagens cobrindo uma faixa de oscilação de 360° foram coletadas com variação angular de 0,5°. Os conjuntos *Bl*Cel9G3 e *Bl*Cel9-G6 apresentaram parâmetros de cela similares à *Bl*Cel9-apo. Os resultados de coleta de dados, processamento, refinamento e validação estão sumarizados na [Tabela 5](#).

5.1.3 Modelo estrutural, domínios catalítico e acessório

O modelo cristalográfico da *Bl*Cel9 é formado por dois domínios ([Figura 33](#)): domínio catalítico (resíduos Glu3-Phe446) e módulo acessório de ligação à celulose da família 3c, nomeadamente CBM3c, correspondendo aos resíduos Tyr467-Pro617; esses dois domínios são separados por um *linker* rico em resíduos Ser/Pro/Thr, o qual corresponde aos resíduos Lys448 -Tyr466.

Figura 33 – Estrutura cristalográfica. O modelo estrutural de *Bl*Cel9 corresponde ao domínio catalítico com enovelamento do tipo barril com sítio catalítico em forma de fenda aberta. Adicionalmente, domínio catalítico é conetado ao ao módulo do CBM3c por um *linker* (colorido em vermelho).

Fonte: Elaborada pelo autor.

O domínio catalítico e CBM3c de *Bl*Cel9 são similares aos outros modelos estruturais de enzimas bimodulares apresentam baixo RMSD (átomos do C-alfa) com os modelos dos homólogos previamente descritos ([Figura 34](#)): *Cc*Cel9G de *C. cellulolyticum* (pdb id

1G87, $RMSD = 1,26 \text{ \AA}^{53}$), *TfCel9A* (anteriormente chamada E4) de *Thermobifida fusca* (pdb id 1JS4, $RMSD = 1,19 \text{ \AA}^{45}$), e *CtCel9I* de *Clostridium thermocellum* (pdb id 2XFG, $RMSD = 0,90 \text{ \AA}^{197}$).

Figura 34 – Comparação estrutural de *B/Cel9* com outras enzimas da família. Alinhamento estrutural de *B/Cel9* com outras GHs da família 9. Todas apresentam o sítio catalítico em fenda aberta alinhado com o CBM3c formando uma superfície plana.

Fonte: Elaborada pelo autor.

B/Cel9 tem domínio (Figura 33 e Figura 34) catalítico composto por 15 hélices-alfa conectadas por *loops* que, no geral, formam uma estrutura compacta com enovelamento do tipo barril com as repetições $(\alpha/\alpha)_6$. O bolsão catalítico é formado por mais de 110 resíduos de aminoácidos hidrofóbicos e aromáticos (dados não mostrados); dentre estes dois grupos de aminoácidos, a grande maioria deles é conservada em GHs da família 9, tais como *TfCel9A*, *CcCel9G*, e *CtCel9I*, por exemplo.

Classicamente, no domínio catalítico destas GHs da família 9 o sítio catalítico corresponde a uma superfície em fenda (Figura 35). Esse domínio é formado por *loops* que conectam as hélices do barril. Em outras estruturas cristalográfica bimodulares, como *TfCel9A*, *CcCel9G*, e *CtCel9I*, essa superfície é limitada em uma de suas bases pelo CBM3c e na outra, parte superior, por um *loop* (20 resíduos de aminoácidos, Figura 35). Na Figura 35 podemos observar que este *loop* é conservado nas estruturas cristalográficas de *TfCel9A*, *CcCel9G*, e *CtCel9I*. Entretanto, em *B/Cel9* está ausente. Isso sugere que este *loop* pode ter importante papel no modo de ação destas enzimas, como será discutido

em seguida.

Figura 35 – Topologia do sítio catalítico e mapa de conservação. Acima, a representação em superfície mostra a topologia do sítio catalítico em forma de fenda aberta. Os resíduos foram coloridos de acordo com o sistema de cor do programa ConSurf.¹⁴⁶ Abaixo é mostrado o alinhamento de sequência mostrando a região do *loop* (em triângulo vermelho) presente em *BlCel9*.

Fonte: Elaborada pelo autor.

Também são observados no sítio catalítico a tríade catalítica observada em outras famílias das hidrolases de glicosídeos,²⁰ aqui representada pelos resíduos Asp57, Asp60 e Glu426. A distância entre os O-carboxil dos resíduos Asp57 e Glu426 é de 9 Å, a qual é observada para outras enzimas (variando entre 9–9,5 Å) que promovem a inversão do carbono anamérico durante a catálise.^{198,199} Próximo à entrada do sítio catalítico é observado em íon cálcio octacoordenado por resíduos de aminoácidos localizados na fenda catalítica (Figura 36). Esse íon faz ligação de hidrogênio com os resíduos Ser213 (2,55 Å), Gly214 (2,42 Å), Asp217 (2,59 Å), Glu218 (bicoordenado 2,51 e 2,53 Å), Asp261 (2,43 Å) e com uma molécula de água H₂O-24 (2,55 Å).

Comparando com as outras estruturas similares descritas até o momento, *TfCel9A*, *CcCel9G*, e *CtCel9I*, esse íon é encontrado em condições equivalentes às observadas para *BlCel9*; em todos esses casos os resíduos na esfera de coordenação parecem ser estruturalmente equivalentes, exceto para *CtCel9I* (pdb id 2XFG), a qual apresenta resíduos e

Figura 36 – Ligante cálcio. Duas íons de cálcio são na estrutura de *BICel9*. Essas ligantes interagem com o CD (esquerda) e o CBM3c (direita).

Fonte: Elaborada pelo autor.

água idênticos a *BICel9* envolvidos na interação com Ca^{2+} . No geral, esse sítio para cálcio parece ser conservado na família GH9, como já observado anteriormente em outros trabalhos.^{45,53,197} Entretanto, quando avaliada a influência de cálcio na atividade enzimática de *BICel9*, não foi observado qualquer aumento; isso sugere que esse íon tem um papel mais estrutural do que funcional, como observado em outros trabalhos em que o cálcio tem importante melhora na termoesbilidade enzimática e aumento da afinidade do substrato dentro do sítio catalítico.^{189,190,200}

5.1.3.1 Domínio de ligação à celulose

Domínios de ligação à celulose da família 3 são largamente observados estando associados à GHs bacterianas.²⁰¹ O domínio de ligação à celulose da subfamília 3c, CBM3c, é composto por 9 folhas-beta antiparalelas formadas por 151 resíduos de aminoácidos, apresentando topologia do tipo beta-sandwich e um motivo tipo jelly roll, motivo este observado em estruturas de CBMs da família 3 de homólogos determinadas anteriormente.^{45,53,197,201}

As folhas 1, 2, 4, e 7 formam uma face quase planar rica em resíduos polares e carregados em sua superfície. Essa região planar do domínio CBM3c é perfeitamente alinhada com a cavidade que forma o bolsão catalítico do domínio catalítico, indicando que ambos podem constituir uma unidade funcional conjunta. Essa proposta foi explorada no trabalho de Sakon et al.⁴⁵ para *TfCel9A*,⁴⁵ e posteriormente enfatizado para *CcCel9G*, e *CtCel9I*; mais recentemente essa mesma proposta foi melhor explorada por Liberato e co-autores, onde foi mostrado que o sítio catalítico da trimodular *BICel5B* (composta por 2 CBMs e um domínio catalítico) é formado pelo ajuste induzido destes três domí-

nios durante a clivagem do substrato,³¹ os quais formam uma unidade funcional. No caso da família das GH9, foi proposto anteriormente que esta superfície do CBM3c interage diretamente com as fibras de celulose durante a ação enzimática.⁴⁵ Nesta interação o CBM3c teria a função de carrear cada fibra a ser clivada para o sítio catalítico. O alinhamento entre as estruturas *TfCel9A*, *CcCel9G*, *CtCel9I* e *BlCel9* apresenta uma topologia bem definida: a superfície plana do CBM3c perfeitamente alinhada com o bolsão do sítio catalítico.

Entretanto, acredita-se que nem todas as orientações entre a superfície planar e o domínio catalítico são permitidas. Na parte oposta do CBM3c, as folhas-beta 3, 5, 6, 8 e 9 formam uma face com uma cavidade rasa ([Figura 36](#)), já descrita para CBMs da família 3;²⁰¹ embora não se saiba ainda o papel funcional dessa região no CBM3c, as suposições feitas em outros trabalhos são que ela pode interagir com oligossacarídeos e até mesmo com regiões do *linker* peptídico que conecta o CBM ao CD.¹⁹⁷ Localizado entre a folha-beta 3 e os *loops* Asp507-Ala518 e Ser574-Asp587, um segundo íon cálcio ([Figura 36](#)) é observado octacoordenado pelos resíduos Asp507 (2,34 Å), Glu510 (2,30 Å), Asn583 (2,34 Å), Asn586 (2,33 Å), Asp587 (3,91 Å), e H₂O (2,36 Å); isso também é observado nas estruturas de *TfCel9A*, *CcCel9G*, e *CtCel9I*, as quais provavelmente também têm um papel mais estrutural do que funcional.²⁰¹

5.1.4 Estrutura cristalográfica de *BlCel9* em complexo com celotriose e celoexaose revela o modo de ligação nos subsítios -2 a -7

Para entender a natureza da interação entre enzima e substrato, os cristais da enzima *BlCel9* foram imergidos em solução crio contendo 15% de etilenoglicol junto com celo-oligossacarídeos com DP= 3 e DP=6, respectivamente para as estruturas *BlCel9*-G3 e *BlCel9*-G6 ([Figura 37](#)).

Através da superposição e comparação estrutural com o modelo de *TfCel9A* (PDB id 4tf4⁴⁵), em complexo com ligante (celotetraose + celobiose) posicionado entre os subsítios -5 a +2, é observado que *BlCel9* tem uma molécula de celotriose nas posições -5 a -3 (*BlCel9*-G3) e uma molécula de celoexaose nas posições -7 a -2 (*BlCel9*-G6). Para o complexo de *BlCel9* com celotriose, os mapas de Fourier diferença (2F_c-F_o e F_c-F_o) mostram uma clara e bem definida densidade eletrônica para este ligante em todos os subsítios ocupados por ele. Por outro lado, para a celoexaose, a densidade eletrônica para a glicose na posição -7 é levemente fraca. No entanto, após alguns ciclos de refinamento desta estrutura a densidade eletrônica na posição -7 foi ficando muito mais clara quanto ao posicionamento do ligante nos subsítios -7 a -2, especialmente no -7. Para os dois ligantes (DP=3 e DP=6), a densidade eletrônica mostra que não houve clivagem destes substratos pela enzima.

Comparativamente, tanto celotriose quanto celoexaose são estabilizadas no bolsão

Figura 37 – Ligantes envolvidos na interação com *BICel9*. Resíduos envolvidos na interação com celotriose (superior) e celohexaose (inferior). São mostrados as principais ligações de hidrogênio dos celo-oligossacarídeos posicionados nos subsítios -5 a -3 (celotriose) e -7 a -2 (celohexaose).

Fonte: Elaborada pelo autor.

catalítico, principalmente, por ligações de hidrogênio, por interações hidrofóbicas (Thr e Tyr) e por interações com resíduos carregados (Asp e Arg); isso permite que o ligante tenha maior estabilidade no sítio ativo, principalmente nos subsítios -5 a -2. Analisando o posicionamento destes ligantes e fazendo a comparação com outras estruturas da família GH9 em complexo com oligossacarídeos de celulose, duas observações se destacam: i) a localização de G6, e mapeamento de dois novos subsítios (-7 e -6), que só é permitida por conta da ausência da *loop* (20 resíduos de aminoácidos) na saída do sítio catalítico de *BICel9*, mas é presente nas estruturas cristalográficas *TfCel9A*, *CcCel9G*, e *CtCel9I*; ii) o ambiente das interações, sejam elas ligações de hidrogênio ou interações com resíduos

aromáticos, parece permitir certa preferência por substratos mais curtos nas posições -5 a -3, como é o caso da celotriose em *BlCel9*. Essas observações são importantes para o entendimento do modo de ação destas enzimas e perfil de produtos liberados durante a hidrólise enzimática. Isso será discutido mais adiante.

Tanto para *BlCel9* em complexo com celotriose, quanto formando complexo com celooxaose, nos subsítios -3 a -5 são observadas as mesmas interações envolvendo quatro resíduos aromáticos (Tyr e Trp) e dois resíduos carregados (Asp), os quais são altamente conservados na família das GH9. Para este último complexo, as unidades de glicose permitem mapear as interações correspondente aos subsítios -7 a -2. A [Figura 37](#) mostra a representação esquemática das principais interações entre os resíduos de aminoácidos e os anéis piranosídicos do substrato nas posições -7 a -2 do sítio catalítico de *BlCel9*.

No subsítio -2 a Glc^{-2} interage via ligações de hidrogênio com os resíduos Arg321 (O2 e O3) e Tyr322 (O3); ligações de hidrogênio mediadas por moléculas de água são observadas para os resíduos Tyr209 e Trp212. Entretanto, também é observada uma interação hidrofóbica (edge-to-face ou conformação-T) entre o Trp260 e o anel piranosídico da Glc^{-2} com distância de aproximadamente 3,5 Å.

Glc^{-3} faz ligação de hidrogênio com Tyr232 (O6), Trp256 (O6) e Ser259 (O6); e, além disso, faz ligações de hidrogênio mediadas por água com o Glu315 (O2 e O3). Já Glc^{-4} interage com Asp261 (O2) e Asp262 (O3) via ligação de hidrogênio. A orientação dos Trp nos subsítios -3 e -4 permite interações de empilhamento com os anéis glicosídicos do ligante; considerando essas interações, Glc^{-3} faz interações de empilhamento com Trp212, enquanto Glc^{-4} interage com Trp256. Entretanto, Trp256 está levemente deslocado da sua posição de empilhamento. Essas interações nas posições -3 e -4 sugerem que a orientação do ligante é altamente estável. Embora os subsítios -3 e -4 ofereçam uma grande rede de interações da proteína com o ligante, Glc^{-5} e Glc^{-6} interagem somente com moléculas de água. Por outro lado, no subsítio -7 a Glc^{-7} interage via interação de hidrogênio com os resíduos Arg 255 (O6) e Asn 251(O2), enquanto a Tyr254 interage mais fracamente com Glc^{-7} .

Essa diversidade de interações (empilhamento/hidrofóbicas, ligações de hidrogênio) parece produzir maior estabilidade do ligante em certas regiões do sítio ativo desta enzima, especialmente nas posições em que o ligante interage com os resíduos aromáticos e carregados, o que provavelmente aumenta a ligação dos oligossacarídeos lá; adicionalmente, essas interações podem desempenhar um papel importante no modo de ação, acomodação e adsorção da fibra de celulose durante a hidrólise. Além disso, a ausência do *exo loop* nas posições -6 e -7 parece ser o único elemento estrutural restritivo para a acomodação do substrato na fenda catalítica de outras enzimas homólogas a *BlCel9*. Para isso, uma comparação estrutural entre essas enzimas e *BlCel9* será apresentada a seguir.

5.1.5 Subsítios envolvidos nas interações substrato-proteína na família GH9: relacionando estrutura e função para *BtCel9*

Os modelos estruturais das enzimas da família GH9 determinados até o momento têm o domínio catalítico e/ou domínio catalítico (CD) mais um domínio acessório cristalizados na forma apo ou na presença de ligantes. Comparando os CDs destas estruturas cristalográficas, observa-se que : i) os que apresentam ação endoglucanásica em celulose têm a fenda catalítica mais aberta, ao passo que ii) os com atividade endo/exo apresentam o bolsão catalítico menos aberto e mais restritivo.

Considerando isso, a *CtCbhA* tem sua fenda muito mais estreita do que as outras enzimas até então reportadas no PDB e, além disso, o seu sítio catalítico está restrito somente às posições -2 a +2, tendo ação exclusivamente de uma celobioidrolase.⁵⁴ Essa arquitetura do sítio catalítico destas enzimas se deve à presença de um *loop* que bloqueia o acesso da fibra de celulose às posições posteriores ao subsítio -2, o que impede a hidrólise de substratos de cadeias longas.

Considerando a estrutura das outras enzimas multimodulares desta família (*CcCel9G*, *TfCel9A* e *CtCel9I*), o *exo loop* localizado na posição não redutora no sítio ativo destas enzimas bloqueia o acesso da fibra de celulose às posições -5 a -7 observadas em *BtCel9* (Figura 38); isso provavelmente contribui para a formação de produtos menores ou iguais a celotetraose durante a hidrólise enzimática destas enzimas.

Figura 38 – Exo-loop e permissividade no subsítio -7. É mostrado que o *loop* *BtCel9* está envolvido na permissividade de celo-oligos nas posições -5 a -7.

Fonte: Elaborada pelo autor.

Do ponto de vista estrutural, evolutivo e funcional, a presença de *loops*, em grande parte dos casos, compondo e estruturando o sítio catalítico destas enzimas, é o que provavelmente garante a diversidade de ação dentro da superfamília das GHs, como observado para *TrCel7A* de *Trichoderma reesei*²⁰² e *TfCel6A* de *Thermobifida fusca*.²⁰³ Similarmente, a presença de um *loop* obstruindo a parte final dos subsítios relativos aos produtos em *TfCel6B*¹⁸⁵ e *GeoGH74*,¹⁵¹ define a permissividade de substratos e o modo de ação destas enzimas.

As enzimas desta família caracterizadas estruturalmente, tanto na forma apo quanto formando complexo com celo-oligossacarídeos, interagem com o anel piranosídico em cada subsítio. Todas apresentam pelos menos os subsítios -4 a +2, exceto para *CtCbhA*⁵⁴ com apenas 4 subsítios observados (-2 a +2); esses subsítios são numerados a partir do terminal redutor até o terminal não redutor do substrato, em acordo com a nomenclatura de Davies, Wilson e Henrissat.²⁰⁴ O posicionamento e a localização dos ligantes nestas estruturas, permitem identificar, comparativamente, os subsítios que não foram mapeados na estrutura cristalográfica em *BlCel9*, bem como os resíduos de aminoácidos envolvidos na interação proteína-substrato; e, além disso, a preferência dos subsítios em alocar sacarídeos com certo grau de polimerização em detrimento de outros. Através desta comparação, pode-se observar que *BlCel9* tem um longo bolsão catalítico abrangendo os subsítios -7 a -1 e +1 a +2, dos quais os resíduos nas posições -2 a +1 são estritamente conservados (Figura 39). Acerca destes resíduos, os que podem interagir com ligantes são: Trp130 (+2); His127, His379, Glu426 e Arg381 (+1); Asp60, Asp57, Ala58, Phe64 e Glu426 (-1); Tyr2019, Trp360, Arg321, Asp430, Trp317 e Tyr322 (-2).

Os resíduos nos subsítios -3 e -4 possuem baixa conservação quando consideramos as posições +1 a -2. Considerando os resíduos que não são altamente conservados neste alinhamento de 2077 sequências únicas, os resíduos Tyr322 (-3), Asp261 (-3) e Val263 (-4) interagem com o ligante nos subsítios -3 e -4; o Trp212 (no subsítio -3) e Trp256 (-4) preservam a mesma conformação e interação com o ligante que as observadas para as bimodulares *TfCel9A* (PDB id 4TF4), *CtCel9I* (PDB id 2XFG) e *CcCel9G* (PDB id 1GA2), exceto em *CcCel9G* em que o Trp na posição -3 é substituído por uma Ser.

Enquanto essas enzimas preservam grande similaridade estrutural, forma e interação com os ligantes entre os subsítios -2 a +1, particularmente para *AaCel9A*,²⁰⁵ *CbCel9A*,⁶⁹ *CcCel9G*⁵³ e *TfCel9A*,⁴⁵ os subsítios -2 a +1 formam uma extensa e coesa rede de interações entre proteína e ligante que contribuem fortemente para a manutenção do ligante lá; por outro lado, os resíduos nos subsítios -3 a -4 são pouco conservados.

Resíduos aromáticos que fazem interações de empilhamento com celo-oligossacarídeos são altamente conservados no sítio catalítico de GHs da família 9 (Figura 39); em *TfCel9A* (PDB id 4TF4), *CtCel9I* (PDB id 2XFG) e *CcCel9G* (PDB id 1GA2) esses resíduos são conservados nas posições +1 (Tyr e His), -2 (Trp), -3 (Trp) e -4 (Trp). Em *BlCel9* esses

Figura 39 – Mapa de conservação da sequência e sub-sítios envolvidos na interação com o ligante. É mostrado o posicionamento de celobiose no sítio catalítico *BtCel9*. Os resíduos foram coloridos de acordo com o sistema de cor do programa ConSurf.¹⁴⁶

Fonte: Elaborada pelo autor.

resíduos também são conservados: Trp256 (-4), Trp212 (-3), Trp317 (-2), Tyr422 (+1) e His127 (+1). Em todas essas estruturas as conformações dos resíduos nas posições -4/-3 são similares, enquanto nas posições -2 e +1 os resíduos estão deslocados de suas posições de empilhamento quando na ausência de ligantes; em *BtCel9*-G6 o Trp317 está deslocado de sua posição de empilhamento, assim como o Trp equivalente em *CtCel9I* (PDB id 2XFG), *TfCel9A* (PDB id 1JS4, com subsítio -2 não ocupado) e *CcCel9G* (PDB id 1GA2).

Por outro lado, em *TfCel9A* (PDB id 4TF4, com subsítios -1 e -2 ocupados) o Trp faz uma interação de empilhamento clássica. Já em *TfCel9A* (PDB id 1JS4 e 4TF4) ambos os resíduos Tyr420 e His125 fazem interação de empilhamento com o resíduo de glicose na posição +1 (um em cada face do anel piranosídico) enquanto em *CtCel9I* (PDB id 2XFG), *TfCel9A* (PDB id 1JS4), *CcCel9G* (PDB id 1GA2) e *BtCel9*-G6 (Tyr422) estão deslocados da posição de empilhamento. Em *BtCel9*-G6 essa diferença é mais pronunciada porque a Tyr422 é deslocada de aproximadamente 90° da posição de empilhamento observada em *TfCel9A* (PDB id 4TF4) e passa a interagir com o resíduo catalítico Glu426 com distância de 2,8 Å, enquanto que na posição de empilhamento essa distância seria de 6,5 Å (Figura 40).

Essas observações indicam que os resíduos de aminoácidos nas posições -2 e +1

Figura 40 – Subsítios e suas interações co resíduos aromáticos. São mostrados os resíduos aromáticos que fazem interações de empilhamento com celo-oligossacarídeos altamente conservados nos subsítios -2 a +2.

Fonte: Elaborada pelo autor.

em *B/Cel9* passam por ajustes conformacionais induzidos pela presença do ligante nestes subsítios. Nessa família de enzimas a topologia do sítio ativo e os resíduos de aminoácidos formam uma plataforma constituída por ligações de hidrogênio e interações de empilhamento com os oligossacarídeos de celulose. Esse tipo de ambiente ajuda a direcionar a fibra de celulose no bolsão catalítico afim de a enzima continuar a hidrólise sem ser adsorvida de dentro do sítio ativo. Juntos, esses resultados sugerem que a ação de *B/Cel9* em seu substrato é direcionada principalmente por ligações de hidrogênio e interações hidrofóbicas. Primeiramente, essas interações se dão entre a superfície planar formada pelo CD e CB3c, seguida pelo ajuste do substrato no interior do sítio catalítico para posterior clivagem. Por fim, tem-se o ciclo processivo até que a enzima seja adsorvida da fibra de celulose. Isso implica que G4 é gerado como produto inicial da hidrólise, e também que para substratos com alto grau de polimerização GP e grande área de adsorção, essa enzima terá menor taxa hidrolítica às custas do ganho na processividade. Além disso, parece que a afinidade nos subsítios - causada principalmente por resíduos aromáticos e interações de hidrogênio - é um requisito para a enzima prolongar o tempo em que se mantém adsorvida na fibra. Isso pode estar relacionado ao balanço entre deslizamento e desassociação da fibra no sítio catalítico, e por conseguinte, com a taxa de processividade.

5.1.6 Estudos estruturais por espalhamento de raios X a baixo ângulo

Como mencionado anteriormente, a orientação espacial entre CD e CBM3c nas estruturas de *CtCel9I*, *TfCel9A*, *CcCel9G* e *BtCel9* é muito similar. Um dos pontos marcantes na orientação entre esses dois domínios é a formação da superfície planar que, a priori, é usada como uma extensão do sítio ativo durante a adsorção e hidrólise da cadeia de celulose por essas enzimas. Entretanto, no empacotamento cristalográfico podem ocorrer mudanças conformacionais induzidas pelo efeito da cristalização. Por essa razão, os estudos em solução destas enzimas vão permitir obter mais informações de quais são as conformações e orientações preferenciais em solução adotadas por *BtCel9A*. Para endoglucanases bacterianas processivas, o CBM associado ao CD desempenha um papel importante na atividade e processividade. Aplicamos SAXS aqui para reconstruir o envelope molecular de *BtCel9* e estudar sua flexibilidade em solução. Os parâmetros derivados da coleta, processamento e análise são mostrados na [Tabela 6](#).

Tabela 6 – Parâmetros gerais. Os números entre parênteses são a discrepância obtida a partir dos valores esperados para os pesos moleculares esperados de proteína calculados usando o software SAXSMoW 2.0 (<http://saxs.ifsc.usp.br/>). # Resolução calculada pelo SASRES.

Amostras	Espaço recíproco	Espaceo real		Massa, kDa	χ^2 –parâmetro	Resolução
	R_g , Å	R_g , Å	D_{max} , Å	Calculado	Ótimo χ^2	resolução, Å
<i>BtGH9</i>	$29,1 \pm 0,7$	30,2	101	$78,9 \pm 0,1$	0,7	30 ± 2 Å
2xfg	29,6	30,1	106,2	ND	1,3	ND
1g87	29,3	29,8	104,7	ND	1,5	ND
1js4	28,9	29,1	102,6	ND	2,1	ND
3gzk	22,9	23,3	79,9	ND	3,7	ND
1ut9	23,3	23,8	79,3	ND	4,1	ND
5dgq	23,1	23,7	76,1	ND	5,2	ND

Fonte – Elaborada pelo autor.

De acordo com a análise dos dados, efeitos dependentes de agregação, interações inter-partículas e variação do R_g a baixo q não foram observados na intensidade de espalhamento medidas para *BtCel9* (dados não mostrados). Na aproximação de Guinier, a região de q inicial das curvas de espalhamento exibiu um comportamento linear característico para sistemas monodispersos (dados não mostrados). O raio de giro calculado com base na aproximação de Guinier foi de 29,3 Å. Esses achados também são apoiados por estimativas da massa molecular determinadas a partir dos dados de SAXS. A massa molecular calculada a partir da intensidade de espalhamento foi de aproximadamente 65,8 kDa com um valor de discrepância de apenas 5,4%. A estreita concordância da massa molecular determinada e esperada com base na sequência de aminoácidos (69,6 kDa) indica que *BtCel9* forma monômero em solução.

A função de distribuição de distâncias, $P(r)$, calculada a partir dos dados de espalhamento é característica de uma partícula alongada com uma distância máxima de 101

Å (Figura 41, Tabela 6). O R_g determinado com base na $P(r)$ foi de aproximadamente 29,3 Å, muito próximo do R_g obtido na análise de Guinier. A distribuição da função interatômica é característica de partículas que possuem duas subunidades espacialmente bem separadas exibindo dois máximos correspondentes à separação entre elas. Essa $P(r)$

Figura 41 – Funções $P(r)$ determinadas a partir dos dados experimentais de espalhamento SAXS, e da curva de espalhamento calculada para os modelos cristalográficos de *CtCbhA* (PDB id 1UT9), *PpGlcNase* (PDB ID 5DGQ), *AaCel9A* (PDB ID 3GZK), *CcCel9G* (PDB id 1G87), *TfCel9A* (PDB id 1JS4) e *CtCel9I* (PDB id 1G87).

Fonte: Elaborada pelo autor.

é bastante semelhante às $P(r)$ reconstruídas a partir das estruturas cristalográficas já conhecidas para outras enzimas bi modulares tais como *CcCel9G* de *Clostridium cellu-lolyticum*,⁵³ *TfCel9A* de *Thermobifida fusca*⁴⁵ e *CtCel9I* de *Clostridium thermocellum*.¹⁹⁷ Em todos os casos, os parâmetros gerais derivados das análises de Guinier e da $P(r)$ estão em acordo. Parâmetros estruturais como R_g e D_{max} são mostrados na Tabela 6.

O envelope molecular de *BICel9* foi reconstruído usando os programas DAMMIN e GASBOR. Nove envelopes foram gerados usando os dois algoritmos. O envelope escolhido em cada caso foi selecionado baseado no menor valor de χ (Tabela 6, Figura 42 e Figura 43). A forma alongada de *BICel9*, semelhante a de um girino, é consistente com a sua respectiva função $P(r)$. Cada envelope contém uma região globular, que corresponde ao CD da família GH9, e uma segunda região correspondente ao CBM 3c, ambos conectados pelo *linker*. A estrutura cristalográfica de *BICel9* foi superposta ao envelope reconstruído por SAXS (Figura 42) com χ de 1,1 entre a intensidade experimental de SAXS e a gerada a

partir do modelo de alta resolução. No modelo *ab initio* a distância da separação entre o CD e o CBM 3c é muita pequena, assemelhando-se à conformação adotada pelas estruturas de *CcCel9G* e *TfCel9A*.

Figura 42 – Sobreposição do envelope SAXS de *BICel9* com o modelo cristalográfico. Vista frontal e lateral do envelope molecular de *BICel9* (superfície semitransparente cinza) das medidas de SAXS e o modelo cristalográfico são mostrados (ii, o envelope molecular foi girado em 180° em torno do eixo x).

Fonte: Elaborada pelo autor.

5.1.6.1 *BICel9* é flexível em solução

Uma vez que *BICel9* é constituída por um CD ligado no seu terminal N a um CBM3c através do *linker*, é esperado algum grau significativo de flexibilidade para esta enzima. O gráfico Kratky adimensional foi útil para a avaliação qualitativa desta flexibilidade em solução.^{74,104,206–208} Nessa análise, o gráfico de Kratky é independente do tamanho e massa molecular da molécula; neste caso, para uma molécula bem enovelada e com uma forma bem definida, os limites $q \cdot R_g$ tende a $\sqrt{3}$ e $(qR_g)^2 I(q)/I(0)$ tende a 1,104.^{108,208} Os desvios desses limites indicam flexibilidade ou instabilidade conformacional.^{108,208}

De fato, a análise do Kratky adimensional das intensidades experimentais de SAXS sugere que *BICel9* é uma partícula compacta que exhibe uma forma bem definida em solução (Figura 43).

Figura 43 – Análise de flexibilidade no pH ótimo. Intensidade experimental de SAXS (a) de *BICel9* em solução e gráfico de Kratky (b)

Fonte: Elaborada pelo autor.

No entanto, no gráfico de Kratky adimensional é conservado um máximo distinto em 2,1 nas coordenadas $R.q$, que é ligeiramente deslocado do valor ideal de $q.R_g = 1,73$. Assim, levando em conta que o *linker* de *BICel9* é relativamente longo (20 resíduos), *BICel9* tem certo grau de flexibilidade em solução que pode ser imprimido pela dinâmica de interações entre os domínios CD e CBM3c. Essa flexibilidade observada pode facilitar o reconhecimento da face formada pelo CD-CBM3c e a superfície da fibra de celulose durante o evento catalítico.

A Figura 44 mostra a comparação entre o envelope de *BICel9* e as estruturas cristalográficas de *TfCel9A* (PDB id 1JS4) e *AaCel9A* de *Alicyclobacillus acidocaldarius* (PDB id 3GZK), sugerindo que os domínios CD e CMB de *BICel9* numa conformação simi-

lar às observadas nos modelos cristalográficos tanto de *B*lCel9 quanto de outras enzimas bimodulares.

Figura 44 – Comparação da forma de *B*lCel9 com a estrutura de outras GHs da família. As estruturas das duas endoglucanases Cel9A modulares de *Alicyclobacillus acidocaldarius* (PDB id 3GZK) e da E4 endo-processiva de *Thermomonospora fusca* (PDB id 1JS4) são comparadas com o modelo SAXS derivado de *B*lCel9.

Fonte: Elaborada pelo autor.

Para esta arquitetura, foi proposto que o CD e CBM3c formam uma superfície cooperativa (os módulos ficam justapostos) para a correta acomodação e reconhecimento da fibra de celulose.^{197,209,210} Coletivamente, nossas análises de SAXS para *B*lCel9 suportam a noção de que *B*lCel9 tem uma forma alongada e bem definida em solução, assemelhando-se muito a outras GH9.

5.1.6.2 Estudos de flexibilidade induzida por pH e temperatura

Chama atenção aqui o ajuste *B*lCel9 com o modelo cristalográfico obtido aqui, indicando que essa enzima mantém em solução um empacotamento muito similar ao cristalográfico. É conhecido que para sistemas monodispersos compostos de partículas flexíveis, o gráfico de Kratky não pode ter a elevação da linha de base a alto q . *B*lCel9 não apresenta uma alta elevação da linha de base quando avaliado somente o pH 7 (Figura 45).

Figura 45 – Flexibilidade de *BICel9* avaliada em condições de variação da temperatura e pH. No gráfico de Kratky é possível observar que *BICel9* apresenta aumento na flexibilidade para pHs mais básicos (A). Entretanto, aumento na taxa de flexibilidade parece ter maior dependência da temperatura, como observado em B e C.

Fonte: Elaborada pelo autor.

Entretanto, quando comparado os pHs 5 e 9, há mudanças no perfil da linha de base no gráfico de Kratky. Estas mudanças foram agrupadas em dois grupos levando em conta o fato de a linha de base da curva no gráfico de Kratky a alto q . Nos pHs 5 a 7, o gráfico de Kratky é bem parecido (grupo 1), tendo a linha de base a alto q quase constante e R_g próximos. Já os pHs 8 e 9 são bem parecidos (grupo 2) no perfil do gráfico de Kratky e os valores dos R_g próximos, com a linha de base apresentando uma elevação (Figura 45). Focando nas análises estes dois grupos, é possível observar que *BICel9* apresenta maior elevação de linha de base no pH9 do que no pH 7 (Figura 45A).

Levando em conta que no pH7 *BICel9* tem maior atividade e maior estabilidade térmica, procuramos entender quais seriam os efeitos causados pelo aumento da temperatura sobre a forma e envelope. Quando avaliadas estas mudanças induzidas por temperatura, vê-se que entre 20 a 40 °C quase não há mudanças no R_g , Figura 45A-B. Entretanto, de

50 a 60 °C são observadas mudanças no R_g (Figura 45B), indicando mudanças conformacionais desta molécula pelo efeito da temperatura. Por outro lado, o fato interessante aqui está relacionado à mobilidade entre os domínios catalítico e CBM3c nas temperaturas de 30 e 40 °C (Figura 45C).

Tomando em conjunto, esses resultados sugerem que a temperatura influencia o aumento da flexibilidade enzimática. Inicialmente induzem pequenas movimentações entre os domínios e com o aumento da temperatura essas mudanças são mais globais. Provavelmente a flexibilidade induzida tanto por pH quanto por temperatura provoque maiores movimentações no CBM3c do que no domínio catalítico de tal modo que o CBM3c se ajuste para manter a planaridade entre a superfície planar e o bolsão catalítico. Essa hipótese é corroborada pelos estudos dinâmica molecular da *TfCel9A*,²⁰⁹ onde foi demonstrado que a temperatura influencia a plasticidade e induz o aumento da coletividade, aproximando os domínios. Essa aproximação é crucial para a atividade enzimática. O CBM3c, seria então, um modulador da função do domínio catalítico, mas a maneira que isto acontece não é conhecido. Mais recentemente novos achados para o papel dos CBMs estão sendo observados, como para *BtCel5B*,³¹ em que o CBM46 funciona cooperativamente com o domínio catalítico para realizar a hidrólise.

6 CONCLUSÕES E PERSPECTIVAS

Os recentes avanços da informação genômica sobre microrganismos capazes de hidrolisar a biomassa lignocelulósica indicam que uma grande diversidade no modo de ação e arquitetura das Hidrolases de Glicosídeos (GHs). GHs com estrutura resolvida representam menos de 2% de todas as já conhecidas. Ao lado disso, questões como eficiência enzimática, especificidade por substrato e sensibilidade a inibidores são questões ainda abertas nos estudos dessas enzimas. A busca por novas enzimas ignocelulósicas é, de fato, um dos grandes desafios na rota tecnológica para a produção de etanol de segunda geração e produção de produtos de valor comercial. O objetivo deste trabalho foi realizar os estudos estrutural e funcional das hidrolases de glicosídeos *BleCel9*, *BleCel48* e *XcGH74*.

Foi observado que *XcGH74* é altamente específica para XyG, liberando como oligos do XyG, principalmente XX e XG. Esse padrão de clivagem de XyG indica que esta enzima é capaz de agir tanto na quebra de XyGs longos quanto de XyGs curtos. Para melhor entender como se dá o reconhecimento e especificidade desta enzima pelo XyG, a estrutura cristalográfica desta enzima foi determinada na presença do produto da hidrólise (XG). Viu-se que *XcGH74* cliva XyG preferencialmente entre motivos X-X; no entanto, não hidrolisa entre os motivos L-L. A razão para isso é justamente a falta de permissividade no subsítio -1 para acomodar a ramificação L; por outro lado uma ramificação de Xyl nessa posição é um pré-requisito para o reconhecimento do substrato.

Além disso, as duas celulasas *BleCel48* e *BleCel9* mostraram maior atividade PASC e BC, liberando celobiose como principal produto. *BleCel48* é altamente processiva em BC. Além disso, essa enzima parece ter certa especificidade no reconhecimento de celooligosacarídeo e também possivelmente pode realizar endo iniciação durante a hidrólise de PASC. Estruturalmente, tentou-se correlacionar estas observações com a topologia do sítio catalítico, bem como os resíduos de aminoácidos que definem o modo de ação desta enzima. Por fim, a enzima *BleCel9* foi caracterizada como uma endoglucanase. É composta por dois domínios, um catalítico e outro acessório, justapostos em uma conformação estendida e formam uma estrutura relativamente rígida em solução, agindo cooperativamente no reconhecimento da celulose. Junto a isso, os resultados de SAXS evidenciam que esta conformação parece ser a preferencial em solução.

Atualmente, estudos de dinâmica molecular estão sendo realizados para estas duas últimas enzimas para testar algumas hipóteses sobre o modo de ação, processividade e função dos resíduos aromáticos no reconhecimento da celulose, bem como de flexibilidade tomando como partida os resultados de SAXS.

Neste tempo de pesquisa durante o doutorado, o presente candidato publicou os

seguinte artigos (ver anexos):

- Evandro Ares de Araújo, Mário de Oliveira Neto and Igor Polikarpov. The two-domain enzyme *BlCel9* from *Bacillus licheniformis* is active against cellulosic substrates and has an elongated molecular shape: functional characterization and low-resolution SAXS structure. *Applied Microbiology and Biotechnology*.
- EA de Araújo, LR Manzone, V Piiadov, MAS Kadowaki, I Polikarpov. Biochemical characterization, low-resolution SAXS structure and an enzymatic cleavage pattern of *BlCel48* from *Bacillus licheniformis*. *International journal of biological macromolecules*, 2018.
- V.Piiadov, E.Ares de Araújo, M.Oliveira Neto, A.F.Craievich and I.Polikarpov. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data in relative scale. *Protein Science*, 2018.
- DE Evangelista, EA de Araújo, MO Neto, MAS Kadowaki, I Polikarpov. Biochemical characterization and low-resolution SAXS structure of an exo-polygalacturonase from *Bacillus licheniformis*. *New biotechnology*, 2017.
- Marcelo V Liberato, Rodrigo L Silveira, Érica T Prates, Evandro A De Araujo, Vanessa OA Pellegrini, Cesar M Camilo, Marco A Kadowaki, Mario de O Neto, Alexander Popov, Munir S Skaf, Igor Polikarpov. Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Scientific reports*, 2016.
- Hevila Brognaro, Vitor Medeiros Almeida, Evandro Ares de Araujo, Vasily Piyadov, Maria Auxiliadora Morim Santos, Sandro Roberto Marana, Igor Polikarpov. Biochemical Characterization and Low-Resolution SAXS Molecular Envelope of GH1 beta-Glycosidase from *Saccharophagus degradans*. *Molecular biotechnology*, 2016.
- F Segato, GL Berto, E Ares de Araújo, JR Muniz, I Polikarpov. Expression, purification, crystallization and preliminary X-ray diffraction analysis of *Aspergillus terreus* endo-beta-1, 4-glucanase from glycoside hydrolase family 12. *Acta Crystallographica Section F*, 2014.

Manuscritos em fase de escrita:

- Structural insights into xyloglucan specificity of a novel endo-xyloglucanase *XcGH74* from *Xanthomonas campestris*: molecular basis of unusual preference for xylosyl decoration at the subsite -1.
- Structural, functional and dynamics relationship of *BlCel48* exo-glucanase from *Bacillus licheniformis*.

- Structural and functional studies of a multimodular endoglucanase *B/Ce9*: processive degradation of cellulose is governed by collective interdomain motions between CBM and CD.

REFERÊNCIAS

- 1 BAR-ON, Y. M.; PHILLIPS, R.; MILO, R. The biomass distribution on earth. **Proceedings of the National Academy of Sciences**, v. 115, n. 25, p. 6506–6511, 2018.
- 2 LYND, L. R. et al. Cellulosic ethanol: status and innovation. **Current Opinion in Biotechnology**, v. 45, p. 202–211, 2017.
- 3 RAGAUSKAS, A. J. et al. The path forward for biofuels and biomaterials. **Science**, v. 311, n. 5760, p. 484–489, 2006.
- 4 CHU, S.; MAJUMDAR, A. Opportunities and challenges for a sustainable energy future. **nature**, v. 488, n. 7411, p. 294, 2012.
- 5 GOLDEMBERG, J. Ethanol for a sustainable energy future. **Science**, v. 315, n. 5813, p. 808–810, 2007.
- 6 BEZERRA, T. L.; RAGAUSKAS, A. J. A review of sugarcane bagasse for second-generation bioethanol and biopower production. **Biofuels, Bioproducts and Biorefining**, v. 10, n. 5, p. 634–647, 2016.
- 7 SOUZA, A. P. de et al. Sugarcane as a bioenergy source: history, performance, and perspectives for second-generation bioethanol. **BioEnergy Research**, v. 7, n. 1, p. 24–35, 2014.
- 8 COSGROVE, D. J. Re-constructing our models of cellulose and primary cell wall assembly. **Current Opinion in Plant Biology**, v. 22, p. 122–131, 2014.
- 9 MARRIOTT, P. E.; GÓMEZ, L. D.; MCQUEEN-MASON, S. J. Unlocking the potential of lignocellulosic biomass through plant science. **New Phytologist**, v. 209, n. 4, p. 1366–1381, 2016.
- 10 LIMA, M. A. et al. Evaluating the composition and processing potential of novel sources of brazilian biomass for sustainable biorenewables production. **Biotechnology for Biofuels**, v. 7, n. 1, p. 10, 2014.
- 11 HIMMEL, M. E. et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. **Science**, v. 315, n. 5813, p. 804–807, 2007.
- 12 RASMUSSEN, H. et al. New degradation compounds from lignocellulosic biomass pretreatment: routes for formation of potent oligophenolic enzyme inhibitors. **Green Chemistry**, v. 19, n. 2, p. 464–473, 2017.
- 13 ALESSI, A. M. et al. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. **Biotechnology for Biofuels**, v. 11, n. 1, p. 166, 2018.
- 14 MORAES, E. C. et al. Lignolytic-consortium omics analyses reveal novel genomes and pathways involved in lignin modification and valorization. **Biotechnology for Biofuels**, v. 11, n. 1, p. 75, 2018.

- 15 PAYNE, C. M. et al. Fungal cellulases. **Chemical reviews**, v. 115, n. 3, p. 1308–1448, 2015.
- 16 MEDIE, F. M. et al. Genome analyses highlight the different biological roles of cellulases. **Nature Reviews Microbiology**, v. 10, n. 3, p. 227, 2012.
- 17 ARTZI, L.; BAYER, E. A.; MORAÏS, S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. **Nature Reviews Microbiology**, v. 15, n. 2, p. 83, 2017.
- 18 STEWART, R. D. et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. **Nature Communications**, v. 9, n. 1, p. 870, 2018.
- 19 KAOUTARI, A. E. et al. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. **Nature reviews Microbiology**, v. 11, n. 7, p. 497, 2013.
- 20 DAVIES, G.; HENRISSAT, B. Structures and mechanisms of glycosyl hydrolases. **Structure**, v. 3, n. 9, p. 853–859, 1995.
- 21 HENRISSAT, B.; TEERI, T.; WARREN, R. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. **FEBS Letters**, v. 425, n. 2, p. 352–354, 1998.
- 22 CONSORTIUM, C. Ten years of cazypedia: a living encyclopedia of carbohydrate-active enzymes. **Glycobiology**, v. 28, n. 1, p. 3–8, 2017.
- 23 BROEKER, J. et al. The hemicellulose-degrading enzyme system of the thermophilic bacterium *Clostridium stercorarium*: comparative characterisation and addition of new hemicellulolytic glycoside hydrolases. **Biotechnology for Biofuels**, v. 11, n. 1, p. 229, 2018.
- 24 TANDRUP, T. et al. Recent insights into lytic polysaccharide monooxygenases (LPMOs). **Biochemical Society Transactions**, v. 46, n. 6, p. 1431–1447, 2018.
- 25 AGGER, J. W. et al. Discovery of lpmo activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. **Proceedings of the National Academy of Sciences**, v. 111, n. 17, p. 6287–6292, 2014.
- 26 VAAJE-KOLSTAD, G. et al. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. **Science**, v. 330, n. 6001, p. 219–222, 2010.
- 27 SONG, B. et al. Real-time imaging reveals that lytic polysaccharide monooxygenase promotes cellulase activity by increasing cellulose accessibility. **Biotechnology for Biofuels**, v. 11, n. 1, p. 41, 2018.
- 28 HAYASHI, T.; KAIDA, R. Functions of xyloglucan in plant cells. **Molecular Plant**, v. 4, n. 1, p. 17–24, 2011.
- 29 MCQUEEN-MASON, S. J.; COSGROVE, D. J. Expansin mode of action on cell walls (analysis of wall hydrolysis, stress relaxation, and binding). **Plant Physiology**, v. 107, n. 1, p. 87–100, 1995.
- 30 SANTOS, C. R. dos et al. Structural basis for xyloglucan specificity and α -d-xylp (1,6)-d-glcp recognition at the-1 subsite within the GH5 family. **Biochemistry**, v. 54, n. 10, p. 1930–1942, 2015.

-
- 31 LIBERATO, M. V. et al. Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. **Scientific Reports**, v. 6, p. 23473, 2016.
- 32 VARNAI, A. et al. Carbohydrate-binding modules of fungal cellulases: occurrence in nature, function, and relevance in industrial biomass conversion. **Advances in Applied Microbiology**, v. 88, p. 103–165, 2014.
- 33 STÅHLBERG, J. et al. Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase i from *Trichoderma reesei*. **Journal of Molecular Biology**, v. 264, n. 2, p. 337–349, 1996.
- 34 PAYNE, C. M. et al. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. **Proceedings of the National Academy of Sciences**, v. 110, n. 36, p. 14646–14651, 2013.
- 35 BORASTON, A. B. et al. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. **Biochemical Journal**, v. 382, n. 3, p. 769–781, 2004.
- 36 BYRT, C. S.; CAHYANEGARA, R.; GROF, C. P. Plant carbohydrate binding module enhances activity of hybrid microbial cellulase enzyme. **Frontiers in Plant Science**, v. 3, p. 254, 2012.
- 37 CUSKIN, F. et al. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. **Proceedings of the National Academy of Sciences**, v. 109, n. 51, p. 20889–20894, 2012.
- 38 GAO, D. et al. Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. **Proceedings of the National Academy of Sciences**, v. 110, n. 27, p. 10922–10927, 2013.
- 39 MELLO, B. L.; POLIKARPOV, I. Family 1 carbohydrate binding-modules enhance saccharification rates. **AMB Express**, v. 4, n. 1, p. 36, 2014.
- 40 ABUJA, P. M. et al. Domain structure of cellobiohydrolase II as studied by small angle X-ray scattering: close resemblance to cellobiohydrolase I. **Biochemical and Biophysical Research Communications**, v. 156, n. 1, p. 180–185, 1988.
- 41 SCHMUCK, M. et al. Investigation of cellobiohydrolase from *Trichoderma reesei* by small angle x-ray scattering. **Biotechnology Letters**, v. 8, n. 6, p. 397–402, 1986.
- 42 OSSOWSKI, I. von et al. Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase. **Biophysical Journal**, v. 88, n. 4, p. 2823–2832, 2005.
- 43 LIMA, L. H. et al. Small-angle x-ray scattering and structural modeling of full-length: cellobiohydrolase i from *Trichoderma harzianum*. **Cellulose**, v. 20, n. 4, p. 1573–1585, 2013.
- 44 TEXTOR, L. C. et al. Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from *Trichoderma harzianum*: deciphering the structural features of cellobiohydrolase catalytic activity. **FEBS Journal**, v. 280, n. 1, p. 56–69, 2013.

- 45 SAKON, J. et al. Structure and mechanism of endo/exocellulase e4 from *Thermomonospora fusca*. **Nature Structural and Molecular Biology**, v. 4, n. 10, p. 810, 1997.
- 46 ATTIA, M. et al. Functional and structural characterization of a potent GH74 endo-xyloglucanase from the soil saprophyte *Cellvibrio japonicus* unravels the first step of xyloglucan degradation. **FEBS Journal**, v. 283, n. 9, p. 1701–1719, 2016.
- 47 SCHELLER, H. V.; ULVSKOV, P. Hemicelluloses. **Annual Review of Plant Biology**, v. 61, n. 1, p. 263–289, 2010.
- 48 TUOMIVAARA, S. T. et al. Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature. **Carbohydrate Research**, v. 402, n. 1, p. 56–66, 2015.
- 49 FRY, S. C. et al. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. **Physiologia Plantarum**, v. 89, n. 1, p. 1–3, 1993.
- 50 MARTINEZ-FLEITES, C. et al. Crystal structures of *Clostridium thermocellum* xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. **Journal of Biological Chemistry**, v. 281, n. 34, p. 24922–24933, 2006.
- 51 YAOI, K. et al. The crystal structure of a xyloglucan-specific endo- β -1, 4-glucanase from *Geotrichum sp.* m128 xyloglucanase reveals a key amino acid residue for substrate specificity. **FEBS Journal**, v. 276, n. 18, p. 5094–5100, 2009.
- 52 _____. Tandem repeat of a seven-bladed β -propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. **Structure**, v. 12, n. 7, p. 1209–1217, 2004.
- 53 MANDELMAN, D. et al. X-ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. **Journal of Bacteriology**, v. 185, n. 14, p. 4127–4135, 2003.
- 54 SCHUBOT, F. D. et al. Structural basis for the exocellulase activity of the cellobiohydrolase cbha from *Clostridium thermocellum*. **Biochemistry**, v. 43, n. 5, p. 1163–1170, 2004.
- 55 KESAVULU, M. M. et al. Structure of the catalytic domain of the *Clostridium thermocellum* cellulase CelT. **Acta Crystallographica D**, v. 68, n. 3, p. 310–320, 2012.
- 56 URBANOWICZ, B. R. et al. A tomato endo- β -1, 4-glucanase, SlCel9C1, represents a distinct subclass with a new family of carbohydrate binding modules (CBM49). **Journal of Biological Chemistry**, v. 282, n. 16, p. 12066–12074, 2007.
- 57 MASTER, E. R. et al. Recombinant expression and enzymatic characterization of PttCel9A, a kor homologue from *Populus tremula x tremuloides*. **Biochemistry**, v. 43, n. 31, p. 10080–10089, 2004.
- 58 YOSHIDA, K.; KOMAE, K. A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. **Plant and Cell Physiology**, v. 47, n. 11, p. 1541–1554, 2006.
- 59 OHMIYA, Y. et al. Evidence that endo-1, 4- β -glucanases act on cellulose in suspension-cultured poplar cells. **Plant Journal**, v. 24, n. 2, p. 147–158, 2000.

-
- 60 ARIMORI, T. et al. Crystal structure of endo-1, 4- β -glucanase from *Eisenia fetida*. **Journal of Synchrotron Radiation**, v. 20, n. 6, p. 884–889, 2013.
- 61 GUIMARÃES, B. G. et al. The crystal structure and catalytic mechanism of cellobiohydrolase cels, the major enzymatic component of the *Clostridium thermocellum* cellulosome. **Journal of Molecular Biology**, v. 320, n. 3, p. 587–596, 2002.
- 62 KOSTYLEV, M.; WILSON, D. B. Determination of the catalytic base in family 48 glycosyl hydrolases. **Applied and Environmental Microbiology**, v. 77, n. 17, p. 6274–6276, 2011.
- 63 DIVNE, C. et al. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. **Journal of Molecular Biology**, v. 275, n. 2, p. 309–325, 1998.
- 64 KOSTYLEV, M.; WILSON, D. A distinct model of synergism between a processive endocellulase (TfCel9A) and an exocellulase (TfCel48A) from *Thermobifida fusca*. **Applied and Environmental Microbiology**, v. 80, n. 1, p. 339–344, 2014.
- 65 SUKHARNIKOV, L. O. et al. Sequence, structure, and evolution of cellulases in the glycoside hydrolase family 48. **Journal of Biological Chemistry**, v. 287, n. 49, p. 41068–41077, 2012.
- 66 SU, X.; MACKIE, R. I.; CANN, I. K. Biochemical and mutational analyses of a multi-domain cellulase/mannanase from *Caldicellulosiruptor bescii*. **Applied and Environmental Microbiology**, v. 78, n. 7, p. 2230–2240, 2012.
- 67 ZVERLOV, V. et al. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile *Anaerocellum thermophilum* with separate glycosyl hydrolase family 9 and 48 catalytic domains. **Microbiology**, v. 144, n. 2, p. 457–465, 1998.
- 68 YI, Z. et al. Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by *Caldicellulosiruptor bescii* during growth on crystalline cellulose. **PloS One**, v. 8, n. 12, p. e84172, 2013.
- 69 BRUNECKY, R. et al. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. **Science**, v. 342, n. 6165, p. 1513–1516, 2013.
- 70 SHOEMAKER, S. et al. Molecular cloning of exo-cellobiohydrolase i derived from *Trichoderma reesei* strain 127. **Nature Biotechnology**, v. 1, n. 8, p. 691, 1983.
- 71 BAKER, J. O. et al. A new thermostable endoglucanase, *Acidothermus cellulolyticus* e1. **Applied Biochemistry and Biotechnology**, v. 45, n. 1, p. 245–256, 1994.
- 72 JASKOLSKI, M.; DAUTER, Z.; WLODAWER, A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. **The FEBS journal**, v. 281, n. 18, p. 3985–4009, 2014.
- 73 DAUTER, Z.; JASKOLSKI, M.; WLODAWER, A. Impact of synchrotron radiation on macromolecular crystallography: a personal view. **Journal of synchrotron radiation**, v. 17, n. 4, p. 433–444, 2010.

- 74 GLATTER, O.; KRATKY, O. **Small angle X-ray scattering**. New York: Academic Press, 1982.
- 75 BLUNDELL, T.; JOHNSON, L. **Protein crystallography**. New York: Academic Press, 1976.
- 76 GAVIRA, J. A. Current trends in protein crystallization. **Archives of Biochemistry and Biophysics**, v. 602, p. 3–11, 2016.
- 77 BERGFORS, T. **Protein Crystallization**. 2nd. ed. La Jolla: International University Lie, 2009. v. 8. 474 p. (IUL biotechnology series, v. 8).
- 78 HOLCOMB, J. et al. Protein crystallization: eluding the bottleneck of x-ray crystallography. **AIMS Biophysics**, v. 4, n. 4, p. 557, 2017.
- 79 D'ARCY, A. et al. Microseed matrix screening for optimization in protein crystallization: what have we learned? **Acta Crystallographica F**, v. 70, n. 9, p. 1117–1126, 2014.
- 80 MCPHERSON, A. **Introduction to macromolecular crystallography**. Hoboken, New Jersey: John Wiley & Sons, 2011.
- 81 ROSSMANN, M. G. The molecular replacement method. **Acta Crystallographica A**, v. 46, n. 2, p. 73–82, 1990.
- 82 READ, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. **Acta Crystallographica D**, v. 57, n. 10, p. 1373–1382, 2001.
- 83 ADAMS, P. D. et al. Recent developments in phasing and structure refinement for macromolecular crystallography. **Current Opinion in Structural Biology**, v. 19, n. 5, p. 566–572, 2009.
- 84 EVANS, P.; MCCOY, A. An introduction to molecular replacement. **Acta Crystallographica D**, v. 64, n. 1, p. 1–10, 2008.
- 85 MCCOY, A. J. Solving structures of protein complexes by molecular replacement with phaser. **Acta Crystallographica D**, v. 63, n. 1, p. 32–41, 2007.
- 86 GARMAN, E. F. Developments in x-ray crystallographic structure determination of biological macromolecules. **Science**, v. 343, n. 6175, p. 1102–1108, 2014.
- 87 SIMPKIN, A. J. et al. SIMBAD: a sequence-independent molecular-replacement pipeline. **Acta Crystallographica D**, v. 74, n. 7, p. 595–605, 2018.
- 88 TAYLOR, G. L. Introduction to phasing. **Acta Crystallographica D**, v. 66, n. 4, p. 325–338, 2010.
- 89 DAUTER, Z. SAD/MAD phasing. In: READ, R.; URZHUMTSEV, A. G.; LUNIN, V. Y. (Ed.). **Advancing methods for biomolecular crystallography**. Berlin: Springer, 2013. p. 135–149.
- 90 USÓN, I.; SHELDRIK, G. M. An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. **Acta Crystallographica D**, v. 74, n. 2, p. 106–116, 2018.

-
- 91 FEIGIN, L.; SVERGUN, D. **Structure analysis by small-angle X-ray and neutron scattering**. Berlin: Springer Science+Business Media, 1987.
- 92 SVERGUN, D. I.; KOCH, M. H. J. Small-angle scattering studies of biological macromolecules in solution. **Reports on Progress in Physics**, v. 66, n. 10, p. 1735, 2003.
- 93 BERNADO, P.; SVERGUN, D. I. Structural analysis of intrinsically disordered proteins by small-angle x-ray scattering. **Molecular Biosystems**, v. 8, n. 1, p. 151–167, 2012.
- 94 PINGALI, S. V. et al. Small-angle neutron scattering reveals pH-dependent conformational changes in *Trichoderma reesei* cellobiohydrolase I: implications for enzymatic activity. **Journal of Biological Chemistry**, v. 286, n. 37, p. 32801–32809, 2011.
- 95 BARBOSA, L. R. et al. The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: a small-angle x-ray scattering study. **Biophysical Journal**, v. 98, n. 1, p. 147–157, 2010.
- 96 GUINIER, A. La diffraction des rayons x aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques. **Annales de Physique**, v. 11, n. 12, p. 161–237, 1939.
- 97 DEBYE, P. Scattering of x-rays [zerstreuung von roentgenstrahlen]. **Annalen der Physik**, v. 351, n. 6, p. 809–823, 1915.
- 98 GLATTER, O. A new method for the evaluation of small-angle scattering data. **Journal of Applied Crystallography**, v. 10, n. 5, p. 415–421, 1977.
- 99 FRANKE, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. **Journal of Applied Crystallography**, v. 50, n. 4, p. 1212–1225, 2017.
- 100 SVERGUN, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **Journal of Applied Crystallography**, v. 25, n. 4, p. 495–503, 1992.
- 101 FISCHER H. D. O. N., O. N. M. D. N. H. B. P. I. . C. A. F. Determination of the molecular weight of proteins in solution from a single small-angle x-ray scattering measurement on a relative scale. **Journal of Applied Crystallography**, v. 43, n. 1, p. 101–109, 2010.
- 102 PIIADOV, V. et al. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental saxs data measured on a relative scale. **Protein Science**, v. 28, n. 2, p. 454–463, 2018.
- 103 DONIACH, S. Changes in biomolecular conformation seen by small angle x-ray scattering. **Chemical Reviews**, v. 101, n. 6, p. 1763–1778, 2001.
- 104 RAMBO, R. P.; TAINER, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by sas using the porod-debye law. **Biopolymers**, v. 95, n. 8, p. 559–571, 2011.

- 105 MERTENS, H. D.; SVERGUN, D. I. Structural characterization of proteins and complexes using small-angle x-ray solution scattering. **Journal of Structural Biology**, v. 172, n. 1, p. 128–141, 2010.
- 106 BERNADÓ, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. **European Biophysics Journal**, v. 39, n. 5, p. 769–780, 2010.
- 107 DURAND, D. et al. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. **Journal of Structural Biology**, v. 169, n. 1, p. 45–53, 2010.
- 108 BURGER, V. M.; ARENAS, D. J.; STULTZ, C. M. A structure-free method for quantifying conformational flexibility in proteins. **Scientific Reports**, v. 6, p. 29040, 2016.
- 109 JEFFRIES, C. M. et al. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. **Nature Protocols**, v. 11, n. 11, p. 2122, 2016.
- 110 TREWHELLA, J. et al. 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. **Acta Crystallographica D**, v. 73, n. 9, p. 710–728, 2017.
- 111 KORASICK, D. A.; TANNER, J. J. Determination of protein oligomeric structure from small-angle X-ray scattering. **Protein Science**, v. 27, n. 4, p. 814–824, 2018.
- 112 MATHEW, E.; MIRZA, A.; MENHART, N. Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. **Journal of Synchrotron Radiation**, v. 11, n. 4, p. 314–318, 2004.
- 113 BROOKES, E. et al. Fibrinogen species as resolved by HPLC-SAXS data processing within the ultrascan solution modeler (us-somo) enhanced sas module. **Journal of Applied Crystallography**, v. 46, n. 6, p. 1823–1833, 2013.
- 114 ACERBO, A. S.; COOK, M. J.; GILLILAN, R. E. Upgrade of macchess facility for x-ray scattering of biological macromolecules in solution. **Journal of Synchrotron Radiation**, v. 22, n. 1, p. 180–186, 2015.
- 115 GRAEWERT, M. A. et al. Automated pipeline for purification, biophysical and x-ray analysis of biomacromolecular solutions. **Scientific Reports**, v. 5, p. 10734, 2015.
- 116 RYAN, T. M. et al. An optimized sec-SAXS system enabling high x-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. **Journal of Applied Crystallography**, v. 51, n. 1, p. 97–111, 2018.
- 117 CAMILO, C. M.; POLIKARPOV, I. High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC). **Protein Expression and Purification**, v. 99, p. 35–42, 2014.
- 118 de Araujo, E. A. **Caracterização estrutural e funcional de uma xiloglucanase de *Xanthomonas campestris***. 2013. 92 f. Dissertação (Mestrado em Biotecnologia) — Centro de Ciências Exatas e de Tecnologia, Universidade Federal de São Carlos, São Carlos, 2013.

-
- 119 STUDIER, F. W. Protein production by auto-induction in high density shaking cultures. **Protein Expression and Purification**, v. 41, n. 1, p. 207–34, 2005.
- 120 LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680–685, 1970.
- 121 ZHANG, Y.-H. P.; HIMMEL, M. E.; MIELENZ, J. R. Outlook for cellulase improvement: screening and selection strategies. **Biotechnology Advances**, v. 24, n. 5, p. 452–481, 2006.
- 122 MCILVAINE, T. A buffer solution for colorimetric comparison. **Journal of Biological Chemistry**, v. 49, n. 1, p. 183–186, 1921.
- 123 GHOSE, T. K. Measurement of cellulase activities. **Pure and Applied Chemistry**, v. 59, n. 2, p. 257–268, 1987.
- 124 BORNE, R. et al. Unraveling enzyme discrimination during cellosome assembly independent of cohesin–dockerin affinity. **FEBS Journal**, v. 280, n. 22, p. 5764–5779, 2013.
- 125 BUCKERIDGE, M. S. et al. Xyloglucan structure and postgerminative metabolism in seeds of *Copaifera langsdorffii* from savanna and forest populations. **Physiologia Plantarum**, v. 86, n. 1, p. 145–151, 1992.
- 126 DAMÁSIO, A. R. L. et al. Functional characterization and oligomerization of a recombinant xyloglucan-specific endo- β -1,4-glucanase (GH12) from *Aspergillus niger*. **Biochimica et Biophysica Acta**, v. 1824, n. 3, p. 461–467, 2012.
- 127 HAMMERSLEY, A. FIT2D: a multi-purpose data reduction, analysis and visualization program. **Journal of Applied Crystallography**, v. 49, n. 2, p. 646–652, 2016.
- 128 FRANKE, D.; SVERGUN, D. I. Dammif, a program for rapid *ab-initio* shape determination in small-angle scattering. **Journal of Applied Crystallography**, v. 42, n. 2, p. 342–346, 2009.
- 129 VOLKOV, V. V.; SVERGUN, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. **Journal of Applied Crystallography**, v. 36, n. 3, p. 860–864, 2003.
- 130 SVERGUN, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. **Biophysical Journal**, v. 76, n. 6, p. 2879–2886, 1999.
- 131 KOZIN, M. B.; SVERGUN, D. I. Automated matching of high-and low-resolution structural models. **Journal of Applied Crystallography**, v. 34, n. 1, p. 33–41, 2001.
- 132 TUUKKANEN, A. T.; KLEYWEGT, G. J.; SVERGUN, D. I. Resolution of *ab initio* shapes determined from small-angle scattering. **IUCrJ**, v. 3, p. 440–447, 2016.
- 133 SCHNEIDMAN-DUHOVNY, D.; HAMMEL, M.; SALI, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. **Nucleic Acids Research**, v. 38, p. W540–4, 2010.

- 134 MCPHERSON, A. Protein crystallization. In: WLODAWER, A.; DAUTER, Z.; JASKOLSKI, M. (Ed.). **Protein Crystallography**. Berlin: Springer, 2017, (Methods in molecular biology, v. 1607). p. 17–50.
- 135 GUIMARÃES, B. G. et al. The MX2 macromolecular crystallography beamline: a wiggler X-ray source at the LNLS. **Journal of Synchrotron Radiation**, v. 16, n. 1, p. 69–75, 2009.
- 136 KABSCH, W. et al. XDS. **Acta Crystallographica D**, v. 66, Pt 2, p. 125–132, 2010.
- 137 EVANS, P. Scaling and assessment of data quality. **Acta Crystallographica D**, v. 62, n. 1, p. 72–82, 2006.
- 138 ADAMS, P. D. et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. **Acta Crystallographica D**, v. 66, n. 2, p. 213–221, 2010.
- 139 MCCOY, A. J. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, n. 4, p. 658–674, 2007.
- 140 BRUNECKY, R. et al. Natural diversity of glycoside hydrolase family 48 exoglucanases: insights from structure. **Biotechnology for Biofuels**, v. 10, n. 1, p. 274, 2017.
- 141 EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica D**, v. 66, n. 4, p. 486–501, 2010.
- 142 WILLIAMS, C. J. et al. Molprobity: more and better reference data for improved all-atom structure validation. **Protein Science**, v. 27, n. 1, p. 293–315, 2018.
- 143 EDGAR, R. C. Muscle: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.
- 144 ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new endscript server. **Nucleic Acids Research**, v. 42, n. Web Server issue, p. W320–4, 2014.
- 145 BIEGERT, A.; SÖDING, J. Sequence context-specific profiles for homology searching. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 106, n. 10, p. 3770–3775, 2009.
- 146 ASHKENAZY, H. et al. Consurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic Acids Research**, v. 44, n. W1, p. W344–W350, 2016.
- 147 FU, L. et al. Cd-hit: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, v. 28, n. 23, p. 3150–3152, 2012.
- 148 RIVOIRE, O.; REYNOLDS, K. A.; RANGANATHAN, R. Evolution-based functional decomposition of proteins. **PLoS Computational Biology**, v. 12, n. 6, p. e1004817, 2016.

-
- 149 DESMET, T. et al. An investigation of the substrate specificity of the xyloglucanase Cel74A from *Hypocrea jecorina*. **FEBS Journal**, v. 274, n. 2, p. 356–363, 2007.
- 150 HASPER, A. A. et al. EglC, a new endoglucanase from *Aspergillus niger* with major activity towards xyloglucan. **Applied and Environmental Microbiology**, v. 68, n. 4, p. 1556–1560, 2002.
- 151 YAOI, K. et al. The structural basis for the exo-mode of action in GH74 oligoxyloglucan reducing end-specific cellobiohydrolase. **Journal of Molecular Biology**, v. 370, n. 1, p. 53–62, 2007.
- 152 YAOI, K.; MITSUISHI, Y. Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. **Journal of Biological Chemistry**, v. 277, n. 50, p. 48276–48281, 2002.
- 153 ISHIDA, T. et al. Substrate recognition by glycoside hydrolase family 74 xyloglucanase from the basidiomycete *Phanerochaete chrysosporium*. **FEBS Journal**, v. 274, n. 21, p. 5727–5736, 2007.
- 154 BAUER, S. et al. Cloning, expression, and characterization of an oligoxyloglucan reducing end-specific xyloglucanobiohydrolase from *Aspergillus nidulans*. **Carbohydrate Research**, v. 340, n. 17, p. 2590–2597, 2005.
- 155 IRWIN, D. C. et al. Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. **European Journal of Biochemistry**, v. 270, n. 14, p. 3083–3091, 2003.
- 156 YAOI, K. et al. Cloning and characterization of two xyloglucanases from *Paenibacillus* sp. strain KM21. **Applied and Environmental Microbiology**, v. 71, n. 12, p. 7670–7678, 2005.
- 157 ENKHBAATAR, B. et al. Identification and characterization of a xyloglucan-specific family 74 glycosyl hydrolase from *Streptomyces coelicolor* A3(2). **Applied and Environmental Microbiology**, v. 78, n. 2, p. 607–611, 2012.
- 158 ICHINOSE, H. et al. Characterization of an endo-processive-type xyloglucanase having a β -1,4-glucan-binding module and an endo-type xyloglucanase from *Streptomyces avermitilis*. **Applied and Environmental Microbiology**, v. 78, n. 22, p. 7939–7945, 2012.
- 159 KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870–1874, 2016.
- 160 MATSUZAWA, T.; YAOI, K. GH74 xyloglucanases: structures and modes of activity. **Trends in Glycoscience and Glycotechnology**, v. 28, n. 162, p. J63–J70, 2016.
- 161 HENRISSAT, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemical Journal**, v. 280, p. 309–316, 1991.

- 162 FENG, T. et al. Characterisation of a novel endo-xyloglucanase (*XcXGHA*) from *Xanthomonas* that accommodates a xylosyl-substituted glucose at subsite -1. **Applied Microbiology and Biotechnology**, v. 98, n. 23, p. 9667–9679, 2014.
- 163 YAOI, K.; MITSUISHI, Y. Purification, characterization, cDNA cloning, and expression of a xyloglucan endoglucanase from *Geotrichum sp.* M128. **FEBS Letters**, v. 560, n. 1-3, p. 45–50, 2004.
- 164 MATSUZAWA, T.; SAITO, Y.; YAOI, K. Key amino acid residues for the endo-processive activity of GH74 xyloglucanase. **FEBS Letters**, v. 588, n. 9, p. 1731–1738, 2014.
- 165 KRUUS, K. et al. Exoglucanase activities of the recombinant *Clostridium thermocellum* cels, a major cellulosome component. **Journal of Bacteriology**, v. 177, n. 6, p. 1641–1644, 1995.
- 166 BERGER, E. et al. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. **FEMS Microbiology Letters**, v. 268, n. 2, p. 194–201, 2007.
- 167 CHAUVAUX, S. et al. Calcium-binding affinity and calcium-enhanced activity of *Clostridium thermocellum* endoglucanase D. **Biochemical Journal**, v. 265, n. 1, p. 261–265, 1990.
- 168 HORN, S. J. et al. Measuring processivity. **Methods in Enzymology**, v. 510, p. 69–95, 2012.
- 169 KOSTYLEV, M. et al. Cel48A from *Thermobifida fusca*: structure and site directed mutagenesis of key residues. **Biotechnology and Bioengineering**, v. 111, n. 4, p. 664–673, 2014.
- 170 KADOWAKI, M. A. S. et al. Biochemical and structural insights into a thermostable cellobiohydrolase from *Myceliophthora thermophila*. **FEBS Journal**, v. 285, n. 3, p. 559–579, 2018.
- 171 KURAŠIN, M.; VÄLJAMÄE, P. Processivity of cellobiohydrolases is limited by the substrate. **Journal of Biological Chemistry**, v. 286, n. 1, p. 169–177, 2011.
- 172 IGARASHI, K. et al. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. **Science**, v. 333, n. 6047, p. 1279–1282, 2011.
- 173 _____. Two-way traffic of glycoside hydrolase family 18 processive chitinases on crystalline chitin. **Nature Communications**, v. 5, p. 3975, 2014.
- 174 NAKAMURA, A. et al. Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose. **Journal of the American Chemical Society**, v. 136, n. 12, p. 4584–4592, 2014.
- 175 _____. Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase *Trichoderma reesei* Cel6A and its domains on crystalline cellulose. **Journal of Biological Chemistry**, v. 291, n. 43, p. 22404–22413, 2016.

- 176 ZHANG, Y. H.; LYND, L. R. Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. **Biomacromolecules**, v. 6, n. 3, p. 1510–1515, 2005.
- 177 ZHANG, Y. H. P.; LYND, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. **Biotechnology and Bioengineering**, v. 88, n. 7, p. 797–824, 2004.
- 178 PARK, S. et al. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. **Biotechnology for Biofuels**, v. 3, 2010.
- 179 URIU, M. B. **Caracterização do modo de ação e do sinergismo entre duas celulasas (famílias GH9 E GH48) de *Bacillus licheniformis***. 2017. 70 f. Dissertação (Mestrado em Bioenergia) — Setor de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2017.
- 180 KESHK, S. M. Bacterial cellulose production and its industrial applications. **Journal of Bioprocessing & Biotechniques**, v. 4, p. 150, 2014.
- 181 PARSIEGLA, G. et al. Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. **Biochemistry**, v. 39, n. 37, p. 11238–11246, 2000.
- 182 _____. The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. **EMBO Journal**, v. 17, n. 19, p. 5551–5562, 1998.
- 183 HOLM, L.; ROSENSTRÖM, P. Dali server: conservation mapping in 3D. **Nucleic Acids Research**, v. 38, p. W545–549, 2010.
- 184 CHEN, M. et al. Strategies to reduce end-product inhibition in family 48 glycoside hydrolases. **Proteins**, v. 84, n. 3, p. 295–304, 2016.
- 185 WU, M. et al. Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. **Journal of Biological Chemistry**, v. 288, n. 46, p. 33107–33117, 2013.
- 186 KNOTT, B. C. et al. The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. **Journal of the American Chemical Society**, v. 136, n. 1, p. 321–329, 2014.
- 187 OLSEN, J. P. et al. Mechanism of product inhibition for cellobiohydrolase Cel7A during hydrolysis of insoluble cellulose. **Biotechnology and Bioengineering**, v. 113, n. 6, p. 1178–1186, 2016.
- 188 COLUSSI, F. et al. Probing substrate interactions in the active tunnel of a catalytically deficient cellobiohydrolase (Cel7). **Journal of Biological Chemistry**, v. 290, n. 4, p. 2444–2454, 2015.
- 189 KIM, S. J. et al. Enhanced thermostability of mesophilic endoglucanase Z with a high catalytic activity at active temperatures. **International Journal of Biological Macromolecules**, v. 86, p. 269–276, 2016.

- 190 ZHANG, C.; ZHANG, W.; LU, X. Expression and characteristics of a Ca^{2+} -dependent endoglucanase from *Cytophaga hutchinsonii*. **Applied Microbiology and Biotechnology**, v. 99, n. 22, p. 9617–9623, 2015.
- 191 KATAEVA, I. A. et al. Domain coupling in a multimodular cellobiohydrolase CbhA from *Clostridium thermocellum*. **FEBS Letters**, v. 579, n. 20, p. 4367–4373, 2005.
- 192 LEIS, B. et al. Comparative characterization of all cellosomal cellulases from *Clostridium thermocellum* reveals high diversity in endoglucanase product formation essential for complex activity. **Biotechnology for Biofuels**, v. 10, n. 1, p. 240, 2017.
- 193 RAVACHOL, J. et al. Characterization of all family-9 glycoside hydrolases synthesized by the cellosome-producing bacterium *Clostridium cellulolyticum*. **Journal of Biological Chemistry**, v. 289, n. 11, p. 7335–7348, 2014.
- 194 YOUNESI, F. S. et al. Deleting the Ig-like domain of *Alicyclobacillus acidocaldarius* endoglucanase Cel9A causes a simultaneous increase in the activity and stability. **Molecular Biotechnology**, v. 58, n. 1, p. 12–21, 2016.
- 195 ZHANG, F. et al. Cloning, expression and characterization of an alkaline thermostable GH endoglucanase from *Thermobifida halotolerans* YIM 90462T. **Bioresource Technology**, v. 102, n. 21, p. 10143–10146, 2011.
- 196 YANG, M. et al. Synergistic cellulose hydrolysis dominated by a multi-modular processive endoglucanase from *Clostridium cellulosi*. **Frontiers in Microbiology**, v. 7, p. 932, 2016.
- 197 PETKUN, S. et al. Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker. **PeerJ**, v. 3, p. e1126, 2015.
- 198 HENRISSAT, B. et al. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. **Proceedings of the National Academy of Sciences**, v. 92, n. 15, p. 7090–7094, 1995.
- 199 KOSHLAND, D. E. Stereochemistry and the mechanism of enzymatic reactions. **Biological Reviews**, v. 28, n. 4, p. 416–436, 1953.
- 200 WANG, H. J. et al. Polarity alteration of a calcium site induces a hydrophobic interaction network and enhances Cel9A endoglucanase thermostability. **Applied and Environmental Microbiology**, v. 82, n. 6, p. 1662–1674, 2016.
- 201 TORMO, J. et al. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. **EMBO Journal**, v. 15, n. 21, p. 5739–5751, 1996.
- 202 VON OSSOWSKI, I. et al. Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. a comparison with *Phanerochaete chrysosporium* Cel7D. **Journal of Molecular Biology**, v. 333, n. 4, p. 817–829, 2003.
- 203 MEINKE, A. et al. Enhancement of the endo- β -1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. **Journal of Biological Chemistry**, v. 270, n. 9, p. 4383–4386, 1995.

-
- 204 DAVIES, G. J.; WILSON, K. S.; HENRISSAT, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. **Biochemical Journal**, v. 321, Pt 2, p. 557–559, 1997.
- 205 ECKERT, K. et al. Crystal structures of A. *Acidocaldarius* endoglucanase Cel9A in complex with cello-oligosaccharides: strong – 1 and – 2 subsites mimic cellobiohydrolase activity. **Journal of Molecular Biology**, v. 394, n. 1, p. 61–70, 2009.
- 206 HAMMEL, M. Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). **European Biophysics Journal**, v. 41, n. 10, p. 789–799, 2012.
- 207 PUTNAM, C. D. et al. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. **Quarterly Reviews of Biophysics**, v. 40, n. 3, p. 191–285, 2007.
- 208 RECEVEUR-BRECHOT, V.; DURAND, D. How random are intrinsically disordered proteins? a small angle scattering perspective. **Current Protein & Peptide Science**, v. 13, n. 1, p. 55–75, 2012.
- 209 BATISTA, P. R. et al. High temperatures enhance cooperative motions between CBM and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics. **Physical Chemistry Chemical Physics**, v. 13, n. 30, p. 13709–13720.
- 210 OLIVEIRA, O. V. et al. Interaction between the CBM of Cel9A from *Thermobifida fusca* and cellulose fibers. **Journal of Molecular Recognition**, v. 22, n. 1, p. 38–45, 2009.

Anexos

ANEXO A – ARTIGOS PUBLICADOS

- A.1 Biochemical characterization and low-resolution SAXS structure of two-domain endoglucanase *BICel9* from *Bacillus licheniformis***

Biochemical characterization and low-resolution SAXS structure of two-domain endoglucanase *B/Cel9* from *Bacillus licheniformis*

Evandro Ares de Araújo¹ · Mário de Oliveira Neto² · Igor Polikarpov¹

Received: 14 July 2018 / Revised: 3 October 2018 / Accepted: 7 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Lignocellulose feedstock constitutes the most abundant carbon source in the biosphere; however, its recalcitrance remains a challenge for microbial conversion into biofuel and bioproducts. *Bacillus licheniformis* is a microbial mesophilic bacterium capable of secreting a large number of glycoside hydrolase (GH) enzymes, including a glycoside hydrolase from GH family 9 (*B/Cel9*). Here, we conducted biochemical and biophysical studies of recombinant *B/Cel9*, and its low-resolution molecular shape was retrieved from small angle X-ray scattering (SAXS) data. *B/Cel9* is an endoglucanase exhibiting maximum catalytic efficiency at pH 7.0 and 60 °C. Furthermore, it retains 80% of catalytic activity within a broad range of pH values (5.5–8.5) and temperatures (up to 50 °C) for extended periods of time (over 48 h). It exhibits the highest hydrolytic activity against phosphoric acid swollen cellulose (PASC), followed by bacterial cellulose (BC), filter paper (FP), and to a lesser extent carboxymethylcellulose (CMC). The HPAEC-PAD analysis of the hydrolytic products demonstrated that the end product of the enzymatic hydrolysis is primarily cellobiose, and also small amounts of glucose, cellobiose, and cellobiose are produced. SAXS data analysis revealed that the enzyme adopts a monomeric state in solution and has a molecular mass of 65.8 kDa as estimated from SAXS data. The *B/Cel9* has an elongated shape composed of an N-terminal family 3 carbohydrate-binding module (CBM3c) and a C-terminal GH9 catalytic domain joined together by 20 amino acid residue long linker peptides. The domains are closely juxtaposed in an extended conformation and form a relatively rigid structure in solution, indicating that the interactions between the CBM3c and GH9 catalytic domains might play a key role in cooperative cellulose biomass recognition and hydrolysis.

Keywords GH9 family · Family 3 carbohydrate-binding module (CBM3c) · Endoglucanase · Small angle X-ray scattering (SAXS)

Introduction

Plant biomass constitutes the most abundant, natural, renewable, and biodegradable carbon source in the biosphere. Plant cell walls form a complex matrix composed mainly of cellulose, hemicellulose, and lignin (Keegstra 2010; Cosgrove 2014; Marriott et al. 2016) which comprise approximately

75% of a dry plant biomass (Marriott et al. 2016). Cellulose is the main polysaccharide of the plant cell walls (Cosgrove and Jarvis 2012; Marriott et al. 2016) formed by a linear poly-β(1,4)-D-glucopyranose chains (Somerville et al. 2004; Gilbert 2010; Cosgrove and Jarvis 2012; Ding et al. 2012). The high crystallinity and its intrinsic recalcitrance make this biopolymer insoluble in most solvents, hindering the processing of vegetable biomass (Lynd et al. 2002; Gilbert 2010; Igarashi 2013). However, pretreated lignocellulosic biomass can be biochemically converted into soluble sugars using enzymatic technology, which has become the preferable biotechnological route to upgrade plant biomass into biofuels, green chemicals, and sustainable materials (Himmel et al. 2007; Soccol et al. 2010; Klein-Marcuschamer et al. 2012; Gilbert 2014; Donohoe and Resch 2015; Kafle et al. 2015; Kumar and Sharma 2017).

Lignocellulosic biomass enzymatic hydrolysis is carried out by a diversity of enzymes that differ structurally and mechanistically (Henrissat and Davies 1997) and

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00253-018-9508-1>) contains supplementary material, which is available to authorized users.

✉ Igor Polikarpov
ipolikarpov@ifsc.usp.br

¹ Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador Saocarlense 400, São Carlos, SP 13560-970, Brazil

² Departamento de Física e Biofísica, Universidade Estadual Paulista “Júlio de Mesquita Filho”, R. Prof. Dr. Antonio Celso Wagner Zanin 689, Jardim São José, Botucatu, SP 18618-970, Brazil

A.2 SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale

SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale

Vassili Piiadov,¹ Evandro Ares de Araújo,¹ Mario Oliveira Neto,² Aldo Felix Craievich,³ and Igor Polikarpov ^{1*}

¹Institute of Physics of São Carlos, University of São Paulo, São Carlos, São Paulo, Brazil

²Institute of Biosciences, University Estadual Paulista, Botucatu, São Paulo, Brazil

³Institute of Physics, University of São Paulo, São Paulo, São Paulo, Brazil

Received 2 August 2018; Accepted 8 October 2018

DOI: 10.1002/pro.3528

Published online 13 December 2018 proteinscience.org

Abstract: Knowledge of molecular weight, oligomeric states, and quaternary arrangements of proteins in solution is fundamental for understanding their molecular functions and activities. We describe here a program SAXSMoW 2.0 for robust and quick determination of molecular weight and oligomeric state of proteins in dilute solution, starting from a single experimental small-angle scattering intensity curve, $I(q)$, measured on a relative scale. The first version of this calculator has been widely used during the last decade and applied to analyze experimental SAXS data of many proteins and protein complexes. SAXSMoW 2.0 exhibits new features which allow for the direct input of experimental intensity curves and also automatic modes for quick determinations of the radius of gyration, volume, and molecular weight. The new program was extensively tested by applying it to many experimental SAXS curves downloaded from the open databases, corresponding to proteins with different shapes and molecular weights ranging from ~10 kDa up to about ~500 kDa and different shapes from globular to elongated. These tests reveal that the use of SAXSMoW 2.0 allows for determinations of molecular weights of proteins in dilute solution with a median discrepancy of about 12% for globular proteins. In case of elongated molecules, discrepancy value can be significantly higher. Our tests show discrepancies of approximately 21% for the proteins with molecular shape aspect ratios up to 18.

Keywords: SAXS; proteins; molecular weight; on-line calculator

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

Grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico 140667/2015-6158752/2015-5303988/2016-9405191/2015-4440977/2016-9; Grant sponsor: Fundação de Amparo à Pesquisa do Estado de São Paulo 2014/00769-5/2015/13684-0.

*Correspondence to: I. Polikarpov, Institute of Physics of São Carlos, University of São Paulo São Carlos São Paulo, Brazil. Email: ipolikarpov@ifsc.usp.br

V. Piiadov and E. Ares de Araújo contributed equally to this work.

Introduction

Small-angle X-ray scattering (SAXS) is an experimental technique frequently applied to low-resolution structural studies of macromolecules embedded in a homogeneous liquid medium, over a molecular size scale within the 1–100 nm range. The SAXS method allows for investigations of both, well-structured and disordered macromolecules in solution, neither requiring crystallization procedures nor highly elaborate sample preparations. The experimental SAXS intensity associated to a set of proteins of the same

A.3 Biochemical characterization, low-resolution SAXS structure and an enzymatic cleavage pattern of *B/Cel48* from *Bacillus licheniformis*

Biochemical characterization, low-resolution SAXS structure and an enzymatic cleavage pattern of BICel48 from *Bacillus licheniformis*

Evandro Ares de Araújo, Lívia Regina Manzone, Vassili Piiadov, Marco Antonio Seiki Kadowaki, Igor Polikarpov *

Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense, 400, São Carlos, SP CEP 13560-970, Brazil

ARTICLE INFO

Article history:

Received 19 November 2017

Received in revised form 17 December 2017

Accepted 25 December 2017

Available online 30 December 2017

Keywords:

Bacillus licheniformis

Cellulase

GH48 family

SAXS

ABSTRACT

Economic sustainability of modern biochemical technologies for plant cell wall transformations in renewable fuels, green chemicals, and sustainable materials is considerably impacted by the elevated cost of enzymes. Therefore, there is a significant drive toward discovery and characterization of novel carbohydrate-active enzymes. Here, the BICel48 cellulase from *Bacillus licheniformis*, a glycoside hydrolase family 48 member (GH48), was functionally and biochemically characterized. The enzyme is catalytically stable in a broad range of temperatures and pH conditions with its enzymatic activity at pH 5.0 and 60 °C. BICel48 exhibits high hydrolytic activity against phosphoric acid swollen cellulose (PASC) and bacterial cellulose (BC) and significantly lower activity against carboxymethylcellulose (CMC). BICel48 releases predominantly cellobiose, and also small amounts of cellobiose and cellobiose as products from PASC hydrolysis. Small-angle X-ray scattering (SAXS) data analysis revealed a globular molecular shape and monomeric state of the enzyme in solution. Its molecular mass estimated based on SAXS data is ~77.2 kDa. BICel48 has an $(\alpha\alpha)_6$ -helix barrel-fold, characteristic of GH48 members. Comparative analyses of homologous sequences and structures reveal the existence of two distinct loops in BICel48 that were not present in other structurally characterized GH48 enzymes which could have importance for the enzyme activity and specificity.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The plant cell walls are constituted mainly by a complex cross-linked matrix composed of polyphenolics, fibers, sugars, and some proteins and polysaccharides [1–3]. Cellulose, a major component of plant cell walls polysaccharides, is an abundant natural, renewable and biodegradable linear homopolymer formed by poly- β (1,4)-D-glucopyranose [3–6]. Extensive hydrogen bonding, high crystallinity, and recalcitrance of cellulose make this polymer insoluble in most solvents hindering the processing of plant biomass [5,7].

At present, biochemical conversion of lignocellulosic biomass using enzymatic technology became a preferential biotechnological route for plant biomass depolymerization [8–14]. Enzymatic hydrolysis is carried out by a diversity of enzymes classified in CAZy (Carbohydrate-Active EnZymes) database, which can differ structurally and mechanistically [15,16]. To overcome recalcitrance of lignocellulosic substrates, synergic action of several enzymes is required for its efficient hydrolysis [7,10,17–20]. Bacterial biomass degrading systems play an important role in the breakdown of plant cell walls, including cellulases from glycoside hydrolase family 48 (GH48) [21]. GH48 contains mostly bacterial

enzymes, with CelS from *Clostridium thermocellum* [22], Celf from *Clostridium cellulolyticum* [23], cellulase from *Caldocellum saccharolyticum* [24] and cellobiohydrolase B from *Cellulomonas fimi* [25] being its founding members. GH48 includes mainly, but not always, reducing end-acting cellobiohydrolases, which make part of complex bacterial cellulose degrading systems.

These enzymes are capable of synergistically enhancing activities of GH9 processive *endo*-cellulases in the process of crystalline cellulose hydrolysis, as has been described for TjCel48A and TjCel9A from *Thermobifida fusca* [26], CbCel9A and CbCel48A from *Caldicellulosiruptor bescii* [27], and CcCel9B and CcCel48A from *Clostridium cellulosi* [20]. Unlike GH9 enzymes, GH48 genes are frequently present as single copies in the genomes of cellulolytic organisms, such as *Bacillus* species, for example [28].

A gram-positive bacterium, *Bacillus licheniformis* is a facultative anaerobe, which is widely distributed in the environment [29]. *B. licheniformis* has a number of important biotechnological, agricultural and industrial applications [30,31]. It is used for production of antibiotics, chemicals [30] and some of the industrially important hydrolytic enzymes including penicillinase, pentosanases, proteases, α -amylases, glucoamylase, pectinases and several cellulase enzymes [32–34]. In this work, we report the heterologous expression, purification and biochemical characterization of BICel48 from *Bacillus licheniformis* (ATCC 14580). We also built and analyzed its low resolution SAXS molecular envelope and its 3D homologous model.

* Corresponding author at: Universidade de São Paulo, Departamento de Física e Ciência Interdisciplinar, Instituto de Física de São Carlos, Av. Trabalhador São-carlense, 400, São Carlos, SP CEP 13560-970, Brazil.

E-mail address: ipolikarpov@ifsc.usp.br (I. Polikarpov).

A.4 Biochemical characterization and low-resolution SAXS structure of an exo-polygalacturonase from *Bacillus licheniformis*

Contents lists available at ScienceDirect

New BIOTECHNOLOGY

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

Full length Article

Biochemical characterization and low-resolution SAXS structure of an *exo*-polygalacturonase from *Bacillus licheniformis*

Danilo Elton Evangelista^a, Evandro A. de Araújo^a, Mario Oliveira Neto^b,
Marco Antonio Seiki Kadowaki^a, Igor Polikarpov^{a,*}

^a Instituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador São-carlense 400, CEP 13566-590 São Carlos, SP, Brazil

^b Departamento de Física e Biofísica, Instituto de Biociências, Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), CEP 18618-970 Botucatu, São Paulo, Brazil

ARTICLE INFO

Keywords:

Pectinase
Exo-polygalacturonase
Biochemical characterization
Thermal and pH stability
SAXS molecular envelope

ABSTRACT

Among the structural polymers present in the plant cell wall, pectin is the main component of the middle lamella. This heterogeneous polysaccharide has an α -1,4 galacturonic acid backbone, which can be broken by the enzymatic action of pectinases, such as *exo*-polygalacturonases, that sequentially cleave pectin from the non-reducing ends, releasing mono or di-galacturonic acid residues. Constant demand for pectinases that better suit industrial requirements has motivated identification and characterization of novel enzymes from diverse sources. *Bacillus licheniformis* has been used as an important source for bioprospection of several industrial biomolecules, such as surfactants and enzymes, including pectate lyases. Here we cloned, expressed, purified, and biochemically and structurally characterized an *exo*-polygalacturonase from *B. licheniformis* (BExoPG). Its low-resolution molecular envelope was derived from experimental small-angle scattering data (SAXS). Our experimental data revealed that BExoPG is a monomeric enzyme with optimum pH at 6.5 and optimal temperature of approximately 60 °C, at which it has considerable stability over the broad pH range from 5 to 10. After incubation of the enzyme for 30 min at pH ranging from 5 to 10, no significant loss of the original enzyme activity was observed. Furthermore, the enzyme maintained residual activity of greater than 80% at 50 °C after 15 h of incubation. BExoPG is more active against polygalacturonic acid as compared to methylated pectin, liberating mono galacturonic acid as a unique product. Its enzymatic parameters are $V_{\max} = 4.18 \mu\text{M}\cdot\text{s}^{-1}$, $K_m = 3.25 \text{ mgmL}^{-1}$ and $k_{\text{cat}} = 2.58 \text{ s}^{-1}$.

Introduction

The plant cell wall is built up of a number of intimately intertwined structural polymers forming complex heterogeneous structures. The middle lamella is mostly composed of pectin, a heterogeneous polysaccharide formed of methyl esterified galacturonic acid units, linked by α -1,4 glycoside bonds [1]. Methyl groups of the backbone chain are removed by the enzymatic action of pectin methylsterase (PME) and the α -1,4 glycosidic bonds are cleaved by the catalytic action of two different classes of pectinases: pectate lyases (PL, EC 4.2.2.) that use a *trans*-elimination reaction for cleavage, thus releasing unsaturated sugars; and the polygalacturonases (PGs), which employ hydrolytic reactions leading to release of saturated sugars [1,2]. Among PGs, the *endo*-PG (EC 3.2.1.15) randomly hydrolyzes internal sites of the pectin chain, while the *exo*-PG promotes a sequential cleavage from non-reducing ends cleaving off mono (EC 3.2.1.67) or di-galacturonic acid residues (EC 3.2.1.82) [2]. Pectinases have a wide range of industrial

applications, e.g., in extraction and clarification of fruit juices, oil extraction, pretreatment of wastewaters containing pectin, and retting and degumming of crop fibers, to list a few [1,3,4]. Some of the applications, such as fruit juice clarification and winemaking, require acid pectinases that are generally obtained from fungi [1,3]. On the other hand, degumming and retting of natural fibers and textile scouring demand applications of alkaline pectinases, which are commonly obtained from bacterial sources [1,3].

Driven by constant biotechnological demands in pectinolytic enzymes, there is a need for identification and characterization of novel pectinases with biochemical properties that could be useful in industrial settings. *Bacillus licheniformis* has been reported as an important microbial producer of several industrial enzymes. This microorganism is a gram-positive, mesophilic bacterium, commonly found in the soil, growing at temperatures between 15 and 68 °C [5]. Superior performance in secretion of alkaline serine protease transformed *B. licheniformis* in one of the most important bacterial sources of these enzymes,

* Corresponding author.

E-mail address: ipolikarpov@ifsc.usp.br (I. Polikarpov).

<http://dx.doi.org/10.1016/j.nbt.2017.10.001>

Received 8 April 2017; Received in revised form 7 September 2017; Accepted 5 October 2017

Available online 06 October 2017

1871-6784/ © 2017 Elsevier B.V. All rights reserved.

A.5 Biochemical Characterization and Low-Resolution SAXS Molecular Envelope of GH1 β -Glycosidase from *Saccharophagus degradans*

Biochemical Characterization and Low-Resolution SAXS Molecular Envelope of GH1 β -Glycosidase from *Saccharophagus degradans*

Hevila Brognaro¹ · Vitor Medeiros Almeida² · Evandro Ares de Araujo¹ · Vasily Piyadov¹ · Maria Auxiliadora Morim Santos¹ · Sandro Roberto Marana² · Igor Polikarpov¹

Published online: 26 September 2016
© Springer Science+Business Media New York 2016

Abstract The marine bacteria *Saccharophagus degradans* (also known as *Microbulbifer degradans*), are rod-shaped and gram-negative motile γ -proteobacteria, capable of both degrading a variety of complex polysaccharides and fermenting monosaccharides into ethanol. In order to obtain insights into structure–function relationships of the enzymes, involved in these biochemical processes, we characterized a *S. degradans* β -glycosidase from glycoside hydrolase family 1 (SdBgl1B). SdBgl1B has the optimum pH of 6.0 and a melting temperature T_m of approximately 50 °C. The enzyme has high specificity toward short D-glucose saccharides with β -linkages with the following preferences β -1,3 > β -1,4 $\gg$ β -1,6. The enzyme kinetic

parameters, obtained using artificial substrates *p*- β -NPGlu and *p*- β -NPFuc and also the disaccharides cellobiose, gentiobiose and laminaribiose, revealed SdBgl1B preference for *p*- β -NPGlu and laminaribiose, which indicates its affinity for glucose and also preference for β -1,3 linkages. To better understand structural basis of the enzyme activity its 3D model was built and analysed. The 3D model fits well into the experimentally retrieved low-resolution SAXS-based envelope of the enzyme, confirming monomeric state of SdBgl1B in solution.

Keywords Glycosidase hydrolase family 1 · *Saccharophagus degradans* · Enzymatic characterization · Modelling · SAXS structure

✉ Igor Polikarpov
ipolikarpov@ifsc.usp.br

Hevila Brognaro
hbrognaro@usp.br

Vitor Medeiros Almeida
vitorma3@gmail.com

Evandro Ares de Araujo
evandro.ufpa.fisica@gmail.com

Vasily Piyadov
piyadov@usp.br

Maria Auxiliadora Morim Santos
santosma@ifsc.usp.br

Sandro Roberto Marana
srmarana@iq.usp.br

¹ Instituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador São Carlense 400, São Carlos, SP 13566-590, Brazil

² Instituto de Química, Universidade de São Paulo, Avenida Prof. Lineu Prestes, 748, Bloco 10, Sala 1054, São Paulo, SP 05508-900, Brazil

Introduction

Deconstruction of lignocellulosic biomass is one of the most important natural recycling processes of plant materials on Earth. Large amounts of biomass available for bioconversion promoted ever-growing interest in the development of second generation biofuel technologies, green chemistry and sustainable materials technologies [1]. To achieve this, distinct biomass and/or agricultural wastes have to be transformed and hydrolyzed enzymatically to produce platform molecules (such as glucose, for example), which can be further converted into liquid biofuels and sustainable chemical products [2]. However identification, characterization and optimizations of the enzymes useful in such transformations still remain a bottleneck for their utilization in the bioprocesses [3, 4]. Therefore, during the last years, a number of suitable microorganisms and enzymes useful for this purpose have been extensively studied.

A.6 Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism

SCIENTIFIC REPORTS

OPEN

Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism

Received: 21 October 2015

Accepted: 25 February 2016

Published: 01 April 2016

Marcelo V. Liberato^{1,*}, Rodrigo L. Silveira^{2,*}, Érica T. Prates², Evandro A. de Araujo¹, Vanessa O. A. Pellegrini¹, Cesar M. Camilo¹, Marco A. Kadowaki¹, Mario de O. Neto³, Alexander Popov⁴, Munir S. Skaf² & Igor Polikarpov¹

Glycoside hydrolases (GHs) play fundamental roles in the decomposition of lignocellulosic biomaterials. Here, we report the full-length structure of a cellulase from *Bacillus licheniformis* (BICel5B), a member of the GH5 subfamily 4 that is entirely dependent on its two ancillary modules (Ig-like module and CBM46) for catalytic activity. Using X-ray crystallography, small-angle X-ray scattering and molecular dynamics simulations, we propose that the C-terminal CBM46 caps the distal N-terminal catalytic domain (CD) to establish a fully functional active site via a combination of large-scale multidomain conformational selection and induced-fit mechanisms. The Ig-like module is pivoting the packing and unpacking motions of CBM46 relative to CD in the assembly of the binding subsite. This is the first example of a multidomain GH relying on large amplitude motions of the CBM46 for assembly of the catalytically competent form of the enzyme.

The production of biofuels from renewable sources is an important element of the global strategy for generating sustainable energy with reduced environmental impact. Current technologies for obtaining liquid biofuels and green chemicals rely on the enzymatic digestion of lignocellulosic biomass from a variety of feedstocks¹. Plant biomass—the most abundant source of carbohydrates on Earth—is primarily composed of cellulose microfibrils surrounded by a hydrated heteropolymeric matrix of hemicellulose and lignin. Plant biomass may be subjected to thermo-chemical pretreatments and enzymatic reactions to produce soluble fermentable sugars.

The canonical model of hydrolytic degradation of cellulose requires at least three classes of enzymes¹. Cellobiohydrolases (CBHs) processively cleave the glycosidic bonds at the reducing and non-reducing ends of cellulose chains in crystalline regions to produce cellobiose. Endoglucanases (EGs) introduce random cuts in the amorphous regions of cellulose and create new chain extremities for CBH attack; thus, these enzymes act synergistically. The released cellobiose molecules are then enzymatically converted into glucose by β -glucosidases.

The molecular architecture of glycoside hydrolases (GHs) frequently consists of a catalytic domain (CD), where hydrolysis occurs, and one or more ancillary modules (AMs), which are usually connected by less structured linkers. The most common type of AMs are carbohydrate-binding modules (CBMs), which are able to recognize and bind specific carbohydrate chains². Generally distinct and independent structural domains, the CBMs facilitate carbohydrate hydrolysis by increasing the local concentration of enzymes at the surface of insoluble substrates, thereby targeting the CD component to its cognate ligands. CBMs might also disrupt the crystalline structure of cellulose microfibrils, although the underlying mechanism remains poorly understood. Thus, CBMs enhance the accessibility of CDs to carbohydrate chains to improve enzymatic activity, making them important candidates for the development of effective biomass-degrading enzymes in industrial settings.

Although there are examples of active GHs that lack AMs^{3,4}, the majority of the enzymes depend on AMs for activity². In several cases, CBMs were shown to extend and complement the CD substrate-binding site in

¹São Carlos Institute of Physics, University of São Paulo, São Carlos 13566-590 São Paulo, Brazil. ²Institute of Chemistry, University of Campinas, Campinas 13084-862, São Paulo, Brazil. ³Institute of Bioscience, State University of São Paulo, Botucatu 18618-970, São Paulo, Brazil. ⁴European Synchrotron Radiation Facility, Grenoble CS40220, France. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.S.S. (email: skaf@iqm.unicamp.br) or I.P. (email: ipolikarpov@ifsc.usp.br)

A.7 Expression, purification, crystallization and preliminary X-ray diffraction analysis of *Aspergillus terreus* endo- β -1,4-glucanase from glycoside hydrolase family 12

Fernando Segato,[‡] Gabriela L. Berto,[‡] Evandro Ares de Araújo, João Renato Muniz and Igor Polikarpov*

Instituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador Saocarlense 400, 13566-590 São Carlos-SP, Brazil

[‡] These authors contributed equally.

Correspondence e-mail: ipolikarpov@ifsc.usp.br

Received 13 October 2013

Accepted 30 December 2013

Expression, purification, crystallization and preliminary X-ray diffraction analysis of *Aspergillus terreus* endo- β -1,4-glucanase from glycoside hydrolase family 12

Endoglucanases are important enzymes that are involved in the modification and degradation of cellulose. Filamentous fungi such as *Aspergillus terreus* are effective biomass degraders in nature owing to their capacity to produce an enzymatic arsenal of glycoside hydrolases, including endoglucanase from glycoside hydrolase family 12 (GH12). The *A. terreus* GH12 endoglucanase was cloned and overexpressed in *A. nidulans*, purified and crystallized. A single crystal was obtained from a solution consisting of 2 M ammonium sulfate, 5% (v/v) 2-propanol. X-ray diffraction data were collected to a resolution of 1.85 Å using synchrotron radiation and a preliminary molecular-replacement solution was obtained in the trigonal space group $P3_221$. The unit-cell parameters were $a = b = 103.24$, $c = 48.96$ Å.

1. Introduction

Concerns over global climate change, increasing demand for energy and the reduction in oil supplies have fostered the need for the development of alternative and renewable energy resources which can replace fossil fuels. Biomass is an abundant feedstock, which can be converted into different types of energy sources and chemicals. It consists of three main components, cellulose, hemicellulose and lignin, the first two of which are composed of carbohydrates that can be depolymerized and fermented to ethanol and/or converted to a number of chemical products (Himmel *et al.*, 2007). The main technological barrier to biomass utilization is the absence of low-cost technologies to decrease the recalcitrance of cellulosic biomass. The most promising strategy to overcome this barrier involves the use of mixtures of microbial cellulolytic enzymes to break down biomass (Lynd *et al.*, 2002; Wang *et al.*, 2011).

Cellulases are glycoside hydrolases found in at least 12 enzyme families (<http://www.cazy.org>) which can be broadly separated into endoglucanases (EC 3.2.1.4), which randomly hydrolyze internal glycoside bonds, exoglucanases (EC 3.2.1.91/176), which split off cellobiose from cellulose termini, and β -glucosidases (EC 3.2.1.21), which hydrolyze cellobiose and oligomers to produce glucose (Zhang & Lynd, 2004). The endoglucanases from GH (glycoside hydrolase) family 12 (GH12) hydrolyze the β -1,4-glycosidic bond in cellulose (Henrissat, 1991) and are widely used in industrial processes and in animal feed to improve the digestibility of glucans (Picart *et al.*, 2012).

Aspergillus terreus is an important filamentous fungus with significant commercial relevance in industrial biotechnology. This fungus is frequently isolated from agricultural wastes and cellulosic materials, and *A. terreus* enzymes have been extensively studied and utilized in various bioprocesses (Han *et al.*, 2010; Chidi *et al.*, 2008). Furthermore, *A. terreus* has been shown to produce secondary metabolites of significant industrial interest using biomass as a substrate (Jahromi *et al.*, 2012). In this work, we report the cloning, heterologous overexpression, purification and crystallization of an endoglucanase member of GH family 12 from *A. terreus* strain NIH2624.

© 2014 International Union of Crystallography
All rights reserved