

RODRIGO LEANDRO SILVEIRA

ASPECTOS MOLECULARES DA DEGRADAÇÃO DE BIOMASSA LIGNOCELULÓSICA: DINÂMICA DE ENZIMAS E NANOARQUITETURA DE PAREDES CELULARES DE PLANTAS

CAMPINAS 2014

ii



UNIVERSIDADE ESTADUAL DE CAMPINAS INSTITUTO DE QUÍMICA

RODRIGO LEANDRO SILVEIRA

ASPECTOS MOLECULARES DA DEGRADAÇÃO DE BIOMASSA LIGNOCELULÓSICA: DINÂMICA DE ENZIMAS E NANOARQUITETURA DE PAREDES CELULARES DE PLANTAS

ORIENTADOR: PROF. DR. MUNIR SALOMÃO SKAF

TESE DE DOUTORADO APRESENTADA AO INSTITUTO DE QUÍMICA DA UNICAMP PARA OBTENÇÃO DO TÍTULO DE DOUTOR EM CIÊNCIAS

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA POR RODRIGO LEANDRO SILVEIRA, E ORIENTADA PELO PROF. DR. MUNIR SALOMÃO SKAF.

ASSINATURA DO ORIENTADOR

CAMPINAS 2014 Ficha catalográfica Universidade Estadual de Campinas Biblioteca do Instituto de Química Simone Lucas Gonçalves de Oliveira - CRB 8/8144

Silveira, Rodrigo Leandro, 1986-

Si39a Aspectos moleculares da degradação de biomassa lignocelulósica : dinâmica de enzimas e nanoarquitetura de paredes celulares de plantas / Rodrigo Leandro Silveira. – Campinas, SP : [s.n], 2014.

Orientador: Munir Salomão Skaf Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Química.

1. Glicosídeo hidrolases. 2. Expansinas. 3. Parede celular. 4. Dinâmica molecular. 5. Teoria de equações integrais. I. Skaf, Munir Salomão. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título outro idioma: Molecular aspects of the lignocellulosic biomass degradation : dynamics of enzymes and plant cell wall nanoarchitecture Palavras-chave em inglês: Glycoside hydrolases Expansins Cell wall Molecular dynamics Integral equation theory Área de concentração: Físico-Química Titulação: Doutor em Ciências Banca examinadora: Munir Salomão Skaf [orientador] Richard John Ward Roberto Dias Lins Neto

Nelson Henrique Morgon Data da defesa: 12-05-2014

Camila Alves de Rezende

Programa de Pós-Graduação: Química

"Imagination is more important than knowledge. For knowledge is limited to all we know and understand, while imagination embraces the entire world."

"The value of a college education is not the learning of many facts, but the training of the mind to think."

Albert Einstein

Agradecimentos

Agradeço a todas as pessoas e instituições que contribuiram para meu crescimento profissinal e pessoal ao longo dos anos de graduação e pós-graduação. Em particular, sou grato:

- à minha família, pelo apoio;
- ao Prof. Munir Skaf, pela orientação;
- a todos os membros do grupo, pela amizade;
- ao Dr. Andriy Kovalenko, pela orientação no exterior;
- ao Dr. Stanislav Stoyanov, pela co-orientação no exterior;
- à Fapesp, pelo financiamento no Brasil;
- à Capes, pelo financiamento no exterior;
- ao Instituto de Química da Unicamp e à Unicamp;
- ao National Institute for Nanotechnology (NINT) e à University of Alberta (UofA), pela admissão no exterior;
- aos membros da comissão examinadora.

Curriculum Vitae

• Informações pessoais

Nome:	Rodrigo Leandro Silveira
Data de nascimento:	03/06/1986
Currículo Lattes:	http://lattes.cnpq.br/8170241379248486

Graduação

2004-2008 Bacharelado em Engenharia de Alimentos Universidade Federal de Lavras

• Mestrado

2008-2010

Mestrado em Química (área de concentração: físico-química) Dissertação: Simulações de dinâmica molecular do receptor ativado de proliferadores de peroxissomos isoforma γ Orientador: Prof. Dr. Munir S. Skaf Instituto de Química, Universidade Estadual de Campinas Financiamento: CNPq

Doutorado

2010-2014

Doutorado em Química (área de concentração: físico-química) Tese: Aspectos moleculares da degradação de biomassa lignocelulósica: dinâmica de enzimas e nanoarquitetura de parede celular de plantas Orientador: Prof. Dr. Munir S. Skaf Instituto de Química, Universidade Estadual de Campinas Financiamento: Fapesp

• Doutorado-sanduíche

02/2012-10/2012 Orientador: Dr. Andriy Kovalenko Co-orientador: Dr. Stanislav R. Stoyanov National Institute for Nanotechnology, University of Alberta, Edmonton, Canadá Financiamento: Capes

• Prêmios e distinções

- 1. (2010) Primeiro colocado na seleção para o ingresso ao doutorado, Instituto de Química, Unicamp.
- (2008) Terceiro colocado na seleção para o ingresso ao mestrado, Instituto de Química, Unicamp.
- (2008) Mérito acadêmico de primeiro lugar entre os formandos de Engenharia de Alimentos de 07/2008, Universidade Federal de Lavras.

Trabalhos publicados em periódicos de circulação internacional

- R. L. Silveira, S. R. Stoyanov, S. Gusarov, M. S. Skaf, A. Kovalenko. Plant biomass recalcitrance: effect of hemicellulose composition on nanoscale forces that control the cell wall strength. J. Am. Chem. Soc. 135, 19048-19051, 2013.
- L. C. Textor, F. Colussi, R. L. Silveira, V. Serpa, F. M. Squina, N. Pereira Jr., M. S. Skaf, I. Polikarpov. Cellobiohydrolase I from *Trichoderma harzianum*: deciphering structural features of cellobiohydrolase catalytic activity. *FEBS J.* 280, 56-69, 2013.

- E. T. Prates, I. Stanković, R. L. Silveira, M. V. Liberato, F. H. Silva, N. Pereira Jr., I. Polikarpov, M. S. Skaf. X-ray structure and molecular dynamics simulations of endoglucanase 3 from *Trichoderma harzianum*: structural organization and substrate recognition by endoglucanases that lack cellulose binding module. *Plos One* 8, e59069, 2013.
- R. L. Silveira, J. Martínez, M. S. Skaf, L. Martínez. Enzyme microheterogeneous hydration and stabilization in supercritical carbon dioxide. *J. Phys. Chem. B* 116, 5671-5678, 2012.
- M. V. Liberato, A. S. Nascimento, S. Ayers, J. Z. Lin, A. Cvoro, R. L. Silveira, L. Martínez, P. C. T. Souza, D. Saidemberg, T. Deng, A. A. Amato, M. Togashi, W. A. Hsueh, K. Phillips, M. S. Palma, F. A. R. Neves, M. S. Skaf, P. Webb, I. Polikarpov. Medium chain fatty acids are selective peroxisome activated receptor (PPAR) γ activators and Pan-PPAR partial agonists. *Plos One* 7, e36297, 2012.
- A. C. Puhl, A. Bernardes, R. L. Silveira, J. Yuan, J. L. O. Campos, D. Saidemberg, M. S. Palma, A. Cvoro, S. D. Ayers, P. Webb, P. S. Reinach, M. S. Skaf, I. Polikarpov. Mode of PPARγ activation by luteolin. *Mol. Pharmacol.*, 81, 788-799, 2012.
- A. A. Amato, S. Rajagopalan, G. Z. Lin, B. M. Carvalho, A. C. M. Figueira, J. Lu, S. D. Ayers, M. Mottin, R. L. Silveira, P. C. T. Souza, R. H. V. Mourão, M. J. A. Saad, M. Togashi, L. A. Simeoni, D. S. P. Abdalla, M. S. Skaf, I. Polikarpov, M. C. A. Lima, S. L. Galdino, R. G. Brennan, J. D. Baxter, I. R. Pitta, P. Webb, K. J. Phillips, F. A. R. Neves. GQ-16, a novel peroxisome proliferator-activated receptor γ (PPARγ) ligand, promotes insulin sensitization without weight gain. J. Biol. Chem. 287, 28169-28179, 2012.
- L. Bleicher, E. T. Prates, T. C. F. Gomes, R. L. Silveira, A. S. Nascimento, A. L. Rojas, A. Golubev, L. Martínez, M. S. Skaf, I. Polikarpov. Molecular basis of the thermostability and thermophilicity of laminarinases: x-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *J. Phys. Chem. B* **115**, 7940-7949, 2011.
- A, Hansson, P. C. T. Souza, R. L. Silveira, L. Martínez, M. S. Skaf. CHARMM force field parametrization of rosiglitazone. *Int. J. Quantum Chem.* 111, 1346-1354, 2011.

• Apresentações em congressos

- Silveira, R. L., Stoyanov, S. R., Kovalenko, A., Skaf, M. S. Computational studies of molecular aspects of the lignocellulosic biomass degradation. I Workshop - CEPID Center for Computational Engineering & Sciences. Campinas, São Paulo, 2014.
- Silveira, R. L., Stoyanov, S. R., Kovalenko, A., Skaf, M. S. Lignocellulosic recalcitrance: the nanoscale forces that control cell wall strength. Gordon Research Conference: Cellulosomes, cellulases & other carbohydrate modifying enzymes. Andover, New Hampshire, Estados Unidos, 2013.
- Silveira, R. L., Stoyanov, S. R., Gusarov, S., Skaf, M. S., Kovalenko, A. Molecular aspects of cell walls investigated by the 3D-RISM-KH integral equation theory. IV Workshop of Molecular Modeling. Itacuruçá, Rio de Janeiro, 2013.
- Silveira, R. L., Stoyanov, S. R., Gusarov, S., Skaf, M. S., Kovalenko, A. Multiscale modeling of the microstructure-recalcitrance relatioship in plant cell walls. 95th Canadian Chemistry Conference and Exhibition. Calgary, Alberta, Canadá, 2012. (Apresentação oral)
- Silveira, R. L., Stoyanov, S. R., Gusarov, S., Skaf, M. S., Kovalenko, A. Plant biomass for biofuel production: computational studies of hydrolases and microstructure. Advances in sustainable chemistry symposium. Edmonton, Alberta, Canadá, 2012.
- Silveira, R. L., Skaf, M. S. Investigation of molecular aspects of cellobiohydrolases by molecular simulations. Coarse grain methods for biomolecular simulations, Montevideo, Uruguai, 2011. (apresentação oral)
- Silveira, R. L., Skaf, M. S. Salt bridges control the kinetic mechanism of aldose reductase: evidence from molecular dynamics simulations. XVI Simpósio Brasileiro de Química Teórica, Ouro Preto, Minas Gerais, 2011.
- Silveira, R. L., Martínez, L., Polikarpov, I. e Skaf, M. S. Estudo computacional do complexo PPARγ+ligante por dinâmica molecular de equilíbrio. XV Simpósio Brasileiro de Química Teórica, Poços de Caldas, Minas Gerais, 2009.
- Silveira, R. L., Souza, S. M. e Costa Jr., A. T. Determinação do coeficiente de autodifusão do argônio por simulações de dinâmica molecular. XXI Congresso de iniciação científica, Universidade Federal de Lavras, Lavras, Minas Gerais, 2008.

Resumo

A produção de bioetanol a partir da biomassa lignocelulósica integra processos físico-químicos e enzimáticos que comprometem sua viabilidade econômica. A biomassa possui uma estrutura recalcitrante composta de celulose, hemicelulose Tal estrutura, bem como os mecanismos de ação das enzimas, não e lignina. são bem compreendidos. Nesta tese, simulações de dinâmica molecular e a teoria mecânico-estatística 3D-RISM foram utilizadas para investigar aspectos moleculares da degradação de biomassa, incluindo: (1) dinâmica estrutural de celulases; (2) base molecular da termofilicidade de laminarinases; (3) disrupção não-hidrolítica de biomassa por expansinas; (4) nanoarquitetura da parede celular primária; e (5) forças termodinâmicas em paredes celulares secundárias. No tópico (1), observou-se que a acessibilidade de celulases ao substrato pode ser modulada por alterações na estrutura primária, com consequências para a atividade enzimática. Observou-se também que a inibição por produto está relacionada a alterações conformacionais de resíduos próximos ao sítio de ligação. Adicionalmente, alterações na dinâmica intrínseca das enzimas ocorrem conforme a etapa do processo de hidrólise. No tópico (2), os resultados mostraram que a conformação do sítio de ligação ao substrato de laminarinases é sensível a variações de temperatura. No tópico (3), observou-se que a expansina pode transladar sobre a superfície da celulose e induzir torções em cadeias de glucano, sugerindo a possibilidade de romper ligações de hidrogênio celulose-celulose e/ou celulose-xiloglucano como um zíper. No tópico (4), observou-se que a agregação de nanofibrilas de celulose se dá através de suas faces hidrofílicas e que a presença de hemicelulose estabiliza tal agregação. No tópico (5), os resultados mostraram que as forças de coesão da parede celular secundária são de natureza entrópica e que a composição química da lignina modula as interações lignina-lignina e lignina-hemicelulose.

Abstract

Biofuel production from lignocellulosic biomass involves physico-chemical and enzymatic processes that challenge its economic viability. The lignocellulosic biomass is recalcitrant against degradation and is made up of cellulose, hemicellulose and lignin. This structure and the enzyme mechanisms are not fully understood. In this thesis, molecular dynamics simulations and the statistical mechanical theory 3D-RISM were employed to assess molecular aspects of the biomass degradation, including: (1) structural dynamics of cellulases; (2) molecular basis of the thermophilicity of laminarinases; (3) non-hydrolytic disruption of biomass by expansins; (4) primary cell wall nanoarchitecture; and (5) thermodynamic forces in secondary cell walls. In the topic (1), we observed that cellulase substrate accessibility can be modulated through changes in its primary structure, with consequences to the enzymatic activity. Moreover, the product inhibition is related to conformational changes of residues located close to the substrate binding site. In addition, changes of the intrinsic dynamics allow cellulases change their function according to the hydrolysis step. In the topic (2), we show that the substrate binding site conformation of laminarinases is sensitive to temperature variations. In the topic (3), we observed that the expansin can translade over the cellulose surface and induce torsions in free glucan chains, suggesting the possibility of disruption of cellulose-cellulose and cellulose-xyloglucan hydrogen bonds as a ziper. In the topic (4), the results showed that aggregation of cellulose nanofibrils takes place through their hydrophilic face and that hemicellulose plays roles in stabilizing such aggregation. In the topic (5), we observed that the cohesion forces within the secondary cell wall are of entropic origin and that the lignin chemical composition modulates the lignin-lignin and lignin-hemicellulose interactions.

Sumário

Li	sta de A	Abreviatu	iras	XXV
Li	sta de H	Figuras		xxvii
Li	sta de 7	F abelas		xliii
I	Introd	lução		1
1	Intro	odução		3
2	Degr	radação d	e biomassa lignocelulósica	7
	2.1	A estrut	ura microscópica da biomassa: parede celular	. 8
		2.1.1	Celulose	. 9
		2.1.2	Hemicelulose	. 13
		2.1.3	Lignina	. 14
	2.2	Pré-trat	amentos	. 16
	2.3	Hidrólis	e enzimática	. 18
		2.3.1	Mecanismo de hidrólise	. 19
		2.3.2	Hidrólise da celulose: sinergismo clássico	. 21
		2.3.3	Além do sinergismo clássico	. 22
II	Metod	lologias c	omputacionais	25
3	Dinâ	imica mol	lecular de biomoléculas	27

	3.1	Algoritm	os	29
	3.2	Campo d	le força	31
	3.3	Condiçõe	s de contorno periódicas	34
	3.4	Aspectos	práticos	38
	3.5	Análises	-	39
		3.5.1	Flutuações estruturais	39
		3.5.2	Correlações dinâmicas cruzadas	40
		3.5.3	Análise de componentes principais	41
4	Teor	ia de Equ	ações Integrais	45
	4.1	Funções o	de distribuição de líquidos monoatômicos	46
	4.2	Equação	de Ornstein-Zernike molecular e RISM	49
	4.3	3D-RISM	I: sistemas soluto-solvente	51
	4.4	Relações	de fechamento	53
				54
	4.5	Termodir	namica de solvatação em 3D-RISM/KH	94
111	4.5 Dinân	Termodir nica molec	ular de hidrolases e proteínas disruptivas	54 59
III 5	4.5 Dinân Dinâ	Termodir nica molec amica estru	namica de solvatação em 3D-RISM/KH	59 61
111 5	4.5 Dinân Dinâ 5.1	Termodir nica molec amica estru Introduçã	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases	54 59 61 61
III 5	4.5 Dinân Dinâ 5.1	Termodir nica molec amica estru Introduçă 5.1.1	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão	 54 59 61 61 61
111 5	4.5 Dinân Dinâ 5.1	Termodir nica molec amica estru Introduçã 5.1.1 5.1.2	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão	 54 59 61 61 61 65
III 5	4.5 Dinân Dinâ 5.1 5.2	Termodir nica molec amica estru Introduçã 5.1.1 5.1.2 A Celobio	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum	 54 59 61 61 61 65 67
III 5	4.5 Dinân Dinâ 5.1 5.2	Termodir nica molec imica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum	 54 59 61 61 61 65 67 67
III 5	4.5 Dinân Dinâ 5.1 5.2	Termodir nica molec amica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1 5.2.2	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum Introdução Metodologia	 54 59 61 61 61 65 67 67 68
111 5	4.5 Dinân Dinâ 5.1 5.2	Termodir nica molec amica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1 5.2.2 5.2.3	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum Introdução Resultados	59 61 61 61 65 67 67 68 70
111 5	4.5 Dinân Dinâ 5.1 5.2	Termodir nica molec amica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1 5.2.2 5.2.3 5.2.3 5.2.4	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum Introdução Resultados Discussão	59 61 61 61 65 67 67 68 70 75
111 5	4.5 Dinân Dinâ 5.1 5.2	Termodir nica molec imica estru Introduçã 5.1.1 5.1.2 A Celobio 5.2.1 5.2.2 5.2.3 5.2.3 5.2.4 Dinâmica	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum Introdução Metodologia Discussão a intrínseca da TrCel7A	59 61 61 61 65 67 67 68 70 75 78
111 5	4.5 Dinân Dinâ 5.1 5.2 5.3	Termodir nica molec imica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1 5.2.2 5.2.3 5.2.3 5.2.4 Dinâmica 5.3.1	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum Introdução Metodologia Discussão a intrínseca da TrCel7A	59 61 61 61 65 67 68 70 75 78 78
111 5	 4.5 Dinân Dinâ 5.1 5.2 5.3 	Termodir nica molec imica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1 5.2.2 5.2.3 5.2.4 Dinâmica 5.3.1 5.3.2	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão ão Estrutura e função Ciclo catalítico ohidrolase Cel7A de Trichoderma harzianum Introdução Resultados Discussão a intrínseca da TrCel7A Metodologia Metodologia	59 61 61 61 65 67 68 70 75 78 78 81
111 5	 4.5 Dinân Dinâ 5.1 5.2 5.3 	Termodir nica molec amica estru Introduçã 5.1.1 5.1.2 A Celobia 5.2.1 5.2.2 5.2.3 5.2.4 Dinâmica 5.3.1 5.3.2 5.3.3	ular de hidrolases e proteínas disruptivas utural de celobiohidrolases ão ão Estrutura e função Ciclo catalítico Ohidrolase Cel7A de Trichoderma harzianum Introdução Metodologia a intrínseca da TrCel7A Metodologia A metodologia <t< td=""><td>59 61 61 61 65 67 67 68 70 75 78 78 81 81</td></t<>	59 61 61 61 65 67 67 68 70 75 78 78 81 81

	5.4	Base estrut	ural da inibição da TrCel7A por celobiose	91
		5.4.1	Introdução	91
		5.4.2	Metodologia	92
		5.4.3	Resultados e discussão	93
	5.5	Consideraç	ões finais	96
6	Base	molecular	da termofilicidade de laminarinases	97
	6.1	Introdução		97
	6.2	Metodologi	a	100
	6.3	Resultados	e discussão	101
		6.3.1	Termoestabilidade	101
		6.3.2	Termofilicidade	103
		6.3.3	Populações dos estados do sítio de ligação	106
	6.4	Consideraç	ões finais	108
7	Disru	ıpção não-h	idrolítica de biomassa por expansinas	111
	7.1	Introdução		111
				111
		7.1.1	Expansina EXLX1 de <i>Bacillus subtilis</i>	111 113
	7.2	7.1.1 Metodologi	Expansina EXLX1 de <i>Bacillus subtilis</i>	111 113 116
	7.2	7.1.1 Metodologi 7.2.1	Expansina EXLX1 de Bacillus subtilis	111113116116
	7.2	7.1.1 Metodologi 7.2.1 7.2.2	Expansina EXLX1 de Bacillus subtilis	 111 113 116 116 118
	7.2 7.3	7.1.1Metodologi7.2.17.2.2Resultados	Expansina EXLX1 de Bacillus subtilis	 111 113 116 116 118 119
	7.2 7.3	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 	Expansina EXLX1 de Bacillus subtilis	 111 113 116 116 118 119 119
	7.2 7.3	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 7.3.2 	Expansina EXLX1 de Bacillus subtilis	111 113 116 116 118 119 119 121
	7.2	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 7.3.2 7.3.3 	Expansina EXLX1 de <i>Bacillus subtilis</i>	111 113 116 116 118 119 119 121
	7.2	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 7.3.2 7.3.3 	Expansina EXLX1 de <i>Bacillus subtilis</i>	 111 113 116 116 118 119 119 121 122
	7.2	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 7.3.2 7.3.3 7.3.4 	Expansina EXLX1 de Bacillus subtilis a Construção dos sistemas Detalhes computacionais Detalhes computacionais Adsorção da EXLX1 sobre celulose cristalina Translação longitudinal da EXLX1 sobre a celulose Centro de torção: interação entre EXLX1 e oligos- sacarídeo Mutante Asp82Asn	 111 113 116 116 118 119 119 121 122 122 127
	7.2	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 7.3.2 7.3.3 7.3.4 7.3.5 	Expansina EXLX1 de Bacillus subtilis . a . Construção dos sistemas . Detalhes computacionais . Adsorção da EXLX1 sobre celulose cristalina . Adsorção da EXLX1 sobre celulose cristalina . Translação longitudinal da EXLX1 sobre a celulose Centro de torção: interação entre EXLX1 e oligos-sacarídeo Mutante Asp82Asn . . Mutante Tyr73Ala . .	 111 113 116 116 118 119 119 121 122 127 128
	7.27.37.4	 7.1.1 Metodologi 7.2.1 7.2.2 Resultados 7.3.1 7.3.2 7.3.3 7.3.4 7.3.5 Discussão 	Expansina EXLX1 de Bacillus subtilis a Construção dos sistemas Detalhes computacionais Adsorção da EXLX1 sobre celulose cristalina Adsorção da EXLX1 sobre celulose cristalina Translação longitudinal da EXLX1 sobre a celulose Centro de torção: interação entre EXLX1 e oligos-sacarídeo Mutante Asp82Asn Mutante Tyr73Ala	 111 113 116 116 118 119 119 121 122 127 128 133

IV	Aspec	tos molec	culares de paredes celulares de plantas	139
8	Nan	oarquitet	ura da parede celular primária	141
	8.1	Introdu	ção	. 141
	8.2	Metodol	logia	. 142
		8.2.1	Modelos moleculares e procedimentos gerais \ldots	. 142
		8.2.2	Detalhes computacionais	. 147
	8.3	Resultad	$\log \ldots \ldots$. 149
		8.3.1	Auto-agregação de nanofibrilas de celulose	. 149
		8.3.2	Papel da hemicelulose nas forças de agregação	. 151
		8.3.3	Especificidade das interações celulose-hemicelulose	. 154
	8.4	Discussã	ão	. 157
	8.5	Conside	rações finais	. 160
9	Forç	as termo	dinâmicas em paredes celulares secundárias	161
	9.1	Introduc	ção	. 161
	9.2	Metodol	logia	. 164
		9.2.1	Modelos moleculares e procedimentos gerais \ldots	. 164
		9.2.2	Detalhes computacionais	. 165
	9.3	Resultad	dos e discussão	. 167
		9.3.1	Interações de monolignóis com lignina e hemicelule	ose 167
		9.3.2	Efeito da conformação da hemicelulose e lignina .	. 170
		9.3.3	Forças termodinâmicas de solvatação	. 172
		9.3.4	Aspectos estruturais das interações \ldots	. 176
		9.3.5	Ligações de hidrogênio	. 178
	9.4	Conside	rações finais	. 181
V	Conclu	usões		183
10	Cone	clusões e	comentários finais	185
\mathbf{A}	Nom	enclatura	a e estrutura dos aminoácidos	189

191

Lista de Abreviaturas

3D-RISM	RISM tridimensional
AFEX	expansão de fibras por amônia
aNLam	laminarinase de alcaliphilic Nocardipsis sp.
apo-TrCel7A	TrCel7A na ausência de substrato
СВН	celobiohidrolase
CBM	módulo de ligação ao carboidrato
CCD	domínio catalítico
DRISM	dielectrically consistent RISM
EG	endoglucanase
EXLX1	expansina de Bacillus subtilis
G	guaiacil
GH	glicosídeo hidrolase
Н	p-hidroxifenil
holo-TrCel7A	TrCel7A na presença de substrato
KH	relação de fechamento de Kovalenko-Hirata
LCC	complexo lignina-carboidrato
LPMO	lytic polysaccharide monooxygenase
OZ	Ornstein-Zernike
PcCel7D	celobiohidrolase Cel7D de Phanerochaete chrysosporium

PcLam	laminarinase de Phanerochaete chrysosporium
PCA	análise de componentes principais
PME	particle mesh Ewald
PMF	potencial de força média
RISM	reference interaction site model
RmLam	laminarinase de Rhodothermus marinus
rmsd	raiz quadrada do deslocamento médio quadrático
rmsf	raiz quadrada da flutuação média quadrática
S	siringil
SANS	espalhamento de nêutrons a baixo ângulo
SAXS	espalhamento de raios X a baixo ângulo
SFED	densidade de energia livre de solvatação
TeCel7B	celobiohidrolase Cel7B de <i>Talaromyces emersonii</i>
TrCel7A	celobiohidrolase Cel7A de Trichoderma reesei
ThCel7A	celobiohidrolase Cel7A de Trichoderma harzianum
TrCel7A-SP	TrCel7A na presença de substrato e produto

Lista de Figuras

2.1	Arquitetura da parede celular secundária. Fibras de celulose são	
	envolvidas por uma matriz amorfa de lignina e hemicelulose, que	
	podem formar os complexos lignina-carboid rato (LCC). Adaptado	
	de [4]	9
2.2	Estrutura da cristalina da celulos e $\mathrm{I}\beta,$ mostrando (a) ligações de	
	hidrogênio (linhas tracejadas) intramoleculares e entre cadeias de	
	glucano e (b) o empacotamento hidrofóbico de camadas de glucano.	11
2.3	(a) Celulose sintases arranjadas em grupos de diferentes isofor-	
	mas ($\alpha 1$, $\alpha 2$ e $\beta 1$). (b) Agrupamento de 36 celulose sintases em	
	forma hexagonal, formando uma roseta, que sintetiza uma fibrila	
	elementar de celulose. (c) Arranjo de rosetas, unidade que forma	
	uma macrofibrila de celulose. Adaptado de [14]	12
2.4	(a) Estrutura de uma fibrila elementar de celulose. (b) Macrofi-	
	brila de celulose formada por 7 fibrilas elementares	12
2.5	Estrutura representativa de cada grupo de hemicelulose: xiloglu-	
	cano, xilano e manano. Adaptado de [21] com permissão da John	
	Wiley and Sons.	13
2.6	(a) Síntese e estrutura da lignina. Adaptado de [26]. (b) Estrutura	
	de um complexo lignina-carboid rato (LCC). Adaptado de [4]. $\ .$.	15
2.7	Mecanismo de hidrólise com inversão da configuração do carbono	
	anomérico	20
2.8	Mecanismo de hidrólise com retenção da configuração do carbono	
	anomérico.	21

2.9	Sinergismo clássico entre endoglucanases (EG), celobiohidrolases (CBH I e CBH II) e β -glicosidades (BG). Os círculos vazios repre-
	sentam resíduos de glicose na estrutura da celulose, e os círculos
	cheios representam extremidades redutoras. Adaptado de [68]
	com permissão da Elsevier
3.1	Termos da parte covalente do campo de força: (a) estiramento de
	ligações; (b) distorção angular; (c) torção diedral. Adaptado de [98]. 32
3.2	Potenciais de Lennard-Jones e de Coulomb
3.3	Condições de contorno periódicas. Quando um átomo cruza as fronteiras da caixa de simulação, um outro átomo entra pela face
3.4	Decomposição de cargas pontuais em densidades de carga. Adap-
	tado de $[98]$
4.1	(a) A função $g(r)$ é tal que $4\pi r^2 \rho g(r) dr$ é o número de átomos pre-
	sentes na calota esférica em cinza em torno de um átomo central
	(círculo vazio). (b) Forma de uma função de distribuição radial
	para argônio líquido, mostrando a estruturação dos átomos em
	torno de um átomo central. A distância r é plotada em unidades
	de σ (parâmetro do potencial de Lennard-Jones) 47
5.1	Modelo tridimensional da estrutura da TrCel7A, mostrando o
	CCD, o CBM e o <i>linker</i> glicosilado sobre a superfície da celulose,
	com uma cadeia de glucano ligada ao CCD para ser hidrolisada.
	Adaptado de [131]
5.2	Estrutura cristalográfica do CCD da TrCel7A (PDB: 8CEL [134]).
	As folhas β (azul) formam um canal acima do qual há vários loops
	(branco) e α -hélices (amarelo) que formam um túnel. Neste túnel,
	há resíduos polares e 4 triptofanos (vermelho), onde se liga uma
	catalíticos (alaraniado) 63

LISTA DE FIGURAS

5.3	Superfície molecular do CCD da TrCel7A, mostrando a topologia de túnel criada pelos loops.	63
5.4	Estrutura tridimensional do CBM (PDB: 1CBH [136]) da Tr-Cel7A, mostrando a superfície aromática de ligação à celulose	64
5.5	Colapso do <i>linker</i> glicosilado sobre a superfície da celulose observado em 1 μ s de dinâmica molecular a partir da estrutura ilustrada na figura 5.1. Adaptado da referência [131].	65
5.6	Ciclo catalítico da TrCel7A, mostrando o processo de complexação da cadeia polissacarídica, ciclos processivos, expulsão da celobi- ose e dessorção da enzima. Adaptado da referência [144] com permissão da Elsevier.	66
5.7	Estruturas cristalográficas da TrCel7A e ThCel7A superpostas. Entre as poucas diferenças entre as enzimas, observa-se a troca do resíduo Tyr371 (TrCel7A) pelo resíduo Ala384 (ThCel7A).	71
5.8	(a) Evolução temporal do rmsd dos resíduos Tyr247(h) (ThCel7A, em verde) e Tyr260(r) (TrCel7A, em azul) em relação às suas estruturas médias. O resíduo Tyr260(h) é mais móvel que o Tyr247(r), e ambos têm acesso a valores de rmsd corresponden- tes a dois estados (1 e 2). (b) Configurações representativas dos estados 1 e 2, na região entre os loops 5 e 6. Devido à presença da Ala384(h), o resíduo Tyr260(h) é móvel e pode abrir a região entre os loops 5 e 6 na ThCel7A. O resíduo Tyr247(r), por outro lado, é menos móvel e volumoso, o que dificulta a abertura dos loops na TrCel7A.	72
5.9	(a) Flutuações quadrática médias (msd) dos carbonos α da Th- Cel7A e TrCel7A ao longo dos 20 primeiros modos de movimento extraídos da análise de componentes principais. Observa-se que a ThCel7A exibe maiores flutuações que a TrCel7A nos dois primei- ros modos. (b) Soma cumulativa e normalizada das flutuações ao longo dos 20 primeiros modos, que capturam aproximadamente 60% do todo o movimento das orgimas	72
		10

5.10	Flutuações médias dos carbonos α (rmsf) da ThCel7A e da Tr- Cel7A. O painel superior mostra o rmsf para cada um dos resíduos e o painel inferior mostra estruturas superpostas obtidas das si- mulações e coloridas conforme a mobilidade de cada resíduo (ver escala de cores de rmsf na figura)	74
5.11	Matriz de correlações cruzadas para ThCel7A e TrCel7A. Para ambas as enzimas, observa-se movimentos correlacionados entre os pares de loops 4/6, 5/6, 4/7 e 5/7. De azul a vermelho, as cores representam correlações negativas e positivas, respectivamente	75
5.12	Superfície molecular da TrCel7A mostrando o resíduo Trp40 lo- calizado na entrada do túnel.	79
5.13	 (a) Matriz de correlações cruzadas da <i>apo</i>-TrCel7A. Pares de resíduos que exibem correlações (b) positivas e (c) negativas de módulo maior que 0,7 são conectados. Observa-se que os loops vizinhos são positivamente correlacionados e os loops opostos são negativamente correlacionados. 	83
5.14	Dinâmica essencial da <i>apo</i> -TrCel7A. (a) Flutuações quadráticas médias ao longo dos 20 primeiros modos de movimento. (b) Direção do modo de movimento dominante, indicando flutuações do tipo abre-fecha. (c) Distribuição da coordenada conformacio- nal ao longo do movimento de abre-fecha, mostrando a existência de interações que estabilizam conformações abertas e fechadas. (d) Estrutura da <i>apo</i> -TrCel7A correspondente ao máximo e mínimo da coordenada conformacional do movimento de abre-fecha	84
5.15	 (a) Matriz de correlações cruzadas da <i>holo</i>-TrCel7A. Pares de resíduos que exibem correlações (b) positivas e (c) negativas de módulo maior que 0,7 são conectados. Observa-se que a ligação do substrato extingue (b) parcialmente as correlações positivas entre loops vizinhos e extigue (c) completamente as correlações 	
	negativas entre loops opostos	85

5.16	Dinâmica essencial da <i>holo</i> -TrCel7A. (a) Flutuações quadráticas médias ao longo dos 20 primeiros modos de movimento. (b) Direção do modo de movimento dominante, indicando flutuações dos loops 1 e 4 em direções paralelas ao túnel catalítico. (c) Dis- tribuição da coordenada conformacional ao longo do movimento dos loops 1 e 4, mostrando dois estados estáveis de flutuações ao longo do túnel. (d) Estrutura da <i>holo</i> -TrCel7A correspondente ao máximo e mínimo da coordenada conformacional, destacando os loops 1 e 4	86
5.17	Ilustração da superfície de energia livre da TrCel7A antes e após a complexação do substrato, com as possíveis transições ilustradas à direita. A complexação do substrato é provável de ocorrer através do caminho $A \rightarrow B \rightarrow C \rightarrow D$, e não pelo caminho $A \rightarrow D$. Para a dissociação, o inverso parece ser verdadeiro: dissociação preferencial pelo caminho $D \rightarrow A$.	88
5.18	Complexo TrCel7A-substrato-produto (TrCel7A-SP). A celobiose (produto) é mostrada em alaranjado e o substrato em verde. A numeração se refere aos sítios de ligação de cada unidade de glicose da cadeia polissacarídica. Sítios (+) ficam do lado da extremidade redutora (à esquerda do ponto de hidrólise) e sítios (-) ficam do lado da extremidade não redutora (à direita do ponto de hidrólise) [208]	92
5.19	Configurações do sítio de ligação da celobiose vistas na simulação. A configuração à esquerda é semelhante à estrutura cristalográfica. A configuração à direita exibe um túnel sobre o sítio de ligação da celulose, e não é vista na estrutura cristalográfica. A TrCel7A-SP tem acesso a ambos os estados.	93

5.21 (a) Distribuições do rmsd da celobiose em relação a sua estrutura média na *holo*-TrCel7A e TrCel7A-SP, mostrando que nesta última, a celobiose exibe maior flexibilidade em seu sítio de ligação. 95

6.1 Estrutura cristalográfica das laminarinases PcLam (PDB: 2CL2)
[233], aNLam (PDB: 2HYK) [231] e RmLam (PDB: 3ILN) [213].
Como membros da família GH16, estas laminarinases possuem enovelamento do tipo β-jelly-roll, formado por um sanduíche de duas folhas β, dentro do qual reside o núcleo hidrofóbico das proteínas. As folhas β possuem uma concavidade e formam um canal de ligação ao substrato, delimitado por vários loops. 99

- 6.2 Configurações das laminarinases PcLam, aNLam e RmLam tomadas das simulações, mostrando as moléculas de água presentes no núcleo hidrofóbico de cada enzima nas temperaturas de 25°C e 90°C. A 25°C, as moléculas de água presentes no núcleo hidrofóbico são as mesmas presentes na estrutura cristalográfica. Ao aumentar a temperatura, uma ponte salina facilita a entrada de até 6 moléculas de água no núcleo hidrofóbico da mesoestável PcLam. A ausência desta ponte salina na termoestável aNLam e na hipertermoestável RmLam impede a penetração de água em seus núcleos hidrofóbicos, permanecendo, assim, estáveis em altas temperaturas.

6.5	Projeções da malha tridimensional de pontos no plano $xy e yz$. O eixo y aponta em direção ao solvente e o eixo z é paralelo ao canal de ligação ao substrato. A projeção no plano yz permite que os três estados das laminarinases sejam univocamente discriminados.	107
6.6	Padrões da projeção da malha tridimensional sobre o plano yz. O estado obstruído é caracterizado por um padrão contendo uma região vazia, o estado aberto por uma região completamente pre- enchida e o estado túnel por uma região vazia isolada entre os pontos da malha	108
7.1	Experimento de extensão de parede celular por expansinas. Pa- redes celulares com expansinas inativadas termicamente não res- pondem à força externa, mas, quando na presença de expansi- nas exógenas, respondem imediatamente à força externa, apre- sentando aumento significativo da taxa de extensão. Adaptado de [244] com permissão de Nature Publishing Group	112
7.2	Estrutura 3D da expansina EXLX1 (PDB: 3D30 [260]), mos- trando os domínios D1 e D2, e os aminoácidos aromáticos e básicos, presentes em D2, e polares/carregados presentes em D1	114
7.3	Adsorção da expansina EXLX1 sobre a superfície hidrofóbica da celulose cristalina. (a) Distância entre a superfície da fibrila de celulose e os domínios D1 e D2 durante o processo de adsorção, com configurações representativas da simulação. Observa-se que interações do domínio D2 com a superfície da celulose dirige o processo de adsorção da EXLX1. (b) Energias de interação entre a celulose e a EXLX1 intacta (esquerda), somente o domínio D1 (central) e somente o domínio D2 (direita) durante o processo de adsorção. Ocorre uma redução na energia de van der Waals entre a celulose e o domínio D2 durante a adsorção.	120

LISTA DE FIGURAS

7.4	Energias de interação entre a celulose e a EXLX1 intacta (es- querda), somente o domínio D1 (central) e somente o domínio D2 (direita) com a expansina adsorvida sobre a fibrila de celulose. As flutuações das energias em torno de valores médios aproxima- damente constantes indicam que a EXLX1 não está num estado transiente de adaptação sobre a celulose.	121
7.5	Movimento translacional da EXLX1 sobre a superfície da celulose. (a) Evolução temporal da distância entre o carbono α do resíduo aromático Trp125 (domínio D2) e o resíduo de glicose com o qual está em contato inicialmente. Observa-se uma transição abrupta em torno de 80 ns, indicativa da translação da EXLX1 ao longo de uma unidade de glucose sobre a celulose. (b) Posições inicial e final da EXLX1 sobre a superfície da celulose, evidenciando que ocorreu uma translação da proteína.	122
7.6	Conformação da cadeia oligossacarídica ligada à EXLX1. (a) Es- trutura da EXLX1 ligada ao oligossacarídeo colorido conforme suas mobilidades atômicas obtidas durante o curso das simulações. De azul a vermelho, a escala de cor representa flutuações médias de <1,5 Å a >3.0Å. Enquanto o domínio D2 segura o substrato em uma conformação planar, o domínio D1 induz uma torção de 90° na cadeia (indicada pelo retângulo pontilhado). (b) Flutuação média dos resíduos de glicose ao longo da cadeia oligossacarídica. Barras de erro indicam desvio padrão de diferentes simulações. Os números dos resíduos de glicose estão indicados no painel (a). Ambos os domínios D1 e D2 possuem resíduos capazes de estabi- lizar a conformação do substrato.	124
7.7	Interações entre EXLX1 e oligossacarídeo. (a) Energias médias de interação entre o substrato e cada um dos resíduos da EXLX1. (b) Estrutura da EXLX1 mostrando a localização dos resíduos	
	que interagem com o substrato	125

7.8	Ligações de hidrogênio entre o domínio D1 e o oligossacarídeo. (a) Visão detalhada das interações do substrato com os resíduos do domínio D1 responsáveis pela torção. A numeração dos resíduos de glicose corresponde àquelas mostradas na figura 7.6a. (b) Frequência das ligações de hidrogênio entre o substrato e a EXLX1. O resíduo Asp82 se envolve na mais persistente ligação de hi- drogênio com o oligossacarídeo
7.9	Conformação do centro de torção (domínio D1) no mutante Asp82Asn, mostrando o equilíbrio conformacional entre os estados ativo e inativo, onde o par Asn82/Tyr73 assume duas conformações dis- tintas. No estado inativo, o resíduo Asn82 torna-se indisponível para interagir com o substrato
7.10	Ocorrência dos estados ativo e inativo no mutante Asp82Asn. (a) Rmsd do resíduo Asn82 em relação às coordenadas iniciais (após relaxação) da simulação. O rmsd flutua em torno de dois valores, correspondentes ao estado ativo (rmsd ≈ 1 Å) e inativo (rmsd \approx 2 Å) do mutante Asp82Asn. (b) Distribuição dos valores de rmsd do resíduo Asn82, mostrando as probabilidades (área integrada) dos estados ativo (40%) e inativo (60%)
7.11	Flutuações estrutura do mutante Tyr73Ala. (a) Rmsf da EXLX1 nativa e mutante. (b) Estruturas da EXLX1 com os rmsf mape- ados numa escala de cor indo de azul (rmsf $< 0,3$ Å) a vermelho (rmsf $> 1,5$ Å). A mutação aumenta as flutuações em torno do centro de torção da EXLX1
7.12	Distâncias entre os resíduos conservados Asp82 e Thr12 (à es- querda) e suas respectivas distribuições (à direita) na EXLX1 na- tiva e na mutante Tyr73Ala. A mutação causa tal ligação de hidrogênio conservada ser regularmente rompida

8.1	Ramificações de xilano: arabinose, ácido glucurônico e glururonato. 144
8.2	Estrutura da nanofibrila de celulose formada por 4 cadeias de glu-
	cano de grau de polimerização 8 e com ambas as faces hidrofílicas
	e hidrofóbicas expostas. As diferentes cores discriminam as dife-
	rentes cadeias de glucano
8.3	Modelo de parede celular primária. O modelo consiste de duas na-
	nofibrilas de celulose imersas numa solução aquosa de monômeros
	de hemicelulose

132
8.4	Caminhos de desagregação ao longo do qual o PMF foi compu- tado. Em (a) contatos hidrofílicos das fibrilas são rompidos e, em (b), contatos hidrofóbicos	146
8.5	(a) Potencial de força média de desagregação das nanofibrilas de celulose ao longo do caminho de desagregação das faces hidrofílicas e hidrofóbicas. (b) e (c): superfície de contorno da função de distribuição $g(\mathbf{r})$ dos átomos de oxigênio da água ao longo da desagregação (b) hidrofílica e (c) hidrofóbica das nanofibrilas. As superfícies de contorno correspondem a $g(\mathbf{r}) = 2, 5. \ldots \ldots$	150
8.6	Potencial de força média de desagregação hidrofílica e hidrofóbica das nanofibrilas de celulose em soluções de (a) arabinose, (b) ácido glucurônico e (c) glucuronato. As curvas de diferentes cores cor- respondem a concentrações crescentes de hemicelulose, variando da fração molar $x = 0$ a $x = 0, 1$. Em todos os casos, o efeito da hemicelulose é a estabilização do estado agregado das nanofibrilas de celulose	153
8.7	Efeito da natureza química e da concentração das ramificações de hemicelulose na energia livre de agregação (ΔG_{agr}) das faces (a) hidrofóbicas e (b) hidrofílicas da nanofibrilas de celulose. A linha pontilhada cinza corresponde à soma dos ΔG_{agr} para arabinose e acetato, conforme detalhado no texto	154
8.8	Mapas da função de distribuição tridimensional $g(\mathbf{r})$ para os átomos de carbono da (a) arabinose, (b) ácido glucurônico, (c) glucuro- nato e (d) acetato em torno das faces hidrofílicas e hidrofóbicas das nanofibrilas de celulose. O esquema de cores é indicado pe- las esferas coloridas das estruturas à direita. Os valores de $g(\mathbf{r})$ foram escolhidos para clara visualização dos modos de interação hemicelulose-celulose e correspondem a $g(\mathbf{r}) = 1, 4$ para arabinose, ácido glucurônico e acetato, e $g(\mathbf{r}) = 2.0$ para glucuronato.	155
	actuo grucuronico e acetato, e $g(\mathbf{r}) = 2,0$ para grucuronato	100

- 9.1 (a) Unidades monoméricas da lignina, como aparecem no polímero, em que os substituintes R1 e R2 determinam as subunidades H, G ou S. (b) Dímero de lignina destacando a ligação β-O-4' em azul.
 (c) Monômero de lignina usado nos cálculos, aqui denotados de monolignóis. (d) Oligômeros de lignina e hemicelulose. 165
- 9.3 Energia livre de solvatação da hemicelulose (à esquerda) e da lignina (à direita) em função da concentração dos monolignóis H, G e S, para os cinco diferentes confôrmeros de hemicelulose e lignina. As diferentes cores representam as diferentes conformações. As energias livres de solvatação são mostradas com relação ao valor mais alto entre os confôrmeros. É evidente que o efeito da concentração dos monolignóis é independente da conformação dos oligômeros, como pode ser inferido pelas inclinações essencialmente iguais das curvas. Entretanto, os valores das energias de solvatação dependem da conformação e refletem a hidratação da hemicelulose e lignina.

171

9.4	Contribuição parcial da água e dos monolignóis para a energia	
	livre de solvatação da (a) hemicelulose e da (b) lignina. A contri-	
	buição da água é termodinamicamente favorável enquanto que a	
	dos monolignóis é desfavorável	173
9.5	Efeito do estado conformacional da hemicelulose (à esquerda) e	
	da lignina (à direita) nas contribuições parciais da água e dos	
	monolignóis para a energia livre de solvatação. A contribuição	
	dos monolignóis é insensível à conformação tanto da hemicelulose	
	quanto da lignina, enquanto contribuição da água depende do	
	confôrmero	174
9.6	Compressibilidades isotérmicas das soluções de monolignóis ob-	
	tidas pela teoria DRISM/KH. À medida que a concentração de	
	monolignol aumenta, a compressibilidade também aumenta mo-	
	notonicamente. Além disso, as compressibilidades aumentam na	
	ordem H < G < S, indicando que esta é a ordem relativa das	
	hidrofobicidades dos monolignóis.	176
9.7	Distribuição tridimensional da densidade de energia livre de sol-	
	vatação em torno da hemicelulose (em alaranjado), mostrando	
	regiões espaciais que concentram as interações favoráveis com a	
	(a) água e com os monolignóis (b) H, (c), G e (d) S. $\dots \dots \dots$	177
9.8	Distribuição tridimensional da densidade de energia livre de sol-	
	vatação em torno da lignina (em alaranjado), mostrando regiões	
	espaciais que concentram as interações favoráveis com a (a) água	
	e com os monolignóis (b) H, (c), G e (d) S. $\dots \dots \dots \dots$	179
9.9	Visão ampliada da SFED da água (em azul) e do monolignol S	
	(em amarelo) em torno da (a) hemicelulose e (b) lignina. En-	
	quanto a água faz ligações de hidrogênio com a hemicelulose, os	
	monolignóis se distribuem sobre as faces hidrofóbicas disponíveis,	
	reduzindo assim o contato da água com tais faces e, ao mesmo	
	tempo, permitindo a formação de ligações de hidrogênio entre a	
	água e os sítios hidrofílicos da hemicelulose e da lignina	180

9.10	Análogo unidimensional do cálculo da força da ligação de hi-
	drogênio como definida pela equação 9.1. Em (a), é mostrada
	a função de distribuição radial $g(r)$ entre átomos de hidrogênio e
	oxigênio da água pura. O PMF obtido por $PMF = -k_BT \ln g(r)$ é
	mostrado em (b). Observa-se que o pico g^{max} de $g(r)$ corresponde
	ao mínimo do PMF, que fornece a medida da força da ligação de
	hidrogêni o $W.$

Lista de Tabelas

7.1	Comparação das constantes de dissociação (K_d) associadas à in-
	teração expansina-celulose de mutantes em que os resíduos lis-
	tados foram substituídos por alanina [264], com as energias de
	interação expansina-glucano obtidas das simulações. Os dados
	mostram que não há correlação entre as energias de interação com
	as constantes de ligação dos mutantes, sugerindo que fatores não-
	entálpicos são importantes na interação EXLX1-substrato. Os
	valores entre parênteses indicam desvio padrão
8.1	Densidade (em g cm ^{-3}) das soluções aquosas de arabinose, ácido
	glucuronico e acido acetico em diferentes concentrações 149
9.1	Energia livre de solvatação $(\Delta \mu)$ e entropia de solvatação mul- tiplicada pela temperatura $(T\Delta S)$ da hemicelulose e lignina em soluções aquosas dos monolignóis H, G e S na fração molar de 0,01. Os valores são relativos ao monolignol H e expressos em
	kcal/mol

Parte I

Introdução

Capítulo 1 Introdução

A produção de bioetanol a partir da biomassa lignocelulósica de plantas integra processos físico-químicos e enzimáticos que atualmente comprometem sua viabilidade econômica. A biomassa é constituída majoritariamente por celulose, hemicelulose e lignina, componentes tais que formam a parede celular das plantas e se arranjam numa arquitetura recalcitrante (ou seja, com resistência mecânica, química e biológica), em que fibras de celulose se estruturam dentro de uma matriz protetora de hemicelulose e lignina. Nas etapas iniciais da conversão da biomassa, severos tratamentos termoquímicos rompem a barreira ligno-hemicelulósica da parede celular e expõem a celulose ao alcance das enzimas. Por formar fibras empacotadas e cristalinas, a celulose, por sua vez, oferece barreiras físicas ao ataque das enzimas: estas devem atuar numa interface sólido-líquido para se ligar a cadeias individuais de celulose e hidrolisá-las. Assim, as enzimas nativas são ineficientes para uso tecnológico.

Estratégias para aumentar a eficiência dos processos de degradação da biomassa incluem: (i) desenvolvimento racional de novos pré-tratamentos eficientes e pouco severos; (ii) desenvolvimento de plantas transgênicas com paredes celulares mais suscetíveis à degradação; e (iii) engenharia de enzimas mais eficientes e estáveis termicamente. O avanço destas três frentes depende da compreensão tanto dos substratos (estrutura nativa e alterações por pré-tratamentos e enzimas), quanto das enzimas (mecanismos físicos e químicos), além da relação entre ambos, ou seja, de como a estrutura do substrato afeta a atividade enzimática. A natureza física de vários processos envolvidos na degradação da biomassa – por exemplo, a alteração estrutural do substrato que precede a hidrólise enzimática – constitui um desafio para as pesquisas: muitas vezes, experimentos biofísicos, e não bioquímicos, devem ser realizados com materiais de alta complexidade e de estrutura e composição química variáveis. Tais estudos são multidisciplinares e abordam problemas em diferentes escalas de tamanho. Técnicas experimentais sofisticadas, incluindo microscopias eletrônicas e de força atômica de alta velocidade, difração e espalhamento de raios X e espalhamento de nêutrons têm sido empregadas para o estudo de diferentes aspectos da degradação de biomassa lignocelulósica. Na escala atômico-molecular, as técnicas computacionais têm fornecido informações quantitativas e qualitativas de processos envolvidos na degradação, principalmente daqueles difíceis de acessar experimentalmente.

Nesta tese, estudos computacionais de aspectos moleculares da degradação da biomassa lignocelulósica são apresentados. Os tópicos abordam tanto a estrutura e interações da parede celular de plantas (substrato) quanto relações entre dinâmica e função de enzimas e proteínas não-hidrolíticas envolvidas na conversão de biomassa lignocelulósica. Estes estudos representam contribuições científicas básicas que podem ser úteis para o desenvolvimento das três frentes descritas anteriormente. Duas técnicas computacionais distintas foram empregadas: simulações de dinâmica molecular e a teoria mecânico-estatística 3D-RISM. As aplicações desta última técnica foram realizadas em doutorado-sanduíche no grupo do Dr. Andriy Kovalenko, National Institute for Nanotechnology/University of Alberta, Canadá, onde parte da teoria foi desenvolvida e implementada computacionalmente.

A tese está organizada em capítulos independentes, de forma que o leitor poderá se direcionar ao tópico de seu interesse. O Capítulo 2 tem caráter informativo e fornece detalhes do processo de degradação de biomassa. Em seguida, os Capítulos 3 e 4 apresentam os fundamentos teóricos das técnicas computacionais utilizadas. Os capítulos que seguem apresentam os resultados das pesquisas, cobrindo os seguintes tópicos: dinâmica de glicosídeo hidrolases (Capítulos 5 e 6), disrupção não-hidrolítica de biomassa (Capítulo 7), nanoarquitetura de paredes celulares primárias (Capítulo 8) e forças termodinâmicas em paredes celulares secundárias (Capítulo 9). No Capítulo 10, comentários finais e perspectivas futuras são apresentados.

Capítulo 2

Degradação de biomassa lignocelulósica

A alta demanda mundial por combustíveis associada com questões ambientais e de sustentabilidade tem impulsionado o interesse acadêmico, governamental e corporativo no desenvolvimento de tecnologias para a produção de bioetanol de segunda geração – aquele obtido da degradação de resíduos lignocelulósicos, como o bagaço de cana-de-açúcar – como alternativas para o tradicional uso de petróleo [1, 2]. Ao contrário da produção de bioetanol de primeira geração – aquele produzido diretamente da fermentação dos açúcares presentes na matéria-prima –, a produção de bioetanol de segunda geração em larga escala só será possível após o desenvolvimento de tecnologias para vencer a estrutura recalcitrante da parede celular de plantas [2]. Do ponto de vista econômico, o desenvolvimento de tais tecnologias de segunda geração são fundamentais para que a produção de biocombustíveis não afete a cadeia de produção de alimentos [3].

Recalcitrância lignocelulósica é a resistência natural da biomassa à degradação, e emerge de uma convolução de fatores que abrangem desde a estrutura nanoscópica da biomassa ao arranjo de tecidos na planta [2]. Em decorrência disso, a conversão de biomassa lignocelulósica em bioetanol é um processo complexo e lento, envolvendo múltiplos estágios: pré-tratamentos termoquímicos, hidrólise enzimática com um coquetel multifuncional de enzimas e, finalmente, a fermentação por microrganismos [4].

Neste capítulo, os fundamentos da degradação de biomassa lignocelulósica são apresentados, com enfoque nos aspectos de estrutura e da hidrólise da biomassa associados à recalcitrância.

2.1 A estrutura microscópica da biomassa: parede celular

A matéria-prima utilizada para a produção de bioetanol está contida na parede celular das células vegetais da biomassa – espessa camada de 0.1 a 10 μ m, externa à membrana plasmática, que proporciona rigidez, suporte estrutural e proteção contra patógenos às plantas. Assim, o processo de degradação da biomassa consiste na desconstrução e hidrólise dos componentes da parede celular de plantas em açúcares fermentáveis [4].

As paredes celulares se organizam em até três camadas: a lamela média, a parede celular primária e a parede celular secundária. A lamela média é a camada mais externa e funciona como um cimento que adere uma célula vegetal a outra. Por ser rica em pectina (ácidos poligalacturônicos e derivados, facilmente extraídos com agentes quelantes), a lamela média não oferece problemas para a desconstrução da parede celular. As paredes celulares primárias ocorrem em todas as células e estão localizadas numa camada adjacente à lamela média. São compostas de polissacarídeos (celulose, hemicelulose e pectina), de proteínas e outros componentes minoritários. As paredes secundárias, por outro lado, ocorrem em células especializadas – como as de vasos condutores de plantas vasculares – e contêm, além dos componentes presentes na parede primária, uma espessa camada de lignina depositada na matriz polissacarídica. Enquanto as paredes primárias podem sofrer alterações em sua nanoarquitetura durante o crescimento da célula, as paredes secundárias são mecanicamente resistentes e construídas pela célula somente após seu completo amadurecimento [4, 5].

Os componentes majoritários de uma parede celular secundária são celulose (20-50%), hemicelulose (15-35%) e lignina (10-30%). Tal composição varia de espécie para espécie de planta, e também com o tipo de tecido e com as condições ambientes [6, 7]. Na parede celular, estes biopolímeros se arranjam numa arquitetura mecânica, química e biologicamente resistente, que oferece suporte estrutural e proteção para plantas. De forma simplificada, esta arquitetura consiste de fibras de celulose imersas numa matriz amorfa de lignina e hemicelulose (figura 2.1). Os mesmos elementos que permitem às paredes celulares exercerem funções que demandam resistência também originam a recalcitrância à desconstrução nas aplicações tecnológicas [4].



Figura 2.1. Arquitetura da parede celular secundária. Fibras de celulose são envolvidas por uma matriz amorfa de lignina e hemicelulose, que podem formar os complexos lignina-carboidrato (LCC). Adaptado de [4].

2.1.1 Celulose

A celulose – biopolímero mais abundante da natureza – é um políssacarídeo linear formado por resíduos de glicose unidos por ligações glicosídicas $\beta(1 \rightarrow 4)$. Dependendo da planta, o grau de polimerização da celulose pode variar de 100 a 10.000 unidades de glicose [4]. Embora seja quimicamente simples, a celulose apresenta aspectos estruturais únicos, o que permite distingui-la de outras classes de polissacarídeos. As cadeias de celulose se empacotam formando complexos supramoleculares em forma de fibras, que podem conter domínios cristalinos e amorfos. Na fase cristalina, a celulose pode assumir diferentes polimorfos: I α , I β , II, III_I, III_{II}, IV_I e IV_{II}. A celulose nativa, encontrada na natureza, assume os polimorfos I α e I β (além da fase amorfa), sendo a celulose I β a mais comum em plantas. Os outros polimorfos (II, III_I, III_I, IV_I e IV_{II}) são obtidos através de processos de regeneração¹ e mercerização² (celulose I \rightarrow celulose II), tratamento com amônia líquida (celulose I \rightarrow celulose III_I; celulose II \rightarrow celulose III_{II}) e aquecimento (celulose III_I) \rightarrow celulose IV_I; celulose III_{II} \rightarrow celulose IV_{II}) [8].

A estrutura da celulose I β foi determinada por difração de raios X e de nêutrons em 2002 [9], consistindo de cadeias de glucano empacotadas paralelamente em várias camadas, com uma célula unitária monoclínica. As cadeias são mantidas unidas por ligações de hidrogênio intra- e intermoleculares e interações hidrofóbicas intercamadas, conforme ilustrado na figura 2.2. Este empacotamento cristalino origina uma das facetas da recalcitrância lignocelulósica. Uma vez que as celulases hidrolisam cadeias individuais de glucano, toda a rede de ligações de hidrogênio e interações hidrofóbicas do cristal de celulose deve ser rompida no processo de hidrólise. A acessibilidade limitada das cadeias de glucano da celulose, portanto, interfere negativamente na cinética do processo. Simulações computacionais mostraram que a descristalização de cadeias de glucano não é espontânea ($\Delta G > 0$), de forma que as celulases devem realizar trabalho para alçá-las a seus sítios ativos [10]. Por outro lado, a recalcitrância associada à descristalização depende da forma polimórfica da celulose [11], conforme elucidado por simulações computacionais, que mostraram que o trabalho de descristalização aumenta na seguinte ordem: II < I $III_{I} < I\alpha < I\beta$ [12]. Dessa forma, a nanoestrutura da celulose é um dos fatores determinantes de sua suscetibilidade à hidrólise enzimática.

Síntese da celulose

Nas plantas, a celulose é sintetizada por um complexo de enzimas denominadas celulose sintases, localizadas na membrana plasmática das células [13]. Cada celulose sintase sintetiza uma única cadeia de glucano, e diferentes isoformas de celulose sintase ($\alpha 1$, $\alpha 2 \in \beta$) são arranjadas em complexos denominados rosetas. As rosetas são arranjos de celulose sintases organizadas em forma hexagonal, contendo 36 unidades (figuras 2.3a e 2.3b). Dessa forma, uma roseta sintetiza 36 cadeias de

¹Regeneração consiste na solubilização da celulose em determinado solvente seguida da precipitação em água.

 $^{^2 {\}rm Mercerização}$ é o processo de intumescimento da celulose nativa com NaOH; a celulose II aparece com a remoção do agente intumescedor.



a) Ligações de hidrogênio entre cadeias de glucano de uma mesma camada

Figura 2.2. Estrutura da cristalina da celulose I β , mostrando (a) ligações de hidrogênio (linhas tracejadas) intramoleculares e entre cadeias de glucano e (b) o empacotamento hidrofóbico de camadas de glucano.

glucano em paralelo, as quais se agregam em formas também hexagonais, dando origem a uma fibrila elementar de celulose, como ilustrado na figura 2.4a. Dependendo da planta, outros arranjos de roseta podem ocorrer, o que gera fibrilas elementares de formas diferentes [15, 16]. As rosetas, por sua vez, podem se arranjar formando complexos de rosetas (figura 2.3c), os quais sintetizam macrofibrilas de celulose, que são arranjos de várias fibrilas elementares (figura 2.4b) [14]. À medida que outros componentes da parede celular (hemicelulose, lignina, pectina) são sintetizados, as macrofibrilas eventualmente se dissociam, dando origem a fibrilas elementares de celulose revestidas por hemicelulose – as microfibrilas de celulose. A forma predominante das fibrilas elementares (agregadas em macrofibrilas ou isoladas como microfibrilas revestidas de hemicelulose) é importante por modular as propriedades físicas e químicas da parede celular [17]. Além disso, a acessibilidade das enzimas à celulose é dependente da arquitetura das fibrilas, o que, por sua vez, sofre alterações com pré-tratamentos termoquímicos [18].



Figura 2.3. (a) Celulose sintases arranjadas em grupos de diferentes isoformas ($\alpha 1$, $\alpha 2 \in \beta 1$). (b) Agrupamento de 36 celulose sintases em forma hexagonal, formando uma roseta, que sintetiza uma fibrila elementar de celulose. (c) Arranjo de rosetas, unidade que forma uma macrofibrila de celulose. Adaptado de [14].



Figura 2.4. (a) Estrutura de uma fibrila elementar de celulose. (b) Macrofibrila de celulose formada por 7 fibrilas elementares.

2.1.2 Hemicelulose

As hemiceluloses constituem um grupo heterogênio de polissacarídeos ramificados de pentoses (xilose, arabinose) e hexoses (manose, glicose, galactose) que estão presentes na matriz não-celulósica da parede celular [19]. Diferentemente da celulose, as hemiceluloses são polímeros amorfos. Para ser incluído no grupo das hemiceluloses, um polissacarídeo deve possuir cadeia principal com unidades monossacarídicas unidas por ligações $\beta(1 \rightarrow 4)$ na configuração equatorial [20].

As hemiceluloses encontradas em plantas terrestres podem ser classificadas em três grandes grupos [4, 20]: (i) xiloglucanos (cadeia principal de resíduos de glicose com ramificações de xilose e galactose), (ii) xilanos (cadeia principal de xilose com ramificações de arabinose e ácido glucurônico) e (iii) mananos (cadeia principal de manose com ramificações de galactose) (figura 2.5). Embora classificadas em



Figura 2.5. Estrutura representativa de cada grupo de hemicelulose: xiloglucano, xilano e manano. Adaptado de [21] com permissão da John Wiley and Sons.

três grupos, as hemiceluloses apresentam alta variabilidade em relação ao tipo e quantidade de ramificações [22]. Por exemplo, os xilanos podem ter ramificações de arabinose e ácido glucurônico em diferentes proporções e localizações ao longo da cadeia principal; podem ainda apresentar acetilações em pontos aleatórios da cadeia [19]. Estas variações ocorrem em diferentes espécies de plantas, em diferentes tecidos e em diferentes condições do meio, e permitem à planta modular as propriedades físicas da matriz hemicelulósica e sua interação com outros componentes da parede celular.

O entendimento das rotas de síntese da hemicelulose tem permitido a modulação de sua composição química em plantas transgênicas, gerando potencial para a produção de plantas mais suscetíveis à degradação [20, 22, 23]. Embora sejam quimicamente bem caracterizadas, os arranjos estruturais da hemicelulose dentro da parede celular são pouco compreendidos. Por revestir as fibrilas elementares de celulose através de interações não-covalentes, as hemiceluloses constituem uma barreira às celulases. Assim, a hidrólise da celulose requer a remoção de hemicelulose por via termoquímica (pré-tratamentos) ou enzimática (com hemicelulases diversas) [4, 24].

2.1.3 Lignina

A lignina é um polímero fenólico altamente hidrofóbico que se deposita na parede celular secundária. Seu papel biológico inclui proteção contra patógenos, regulação da hidrofobicidade de canais condutores de água e manutenção da integridade estrutural da célula vegetal madura [25, 26]. A lignina é considerada o principal responsável pela recalcitrância da parede celular: vários estudos mostram uma anticorrelação entre a quantidade de lignina presente na biomassa e sua digestibilidade enzimática [3, 17, 27]. Além de ser uma barreira protetora para a celulose, a lignina pode ligar enzimas não-produtivamente, dificultando a degradação da biomassa [28].

Através da via dos fenilpropanóides, três diferentes monolignóis (unidades monoméricas precursoras) – os álcoois p-cumarílico, coniferílico e sinapílico – são sintetizados a partir do aminoácido L-fenilalanina e posteriormente incorporados à lignina através de diferentes tipos de ligações químicas, sendo as ligações éter arílicas (β -O-4') as mais frequentes (figura 2.6) [3, 29]. Ao serem incorporados



Figura 2.6. (a) Síntese e estrutura da lignina. Adaptado de [26]. (b) Estrutura de um complexo lignina-carboidrato (LCC). Adaptado de [4].

na lignina polimérica, os álcoois *p*-cumarílico, coniferílico e sinapílico se transformam, respectivamente, em *p*-hidroxifenil (subunidade H), guaiacil (subunidade G) e siringil (subunidade S), que variam no grau de metoxilação aromática (figura 2.6a). O teor relativo das subunidades H, G e S varia em diferentes espécies e tecidos, o que influencia as propriedades físicas e químicas da lignina. Além disso, a lignina pode se associar covalentemente com as ramificações das hemicelulose, originando os complexos lignina-carboidrato (LCC: *lignin-carbohydrate complex*) (figura 2.6b); estes complexos criam *cross-links* entre hemicelulose e lignina, o que reforça a recalcitrância da parede celular [3].

2.2 Pré-tratamentos

Pré-tratamentos são processos que reduzem a recalcitrância da biomassa, convertendo sua estrutura nativa em estruturas mais suscetíveis ao ataque enzimático [28]. Esta é uma etapa essencial do processo de degradação da biomassa; dificilmente a hidrólise enzimática será rápida e eficiente sem processos que aumentam a acessibilidade da biomassa às enzimas. Devido às elevadas demandas energéticas envolvidas nos pré-tratamentos, um dos desafios atuais é o desenvolvimento de prétratamentos eficientes e ao mesmo tempo de baixo custo, para serem integrados às operações de processamento da biomassa [3].

Dependendo do tipo e severidade do pré-tratamento, várias modificações químicas e físicas podem ocorrer na biomassa, tais como: solubilização de hemicelulose [30], quebra das ligações lignina-hemicelulose dos LCCs [31], modificação e redistribuição da lignina [32–36], alterações estruturais nas fibrilas de celulose [34, 37, 38], rearranjo das micro- e macrofibrilas de celulose [18, 33, 38–40] e aumento da porosidade da parede celular [31, 41]. Todas estas modificações estão associadas, em menor ou maior grau, a uma modulação da recalcitrância lignocelulósica, que é geralmente medida em termos da digestibilidade enzimática da biomassa. Por exemplo, a quebra das ligações dos LCCs é importante para a remoção de lignina e hemicelulose; alterações estruturais na celulose, como redução da cristalinidade, aumentam a acessibilidade das cadeias individuais de glucano às enzimas; e o aumento da porosidade da biomassa facilita a difusão das enzimas até o substrato celulósico [4, 31].

Pré-tratamentos comuns incluem: tratamentos com ácido diluído, tratamentos hidrotérmicos, explosão de vapor, tratamentos com álcali e expansão de fibras por amônia (AFEX) [3, 4, 24, 28]. Estes pré-tratamentos variam quanto à severidade, que é definida em termos da temperatura, do pH e do tempo duração do processo [24, 28]. Em geral, pré-tratamentos ácidos (ácido diluído, hidrotérmico e explosão de vapor¹) resultam na hidrólise e solubilização da hemicelulose, enquanto somente uma pequena fração da lignina é removida [3, 30]. Alterações na distribuição da lignina são, contudo, evidentes durante estes pré-tratamentos [32, 34]. Em pH neutro (tratamento hidrotérmico e explosão de vapor em pH controlado), somente a hemicelulose é removida. Em pH alcalino (tratamentos com álcali e outros), grande parte da lignina é removida junto com a hemicelulose, com exceção do pré-tratamento AFEX, que não remove nem a hemicelulose nem a lignina, mas causa outras alterações na biomassa, como a transformação da celulose I β em celulose III_I, aumentando a taxa de hidrólise enzimática [30, 42].

Mais recentemente, líquidos iônicos – sais de cátions orgânicos com ponto de fusão abaixo de 100 °C – têm sido empregados para solubilizar a celulose e fracionar os componentes da matriz não-celulósica da parede celular [4, 43–46]. Simulações computacionais mostraram que, ao contrário do que se passa em meios aquosos, a descristalização de cadeias de celulose é espontânea em líquidos iônicos ($\Delta G < 0$) [10, 47]. Entretanto, problemas com a recuperação do solvente tornam o processo inviável até então. Outras misturas de solventes mostraram ser capazes de descristalizar a celulose e fracionar a lignina e a hemicelulose. A dissolução instantânea da celulose em soluções de eletrólitos orgânicos, por exemplo, foi reportada em 2011 [48].

Devido à severidade das condições químicas (pH extremos) e termodinâmicas (altas temperaturas) requeridas para o rompimento da parede celular, os prétratamentos frequentemente resultam na formação de produtos secundários a partir de lignina e carboidratos (furfurais, pseudo-lignina) [4, 28, 46, 49, 50]. Tais produtos inibem a atividade enzimática e a fermentação por microrganismos. Até o momento, nenhum pré-tratamento é capaz, por si só, de promover todas as alterações desejadas na parede celular (redução da cristalinidade da celulose e remoção da hemicelulose/lignina sem a geração de produtos secundários de degradação). A hemicelulose é o componente mais sensível aos pré-tratamentos, podendo ser removida mesmo em condições brandas [51]. Em condições mais severas – necessárias para remoção da lignina ou alterações morfológicas na celulose – a hemicelulose

¹Nos tratamentos hidrotérmicos e de explosão de vapor, o pH se torna ácido devido à liberação de compostos ácidos da hemicelulose [3, 28].

pode dar origem a produtos secundários de degradação. Uma alternativa para este problema é a utilização de pré-tratamentos em mais de um estágio. Nesse tipo de estratégia, a hemicelulose é, num primeiro estágio, solubilizada em condições brandas e recuperada; num segundo estágio, a fração sólida resultante é submetida a tratamentos mais severos para a remoção da lignina e descristalização e/ou intumescimento da celulose. Assim, maximiza-se a recuperação de açúcares fermentáveis da hemicelulose enquanto aumenta-se a suscetibilidade da celulose ao ataque enzimático.

Apesar de esforços constantes estarem sendo feitos para a otimização e desenvolvimento de novos pré-tratamentos, há pouco entendimento destes processos do ponto de vista mecanístico, o que é em grande parte associado à falta de entendimento de aspectos moleculares da parede celular nativa. Com isso, o desenvolvimento de pré-tratamentos se baseia em estratégias essencialmente empíricas. Nos últimos anos, estudos experimentais e computacionais têm começado a elucidar processos moleculares que ocorrem durante os pré-tratamentos [18, 39, 40].

Paralelamente ao desenvolvimento e otimização de pré-tratamentos, existem tentativas de desenvolver plantas transgênicas com alterações químicas e estruturais em suas paredes celulares para, assim, reduzir a necessidade de pré-tratamentos severos [52]. O entendimento das rotas de síntese da hemicelulose e lignina tem permitido gerar alterações na composição química destes componentes [21, 23, 25, 53, 54]. Além disso, mutações na celulose sintase podem resultar na síntese de fibrilas de celulose com menor grau de cristalinidade, aumentando sua digestibilidade enzimática [55]. Recentemente, a estrutura cristalográfica de uma celulose sintase bacteriana contendo uma cadeia de celulose nascente foi resolvida, permitindo a ampliação do conhecimento ligado à sintese de celulose [56]. Assim, o desenvolvimento de paredes celulares geneticamente modificadas parece constituir um dos pilares da produção de biocombustíveis em larga escala esperada para o futuro.

2.3 Hidrólise enzimática

A hidrólise enzimática segue a etapa de pré-tratamento e envolve a ação de um coquetel multifuncional de enzimas, capaz de hidrolisar ligações glicosídicas dos diferentes carboidratos presentes na biomassa pré-tratada (principalmente celulose e hemicelulose oligomérica) [4, 57]. Agindo em conjunto, enzimas deste coquetel transformam os carboidratos em suas formas monoméricas fermentáveis. Em relação à degradação química de biomassa, a hidrólise enzimática oferece vantagens como maiores rendimentos, alta seletividade e baixos gastos energéticos [57].

Organismos que degradam biomassa lignocelulósica naturalmente secretam uma variedade de enzimas e proteínas auxiliares. Em particular, o sistema enzimático secretado pelo fungo *Trichoderma reesei* é o mais utilizado industrialmente devido à sua elevada taxa de expressão e à sua eficiência relativamente alta em degradar celulose [57–59]. Dada a sua importância, as enzimas de *Trichoderma reesei* têm recebido atenção especial da comunidade científica [4].

As enzimas que hidrolisam ligações glicosídicas são coletivamente denominadas de glicosídeo hidrolases (GH) (EC 3.2.1.*x*) [60]. Associada à diversidade de combinações químicas e variações estereoquímicas de carboidratos, há na natureza um conjunto enorme de GHs, com diferentes especificidades químicas e estruturais [60, 61]. Com base na similaridade de sequências primárias – e, portanto, estrutural –, as GHs são classificadas em famílias [61]. Atualmente, existem 133 famílias de GHs, segundo o banco de dados CAZy (*Carbohydrate-Active enZymes*; www.cazy.org) [62]. Com frequência, enzimas de famílias distintas possuem similaridades em suas estruturas terciárias, resíduos catalíticos e mecanismo de hidrólise, sugerindo evolução a partir de ancestrais comuns. As famílias que compartilham estas três características são agrupadas em clãs [63]. A classificação das GHs em famílias e clãs permite que predições de mecanismos e estruturas de GHs recémdescobertas sejam feitas. Informações sobre todas estas famílias estão compiladas e atualizadas na enciclopédia eletrônica *cazypedia*: www.cazypedia.org.

2.3.1 Mecanismo de hidrólise

A hidrólise enzimática de ligações glicosídicas ocorre através de catálise ácidobase geral, com a participação de um resíduo doador de prótons e um resíduo que funciona como base/nucleófilo. Geralmente, resíduos de Asp ou Glu em seus estados protonados e desprotonados assumem estes papéis. As GHs apresentam dois mecanismos majoritários: com inversão ou retenção do carbono anomérico [64]. O tipo de mecanismo é, em geral, conservado em enzimas de uma mesma família [65], com poucas exceções [66, 67].

O mecanismo com inversão consiste numa única etapa de deslocamento simples. Enquanto a base geral desprotona uma molécula de água, esta age como nucleófilo, atacando o carbono anomérico. Ao mesmo tempo, o ácido geral protona a ligação glicosídica. A reação envolve um estado de transição com estrutura semelhante a um íon oxocarbênio, o qual evolui para o substrato hidrolisado com a configuração do carbono anomérico invertida (figura 2.7).



Figura 2.7. Mecanismo de hidrólise com inversão da configuração do carbono anomérico.

O mecanismo com retenção ocorre em duas etapas (deslocamento duplo). Na primeira etapa, o ácido geral protona o oxigênio da ligação glicosídica enquanto o resíduo nucleófilo faz um ataque nucleofílico ao carbono anomérico. Ocorre então a formação de um estado de transição de estrutura semelhante a um íon oxocarbênio, o qual evolui para um complexo glicosídico enzima-substrato. Esta primeira etapa é denominada glicosilação. Na etapa seguinte, ocorre a entrada de uma molécula de água próxima ao resíduo ácido desprotonado, o qual passa a funcionar como base geral que recebe um próton da água. Ao mesmo tempo, a água faz um ataque nucleofílico ao carbono anomérico e ocorre a dissociação da ligação enzima-substrato. Passando por outro estado de transição semelhante a um íon oxocarbênio, o sistema evolui para o estado hidrolisado, com o carbono anomérico do produto na mesma configuração do reagente (figura 2.8). O tipo de mecanismo adotado pela enzima depende da configuração relativa dos resíduos catalíticos. No mecanismo com inversão, uma molécula de água nucleofílica deve estar entre a base geral e o substrato (figura 2.7), enquanto que, no mecanismo com retenção, o ataque nucleofílico é realizado pelo próprio resíduo catalítico (figura 2.8). Com isso, a distância entre os resíduos catalíticos é um indicador do tipo de mecanismo (\approx 5,5 Å para retenção e \approx 10 Å para inversão) [60].



Figura 2.8. Mecanismo de hidrólise com retenção da configuração do carbono anomérico.

2.3.2 Hidrólise da celulose: sinergismo clássico

A hidrólise enzimática da celulose envolve a ação de duas classes de enzimas: as endo- β -1,4-glucanases (EC 3.2.1.4) e as exo- β -1,4-glucanases (EC 3.2.1.91). As endoglucanases (EG) hidrolisam ligações glicosídicas da celulose em pontos aleatórios das cadeias polissacarídicas, gerando extremidades redutoras e não-redutoras ao longo da cadeia. As exoglucanases, ou celobiohidrolases (CBH), hidrolisam ligações glicosídicas nas extremidades das cadeias polissacarídicas, liberando celobiose como produto. Juntas, as endoglucanases e celobiohidrolases despolimerizam a celulose até celobiose. Para completar o processo, a celobiose é degradada a glicose através das β -1,4-glucosidases (EC 3.2.1.21) [4]. Estes processos estão ilustrados na figura 2.9.

O sistema celulolítico secretado pelo fungo T. recsei é composto por oito endoglucanases e duas celobiohidrolases (CBH I e CBH II) [59]. Enzimas destas duas classes agem sinergisticamente na degradação da celulose: as extremidades redutoras e não-redutoras geradas pelas EGs se tornam subtratos para as CBHs. A CBH I ataca as extremidades redutoras da celulose enquanto a CBH II ataca as extremidades não-redutoras [69]. À medida que a concentração de celobiose aumenta no meio, esta inibe a atividade das celulases [70]. As β -glucosidades, por transformarem a celobiose em glicose, potencializam, assim, a ação das celulases [4, 70]. Além disso, as endoglucanases agem especificamente nos domínios amorfos da celulose, enquanto as celobiohidrolases agem nos domínios cristalinos. Por isso, as CBHs, EGs e β -glucosidases exibem sinergismo na degradação de celulose – conhecido como sinergismo clássico.

2.3.3 Além do sinergismo clássico

Além do sinergismo exibido pelas celulases, nos últimos anos começaram a ser descobertas proteínas com atividades auxiliares, que potencializam a ação das celulases. Estas proteínas pertencem a duas categorias: enzimas com atividades auxiliares (AA) e proteínas disruptivas não-hidrolíticas.

Proteínas disruptivas não-hidrolíticas

As proteínas disruptivas não-hidrolíticas potencializam a atividade de celulases por supostamente causarem a amorfização da estrutura da celulose e afrouxamento da arquitetura da parede celular, aumentando a acessibilidade do substrato às enzimas [71–73]. Quando misturadas com celulases, causam um aumento da liberação de açúcares solúveis. Na ausência de celulases, porém, não dão indícios de atividade



Figura 2.9. Sinergismo clássico entre endoglucanases (EG), celobiohidrolases (CBH I e CBH II) e β -glicosidades (BG). Os círculos vazios representam resíduos de glicose na estrutura da celulose, e os círculos cheios representam extremidades redutoras. Adaptado de [68] com permissão da Elsevier.

hidrolítica. Proteínas dessa classe ocorrem em plantas, bactérias e fungos, e são conhecidas por nomes como expansinas, suoleninas e looseninas [74–77].

Observações qualitativas e semi-quantitativas sugerem que o efeito da atuação

de proteínas disruptivas pode se manifestar em diferentes escalas, incluindo dispersão de fibras de celulose adjacentes, enfraquecimento de paredes celulares e descristalização de cadeias de celulose [78]. Entretanto, o mecanismo de atuação preciso destas proteínas é desconhecido.

Enzimas com atividades auxiliares (AA)

As enzimas com atividades auxiliares (AAs) são monooxigenases de clivagem de polissacarídeos (LPMO: *lytic polysaccharide monooxygenase*) que empregam cátions Cu(II) para clivar ligações glicosídicas através de um mecanismo oxidativo. Proteínas dessa classe foram inicialmente classificadas, no banco de dados CAZy, como módulos de ligação ao carboidrato da família 33 (CBM33) e membros da família GH61 de glicosídeo hidrolases [79]. Com a descoberta de seus mecanismos oxidativos, que introduziu um novo paradigma de degradação de celulose, estas enzimas foram reclassificadas nas novas classes de AAs [80]. As AAs clivam ligações glicosídicas na superfície da celulose, sem a necessidade de descristalizar cadeias individuais de glucano [79, 81]. Assim, introduzem extremidades em cadeias de glucano em pontos aleatórios da superfície da celulose e, como consequência, potencializam a ação das celulases [82]. Diferentemente das proteínas disruptivas não-hidrolíticas, o entendimento mecanístico das AAs tem avançado rapidamente nos últimos anos [81, 83–85], o que certamente é devido à natureza bioquímica de seus efeitos na biomassa – ao contrário dos efeitos biofísicos das proteínas disruptivas.

Parte II

Metodologias computacionais

Capítulo 3 Dinâmica molecular de biomoléculas

A dinâmica das biomoléculas está intimamente ligada a suas funções e, portanto, deve ser caracterizada para o entedimento de processos moleculares. No caso de proteínas, o estado nativo, apto a exercer alguma função biomolecular, é caracterizado por um *ensemble* de conformações que, embora sejam globalmente similares, diferem em detalhes, tais como conformações de cadeias laterais e de loops, ou de posições relativas de diferentes domínios. Estes estados conformacionais existem num equilíbrio dinâmico, o que está diretamente ligado aos movimentos funcionais da proteína.

Simulações de dinâmica molecular constituem uma metodologia computacional utilizada para estudar o movimento de partículas de um sistema atômico-molecular qualquer [86, 87] e, em particular, de biomoléculas em solução [88, 89], possibilitando a visualização das transições naturais entre suas várias conformações termicamente acessíveis [90]. Desde a publicação da primeira simulação de dinâmica molecular de proteína, em 1977 [91], o progressivo desenvolvimento de algoritmos, *softwares* e *hardwares* tem permitido a consolidação da dinâmica molecular como uma técnica precisa e acessível para caracterização dinâmica de biomoléculas [92]. Frequentemente, as simulações de dinâmica molecular não somente reproduzem medidas experimentais, mas, por conta da alta resolução espaço-temporal, permitem a visualização de mecanismos e a formulação de novas hipóteses acerca de determinados processos biomoleculares. Hoje em dia, é cada vez mais comum simulações na escala de tempo de vários microssegundos e de até milissegundos, que é a escala de tempo real de vários eventos moleculares [92]. Devido a desenvolvimentos de métodos multiescala que permitem o acomplamento de dinâmica molecular clássica e quântica (QM/MM), hoje é possível estudar reações químicas em ambientes complexos, como no interior de enzimas [93, 94]. Tais desenvolvimentos renderam a Martin Karplus, Michael Levitt e Arieh Warshel o Prêmio Nobel de Química de 2013.

Na dinâmica molecular clássica, em que os elétrons não são levados em conta explicitamente [95], a evolução temporal das posições de cada átomo do sistema é obtida através das leis da mecânica newtoniana. Cada átomo se move sob forças de interação exercidas por todos os outros átomos do sistema. A magnitude dessas forças depende dos detalhes químicos do meio. Após a simulação por um determinado tempo, tem-se acesso às posições (x, y, z) e velocidades (v_x, v_y, v_z) de cada um dos átomos em resolução de femtossegundos. Estes dados permitem, por exemplo, a visualização da dinâmica estrutural de uma proteína e o cálculo de propriedades estruturais, como flutuações médias ou a variação temporal da estrutura secundária. Além disso, através da mecânica estatística [96], as propriedades termodinâmicas do sistema podem ser obtidas, como a temperatura T, que é uma função das velocidades quadráticas médias $\langle v_i^2 \rangle$ dos átomos (teorema de equipartição de energia):

$$T = \frac{1}{3Nk_B} \sum_{i}^{3N} m_i \langle v_i^2 \rangle, \qquad (3.1)$$

e a pressão p, que é função tanto das coordenadas quanto das velocidades/temperatura (teorema do virial):

$$pV = Nk_BT + \frac{1}{3}\sum_{i < j} \langle \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \rangle.$$
(3.2)

Nestas equações, k_B é a constante de Boltzmann, N é o número total de partículas, V é o volume do sistema, m_i é a massa do átomo i, \mathbf{r}_{ij} é o vetor que vai da partícula i até a partícula j e \mathbf{f}_{ij} é a força que a partícula j exerce sobre a partícula i. Podese ainda obter propriedades mais específicas, tais como variações de energia livre associadas a determinados processos, como a complexação de um ligante a seu receptor ou efeitos de mutações locais na afinidade ligante-receptor [88, 97]. Em geral, uma propriedade termodinâmica $\bar{\mathcal{A}}$ pode ser expressa em termos da média temporal de seu equivalente microscópico [87]:

$$\bar{\mathcal{A}} = \frac{1}{t_s} \int_0^{t_s} \mathcal{A}(\mathbf{r}^N, \mathbf{p}^N) \mathrm{d}t, \qquad (3.3)$$

onde $\mathbf{r}^N = \mathbf{r}^N(t)$ é o conjunto das posições (x, y, z) das N partículas do sistema no tempo t, e $\mathbf{p}^N = \mathbf{p}^N(t)$ é o conjunto dos momentos (p_x, p_y, p_z) das partículas no tempo t, e t_s é o tempo de simulação. Para a equação 3.3 ser aplicada, o tempo t_s deve ser longo o suficiente para que a média convirja, o que nem sempre acontece na prática, principalmente em cálculos de energia livre [97].

Neste capítulo, será apresentada uma visão geral dos fundamentos das simulações de dinâmica molecular. Por ser uma metodologia muito bem estabelecida, inúmeras referências fornecem detalhes dos algoritmos e teorias nas quais se baseia o método [86–89, 95, 98].

3.1 Algoritmos

Em dinâmica molecular, resolve-se numericamente as equações de movimento clássicas de cada partícula, de acordo com a Segunda Lei de Newton:

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i},\tag{3.4}$$

onde \mathbf{a}_i é a aceleração do átomo $i \in \mathbf{F}_i$ é a força resultante atuando sobre o átomo i, de massa m_i . As forças são derivadas de uma função energia potencial $V(\mathbf{r}^N)$, comumente denominada *campo de força*, conforme

$$\mathbf{F}_i = -\frac{\mathrm{d}V}{\mathrm{d}\mathbf{r}_i}.\tag{3.5}$$

O problema a ser resolvido é: dadas as posições $\mathbf{r}^{N}(t)$ e velocidades $\mathbf{v}^{N}(t)$ de todas as N partículas no tempo t, quais serão as posições $\mathbf{r}^{N}(t + \delta t)$ e velocidades $\mathbf{v}^{N}(t + \delta t)$ dessas partículas após o intervalo de tempo δt ?

Se $N \geq 3$, este problema não possui solução analítica e métodos aproximados
são necessários. Se δt é escolhido o pequeno o bastante para que as variações nas posições, velocidades e acelerações das partículas sejam pequenas, então os vetores posição $\mathbf{r}_i(t)$ de cada partícula *i* podem ser expandidos numa série até termos de segunda ordem:

$$\mathbf{r}_{i}(t+\delta t) = \mathbf{r}_{i}(t) + \mathbf{v}_{i}\delta t + \frac{1}{2}\mathbf{a}_{i}(t)(\delta t)^{2}$$
(3.6)

Nesta equação, as acelerações $\mathbf{a}_i(t)$ são obtidas diretamente do campo de força, pela combinação das equações 3.4 e 3.5:

$$\mathbf{a}_{i}(t) = -\frac{1}{m_{i}} \frac{\mathrm{d}V(\mathbf{r}^{N}(t))}{\mathrm{d}\mathbf{r}_{i}}.$$
(3.7)

As velocidades em $t + \delta t$, conforme derivado em [95], podem ser obtidas por

$$\mathbf{v}_i(t+\delta t) = \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t+\delta t) + \mathbf{a}(t)}{2}.$$
(3.8)

Como esta expressão envolve acelerações no tempo $t + \delta t$, as velocidades $\mathbf{v}^N(t + \delta t)$ devem ser computadas somente após a obtenção das posições $\mathbf{r}^N(t + \delta t)$, usando

$$\mathbf{a}_{i}(t+\delta t) = -\frac{1}{m_{i}} \frac{\mathrm{d}V(\mathbf{r}^{N}(t+\delta t))}{\mathrm{d}\mathbf{r}_{i}}.$$
(3.9)

Este esquema é conhecido como algoritmo *Velocity-Verlet* e é frequentemente utilizado nas simulações de biomoléculas.

Teoricamente, as trajetórias geradas com este algoritmo irão apresentar conservação de energia E, e o sistema de N partículas permanecerá, durante toda a simulação, com o mesmo volume V. Simulações nestas condições, portanto, geram configurações correspondentes ao *ensemble* NVE, ou microcanônico. Na prática, porém, simulações de biomoléculas são realizadas em temperatura (T) constante, em vez de energia constante. Além disso, em muitos casos a pressão (p) também é mantida constante, em vez do volume. Assim, simulações biomoleculares geram trajetórias correspondentes aos *ensembles* NVT (canônico) ou NpT (isotérmicoisobárico). Nestes casos, algoritmos específicos devem ser utilizados para o controle de temperatura e pressão nas simulações, como o termostato e pistão de Langevin, descritos em [95].

O valor do passo de tempo δt deve ser pequeno o bastante para permitir a descrição dos movimentos mais rápidos do sistema, que são os movimentos de vibração de átomos leves, da ordem de femtossegundos [86]. Ligações químicas envolvendo átomos de hidrogênio, por apresentarem frequência muito alta, são geralmente mantidas rígidas em seus comprimentos de equilíbrio, através de algoritmos específicos [95, 99–101]. Nestes casos, é comum a utilização de $\delta t = 2$ fs. Isso significa que uma simulação de, por exemplo, 100 ns, requer a execução de 50.000.000 passos, envolvendo o cálculo das forças entre todos os pares de átomos (geralmente $10^4 - 10^6$ átomos, a maioria dos quais são moléculas de água) a cada passo. Assim, por conta do baixo valor de δt em relação à escala de tempo dos processos biomoleculares, as simulações de dinâmica molecular são caras computacionalmente, demandando o uso de supercomputadores em arquitetura paralela.

3.2 Campo de força

A função energia potencial $V(\mathbf{r}^N)$, ou campo de força, que descreve as interações entre os diferentes átomos do sistema, determina a qualidade da simulação. Para biomoléculas, as formas funcionais dos campos de força devem ser simples – já que o cálculo das forças demandam a maior parte do custo computacional envolvido – e ao mesmo tempo ser capazes de reproduzir dados experimentais. O conjunto de todos os parâmetros dessas funções são obtidos através de cálculos quânticos e dados empíricos, e otimizados para biomoléculas. Várias parametrizações existem, como os campos de força CHARMM, AMBER e OPLS [102], os quais são continuamente incrementados e otimizados.

O campo de força CHARMM, utilizado nas simulações desta tese, contêm termos de interações entre átomos ligados covalentemente e entre átomos não ligados covalentemente:

$$V(\mathbf{r}^{N}) = V^{\text{ligado}}(\mathbf{r}^{N}) + V^{\text{não-ligado}}(\mathbf{r}^{N}).$$
(3.10)

Os termos de interações de átomos ligados incluem vibrações de ligações químicas (figura 3.1a), distorção angular entre três átomos adjacentes (figura 3.1b) e torções

diedrais envolvendo quatro átomos adjacentes (figura 3.1c). As vibrações e distorções angulares são descritas por potenciais harmônicos, como

$$V_b(r) = k_b(r - r_0)^2, (3.11)$$

em que k_b é a constante de mola e r_0 é o comprimento de equilíbrio da ligação, e

$$V_a(\theta) = k_\theta (\theta - \theta_0)^2, \qquad (3.12)$$

onde k_{θ} é a constante de mola e θ_0 é o ângulo de equilíbrio. As torções diedrais são descritas como série de cossenos, que dá conta da periodicidade das torções:

$$V_t(\phi) = \sum_n k_{\phi_n} [1 + \cos(n\phi - \delta_n)],$$
 (3.13)

onde k_{ϕ_n} é uma constante ligada à altura das barreiras torcionais, n é a multiplicidade do potencial, que reflete o número de mínimos ao longo de uma torção diedral completa e δ_n é o ângulo de fase, que determina a localização dos máximos da barreira torcional.



Figura 3.1. Termos da parte covalente do campo de força: (a) estiramento de ligações; (b) distorção angular; (c) torção diedral. Adaptado de [98].

Além destes três termos, o campo de força CHARMM contém, para alguns grupos de átomos, dois outros potenciais: o de Urey-Bradley e as torções diedrais impróprias. O potencial de Urey-Bradley representa interações entre dois átomos separados por duas ligações químicas (interações 1-3), que é modelado por um potencial harmônico:

$$V_{UB} = k_{UB}(u - u_0)^2, (3.14)$$

onde u é a distância entre os átomos i e k da figura 3.1b. As torções diedrais impróprias são utilizadas para manter a planaridade e quiralidade de grupos de quatro átomos ligados, e são modeladas também por potenciais harmônicos:

$$V_I = k_I (\varphi - \varphi_0)^2, \qquad (3.15)$$

onde φ é o ângulo diedro impróprio formado pelos quatro átomos ligados entre si, e k_I é a constante de força associada às interações.

Os termos não-covalentes (figura 7.4) incluem interações eletrostáticas e de van der Waals. As interações eletrostáticas são descritas pelo potencial de Coulomb

$$V_e(\{r_{ij}\}) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} \frac{q_i q_j}{\epsilon r_{ij}},$$
(3.16)

em que q_i é a carga do átomo i, r_{ij} é a distância entre os átomos i e j, e ϵ é a permissividade elétrica no vácuo multiplicada por 4π . As interações de van der Waals, são modeladas com o potencial de Lennard-Jones:

$$V_{vdW}(\{r_{ij}\}) = \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}}\right)^{6} \right], \quad (3.17)$$

em que ϵ_{ij} é a profundidade do poço de potencial associados aos átomos $i \in j$, e $R_{min,ij}$ é a distância em que o potencial é mínimo. O termo dependente de $1/r_{ij}^{12}$ descreve repulsões de curto alcance associadas à superposição de núvens eletrônicas, e o termo dependente de $1/r_{ij}^6$ descreve as atrações dispersivas de London.

A figura 7.4 ilustra a forma dos potenciais de Lennard-Jones e de Coulomb, e ressalta o fato de que o primeiro exibe um mínimo e decai a zero rapidamente, enquanto o segundo decai lentamente com a distância, e é sempre atrativo ou sempre repulsivo, dependendos das cargas dos átomos.



Figura 3.2. Potenciais de Lennard-Jones e de Coulomb.

3.3 Condições de contorno periódicas

As simulações de dinâmica molecular são realizadas com um número de átomos da ordem de $10^4 - 10^6$, que é muito menor que o número de átomos de um sistema macroscópico, da ordem de 10^{24} . Grande parte desses átomos estarão nas extremidades do sistema e, portanto, sujeitos a efeitos de borda. Para contornar este problema, utiliza-se condições de contorno periódicas (figura 3.3). O domínio das simulações é geralmente uma caixa retangular. Quando um átomo cruza a fronteira dessa caixa, um outro átomo entra pela face oposta da caixa com a mesma velocidade. Assim, o sistema se torna infinitamente periódico em todas as direções e os efeitos de borda são minimizados.

Interações de curto alcance: imagem mínima

Por serem de curto alcance, as interações de van der Waals são truncadas a partir de uma determinada distância, ou seja, as interações são computadas somente para pares de átomos separados por uma distância menor que uma distância de corte (geralmente entre 10 e 12 Å). Interações entre átomos mais distantes que esta distância são de magnitude desprezível e não necessitam ser computadas. Para pa-



Figura 3.3. Condições de contorno periódicas. Quando um átomo cruza as fronteiras da caixa de simulação, um outro átomo entra pela face oposta com a mesma velocidade.

res de átomos próximos a fronteiras opostas da caixa de simulação, embora estejam separados por uma distância maior que a distância de corte, devem interagir com a imagem virtual dos átomos numa caixa de simulação vizinha. Na figura 3.3, embora a distância real dos átomos 1 e 2 seja maior que a distância de corte, estes átomos devem interagir um com outro através de suas imagens virtuais; assim, o átomo 1 deve interagir com o átomo 2', e o átomo 2 deve interagir com o átomo 1'. Essa é a chamada convenção da imagem mínima.

Interações de longo alcance: somas de Ewald

Por serem de longo alcance, as interações eletrostáticas não podem ser truncadas. Um átomo dentro da caixa de simulação interage com todos os outros átomos da caixa e também com todos os átomos virtuais das infinitas caixas periódicas em torno da caixa central [98]. Logo, a energia potencial eletrostática do sistema periódico é

$$V_e = \frac{1}{2} \sum_{\mathbf{n}}' \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{q_i q_j}{\epsilon |\mathbf{r}_{ij} + \mathbf{n}|},$$
(3.18)

em que $\mathbf{n} = (n_x L, n_y L, n_z L)$ é o vetor que determina a posição de cada caixa de simulação, de dimensões $L \times L \times L$, com n_x, n_y e n_z inteiros. O apóstrofo no somatório em \mathbf{n} indica que $i \neq j$ para $\mathbf{n} = 0$, ou seja, uma partícula não pode interagir consigo mesma, mas pode interagir com sua imagem virtual. A soma 3.18 converge muito lentamente e, além disso, é condicionalmente convergente (a convergência depende da ordem em que os termos são somados), o que dificulta sua utilização nas simulações. As somas de Ewald contornam estes problemas através da decomposição da série 3.18 em duas outras séries que convergem mais rapidamente [103]. A figura 3.4 representa as cargas pontuais q_i por barras. A cada uma destas cargas pontuais q_i adiciona-se uma densidade de carga gaussiana, do tipo

$$\rho_i(r) = -q_i \frac{\alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2), \qquad (3.19)$$

tal que

$$\int d\mathbf{r}\rho_i(r) = -q_i, \qquad (3.20)$$

onde α é um parâmetro que regula a largura das gaussianas e interfere na convergência da série. Ou seja, a densidade de carga engloba a mesma carga total de q_i , mas com o sinal oposto. Isso tem o efeito de blindar o potencial gerado pelas cargas pontuais. Para compensar, uma outra densidade de carga gaussiana, equivalente à anterior, mas de mesmo sinal das cargas pontuais, é adicionada. O efeito líquido de adicionar as densidades de carga é, portanto, nulo, e constitui apenas um artifício matemático. A figura 3.4 ilustra esta construção.



Figura 3.4. Decomposição de cargas pontuais em densidades de carga. Adaptado de [98].

A energia potencial associada às cargas blindadas é

$$V_{real} = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j \operatorname{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{\epsilon |\mathbf{r}_{ij} + \mathbf{n}|}.$$
(3.21)

A função erro complementar, $\operatorname{erfc}(r)$, que decai a zero rapidamente, faz com que este potencial seja de curto alcance, de forma que pode ser truncado a partir de uma distância de corte e computado utilizando a convenção da imagem mínima.

O potencial devido às densidades de carga gaussianas computado utilizando a equação de Poisson da eletrostática ($\nabla^2 V(r) = -4\pi\rho(r)/\epsilon$) [104] no espaço recíproco. A solução da equação fornece

$$V_{reciproco} = \frac{2\pi^2}{\epsilon L^3} \sum_{\mathbf{k}}' \sum_{i=1}^N \sum_{j=1}^N q_i q_j \exp(i\mathbf{k} \cdot \mathbf{r}_{ij}) \frac{1}{\mathbf{k}^2} \exp\left(-\frac{\mathbf{k}^2}{4\alpha^2}\right) + \frac{\alpha}{\epsilon \pi^{1/2}} \sum_{i=1}^N q_i^2, \qquad (3.22)$$

onde $\mathbf{k} = 2\pi \mathbf{n}/L$ são os vetores recíprocos e $i = \sqrt{-1}$. Embora este potencial seja de longo alcance, a série converge muito mais rapidamente que a série 3.18 [103]. Logo, a energia potencial eletrostática se escreve como

$$V_{e} = V_{real} + V_{reciproco}$$

$$= \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_{i}q_{j}\operatorname{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{\epsilon |\mathbf{r}_{ij} + \mathbf{n}|} + \frac{2\pi^{2}}{\epsilon L^{3}} \sum_{\mathbf{k}} \sum_{i=1}^{N} \sum_{j=1}^{N} q_{i}q_{j} \exp(i\mathbf{k} \cdot \mathbf{r}_{ij}) \frac{1}{\mathbf{k}^{2}} \exp\left(-\frac{\mathbf{k}^{2}}{4\alpha^{2}}\right) + -\frac{\alpha}{\epsilon \pi^{1/2}} \sum_{i=1}^{N} q_{i}^{2}.$$
(3.23)

A soma de Ewald, como apresentada aqui, foi desenvolvida para tratar sistemas periódicos em física do estado sólido, e não para estudos computacionais. Nas implementações computacionais, utiliza-se variações que empregam transformadas de Fourier discretas. Um dos algoritmos, em que as densidades de carga são discretizadas numa malha tridimensional, é o *particle mesh* Ewald (PME), frequentemente utilizado em simulação biomolecular [105].

3.4 Aspectos práticos

A simulação de dinâmica molecular deve partir de condições iniciais bem definidas. Para proteínas, as coordenadas iniciais são obtidas experimentalmente por cristalografia de raios X ou ressonância magnética nuclear. Na maior parte dos casos, o experimento de difração de raios X não resolve átomos de hidrogênio e, portanto, estes devem ser adicionados à proteína. A biomolécula deve então ser imersa numa caixa contendo moléculas de água em excesso, além de íons como Na⁺ e Cl⁻ em concentrações adequadas para o estudo em questão. Tendo as coordenadas espaciais de todos os átomos, as velocidades iniciais são atribuídas aleatoriamente, segundo a distribuição de Maxwell-Boltzmann correspondente à temperatura desejada [89].

As coordenadas iniciais do sistema devem ser compatíveis com o campo de força. Isso significa que possíveis superposições entre dois átomos, ou ligações químicas com comprimentos muito diferentes dos comprimentos de equilíbrio ditado pelo campo de força, irão gerar forças muito intensas e causar uma instabilidade na propagação numérica da trajetória. Para evitar isso, utiliza-se de métodos de minimização de energia para aliviar as distorções do sistema em relação a um mínimo local da superfície de energia potencial. Métodos de otimização como *Steepest descent* ou Gradientes Conjugados são frequentemente utilizados em simulações de biomoléculas [88].

Com o início da simulação, o sistema entrará num regime transiente, durante o qual ocorrerá a relaxação para a condição termodinâmica de interesse, como temperatura e pressão. Assim, os instantes iniciais devem ser monitorados até que as propriedades do sistema comecem a flutuar em torno de valores médios constantes. A partir daí, a trajetória pode ser considerada para análises diversas (seção 3.5).

O tempo de simulação deve ser consistente com a escala de tempo dos processos moleculares. Em proteínas, movimentos de cadeias laterais dos aminoácidos ocorrem na escala de tempo de 10^{-15} a 10^{-9} s, enquanto movimentos coletivos de loops e movimentos entre domínios podem demorar de 10^{-9} a 10^{-3} s. Obviamente, simulações longas demandam maiores custos computacionais. Embora simulações de proteínas na ordem dos milissegundos já tenham sido reportadas [106], a maior parte dos computadores conseguem atingir a escala de tempo de dezenas de nanossegundos a alguns microssegundos [92]. Dadas as limitações computacionais, diversos métodos que permitem acelerar a escala de tempo dos processos moleculares têm sido desenvolvidos nos últimos anos, conhecidos coletivamente como métodos de amostragem ampliada. Com isso, o estudo da dinâmica biomolecular em longas escalas de tempo torna-se possível com baixo custo computacional [107]. Estes métodos, porém, não foram utilizados nesta tese.

3.5 Análises

Uma vez geradas as trajetórias do sistema, estas podem ser utilizadas para visualização dos movimentos dos átomos, utilizando programas de visualização como o VMD [108], ou para a obtenção de propriedades estruturais e termodinâmicas. Em sistemas bioquímicos, muitas vezes o interesse é estudar a dinâmica estrutural da biomolécula, que é caracterizada por funções que dependendem das coordenadas espaciais do sistema. Abaixo, as análises utilizadas neste tese são descritas.

3.5.1 Flutuações estruturais

A evolução temporal das flutuações de um determinado grupo de átomos Ω , como, por exemplo, o conjunto dos carbonos α de uma proteína ou o conjunto de átomos de um determinado aminoácido ou ligante, pode ser obtida através da raiz quadrada do desvio médio quadrático (rmsd: *root mean squared deviation*):

$$\operatorname{rmsd}_{\Omega}(t) = \left(\frac{1}{N_{\Omega}} \sum_{i \in \Omega} \left[\mathbf{r}_{i}(t) - \mathbf{r}_{i}^{ref}\right]^{2}\right)^{\frac{1}{2}}.$$
(3.24)

Nesta equação, N_{Ω} é o número de átomos do grupo Ω , e \mathbf{r}_i^{ref} são coordenadas de referência em torno da qual as flutuações do grupo Ω são calculadas. A referência

do rmsd pode ser as coordenadas médias, as coordenadas cristalográficas ou as coordenadas de um determinado instante da simulação. Antes de aplicar a equação 3.24, o *ensemble* de estruturas da biomolécula ao longo da trajetória deve ser alinhado, através da remoção dos movimentos translacionais e rotacionais, para que somente o movimento interno da biomolécula seja levado em consideração.

A flutuação média de um dado resíduo da biomolécula pode ser obtido pela média temporal do rmsd deste resíduo. Esta quantidade é denotada por rmsf (*root mean squared fluctuation*), e definida por

$$\operatorname{rmsf}_{\operatorname{res}} = \langle \operatorname{rmsd}_{\operatorname{res}} \rangle = \frac{1}{N_p} \sum_{n=0}^{N_p} \operatorname{rmsd}_{\operatorname{res}}(t_n), \qquad (3.25)$$

onde N_p é o número total de configurações do *ensemble* gerado pela simulação, e t_n é o instante correspondente à *n*-ésima configuração do *ensemble*. Para proteínas, geralmente se plota o rmsf para todos os resíduos, o que fornece o perfil de mobilidade ao longo da estrutura primária da proteína.

3.5.2 Correlações dinâmicas cruzadas

Correlações entre flutuações de diferentes regiões da biomolécula podem ser computadas em termos do coeficiente de correlação dinâmica cruzada, que para o par de átomos i e j (normalmente pares de carbonos α) é definido por

$$c_{ij} = \frac{\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle}{\sqrt{\langle (\Delta \mathbf{r}_i)^2 \rangle} \sqrt{\langle (\Delta \mathbf{r}_j)^2 \rangle}},$$
(3.26)

em que \mathbf{r}_i é o vetor posição do átomo i, e $\Delta \mathbf{r}_i = \mathbf{r}_i - \langle \mathbf{r}_i \rangle$ é seu desvio em relação à sua posição média. Os valores de c_{ij} são usualmente plotados numa matriz $N \times N$ em escala de cores, contendo as correlações cruzadas entre todos os pares de carbonos α de uma proteína, o que mostra informações sobre sua dinâmica intrínseca.

Os valores de c_{ij} variam de -1 a +1. Se as flutuações de i e j são independentes,

então

$$c_{ij} = \frac{\langle \Delta \mathbf{r}_i \rangle \cdot \langle \Delta \mathbf{r}_j \rangle}{\sqrt{\langle (\Delta \mathbf{r}_i)^2 \rangle} \sqrt{\langle (\Delta \mathbf{r}_j)^2 \rangle}} = 0, \qquad (3.27)$$

pois $\langle \Delta \mathbf{r} \rangle = \langle \mathbf{r} - \langle \mathbf{r} \rangle \rangle = \langle \mathbf{r} \rangle - \langle \mathbf{r} \rangle = 0$. Se as flutuações de *i* e *j* são, em média, sincronizadas e ocorrem na mesma direção, então $c_{ij} > 0$ e os movimentos são correlacionados. Se as flutuações são sincronizadas, mas ocorrem em direções opostas, então $c_{ij} < 0$ e os movimentos são anti-correlacionados.

Há dois casos em que o coeficiente de correlação c_{ij} , como calculado pela equação 3.28, é nulo, mas os movimentos de i e j são ainda correlacionados. O primeiro é quando os movimentos são correlacionados, mas ocorrem fora de fase. Por exemplo, enquanto a flutuação em torno da média do átomo i aumenta, a do átomo j diminui, de forma que o produto escalar $\Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j$ é sempre próximo de zero. O outro caso é quando os movimentos são correlacionados e em fase, mas ocorrem em direções ortogonais entre si. A equação 3.28 pode ser reescrita como

$$c_{ij} = \frac{\langle |\Delta \mathbf{r}_i| |\Delta \mathbf{r}_j| \cos \theta_{ij} \rangle}{\sqrt{\langle (\Delta \mathbf{r}_i)^2 \rangle} \sqrt{\langle (\Delta \mathbf{r}_j)^2 \rangle}},$$
(3.28)

em que θ_{ij} é o ângulo entre os vetores $\Delta \mathbf{r}_i \in \Delta \mathbf{r}_j$. Se os movimentos forem perpendiculares, então $\cos \theta_{ij} = 0$ e, logo, $c_{ij} = 0$. Assim, descorrelação implica em $c_{ij} = 0$, mas a recíproca nem sempre é verdadeira.

3.5.3 Análise de componentes principais

A dinâmica de uma proteína apresenta flutuações em diferentes escalas, desde movimentos vibracionais de baixa amplitude até flutuações de larga escala, envolvendo deslocamentos de loops inteiros e movimentos entre domínios. Os movimentos que permitem às proteínas exercerem suas funções, como ligação ao substrato ou catálise, são aqueles de mais alta amplitude, como revelado por inúmeros estudos estruturais [90]. Os movimentos de baixa amplitude funcionam como vínculos que mantêm a integridade estrutural da proteína [109].

A análise de componentes principais (PCA) consiste num conjunto de operações

matemáticas que permitem decompor a trajetória da biomolécula em modos de alta e baixa amplitudes, separando os movimentos coletivos de larga escala dos movimentos vibracionais locais [109]. Assim, esta decomposição filtra o movimento aparentemente aleatório de uma biomolécula, mostrando aqueles que são relevantes para sua função [90, 110]. A trajetória filtrada, isenta de vibrações randômicas de baixa amplitude, é denotada por trajetória essencial, e numerosos estudos mostram que esta trajetória pode ser descrita por um número muito pequeno de modos de movimento, em relação ao número total de graus de liberdade do sistema (3N) [90, 109].

Matematicamente, a análise de componentes principais consiste numa transformação ortogonal que permite representar a trajetória da biomolécula, originalmente no sistema de coordenadas Cartesianas, em um novo sistema de coordenadas coletivas, tal que as flutuações de mais alta amplitude ocorrem ao longo do primeiro eixo (PC1: primeiro componente principal), seguido do segundo eixo (PC2), e assim por diante [90].

O primeiro passo da análise de componentes principais consiste em computar a matriz de covariância \mathbf{C} , de dimensão $3N \times 3N$, entre cada grau de liberdade do sistema:

$$\mathbf{C} = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) (\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle, \qquad (3.29)$$

em que \mathbf{x} é o vetor coluna, de dimensão 3N, que contém as coordenadas cartesianas de todas as N partículas:

$$\mathbf{x} = \begin{bmatrix} x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2 \ \cdots \ x_N \ y_N \ z_N \end{bmatrix}^T.$$
(3.30)

A matriz \mathbf{C} deve então ser diagonalizada por uma transformação ortogonal \mathbf{U} :

$$\mathbf{U}^T \mathbf{C} \mathbf{U} = \mathbf{\Lambda}.\tag{3.31}$$

As colunas da matriz **U** são os autovetores de **C**, $\{\mathbf{u}_i\}$, e Λ é a matriz diagonal dos autovalores de **C**, $\{\lambda_i\}$. Os autovalores representam a flutuação quadrática média ao longo de cada um dos autovetores de **C**, chamados de componentes principais (PC1, PC2 *etc.*). O conjunto dos 3N autovetores de **C**, $\{\mathbf{u}_i\}$, forma uma nova base ortonormal, na qual a trajetória do sistema pode ser representada. Neste nova base $\{\mathbf{u}_i\}$, cada coordenada p_i , chamada coordenada coletiva, é obtida pela projeção do vetor $\mathbf{x} - \langle \mathbf{x} \rangle$ sobre o autovetor correspondente \mathbf{u}_i :

$$p_i(t) = \mathbf{u}_i \cdot (\mathbf{x}(t) - \langle \mathbf{x} \rangle), \qquad (3.32)$$

ou, na forma matricial,

$$\mathbf{p}(t) = \mathbf{U}^T(\mathbf{x}(t) - \langle \mathbf{x} \rangle) \tag{3.33}$$

onde **p** é o vetor coluna contendo as coordenadas coletivas ao longo dos 3N autovetores de **C**. As coordenadas coletivas ao longo dos primeiros autovetores $(p_1(t), p_2(t), ...)$ usualmente indicam as principais alterações conformacionais que ocorrem com a proteína.

A trajetória do sistema pode ser recuperada de volta para o espaço cartesiano através de

$$q_i(t) = \langle q_i \rangle + \sum_{j=1}^{3N} U_{ij} p_j(t),$$
 (3.34)

onde q_i é um elemento qualquer do vetor de coordenadas cartesianas **x**. Mais interessante, a trajetória ao longo de cada autovetor isolado pode ser expressa no sistema cartesiano, permitindo a visualização da trajetória ao longo de cada PC:

$$q_i^j(t) = \langle q_i \rangle + U_{ij} p_j(t). \tag{3.35}$$

Além disso, a trajetória pode ser projetada no subespaço formado pelos M < 3Nprimeiros autovetores utilizados para descrever o movimento essencial da proteína:

$$q_i^{1 \to M}(t) = \langle q_i \rangle + \sum_{j=1}^M U_{ij} p_j(t).$$
 (3.36)

Com a escolha apropriada de M, a trajetória obtida pela equação 3.36 mostra os movimentos funcionais da proteína sem a presença dos ruídos dos movimentos vibracionais locais, e pode ser utilizada para análises estruturais adicionais.

A distribuição dos valores de $p_i(t)$ pode ainda ser utilizada para obter o perfil

de energia livre ao longo de cada PC através de [97]

$$\Delta G(p_i) = -k_B T \ln \frac{P(p_i)}{P_0}, \qquad (3.37)$$

ou, ao longo de dois PCs,

$$\Delta G(p_i, p_j) = -k_B T \ln \frac{P(p_i, p_j)}{P_0}, \qquad (3.38)$$

em que $P(p_i)$ é a distribuição de probabilidades da coordenada p_i , e $P(p_i, p_j)$ é a distribuição conjunta das coordenadas p_i e p_j . P_0 é um valor de referência arbitrário, que pode ser considerado como o valor máximo de $P(p_i)$ ou $P(p_i, p_j)$, de forma que o mínimo de G coincida com zero.

Capítulo 4 Teoria de Equações Integrais

As simulações de dinâmica molecular propagam a trajetória de um sistema ao longo do espaço de fases, que é o espaço das coordenadas \mathbf{r}^N e momentos \mathbf{p}^N de todas as partículas do sistema. Após a simulação por um longo período de tempo, temse uma coleção de pontos neste espaço de fases, e propriedades termodinâmicas $(\bar{\mathcal{A}})$ podem ser obtidas por uma média temporal do equivalente microscópico da propriedade $(\mathcal{A}(\mathbf{r}^N, \mathbf{p}^N))$, como na equação 3.3. Alternativamente, a estes pontos no espaço de fases pode ser atribuída uma densidade de probabilidade $\rho(\mathbf{r}^N, \mathbf{p}^N)$ de que o sistema esteja no estado definido por um dado \mathbf{r}^N e \mathbf{p}^N . Esta coleção de pontos constitui um *ensemble* mecânico-estatístico, e propriedades termodinâmicas podem ser obtidas como média sobre todas as suas configurações:

$$\langle \mathcal{A} \rangle = \iint \mathrm{d} \mathbf{r}^N \mathrm{d} \mathbf{p}^N \mathcal{A}(\mathbf{r}^N, \mathbf{p}^N) \rho(\mathbf{r}^N, \mathbf{p}^N).$$
 (4.1)

Numa simulação suficientemente longa para que o sistema passe arbitrariamente próximo de todos os pontos acessíveis do espaço de fases, tem-se que $\langle \mathcal{A} \rangle = \bar{\mathcal{A}}$ [111]. Assim, o que se faz numa simulação de dinâmica molecular é amostrar o espaço de fases tão bem quanto possível para se ter uma boa estimativa de $\rho(\mathbf{r}^N, \mathbf{p}^N)$.

A teoria de equações integrais objetiva resolver, do ponto de vista termodinâmico-estatístico, o mesmo problema de uma simulação de dinâmica molecular, mas por uma rota diferente: obteção de funções de distribuição diretamente da mecânica-estatística de líquidos [112]. A teoria de equações integrais na forma de 3D-RISM (three-dimensional reference interaction site model), que foi usada nos estudos dos Capítulos 8 e 9, consiste num conjunto de equações integrais cujas soluções, obtidas computacionalmente, são funções de distribuição de sítios atômicos de um dado solvente em torno de um soluto qualquer [113]. Estas funções permitem a obtenção das propriedades termodinâmicas do sistema através de expressões analíticas simples, como descrito neste capítulo.

O fato de ser uma teoria baseada na mecânica estatística, em que médias sobre um *ensemble* são tomadas explicitamente, esta metodologia não exibe os problemas de amostragem limitada das simulações de dinâmica molecular. Assim, propriedades termodinâmicas sensíveis à amostragem, como energia livre e entropia, podem ser facilmente obtidas, evidentemente a custo de aproximações, como descrito adiante. Propriedades dinâmicas estão fora do escopo da teoria de equações integrais. Uma breve descrição da teoria de equações integrais será dada aqui, começando pela teoria de líquidos monoatômicos, passando por líquidos moleculares, até chegar na descrição de sistemas soluto-solvente pela teoria 3D-RISM, também conhecida como teoria molecular da solvatação. A formulação completa é muito envolvida matematicamente para ser desenvolvida aqui, mas pode ser obtida em outras referências [96, 112, 114, 115].

4.1 Funções de distribuição de líquidos monoatômicos

A estrutura de um líquido monoatômico é descrita pela função de distribuição radial g(r), que é uma medida da probabilidade de encontrar um átomo do sistema a uma distância r de um outro átomo central. Se ρ é a densidade numérica de átomos do líquido, então $4\pi r^2 \rho g(r) dr$ é o número de átomos presentes numa calota esférica de espessura dr a uma distância r do átomo central (figura 4.1a), de forma que $\rho g(r)$ é a densidade local do líquido a uma distância r do átomo central. A figura 4.1b mostra a forma típica de g(r) para um líquido monoatômico (argônio líquido), onde os picos representam regiões de maior densidade local em relação a uma distribuição homogênea dos átomos do líquido [116].

Se as interações do líquido de N átomos puderem ser descritas por potenciais



Figura 4.1. (a) A função g(r) é tal que $4\pi r^2 \rho g(r) dr$ é o número de átomos presentes na calota esférica em cinza em torno de um átomo central (círculo vazio). (b) Forma de uma função de distribuição radial para argônio líquido, mostrando a estruturação dos átomos em torno de um átomo central. A distância r é plotada em unidades de σ (parâmetro do potencial de Lennard-Jones).

de pares aditivos, do tipo

$$V(\mathbf{r}^N) = \sum_{i < j} u(r_{ij}), \qquad (4.2)$$

como o potencial de Lennard-Jones (equação 3.17), então toda a termodinâmica do sistema pode ser obtida através de g(r) [96]. Por exemplo, as funções termodinâmicas energia interna U a pressão p são expressas por

$$U(\rho,T) = \frac{N\rho}{2} \int_0^\infty u(r)g(r;\rho,T)4\pi r^2 \mathrm{d}r$$
(4.3)

е

$$p(\rho,T) = \rho k_B T - \frac{\rho^2}{6} \int_0^\infty r u'(r) g(r;\rho,T) 4\pi r^2 \mathrm{d}r, \qquad (4.4)$$

em que k_BT é a constante de Boltzmann multiplicada pela temperatura.

O problema de obter a função g(r) a partir de primeiros princípios é o objeto da teoria de equações integrais. Há várias dessas equações, como a de Kirkwood, de Born-Green-Yvon, entre outras, que podem ser encontradas em textos de mecânica estatística [96, 115]. Dentre as equações integrais, a equação de Ornstein-Zernike (OZ), que pode ser obtida através da teoria do funcional de densidade de líquidos [112, 117], consitui a base da teoria 3D-RISM. Para líquidos monoatômicos a equação de OZ é

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13})h(r_{23}) \mathrm{d}\mathbf{r}_3.$$
(4.5)

Nesta equação, $h(r_{12}) = g(r_{12}) - 1$ é a função de correlação total, e é uma medida da influência total do átomo 1 sobre o átomo 2. A função de correlação total se divide em duas partes: uma direta e outra indireta. A função de correlação direta, $c(r_{12})$, reflete a influência direta do átomo 1 sobre o átomo 2 e, assintoticamente, tende ao potencial de interação $[c(r) \rightarrow -u(r)/k_BT]$. A parte indireta, representada pela integral da equação 4.5, reflete a influência propagada do átomo 1 para o 3, o qual, por sua vez, exerce influência sobre átomo 2, diretamente ou através de outros átomos. Esta influência indireta do átomo 3 é ponderada pela densidade e tomada como média sobre todas as suas posições [96]. É fácil ver que, substituindo $h(r_{23})$ na equação 4.5 por sua correspondente equação de OZ, aparece

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13})c(r_{23})d\mathbf{r}_3 + \rho^2 \iint c(r_{13})c(r_{24})h(r_{34})d\mathbf{r}_3 d\mathbf{r}_4.$$
(4.6)

Continuando o processo, chega-se numa série de integrais múltiplas, que mostra a propagação das contribuições indiretas sobre o átomo 2.

A equação integral de OZ (4.5) envolve duas funções de correlação: h(r) e c(r); portanto, sua solução requer uma outra relação entre c(r) e h(r) para fechar uma equação que dependa somente de h(r). A forma geral da relação de fechamento é

$$h(r) = \exp\left[-u(r)/k_B T + h(r) - c(r) + B(r)\right] - 1, \qquad (4.7)$$

onde B(r) é um funcional representado por uma série infinita de integrais das funções de correlação que, na prática, é impossível de ser computado, fazendo-se necessário a incorporação de aproximações [112, 117, 118]. Há várias aproximações para a relação de fechamento, as quais serão discutidas adiante.

A solução equação integral de OZ 4.5 complementada com a relação de fechamento 4.7 fornece função de distribuição radial do líquido, g(r) = h(r) + 1, a partir da qual todas as funções termodinâmicas do líquido podem ser obtidas.

4.2 Equação de Ornstein-Zernike molecular e RISM

A equação integral de OZ pode ser generalizada para tratar líquidos moleculares. Neste caso, cada molécula do líquido é identificada pela posição de seu centro de massa $\mathbf{r} = (x, y, z)$ e por sua orientação espacial $\mathbf{\Omega} = (\psi, \theta, \phi)$ (ângulos de Euler). A equação de OZ molecular pode ser expressa como

$$h(r_{12}, \boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = c(r_{12}, \boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) + \frac{\rho}{\Omega} \iint c(r_{13}, \boldsymbol{\Omega_1}, \boldsymbol{\Omega_3}) h(r_{32}, \boldsymbol{\Omega_3}, \boldsymbol{\Omega_2}) \mathrm{d}\mathbf{r}_3 \mathrm{d}\boldsymbol{\Omega}_3, \quad (4.8)$$

onde $\Omega = \int d\Omega$ é um fator de normalização geométrico. A aplicação desta equação requer que interações intermoleculares do líquido sejam descritas por potenciais que dependam da distância entre os centros de massa de duas moléculas e de suas orientações relativas no espaço. Embora tais desenvolvimentos tenham sido feitos, possuem aplicações limitadas [117]. Além disso, a equação de OZ molecular apresenta 6 dimensões $(x, y, z, \psi, \theta, \phi)$, o que dificulta sua solução computacional para sistemas químicos complexos [119]. Potenciais baseados em sítios de interação, como os potenciais intermoleculares utilizados em dinâmica molecular, apresentam faixa de aplicabilidade muito mais ampla. Nos modelos de sítios de interação, o potencial de interação entre duas moléculas é:

$$u_{12} = \sum_{\alpha,\gamma} u_{\alpha\gamma}(r) \tag{4.9}$$

onde α são sítios (átomos) da molécula 1 e γ são sítios da molécula 2. Para incorporar este tipo de potencial, Chandler & Andersen introduziram, em 1972, aproximações na equação 4.8 para fazê-la operar com sítios de interação e reduzir sua dimensionalidade de 6 para 1 (distância escalar r entre os sítios atômicos) [120]. Neste desenvolvimento, para cada par de sítios de interação $\alpha \in \gamma$ (em vez de cada par de moléculas) deve haver uma equação de OZ relacionado as funções de correlação. Tais correlações sítio-sítio são obtidas fixando a distância entre cada par de sítio $\alpha \in \gamma$ e fazendo a média das funções de correlação moleculares sobre todas as possíveis orientações das moléculas. Para fazer este processo analiticamente, a função de correlação direta $c(r_{12}, \Omega_1, \Omega_2)$ entre moléculas do líquido deve ser aproximada como a soma das correlações sítio-sítio das moléculas [117]:

$$c(r_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2) = \sum_{\alpha, \gamma} c_{\alpha\gamma} (\mathbf{r}_{\alpha 1} - \mathbf{r}_{\gamma 2})$$
(4.10)

Com esta aproximação, a equação de OZ molecular pode ser modificada para operar com modelos de sítio de interação, fornecendo

$$h_{\alpha\gamma}(r) = \sum_{\mu\nu} \iint \omega_{\alpha\mu} \left(|\mathbf{r}_{1\alpha} - \mathbf{r}_{1\mu}| \right) c_{\mu\nu} \left(|\mathbf{r}_{1\mu} - \mathbf{r}_{2\nu}| \right) \times \\ \times \left[\omega_{\nu\gamma} \left(|\mathbf{r}_{2\nu} - \mathbf{r}_{2\gamma}| \right) + \rho h_{\nu\gamma} \left(|\mathbf{r}_{2\nu} - \mathbf{r}_{2\gamma}| \right) \mathrm{d}\mathbf{r}_{1\alpha} \mathrm{d}\mathbf{r}_{2\gamma}, \tag{4.11}$$

que geralmente aparece na literatura na forma

$$h_{\alpha\gamma}(r) = \sum_{\mu,\nu} \omega_{\alpha\mu}(r) \otimes c_{\mu\nu}(r) \otimes \left[\omega_{\nu\gamma}(r) + \rho h_{\nu\gamma}(r)\right], \qquad (4.12)$$

onde \otimes indica os produtos de funções dentro das integrais (integrais de convolução). Na equação 4.12, as funções $\omega_{\alpha\gamma}(r)$, definidas por

$$\omega_{\alpha\gamma}(r) = \delta_{\alpha\gamma}\delta(r) + (1 - \delta_{\alpha\gamma})\frac{\delta(r - l_{\alpha\gamma})}{4\pi l_{\alpha\gamma}^2},$$
(4.13)

são funções de correlação intramolecular – entre sítios atômicos de uma mesma molécula – que representam a geometria das moléculas do líquido, com distâncias sítio-sítio $l_{\alpha\gamma}$. Nesta equação, $\delta(r)$ é a função delta de Dirac, e $\delta_{\alpha\gamma}$ é a delta de Kronecker. Se $\alpha = \gamma$, então $\delta_{\alpha\gamma} = 1$ e $\omega_{\alpha\gamma}(r) = \delta(r)$, ou seja, a função de correlação de um sítio com ele mesmo é uma função com um pico de intensidade infinita centrado em r = 0, mas cuja integral sobre todo o espaço é igual a 1. Se $\alpha \neq \gamma$, então $\delta_{\alpha\gamma} = 0$ e $\omega_{\alpha\gamma}(r) = \delta(r - l_{\alpha\gamma})/4\pi l_{\alpha\gamma}^2$, que é uma distribuição centrada na distância entre os sítios $\alpha = \gamma$, $r = l_{\alpha\gamma}$, e normalizada pela área da esfera de raio $l_{\alpha\gamma}$.

As equações 4.12 para cada par de sítio $\alpha \in \gamma$, complementada por uma relação

de fechamento do tipo

$$h_{\alpha\gamma}(r) = \exp\left[-u_{\alpha\gamma}(r)/k_BT + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r) + B_{\alpha\gamma}(r)\right] - 1, \qquad (4.14)$$

consistituem a teoria RISM (reference interaction site model) unidimensional. Sua solução fornece as funções de correlação $h_{\alpha\gamma}(r) \in c_{\alpha\gamma}(r)$ para todos os pares de sítios $\alpha \in \gamma$ das moléculas do líquido, que podem ser utilizadas para obter as funções termodinâmicas do sistema. As equações da teoria RISM podem ser facilmente generalizadas para tratar líquidos compostos, com múltiplas espécies químicas [121].

Devido a aproximações na teoria RISM, como apresentada aqui, a descrição de propriedades dielétricas de sistemas contendo espécies polares e carregadas é insatisfatória. Para eliminar estes defeitos, Perkyns & Pettitt introduziram correções analíticas para que a teoria RISM seja capaz de apresentar consistência dielétrica. Nesta versão, a teoria é denominada DRISM (*dieletrically consistent* RISM) [113, 122, 123].

4.3 3D-RISM: sistemas soluto-solvente

No caso de uma mistura soluto-solvente em que a densidade numérica do soluto tende a zero (diluição infinita), as equações de OZ moleculares para todo o sistema podem ser decompostas em equações envolvendo correlações solventesolvente, soluto-solvente e soluto-soluto [117]:

$$h^{vv}(r_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = c^{vv}(r_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) + \frac{\rho^v}{\Omega} \iint c^{vv}(r_{13}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_3) h^{vv}(r_{32}, \boldsymbol{\Omega}_3, \boldsymbol{\Omega}_2) \mathrm{d}\mathbf{r}_3 \mathrm{d}\boldsymbol{\Omega}_3$$

$$(4.15)$$

$$h^{uv}(r_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = c^{uv}(r_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) + \frac{\rho^v}{\Omega} \iint c^{uv}(r_{13}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_3) h^{vv}(r_{32}, \boldsymbol{\Omega}_3, \boldsymbol{\Omega}_2) \mathrm{d}\mathbf{r}_3 \mathrm{d}\boldsymbol{\Omega}_3$$

$$(4.16)$$

$$h^{uu}(r_{12}, \boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = c^{uu}(r_{12}, \boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) + \frac{\rho^v}{\Omega} \iint c^{uv}(r_{13}, \boldsymbol{\Omega_1}, \boldsymbol{\Omega_3}) h^{vu}(r_{32}, \boldsymbol{\Omega_3}, \boldsymbol{\Omega_2}) \mathrm{d}\mathbf{r}_3 \mathrm{d}\boldsymbol{\Omega_3}$$

$$(4.17)$$

O superíndice u se refere ao soluto e v ao solvente, e $\Omega = \int d\Omega$ é um fator geométrico de normalização da integral orientacional. A equação 4.16, que envolve correlações entre soluto e solvente, é de particular interesse, pois descreve processos de solvatação. Sua solução depende do conhecimento *a priori* das correlações solvente-solvente $h^{vv}(r_{32}, \Omega_3, \Omega_2)$, as quais podem ser obtidas pela solução da equação de OZ 4.15, para o solvente puro. No caso de solvente multicomponente, a densidade numérica ρ^v deve ser especificada para cada componente, e um somatório sobre os diferentes componentes químicos deve ser aplicado sobre as integrais duplas das equações acima.

Para converter a equação 4.16 para a forma de sítios de interação, pode-se proceder como na seção anterior, o que dará origem às equações RISM unidimensionais (radiais) para as correlações soluto-solvente. Este procedimento, entretanto, não fornecerá uma boa descrição da solvatação de solutos com formas arbitrárias, não-esféricas. Nestes casos, é desejável ter descrições mais detalhadas do soluto. Sendo assim, em vez de tomar a média sobre os graus de liberdade orientacionais de todo o sistema, como se faz em RISM, pode-se restringir a redução destes graus de liberdade apenas ao solvente, mantendo a descrição tridimensional (3D) do soluto. Com isso, pode-se obter distribuições 3D do solvente em torno de um soluto de forma arbitrária, o que permite a descrição de, por exemplo, interfaces sólido-líquido [114]. Com este procedimento, as equações de OZ soluto-solvente dão origem à teoria de equações integrais 3D-RISM, ou teoria molecular da solvatação:

$$h_{\gamma}(\mathbf{r}) = \sum_{\alpha} \int d\mathbf{r}' c_{\alpha}(\mathbf{r}') \chi_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|)$$
(4.18)

onde $h_{\gamma}(\mathbf{r}) = g_{\gamma}(\mathbf{r}) - 1$ é a função de correlação total 3D entre o sítio γ do solvente e o soluto. A função $\chi_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|)$ é a suscetibilidade entre os sítios α e γ do solvente, definida por

$$\chi_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|) = \omega_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|) + \rho^v h_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|).$$
(4.19)

Nesta equação, $h_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|)$ é a função de correlação intermolecular entre os sítios $\alpha \in \gamma$ do solvente, e $\omega_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|)$ é a função de distribuição intramolecular:

$$\omega_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|) = \frac{\delta(|\mathbf{r} - \mathbf{r}'| - l_{\alpha\gamma})}{4\pi l_{\alpha\gamma}^2},$$
(4.20)

que especifica a geometria das moleculas do solvente, com as distâncias inte-

ratômicas $l_{\alpha\gamma}$, como na equação 4.13. As correlações solvente-solvente $h_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}'|)$ são obtidas pela teoria RISM ou DRISM aplicada ao solvente puro (equações 4.12). Portanto, o *output* da teoria RISM aplicada ao solvente puro é *input* para a teoria 3D-RISM de sistemas soluto-solvente. Isso permite interpretar que a teoria 3D-RISM fornece a perturbação sofrida pelo solvente puro ao ser introduzido no campo de potencial gerado pelo soluto [124]. Finalmente, a equação integral 3D-RISM deve ser complementada por uma relação de fechamento, que assume a forma geral 3D

$$h_{\gamma}(\mathbf{r}) = \exp\left[-u_{\gamma}(\mathbf{r})/k_{B}T + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r}) + B_{\gamma}(\mathbf{r})\right] - 1, \qquad (4.21)$$

em que $u_{\gamma}(\mathbf{r})$ é o potencial de interação do soluto com o sítio γ do solvente.

4.4 Relações de fechamento

As relações de fechamento, conhecidas na literatura como *closure*, envolvem o funcional B, que, analiticamente, é expresso como uma série infinita de integrais que, na prática, é impossível de ser utilizada. Logo, aproximações se fazem necessárias. Uma das aproximações consiste em fazer B = 0, que é a relação de fechamento conhecida como 3D-HNC (*three-dimensional hypernetted chain*):

$$h_{\gamma}(\mathbf{r}) = \exp\left[-u_{\gamma}(\mathbf{r})/k_{B}T + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r})\right] - 1$$
(4.22)

ou

$$g_{\gamma}(\mathbf{r}) = \exp\left[-u_{\gamma}(\mathbf{r})/k_{B}T + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r})\right].$$
(4.23)

Formas análogas se amplicam para os casos esféricos. Esta aproximação é adequada para descrever a estrutura e termodinâmica de líquidos polares [114]. Entretanto, para sistemas fortemente associativos, com fortes atrações sítio-sítio, a 3D-HNC causa divergência na solução das equações 3D-RISM.

Uma segunda aproximação é a 3D-MSA (*three-dimensional mean spherical approximation*), que consiste na linearização da relação de fechamento 3D-HNC:

$$g_{\gamma}(\mathbf{r}) = 1 - u_{\gamma}(\mathbf{r})/k_B T + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r}).$$
(4.24)

Embora esta relação não apresente problemas de convergência, é conhecida por fornecer, para líquidos associativos em alta densidade, funções de distribuição com valores negativos nas regiões de baixa densidade local $[g(\mathbf{r}) < 0]$, que é fisicamente impossível [114]. Kovalenko & Hirata combinaram a 3D-HNC e 3D-MSA; a primeira é aplicada em regiões de baixa densidade local $(g(\mathbf{r}) < 1)$ e a segunda é aplicada nas regiões de alta densidade $(g(\mathbf{r}) > 1)$, onde a 3D-HNC diverge [113, 114, 123]. Esta relação de fechamento, conhecida como 3D-KH, portanto, combina as vantagens da 3D-HNC e 3D-MSA, reduzindo os problemas de convergência. A 3D-KH é expressa por

$$h_{\gamma}(\mathbf{r}) = \begin{cases} \exp(d_{\gamma}(\mathbf{r})) - 1, & \text{se } d_{\gamma}(\mathbf{r}) \leq 0\\ d_{\gamma}(\mathbf{r}), & \text{se } d_{\gamma}(\mathbf{r}) > 0 \end{cases}$$
(4.25)

onde

$$d_{\gamma}(\mathbf{r}) = -\frac{u_{\gamma}(\mathbf{r})}{k_B T} + h_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r}).$$
(4.26)

A proposta da relação de fechamento 3D-KH permitiu a popularização da teoria 3D-RISM e sua aplicação a sistemas de interesse prático, envolvendo complexidades químicas, como proteínas [113].

Complementada com a relação 3D-KH, a teoria 3D-RISM é conhecida como 3D-RISM/KH. Sua solução é obtida computacionalmente discretizando as funções de correlação numa malha 3D finita ao redor do soluto. Transformadas de Fourier rápidas (3D-FFT) são utilizadas para calcular a integral da equação 3D-RISM que, por serem integrais de convolução, se transformam num conjunto de equações algébricas não-lineares. Correções assintóticas devido a efeitos de longo alcance são obtidas analiticamente [114, 125, 126]. Rotinas computacionais para resolver as equações das teorias RISM/DRISM e 3D-RISM estão implementadas no pacote computacional AMBER [127].

4.5 Termodinâmica de solvatação em 3D-RISM/KH

Uma vez resolvida as equações integrais 3D-RISM/KH, as funções de correlação h_{γ} e c_{γ} podem ser utilizadas para obter funções termodinâmicas. A energia livre molar de solvatação, ou potencial químico de excesso do solvente em relação ao

bulk, é expressa por [114]

$$\Delta \mu = k_B T \int \sum_{\gamma} \Phi_{\gamma}(\mathbf{r}) \mathrm{d}\mathbf{r}, \qquad (4.27)$$

onde

$$\Phi_{\gamma}(\mathbf{r}) = \rho_{\gamma} \left[\frac{1}{2} h_{\gamma}^{2}(\mathbf{r}) \Theta(-h_{\gamma}(\mathbf{r})) - c_{\gamma}(\mathbf{r}) - \frac{1}{2} h_{\gamma}(\mathbf{r}) c_{\gamma}(\mathbf{r}) \right]$$
(4.28)

em que Θ é a função degrau de Heaviside ($\Theta(x) = 0$ se x < 0; $\Theta(x) = 1$ se x > 0). A função $\Phi_{\gamma}(\mathbf{r})$ representa a densidade de energia livre de solvatação, SFED (solvation free energy density), que é a função que, quando integrada no espaço, fornece a energia livre de solvatação. Ao ser representada tridimensionalmente, indica regiões do espaço que concentram interações do soluto com o sítio γ do solvente que contribuem mais ou menos para a energia livre de solvatação. Pode ainda ser obtida para uma determinada espécie química do solvente P por

$$SFED_P(\mathbf{r}) = \sum_{\gamma \in P} \Phi_{\gamma}(\mathbf{r}),$$
 (4.29)

onde a soma se estende sobre todos os sítios do componente P do solvente. Nesta forma, a SFED permite estudar as interações locais de um dado componente do solvente com o soluto [128].

A energia livre molar de solvatação pode ser utilizada para obter a entropia molar de solvatação através da relação termodinâmica

$$\Delta S = -\frac{1}{T} \left(\frac{\partial \Delta \mu}{\partial T} \right)_V, \tag{4.30}$$

enquanto as interações soluto-solvente podem ser obtidas por

$$\Delta \epsilon^{uv} = k_B T \sum_{\gamma} \rho_{\gamma} \int g_{\gamma}(\mathbf{r}) u_{\gamma}(\mathbf{r}) d\mathbf{r}, \qquad (4.31)$$

permitindo a obtenção da energia de reorganização do solvente em torno do soluto $\Delta \epsilon^{vv}$ através de

$$\Delta \mu = \Delta \epsilon^{uv} + \Delta \epsilon^{vv} - T \Delta S. \tag{4.32}$$

O volume parcial molar do soluto pode ser computado por [113]

$$\bar{V} = k_B T \chi_T \left(1 - \sum_{\gamma} \rho_{\gamma} \int c_{\gamma}(\mathbf{r}) \mathrm{d}\mathbf{r} \right), \qquad (4.33)$$

onde χ_T é a compressibilidade isotérmica do solvente puro, obtida pela função de correlação direta radial da teoria RISM/KH ou DRISM/KH:

$$\rho k_B T \chi_T = \left(1 - \sum_{\alpha, \gamma} \rho_\alpha \int_0^\infty c_{\alpha\gamma}(r) 4\pi r^2 \mathrm{d}r \right)^{-1} \tag{4.34}$$

A teoria 3D-RISM/KH permite ainda a obtenção do potencial de força média (PMF) atuando sobre sítios do solvente em torno do soluto [113]. O PMF de um conjunto de n partículas é o potencial que fornece a força média atuando sobre essas partículas numa configuração fixa \mathbf{r}^n , com a média tomada sobre o *ensemble* de todas as possíveis configurações das outras N - n partículas. O PMF corresponde ao trabalho médio para trazer as n partículas de uma configuração em que estão infinitamente separadas para a configuração \mathbf{r}^n . Isso permite interpretar o PMF ao longo de um dado parâmetro de ordem como o perfil de energia livre ao longo do caminho representado pelo parâmetro de ordem.

O PMF relaciona-se diretamente com as funções de distribuição [96]. No caso de sistemas soluto-solvente, o PMF sobre o sítio γ do solvente pode ser obtido por

$$PMF_{\gamma}(\mathbf{r}) = -k_B T \ln g_{\gamma}(\mathbf{r}). \tag{4.35}$$

A associação do PMF com o perfil de energia livre permite computar o PMF entre duas moléculas de soluto, $a \in b$, em solução. Se as moléculas estão separadas pela distância r_{ab} e estão nas orientações $\Omega_a \in \Omega_b$, o PMF pode ser obtido por [114, 126]

$$PMF(r_{ab}, \Omega_a, \Omega_b) = u_{ab}(r_{ab}, \Omega_a, \Omega_b) + \mu_{ab}(r_{ab}, \Omega_a, \Omega_b) - \mu_a - \mu_b, \qquad (4.36)$$

onde u_{ab} é a energia de interação entre as moléculas $a \in b$, obtida do campo de

força, μ_{ab} é a energia livre de solvatação da moléculas a e b em solução, obtida pela equação 4.27, e μ_a e μ_b são as energias livres de solvatação das moléculas a e b individuais, ou infinitamente separadas. A utilização da equação 4.36 ao longo de um dado parâmetro de ordem, permite a obtenção do PMF do processo molecular que ocorre ao longo deste parâmetro de ordem.

Parte III

Dinâmica molecular de hidrolases e proteínas disruptivas

Capítulo 5

Dinâmica estrutural de celobiohidrolases

5.1 Introdução

As celobiohidrolases (CBH), ou exo-1,4- β -glucanases (EC 3.2.1.91), hidrolisam processivamente as ligações $\beta(1 \rightarrow 4)$ das extremidades redutoras e não-redutoras das cadeias da celulose. O sistema celulolítico secretado pelo fungo *Trichoderma reesei*, que é o mais estudado, possui uma CBH da família GH6 (TrCel6A) e uma da família GH7 (TrCel7A). Estas duas enzimas representam mais de 50% (w/w) do coquetel secretado por *T. reesei*, sendo, pois, enzimas-chave [4]. Em termos funcionais, estas enzimas são distintas: enquanto a TrCel7A ataca a extremidade redutora da celulose, a TrCel6A ataca a extremidade não-redutora [69]. Assim, ambas são necessárias para a degradação da celulose. Os estudos reportados nesta tese foram realizados com CBHs da família GH7; assim, esta introdução será focada na TrCel7A. Porém, muito do que está descrito adiante se aplica também a CBHs da família GH6 e CBHs de outros organismos.

5.1.1 Estrutura e função

A TrCel7A é uma proteína bimodular, possuindo dois domínios conectados por um longo *linker* flexível e glicosilado (figura 5.1). O domínio N-terminal, denominado domínio catalítico (CCD: *catalytic core domain*), é composto por aproximadamente 430 resíduos e é onde ocorre a hidrólise do substrato. O domínio C-terminal, o módulo de ligação ao carboidrato (CBM: *carbohydrate binding module*), é uma pequena porção da enzima responsável pela ligação da proteína à superfície da celulose [129]. Embora essa seja a estrutura da maioria das CBHs, a estrutura de uma CBH que possui apenas o CCD foi reportada recentemente [130].



Figura 5.1. Modelo tridimensional da estrutura da TrCel7A, mostrando o CCD, o CBM e o *linker* glicosilado sobre a superfície da celulose, com uma cadeia de glucano ligada ao CCD para ser hidrolisada. Adaptado de [131].

Domínio catalítico (CCD)

A primeira estrutura cristalográfica do CCD da TrCel7A, obtida em 1994 [132], revelou que o CCD é formado por um sanduíche de folhas β anti-paralelas, no interior do qual reside um núcleo fortemente hidrofóbico. Acima das folhas β , há um longo canal coberto por loops, que dá à TrCel7A um sítio de ligação ao substrato na forma de túnel (figuras 5.2 e 5.3). Uma vez que o substrato se liga à TrCel7A, os loops do túnel catalítico impedem sua dissociação. Assim, a TrCel7A executa vários ciclos catalíticos sobre uma única cadeia polissacarídica, através de um mecanismo de hidrólise processiva. Estes longos e numerosos loops da TrCel7A são os elementos estruturais que a caracterizam como celobiohidrolase. A endoglucanase Cel7B de *T. reesei* apresenta alta identidade sequencial com a TrCel7A, exceto pelas deleções nos loops que formam o túnel catalítico. Isso confere à Cel7B, e outras endoglucanases, uma estrutura completamente aberta, apta a clivar, de forma não-processiva, ligações glicosídicas através de ataques endoglucolíticos em pontos aleatórios da celulose [133].

Uma série de estruturas cristalográficas de mutantes cataliticamente deficientes da TrCel7A foram obtidas em 1998 na presença de oligossacarídeos [134]. Tais



Figura 5.2. Estrutura cristalográfica do CCD da TrCel7A (PDB: 8CEL [134]). As folhas β (azul) formam um canal acima do qual há vários loops (branco) e α -hélices (amarelo) que formam um túnel. Neste túnel, há resíduos polares e 4 triptofanos (vermelho), onde se liga uma cadeia polissacarídica (verde) para ser hidrolisada pelos resíduos catalíticos (alaranjado).



Figura 5.3. Superfície molecular do CCD da TrCel7A, mostrando a topologia de túnel criada pelos loops.

estruturas revelaram a existência de 10 subsítios de ligação ao longo do túnel catalítico, onde se ligam os monômeros de glicose da cadeia polissacarídica (figura 5.3). As principais interações enzima-substrato são ligações de hidrogênio envolvendo, principalmente, resíduos de Gln e Asn, e interações hidrofóbicas do tipo *stacking* com 4 resíduos de Trp dispostos ao longo do túnel (figura 5.3). Com base nestes estudos de cristalografia, foi sugerido que tais resíduos de Trp seriam importantes na processividade, pois permitiriam que o substrato deslizasse ao longo do túnel catalítico [134].

Módulo de ligação ao carboidrato (CBM)

Módulos de ligação ao carboidrato (CBM) são domínios não-catalíticos presentes em glicosídeo hidrolases, cuja função primária é o reconhecimento e ligação da enzima ao substrato. Como as GHs, os CBMs são classificados em famílias, com base em similaridades sequenciais [135]. O CBM da TrCel7A é um pequeno domínio de 36 resíduos [136], adaptado ao reconhecimento da superfície da celulose cristalina através de três resíduos aromáticos arranjados de forma planar (figura 5.4) [135]. CBMs semelhantes ao da TrCel7A ocorrem em outras GHs que se ligam a superfícies cristalinas de celulose ou de quitina, e são classificados como CBMs do tipo A [135]. Há evidências de que o CBM esteja envolvido em três funções [4, 71]: (1) aumento da concentração da enzima sobre o substrato celulósico; (2) direcionamento da enzima para alguma extremidade de uma cadeia polissacarídica; (3) modificações da estrutura cristalina da celulose. A presença do CBM é importante para a atividade da TrCel7A sobre substratos insolúveis, mas não sobre substratos solúveis [71]. Simulações computacionais mostram que o CBM se liga seletivamente sobre a face hidrofóbica da celulose [137], exibindo mínimos de energia a cada unidade de celobiose da superfície da celulose [138], que coincide com o comprimento de um ciclo processivo da TrCel7A.



Figura 5.4. Estrutura tridimensional do CBM (PDB: 1CBH [136]) da TrCel7A, mostrando a superfície aromática de ligação à celulose.

Linker

Por ser desestruturado [4, 139], não há estruturas cristalográficas resolvidas para o linker, mas estudos de espalhamento de raios X a baixo ângulo (SAXS) revelam seu comprimento em solução [140], que varia dependendo de seu estado de glicosilação e de sua sequência primária, o que, por sua vez, é dependente do microrganismo [141]. Simulações de dinâmica molecular mostram que o linker possui propriedades elásticas em solução [139, 141, 142], e, dependendo da sequência de aminoácidos, pode exibir uma distribuição de comprimentos bimodal [143]. Recentemente, simulações partindo da configuração ilustrada na figura 5.1 mostraram que, na escala de tempo de 1 μ s, o linker glicosilado colapsa sobre a superfície da celulose, sugerindo seu papel de ligação da TrCel7A à celulose (figura 5.5) [131]. Experimentos de calorimetria com CBM isolado e com a construção CBM-linker mostraram que este último se liga à celulose com maior afinidade que o CBM isolado, corroborando os resultados das simulações [131].



Figura 5.5. Colapso do *linker* glicosilado sobre a superfície da celulose observado em 1 μ s de dinâmica molecular a partir da estrutura ilustrada na figura 5.1. Adaptado da referência [131].

5.1.2 Ciclo catalítico

A figura 5.6 ilustra o ciclo catalítico atualmente aceito para TrCel7A e outras CBHs similares. Inicialmente, a CBH se adsorve sobre a superfície hidrofóbica da celulose com a assistência do CBM. O sistema então se difunde sobre a superfície da celulose até encontrar uma cadeia polissacarídica descristalizada. Esta cadeia é então inserida no interior do túnel do CCD para formar o primeiro complexo cataliticamente
ativo. A cadeia polissacarídica se posiciona de forma que os resíduos catalíticos ficam em contato com a segunda ligação glicosídica a partir de sua extremidade redutora (ou não-redutora na TrCel6A). Neste ponto, ocorre a hidrólise da ligação glicosídica e uma molécula de celobiose é formada como produto. A celobiose é expulsa do CCD e, então, a enzima desliza ao longo da cadeia polissacarídica para formar um novo complexo cataliticamente ativo e reiniciar o ciclo. Isso se repete até que, eventualmente, a cadeia polissacarídica se dissocia da enzima e esta se dessorve da superfície da celulose [4].



Figura 5.6. Ciclo catalítico da TrCel7A, mostrando o processo de complexação da cadeia polissacarídica, ciclos processivos, expulsão da celobiose e dessorção da enzima. Adaptado da referência [144] com permissão da Elsevier.

Os detalhes de cada uma dessas etapas não foram ainda completamente elucidados, mas avanços significativos têm sido obtidos nos últimos anos. Muito pouca informação se tem de como o CCD reconhece e insere dentro de seu túnel uma única cadeia polissacarídica da superfície da celulose. Há duas propostas para este mecanismo na literatura. Uma delas é a de que os loops que formam o túnel catalítico se abririam momentaneamente, dessa forma permitindo que a cadeia polissacarídica entre no CCD [145, 146]. Outra proposta é a de que a cadeia polissacarídica entraria no CCD através de uma das extremidades do túnel [147, 148]. No entanto, não há um consenso na literatura sobre qual é o mecanismo real. Os detalhes da reação de hidrólise têm sido extensivamente explorados através de experimentos de mutação sítio-dirigida [134] e simulações híbridas QM/MM [149–152]. Estes estudos mostram que a hidrólise na TrCel7A ocorre com retenção da configuração do carbono anomérico (figura 2.8), envolvendo os resíduos Glu212 (nucleófilo) e Glu217 (ácido geral).

A processividade da TrCel7A tem sido amplamente estudada nos últimos anos, tanto experimentalmente quanto computacionalmente [144]. O fato de haver mínimos no perfil de energia de interação do CBM com a celulose a cada unidade de celobiose sugere que o CBM está envolvido com a processividade [138]. Por outro lado, imagens de microscopia de força atômica de alta velocidade mostram que tanto a TrCel7A intacta quanto seu CCD isolado podem se deslocar em linha reta sobre a celulose [153], sugerindo que a função de processividade está contida no próprio CCD. Posteriormente, simulações mostraram que o perfil de energia livre da penetração de uma cadeia de glucano no túnel do CCD da TrCel7A favorece o processo termodinamicamente [147]. Assim, o papel do CBM e do *linker* na processividade – se é que possuem algum papel ativo neste processo – não é claro, ainda que ambos se movimentem juntamente com o CCD.

Finalmente, a celobiose, quando formada, age como um potente inibidor da TrCel7A [70]. Por não ser expulsa imediatamente após sua formação, a celobiose impede o início de um novo ciclo processivo. Diferentes CBHs exibem diferentes graus de inibição pela celobiose, mas as razões disso não são claras [154, 155]. O desenvolvimento de enzimas menos sensíveis à celobiose depende do entendimento das bases moleculares da inibição por produto.

Neste capítulo, são apresentados três diferentes estudos focados em entender diferentes aspectos do funcionamento das celobiohidrolases.

5.2 A Celobiohidrolase Cel7A de Trichoderma harzianum

5.2.1 Introdução

Trichoderma harzianum é um fungo filamentoso que secreta várias enzimas celu-

lolíticas que, ao contrário das do fungo *Trichoderma reesei*, somente começaram a ser estudadas para aplicações biotecnológicas nos últimos anos. Estudos recentes mostram que *T. harzianum* secreta um coquetel balanceado de celulases, contendo celobiohidrolases, endoglucanases e β -glucosidases, capaz de clivar completamente substratos celulósicos em suas formas monoméricas [156–158]. Enzimas deste coquetel têm sido estudadas bioquímica, biofísica e estruturalmente por colaboradores [140, 159–161], e computacionalmente pelo autor desta tese e outros membros do grupo [162, 163].

A comparação da estrutura primária da CBH Cel7A de *T. harzianum* (ThCel7A) com a TrCel7A releva que estas enzimas exibem 81% de identidade sequencial. Devido a este alto grau de identidade, estas duas enzimas pertencem à mesma família de glicosídeo hidrolases: GH7. Logo, espera-se que a TrCel7A e a ThCel7A compartilhem um elevado grau de similaridade estrutural e, por consequência, funcional. Medidas de atividade enzimática sobre diferentes substratos solúveis, entretanto, mostraram que, em relação à TrCel7A, a ThCel7A tende a apresentar valores de k_{cat} e K_m uma ordem de grandeza maiores, indicando reações de hidrólise mais rápidas, porém menores afinidades pelo substrato [155, 162, 164–166]. Medidas de inibição mostraram que a ThCel7A, como a TrCel7A, é competitivamente inibida pela celobiose, mas em muito menor grau: enquanto a constante de inibição K_i medida para a TrCel7A foi de 20 μ M [166, 167], a da ThCel7A foi de 720 μ M [162], indicando que uma concentração 360 vezes maior de celobiose é necessária alcançar, na ThCel7A, o mesmo nível de inibição da TrCel7A.

Para entender as bases estruturais das diferenças de atividade enzimática e inibição observadas para a TrCel7A e ThCel7A, a estrutura cristalográfica desta última foi obtida com resolução de 1,67 Å por colaboradores e utilizada para estudos comparativos com a TrCel7A através de simulações de dinâmica molecular [162], cujos resultados são reportados nesta seção.

5.2.2 Metodologia

A estrutura inicial da TrCel7A foi obtida do banco de dados PDB (www.rscb.org) [168], depositada com o código 8CEL [134], e a da ThCel7A com o código 2YOK

(depositada no PDB por colaboradores). Inicialmente, todos os ligantes (β -glucano e PEG (polietilenoglicol)) e N-glucosaminas foram removidos para gerar as estruturas *apo*. Átomos de hidrogênio foram adicionados considerando os estados de protonação de resíduos ácidos e básicos em pH=6, estimados com o servidor H++ [169, 170]. As estruturas foram então solvatadas com uma caixa retangular de moléculas de água de tal forma que a mínima distância entre a proteína e a extremidade da caixa fosse pelo menos 16 Å. Para tornar os sistemas eletricamente neutros, íons Na⁺ e Cl⁻ foram adicionados em quantidades tais que a concentração total fosse de aproximadamente 0,1 M e o sistema ficasse eletricamente neutro. O número total de átomos dos sistemas finais foi de aproximadamente 60.000. A construção dos sistemas foi feita utilizando o programa VMD [108].

As simulações foram executadas em pressão e temperatura constantes de 1 atm e 310 K, utilizando o pistão e o termostato de Langevin, respectivamente [171]. O campo de força CHARMM22 [172, 173] foi usado em conjunto com o modelo de água TIP3P/CHARMM [172, 174]. Para as interações de curto alcance, um raio de corte de 12 Å foi utilizado e as interações de longo alcance foram computadas através do algoritmo PME (*particle mesh* Ewald) [105]. Todas as ligações envolvendo átomos de hidrogênio foram mantidas rígidas em seus comprimentos de equilíbrio e um passo de integração de 2.0 fs foi utilizado. Uma série de etapas para relaxar o sistema foram feitas antes de gerar as trajetórias produtivas, utilizadas para análise. Inicialmente, aplicando restrições harmônicas a todos os átomos da proteína, exceto aos hidrogênios, foram executados 1000 passos de minimização de energia utilizando o método dos Gradientes Conjugados (CG) [88] seguido de 200 ps de dinâmica molecular. Numa segunda etapa, apenas os carbonos α foram mantidos restringidos e 1000 passos de CG seguidos de 200 ps de dinâmica molecular foram executados. Por fim, 600 ps de dinâmica molecular foram executados com todos os átomos completamente livres. Após esta etapa inicial, 3 simulações independentes de 60 ns foram executadas para cada sistema. As simulações foram executadas com o programa NAMD [171].

5.2.3 Resultados

Estruturas cristalográficas das celobiohidrolases ThCel7A e TrCel7A

A comparação das estruturas cristalográficas das celobiohidrolases ThCel7A e Tr-Cel7A mostra que a alta identidade sequencial destas enzimas se reflete em características estruturais muito parecidas, como mostra a figura 5.7. Exceto por algumas pequenas diferenças estruturais nos loops 5 e 7 e a deleção do loop 3, as estruturas são idênticas. Nos loops que formam o sítio de ligação, no entanto, há algumas alterações locais causadas pelas cadeias laterais de resíduos. O resíduo Tyr371(r) – resíduos da TrCel7A serão identificados por (r) e da ThCel7A por (h) localizado no loop 6 corresponde ao resíduo Ala384(h). Esta alteração supostamente afeta as interações entre o loop 5 e 6 através dos resíduos correspondentes Tyr247(r) e Tyr260(h). Na TrCel7A, os resíduos Tyr247(r) e Tyr371(r) parecem ser responsáveis por manter os loops 5 e 6 em contato. No caso da ThCel7A, a presença da Ala384(h) no lugar da Tyr371(r) enfraquece as interações entre os loops e poderia conferir maior liberdade para o resíduo Tyr260(h). Embora as estruturas cristalográficas revelem que a troca da Tyr371(r) pela Ala384(h) causa apenas um ligeiro descamento do loop 5 na ThCel7A, o fato das interações Ala384(h)-Tyr260(h) serem mais fracas que as interações correspondentes Tyr371(r)-Tyr247(r) e, além disso, o resíduo Ala384(h) ser menos volumoso que a Tyr371(r), sugere que as flutuações estruturais dos loops 5 e 6 poderiam ser maiores na ThCel7A, resultando num túnel catalítico mais flexível e com maior acesso ao substrato comparado ao túnel da TrCel7A. Para estudar estas flutuações, simulações de dinâmica molecular foram executadas para a ThCel7A e TrCel7A.

Flutuações estruturais locais do túnel catalítico

As simulações mostraram que o fato da Ala384(h) substituir a Tyr371(r) leva a alterações dinâmicas na região dos loop 5 e 6. O resíduo Tyr260(h) pode sofrer várias mudanças conformacionais ao longo do tempo e modular o grau de exposição do túnel catalítico da ThCel7A. A figura 5.8a mostra a evolução temporal do rmsd (raiz quadrada do deslocamento médio quadrático) dos resíduos Tyr260(h)



Figura 5.7. Estruturas cristalográficas da TrCel7A e ThCel7A superpostas. Entre as poucas diferenças entre as enzimas, observa-se a troca do resíduo Tyr371 (TrCel7A) pelo resíduo Ala384 (ThCel7A).

e Tyr247(r) do loop 5, em relação às suas estruturas médias. Os valores de rmsd foram utilizados para discriminar dois estados: estado 1, com rmsd menor que 3,0 Å, e estado 2, com rmsd maior 3,0 Å. A figura 5.8b mostra configurações representativas dos estados 1 e 2 na região dos loops 5 e 6. Os resultados mostram que a cadeia lateral do resíduo Tyr260(h) é substancialmente mais móvel que o correspondente Tyr247(r). Além disso, as interações da Tyr260(h) com a Ala384(h) são regularmente rompidas, fazendo com que a Tyr260(h) frequentemente transite do estado 1 para o 2. Quando isso acontece, o túnel catalítico da ThCel7A se torna exposto (estado 2). Por outro lado, as cadeias laterais dos resíduos Tyr247(r) e Tyr371(r) permanecem sempre em contato e não se observa a abertura dos loops 5 e 6 na TrCel7A. Ainda que o resíduo Tyr247(r) possa sofrer mudanças conformacionais e ter acesso ao estado 2, suas flutuações não são suficientes para romper o contato entre os loops 5 e 6, o que impede o túnel catalítico da TrCel7A de ficar completemente exposto ao solvente. Assim, as simulações mostram que os resíduos Tyr260(h) e Tyr247(r) do loop 6 permanecem num equilíbrio conformacional, cujas consequências estruturais são dependentes dos resíduos Ala384(h) e Tvr371(r), os quais determinam a facilidade com que os loops 5 e 6 se abrem.



Figura 5.8. (a) Evolução temporal do rmsd dos resíduos Tyr247(h) (Th-Cel7A, em verde) e Tyr260(r) (TrCel7A, em azul) em relação às suas estruturas médias. O resíduo Tyr260(h) é mais móvel que o Tyr247(r), e ambos têm acesso a valores de rmsd correspondentes a dois estados (1 e 2). (b) Configurações representativas dos estados 1 e 2, na região entre os loops 5 e 6. Devido à presença da Ala384(h), o resíduo Tyr260(h) é móvel e pode abrir a região entre os loops 5 e 6 na ThCel7A. O resíduo Tyr247(r), por outro lado, é menos móvel e volumoso, o que dificulta a abertura dos loops na TrCel7A.

Flutuações globais

A decomposição das trajetórias das celobiohidrolases ThCel7A e TrCel7A em componentes principais (PCA) revela que a ThCel7A apresenta uma maior flexibilidade global que a TrCel7A. A figura 5.9a mostra as flutuações médias quadráticas (msd) ao longo dos 20 primeiros autovetores de maiores amplitudes. Observa-se que as diferenças nas flutuações dos carbonos α da ThCel7A e TrCel7A são capturadas pelos dois primeiros componentes principais, indicando que movimentos de mais alta amplitude dos loops, que possivelmente estão associados à função regulatória destas enzimas, são diferentes.



Figura 5.9. (a) Flutuações quadrática médias (msd) dos carbonos α da ThCel7A e TrCel7A ao longo dos 20 primeiros modos de movimento extraídos da análise de componentes principais. Observa-se que a ThCel7A exibe maiores flutuações que a TrCel7A nos dois primeiros modos. (b) Soma cumulativa e normalizada das flutuações ao longo dos 20 primeiros modos, que capturam aproximadamente 60% de todo o movimento das enzimas.

A análise das flutuações médias por resíduo (figura 5.10) mostra a origem das diferenças de mobilidade. As folhas β , que constituem o core estrutural das enzimas, são rígidas, apresentando somente flutuações locais. A mobilidade do carbono α dos resíduos Tyr260(h) e Tyr247(r) do loop 5, bem como dos resíduos Ala384(h) e Tyr371(r) do loop 6, diferem por um valor menor que 0,5 Å. Isso significa que os efeitos relatados na seção anterior são devidos às flutuações das cadeias laterais destes resíduos, e não a flutuações da estrutura terciária das enzimas. Estas, por sua vez, ocorrem no loop 4 e no loop 7. Enquanto o loop 4 é mais móvel, o loop 7 é menos móvel na ThCel7A que na TrCel7A. O loop 4 cobre o túnel catalítico e, por isso, sua flexibilidade pode favorecer a formação do complexo enzima-substrato. O loop 7, por outro lado, por estar localizado na saída do túnel catalítico, próximo ao sítio de ligação do produto, poderia estar associado às diferentes constantes de inibição apresentadas pelas duas enzimas.

A figura 5.11 mostra a matriz de correlações cruzadas para a ThCel7A e TrCel7A computadas a partir da trajetória essencial destes sistemas. A trajetória



Figura 5.10. Flutuações médias dos carbonos α (rmsf) da ThCel7A e da TrCel7A. O painel superior mostra o rmsf para cada um dos resíduos e o painel inferior mostra estruturas superpostas obtidas das simulações e coloridas conforme a mobilidade de cada resíduo (ver escala de cores de rmsf na figura).

essencial foi obtida pela projeção das trajetórias concatenadas de cada sistema no subespaço formado pelos 20 primeiros autovetores da análise de PCA. A figura 5.9b mostra a soma cumulativa e normalizada das flutuações quadráticas ao longo desses 20 autovetores. Aproximadamente 60% do movimento é capturado nesse subespaço de dimensão 20. Os outros 40% do movimento estão distribuídos no espaço gerado por todos os outros autovetores, de dimensão em torno de 1200. Assim, os 40% do movimento estão diluídos em modos de baixa amplitude. Observa-se que as enzimas exibem várias regiões correlacionadas; muitas delas, entretanto, envolvem resíduos das folhas β , que naturalmente são correlacionados por participarem da rede de ligações de hidrogênio que forma as estruturas secundárias das enzimas. Por outro lado, há correlações entre regiões funcionalmente importantes das enzimas. Pares de loops localizados em lados opostos do túnel catalítico – loops 4/6, 4/7, 5/6 e 5/7 – exibem correlações. O fato dessas correlações aparecerem tanto para a ThCel7A quanto para a TrCel7A sugere que estes movimentos dependem



Figura 5.11. Matriz de correlações cruzadas para ThCel7A e TrCel7A. Para ambas as enzimas, observa-se movimentos correlacionados entre os pares de loops 4/6, 5/6, 4/7 e 5/7. De azul a vermelho, as cores representam correlações negativas e positivas, respectivamente.

mais da estrutura terciária das enzimas que dos detalhes da estrutura primária, de forma que tais movimentos devem estar relacionados à função intrínseca das celobiohidrolases. Além disso, o fato do loop 7 – localizado próximo ao sítio do produto – se correlacionar com loops localizados sobre o sítio ativo, indica uma possível conexão entre inibição por produto e atividade enzimática.

5.2.4 Discussão

Neste trabalho, estudos comparativos das celobiohidrolases ThCel7A e TrCel7A foram feitos para entender as bases moleculares das diferenças de atividade enzimática entre tais enzimas. Os resultados mostram que o túnel catalítico da ThCel7A é mais flexível que o da TrCel7A. Esta flexibilidade resulta tanto de flutuações locais de cadeias laterais (resíduos Tyr260(h) e Tyr247(r)) quanto de flutuações da estrutura terciária (loop 4). A substituição do resíduo Tyr371(r) pela Ala384(h) na ThCel7A confere liberdade conformacional para a Tyr260(h), resultando num equilíbrio entre estados abertos e fechados da região entre o loop 5 e 6. O loop 4 é, por sua vez, o maior loop envolvido na formação do túnel catalítico e está localizado na suposta entrada do túnel catalítico, de forma que sua flexibilidade pode influenciar a ligação do substrato. Estas variações de flexibilidade do túnel catalítico da ThCel7A e TrCel7A, associadas ao fato de seus sítios ativos e resíduos vizinhos serem idênticos, sugere que a tendência de maiores valores de k_{cat} e K_m da ThCel7A é mais dependente de sua maior acessibilidade ao substrato que dos detalhes químicos da reação de hidrólise.

Estas ideias são corroboradas pela análise das estruturas de outras celobiohidrolases pertencentes à família 7. A celobiohidrolase PcCel7D de *Phanerochaete chrysosporium* [175] e a TeCel7B de *Talaromyces emersonii* [176] exibem algumas características comuns à ThCel7A quando comparadas à TrCel7A, como uma maior abertura na entrada do túnel catalítico e um loop 5 móvel (TeCel7B) ou mais curto (PcCel7D), resultando em maiores acessibilidades ao substrato. Consistentemente, a PcCel7D e a TeCel7B também exibem maoires valores de k_{cat} e K_m em relação à TrCel7A [175, 177]. Dando suporte a estas ideias, a deleção do loop 5 da TrCel7A – que causa a abertura completa da região entre os loops 5 e 6 – resultou em aumentos nos valores de k_{cat} e K_m [155]

Logo após a publicação deste estudo, Payne *et al.* [178] utilizaram a estrutura da ThCel7A para simulações na presença de substrato e mostraram que enquanto uma cadeia de celononaose se liga à TrCel7A com afinidade em torno de -28 kcal/mol, o mesmo substrato se liga à ThCel7A com menor afinidade, em torno de -24 kcal/mol. Ainda, mostraram que a afinidade da celononaose pela PcCel7D é cerca de -15 kcal/mol. Estes dados demonstram que a acessibilidade do túnel determina a afinidade enzima-substrato. Por outro lado, estudos experimentais mostram que CBHs com túnel catalítico mais fechado e rígido possuem maiores valores de processividade intrínseca (número médio de celobiose gerada por cada cadeia de glucano que se liga à enzima) [146, 179], menores velocidades de processividade [179] e menores taxas de dissociação do substrato [180]. Isso permite dizer que a ThCel7A deve ser menos processiva que a TrCel7A, porém deve apresentar maiores velocidades de processividade e maiores taxas de dissociação. A importância das taxas de dissociação é que CBHs são limitadas pelo comprimento de processividade, que é imposto pela complexidade do substrato: quando encontram um obstáculo na superfície da celulose (como hemicelulose, lignina, contatos de outras fibrilas de celulose ou mesmo outras celobiohidrolases), ficam presas improdutivamente até que se dissociem e voltem a se ligar produtivamente ao substrato [180, 181]. Isso faz com que a velocidade de hidrólise se reduza muito rapidamente após o início da catálise. Assim, embora a ThCel7A e outras CBHs mais abertas devam apresentar menores valores de processividade intrínseca que a TrCel7A, as maiores taxas de dissociação devem resultar numa redução da quantidade enzimas ligadas improdutivamente ao substrato.

O fato da ThCel7A ser menos inibida pela celobiose que a TrCel7A e, ao mesmo tempo, exibir sítios de ligação à celobiose estruturalmente idênticos, como inferido pelas estruturas cristalográficas, indica que a inibição por produto deve ser afetada por alterações estruturais e dinâmicas distantes do sítio do produto. A análise de correlações cruzadas mostrou que o loop 7, localizado na saída do túnel catalítico, exibe movimentos correlacionados com os loops 4 e 5, sugerindo uma possível conexão entre elementos que regulam a inibição pelo produto com elementos que regulam a atividade enzimática dentro do túnel catalítico. De fato, observa-se que celobiohidrolases cujas estruturas apresentam túnel catalítico mais aberto que a da TrCel7A exibem também constantes de inibição uma ordem de grandeza maiores: $K_i(\text{TrCel7A}) = 20 \ \mu\text{M} [166, 167], K_i(\text{ThCel7A}) = 360 \ \mu\text{M} [162],$ $K_i(\text{PcCel7D}) = 180 \ \mu\text{M} \ [155] \ \text{e} \ K_i(\text{TeCel7B}) = 180 \ \mu\text{M} \ [177].$ Por outro lado, a celobiohidrolase MaCel7B, de Melanocarpus albomyces, exibe menor constante de inibição $[K_i(MaCel7B) = 6 \ \mu M]$ que a TrCel7A, associado a um maior valor k_{cat} [182]. Consistentemente, o túnel catalítico da MaCel7B exibe mais similaridades com a TrCel7A que com a ThCel7A em termos de acessibilidade ao substrato [183]. A mesma tendência é observada para a TrCel7A contendo a deleção no loop 5 [155]. Logo, a inibição por produto em celobiohidrolases é influenciada por fatores que governam também a atividade enzimática; isso parece ocorrer através do acoplamento dos loops do túnel catalítico (loop 4, 5 e 6) com o loop do sítio do produto (loop 7), como evidenciado pelas análises de correlação entre resíduos.

5.3 Dinâmica intrínseca da TrCel7A

5.3.1 Introdução

A formação do complexo ativo enzima-substrato é uma etapa do ciclo catalítico das celobiohidrolases pouco entendida. De alguma forma, as celobiohidrolases descristalizam cadeias individuais de glucano na superfície da celulose e as alçam para dentro do túnel catalítico. Como mostra as estruturas cristalográficas, as celobiohidrolases possuem um enovelamento que dificulta a dissociação do substrato, uma vez que este é complexado à enzima (figura 5.12). Isso permite que as CBHs funcionem de forma processiva, hidrolisando várias ligações glicosídicas de uma única cadeia de glucano sem que esta se dissocie. Ao mesmo tempo, este tipo de enovelamento faz com que a primeira associação produtiva entre a enzima e o substrato seja extremamente difícil. Com base na análise de estruturas cristalográficas de CBHs ligadas a oligossacarídeos, tem sido proposto que as CBHs inserem a cadeia de glucano dentro do túnel catalítico através de uma de suas entradas [4, 134]. Com efeito, vários estudos sugerem indícios de que isso ocorre. Imagens obtidas por microscopia de força atômica de alta velocidade mostraram que, enquanto a TrCel7A nativa caminha sobre a superfície da celulose cristalina, mutações sobre o resíduo catalítico Glu212 e sobre o resíduo Trp40, localizado na suposta entrada do túnel catalítico (figura 5.12), impediram a atividade processiva da TrCel7A sobre celulose. Isso mostra que tanto a atividade química quanto interações do substrato com o resíduo Trp40 na entrada do túnel são necessárias para a atividade processiva das celobiohidrolases [153]. Em outro estudo, imagens de microcopia eletrônica de transmissão mostraram que, durante a atividade enzimática da TrCel7A, as fibrilas de celulose se afinam em uma de suas extremidades [145], evidenciando a atuação processiva da TrCel7A, bem como sua seletividade por uma das extremidades da celulose, sugerindo que cadeias de glucano são lançadas para dentro do túnel catalítico. O mesmo foi observado por microscopia de força atômica [181]. O papel do resíduo Trp40 da entrada do túnel é ainda suportado por simulações de dinâmica molecular, que mostraram que uma cadeia de glucano, inicialmente posicionada sobre o resíduo Trp40, desliza para dentro do túnel catalítico na TrCel7A nativa.

mas dissocia para o meio no mutante Trp40Ala, o que reflete a menor atividade enzimática deste mutante em celulose cristalina [148]. O potencial de força média para inserir uma cadeia oligossacarídica dentro do túnel da TrCel7A também sugere que o resíduo Trp40 é essencial para a complexação do substrato [147].



Figura 5.12. Superfície molecular da TrCel7A mostrando o resíduo Trp40 localizado na entrada do túnel.

No entanto, existem controvérsias a respeito deste mecanismo. Há dados que suportam a hipótese de que o primeiro contato da celobiohidrolase com a celulose seja do tipo endoglucolítico e, uma vez formado o complexo ativo enzima-substrato, a atividade processiva siga naturalmente. O uso de sondas fluorescentes nas extremidades redutoras da celulose cristalina revelou duas possibilidades de iniciação da hidrólise por celobiohidrolases: (1) exo-iniciação, em que a celobiohidrolase se liga a uma extremidade da cadeia de glucano e (2) endo-iniciação, em que a celobiohidrolase cria extremidades em pontos aleatórios da celulose e se liga a uma delas, conforme sua afinidade por extremidades redutoras ou não-redutoras. A probabilidade de endo-iniciação da TrCel7A aumenta com a concentração da enzima, podendo chegar a mais que 50% em celulose cristalina e a mais que 90% em celulose amorfa. Para a PcCel7D, com loops mais curtos sobre o túnel catalítico, as probabilidades são ainda maiores: > 80% em celulose cristalina e > 90% em celulose amorfa. Para as endoglucanases TrCel5A e TrCel12A, as probabilidades

de endo-iniciação são maiores que 97% em celulose amorfa [146]. Estes dados sugerem que CBHs possuem habilidade de criar seus próprios substratos através de uma atividade endoglucolítica inicial. Medidas de atividade enzimática reportadas em diferentes trabalhos mostram as CBHs exibem, também, atividade endoglucolítica, que é maior para a TrCel6A e menor para a TrCel7A, provavelmente devido ao maior número de loops desta última [145, 184, 185]. Isso necessariamente implica na abertura dos loops das celobiohidrolases e a consequente exposição completa do canal catalítico. A hidrólise de tetrassacarídeos funcionalizados com cromóforos volumosos em ambas as extremidades fornece uma evidência funcional da abertura dos loops em celobiohidrolases [186]. Estruturas cristalográficas na ausência e presença de substrato da CBH Cel6A de Humicula insolens mostrou alterações conformacionais dos loops que cobrem o sítio ativo (mas não a ponto de abrir o túnel catalítico), fornecendo uma indicação estrutural da possibilidade de abertura de loops para a endo-iniciação [187, 188]. Estudos computacionais de diversas CBHs também mostram flutuações locais dos loops [10, 130, 162, 189] e, além disso, mostram que as flutuações dos loops aumentam no pH ótimo das CBHs TrCel6A e TrCel7A, como revelado por simulações de dinâmica molecular em pH constante [190], em concordância com dados de SANS [191]. Apesar da existência de várias evidências de endo-iniciação, tanto experimentais quanto computacionais, a abertura completa do túnel catalítico das celobiohidrolases nunca foi observada.

O processo de complexação com substrato em celobiohidrolases é, portanto, uma questão aberta do ponto de vista molecular. Ao mesmo tempo, de todas as etapas, físicas e químicas, da digestão enzimática da celulose, a complexação com o substrato é a que limita a velocidade do processo de hidrólise [192]. Isso significa que estratégias de engenharia de enzimas devem focar nesta etapa, mas a falta do entendimento mecanístico do processo retarda tais desenvolvimentos. Dada a multiplicidade de funções da celobiohidrolase, é improvável que esta assuma uma única conformação ao longo do ciclo catalítico, como sugerido por análises de estruturas cristalográficas. Evidentemente, as alterações estruturais necessárias para um evento particular do processo catalítico deve acontecer nos loops que cobrem o túnel catalítico. Na ausência de substrato, os loops devem exibir movimentos característicos relacionados à ligação do substrato e tais movimentos devem se alterar na presença do substrato para permitir a enzima realizar outras funções, como processividade, hidrólise de ligações glicosídicas, liberação do produto e a eventual dissociação substrato. Nesta seção, simulações da TrCel7A na ausência (denotada por *apo*-TrCel7A) e na presença (denotada por *holo*-TrCel7A) de substrato foram realizadas para estudar a dinâmica intrínseca da enzima antes e após a complexação com o substrato, e relações entre função e dinâmica nesta celobiohidrolase.

5.3.2 Metodologia

Para estudar a dinâmica da *apo*-TrCel7A, uma das simulações de 60 ns do estudo da seção 5.2 foi estendida até 150 ns. Para a *holo*-TrCel7A, a estrutura depositada no PDB [168] com o código 8CEL [134] foi utilizada. Esta estrutura contém uma cadeia de glucano ocupando todo o túnel catalítico [134]. Os mesmos procedimentos descritos na seção anterior foram utilizados para construir o sistema com substrato. O campo de força CHARMM36 [193] foi usado para modelar as interações envolvendo a cadeia de glucano, e uma simulação de 150 ns foi executada após as etapas iniciais de relaxação descritas na seção 5.2.

5.3.3 Resultados

Dinâmica da apo-TrCel7A

A dinâmica da TrCel7A foi analisada em termos de correlações de movimentos através da matriz de correlação cruzada das flutuações dos carbonos α , obtidas através da trajetória essencial do sistema (movimento ao longo dos 20 primeiros autovetores obtidos com análise de componentes principais). Valores positivos de correlação indicam dois resíduos com movimentos sincronizados e na mesma direção; valores negativos indicam resíduos em movimentos sincronizados e em direções opostas; ausência de correlações indica independência dos movimentos entre dois resíduos. O mapa de correlação cruzada da *apo*-TrCel7A está mostrado na figura 5.13a e não difere significativamente do mapa da figura 5.11, obtido de simulações mais curtas. Para melhor visualização, os dados de correlação foram

representados graficamente atribuindo uma ligação entre os resíduos que exibem correlação positiva (figura 5.13b) e negativa (figura 5.13c) de módulo acima de 0,7. A dinâmica estrutural da apo-TrCel7A é altamente acoplada; alguns dos movimentos correlacionados refletem vínculos geométricos decorrentes da proximidade espacial dos resíduos e outros se originam de acoplamentos estruturais não-triviais da proteína. A 5.13b mostra que as correlações positivas ocorrem, em sua maioria, entre resíduos próximos entre si e reflete os vínculos estruturais da proteína. Isso pode ser observado, por exemplo, na região das folhas β , onde há numerosas ligações de hidrogênio inter-resíduos. Mais interessante, observa-se que os loops localizados no mesmo lado da proteína estão fortemente acoplados. Isso significa que é mais provável que loops vizinhos se movimentem concertadamente, como um corpo rígido, que de forma aleatória e independente um dos outros. Por outro lado, as correlações negativas aparecem entre loops localizados em lados opostos do canal de folhas β . Isso indica que, em média, os loops se movimentam sincronizadamente em direções opostas. Além disso, as correlações são de longo alcance, o que significa que suas origens não são simplesmente os contatos diretos entre resíduos, mas um efeito cooperativo complexo dependente da estrutura terciária da enzima.

Análise de componentes principais – utilizada para decompor o movimento da proteína em diferentes modos coletivos [109] – mostra que as correlações podem ser interpretadas em termos dos movimentos de mais alta amplitude e menor frequência da proteína. A figura 5.14a, mostra que a flutuação associada ao primeiro modo de movimento é muito maior que os modos seguintes, e a figura 5.14b mostra as direções das flutuações ao longo do primeiro modo. Observa-se que loops vizinhos se movimentam na mesma direção e os opostos se movem em direções opostas. Estes resultados mostram que a dinâmica intrínseca da *apo*-TrCel7A é dominada por flutuações do tipo abre-fecha, de tal modo que a acessibilidade da fenda de ligação ao substrato pode mudar. A figura 5.14d mostra as estruturas correspondentes aos dois extremos de flutuação ao longo do primeiro modo de movimento, mostrando o movimento abre-fecha, e a figura 5.14c mostra a existência de dois mínimos ao longo do movimento de abre-fecha, indicando que existem



Figura 5.13. (a) Matriz de correlações cruzadas da *apo*-TrCel7A. Pares de resíduos que exibem correlações (b) positivas e (c) negativas de módulo maior que 0,7 são conectados. Observa-se que os loops vizinhos são positivamente correlacionados e os loops opostos são negativamente correlacionados.

interações que estabilizam estados mais abertos e mais fechados dos loops.

Dinâmica da holo-TrCel7A

Quando aplicadas ao complexo enzima-substrato (*holo*-TrCel7A), as análises de correlação mostram que os movimentos anticorrelacionados dos loops opostos são completamente extinguidos na presença do substrato, enquanto os movimentos correlacionados entre loops vizinhos são apenas parcialmente extinguidos (figura 5.15). Isso indica que a dinâmica dos loops da TrCel7A é fortemente determinada pela presença ou ausência do substrato. A decomposição da trajetória da



Figura 5.14. Dinâmica essencial da *apo*-TrCel7A. (a) Flutuações quadráticas médias ao longo dos 20 primeiros modos de movimento. (b) Direção do modo de movimento dominante, indicando flutuações do tipo abre-fecha. (c) Distribuição da coordenada conformacional ao longo do movimento de abre-fecha, mostrando a existência de interações que estabilizam conformações abertas e fechadas. (d) Estrutura da *apo*-TrCel7A correspondente ao máximo e mínimo da coordenada conformacional do movimento de abre-fecha.

holo-TrCel7A em componentes principais mostra que o modo de mais alta amplitude exibe menor flutuação ($\approx 28 \text{ Å}^2$) quando comparado ao modo de maior amplitude da *apo*-TrCel7A ($\approx 41 \text{ Å}^2$), indicando que o substrato induz uma estabilização dos movimentos dos loops (figura 5.16). No lugar dos movimentos do tipo abre-fecha observado para os loops da *apo*-TrCel7A, os movimentos funcionais da *holo*-TrCel7A se concentram apenas nos loop 1 e 4, que passam a se movimentar paralelamente ao túnel catalítico, conforme mostrado na figura 5.16b. A figura 5.16c mostra que, durante o curso da simulação, os loops 1 e 4 visitam dois estados distintos – estabilizados por interações específicas da proteína com o substrato –,



Figura 5.15. (a) Matriz de correlações cruzadas da *holo*-TrCel7A. Pares de resíduos que exibem correlações (b) positivas e (c) negativas de módulo maior que 0,7 são conectados. Observa-se que a ligação do substrato extingue (b) parcialmente as correlações positivas entre loops vizinhos e extigue (c) completamente as correlações negativas entre loops opostos.

os quais são mostrados na figura 5.16d.

5.3.4 Discussão

Mecanismos de associação e dissociação do substrato

As estruturas cristalográficas da TrCel7A e outras celobiohidrolases apresentam um longo e estreito túnel cujas paredes interagem fortemente com o substrato. Com base nestas estruturas, foi proposta a hipótese de que a cadeia polissacarídica



Figura 5.16. Dinâmica essencial da *holo*-TrCel7A. (a) Flutuações quadráticas médias ao longo dos 20 primeiros modos de movimento. (b) Direção do modo de movimento dominante, indicando flutuações dos loops 1 e 4 em direções paralelas ao túnel catalítico. (c) Distribuição da coordenada conformacional ao longo do movimento dos loops 1 e 4, mostrando dois estados estáveis de flutuações ao longo do túnel. (d) Estrutura da *holo*-TrCel7A correspondente ao máximo e mínimo da coordenada conformacional, destacando os loops 1 e 4.

deveria se ligar à celobiohidrolase através da penetração neste longo e estreito túnel [132, 134]. Entretanto, este seria um evento muito pouco provável, a não ser que a enzima fosse capaz de exercer forças muito específicas sobre a cadeia polissacarídica para inseri-la ao sítio ativo. Uma hipótese alternativa seria a de que a estrutura eventualmente se abriria para a ligação do substrato. Uma vez ligado o substrato, as interações enzima-substrato estabilizariam a conformação fechada do túnel. Os movimentos correlacionados observados para os loops da *apo*-TrCel7A dão suporte a esta segunda hipótese, indicando que loops localizados em lados opostos exibem movimentos anti-correlacionados – se movem sincronizadamente em direções

opostas – e loops localizados no mesmo lado estão correlacionados – se movem sincronizadamente na mesma direção. Segundo a análise de componentes principais, as flutuações de maior amplitude ocorrem ao longo da direção de abre-fecha, mostrando que os movimentos correlacionados dos loops são resultado do modo de movimento dominante da TrCel7A. Estes resultados indicam que a superfície de energia livre sobre a qual a enzima se movimenta oferece baixas penalidades para movimentos do tipo abre-fecha, e altas para movimentos em outras direções, e sugere que a abertura dos loops pode ocorrer numa escala de tempo maior do que a acessada pelas simulações. O fato dos loops vizinhos exibirem flutuações dinâmicas altamente correlacionadas mostra que a abertura simultânea de todos os loops opostos ao longo do canal catalítico não é um evento de probabilidade baixa. Posto que a abertura apenas parcial dos loops seria um estado improdutivo da *apo*-TrCel7A, os movimentos correlacionados reduzem a probabilidade da enzima acessar estados que não a permite exercer sua função que, neste caso, é a ligação ao substrato.

Com a ligação do substrato, os movimentos anti-correlacionados dos loops opostos desaparecem e os loops 1 e 4 passam a se mover paralelamente ao canal do túnel catalítico. Curiosamente, a direção das flutuações da *holo*-TrCel7A coincide com a direção do movimento relativo entre a enzima e o substrato durante a atividade processiva (deslizamento do substrato dentro do túnel), mostrando que a enzima realiza movimentos que favorecem a processividade. A dinâmica da TrCel7A, portanto, se altera conforme a função que deve exercer em cada etapa do processo de hidrólise.

A mudança da dinâmica intrínseca da TrCel7A quando esta passa do estado *apo* ao estado *holo* é reflexo da alteração da superfície de energia livre sobre a qual a enzima se movimenta. A figura 5.17 ilustra, de forma qualitativa, as superfícies de energia livre antes e após a complexação com o subtrato, e representações de diferentes estados da enzima com suas possíveis transições. No estado *apo* (curva vermelha), há dois estados, $A \in B$, correspondentes às conformações fechada e aberta, respectivamente. A simulação da *apo*-TrCel7A explora subestados em torno do ponto A na figura e indica movimentos na direção da transição $A \to B$. A barreira entre $A \in B$ é demasiadamente alta para ser vencida na escala de tempo de 150 ns. No estado B, a enzima pode eventualmente voltar para o estado A, ou pode se ligar ao substrato (estado C), o que induz uma mudança da superfície de ener-



Figura 5.17. Ilustração da superfície de energia livre da TrCel7A antes e após a complexação do substrato, com as possíveis transições ilustradas à direita. A complexação do substrato é provável de ocorrer através do caminho $A \to B \to C \to D$, e não pelo caminho $A \to D$. Para a dissociação, o inverso parece ser verdadeiro: dissociação preferencial pelo caminho $D \to A$.

gia livre. Do estado C (enzima aberta ligada ao substrato), devido às interações enzima-substrato altamente específicas, o sistema pode transitar com facilidade para o estado D, que corresponde à enzima fechada ligada ao substrato – condição em que a enzima se encontra nos cristais utilizados nos experimentos de difração de raios X. Neste ponto, ocorre a formação do primeiro complexo ativo enzimasubstrato e começam os ciclos catalíticos processivos. Eventualmente, o substrato ainda não hidrolisado deve se dissociar da enzima, o que pode ocorrer por dois caminhos: $D \to C \to B$ ou $D \to A$. Embora, segundo as simulações, o caminho de associação com o substrato parece ocorrer através da abertura do túnel catalítico $(A \to B \to C \to D)$, é provável que a dissociação ocorra através da etapa elementar $D \to A$. Em primeiro lugar, as simulações mostram que os movimentos de abre-fecha não dominam os movimentos da holo-TrCel7A, sugerindo que a superfície de energia potencial não é rasa nesta direção e que, portanto, há altas penalidades energéticas associadas a este tipo de movimento. Em segundo, vários experimentos revelam que CBHs que apresentam maiores taxas de processividade apresentam também maiores taxas de dissociação [179, 180]. Isso sugere que fatores que favorecem o movimento do substrato dentro do túnel catalítico favorecem também a dissociação do substrato, o que leva à hipótese de que o substrato deixa o túnel catalítico sem que ocorra a abertura dos loops. A cada uma dessas etapas elementares está associada uma constante de velocidade $(k/k', k_1/k'_1, k_2/k'_2 e k_3/k'_3)$, conforme ilustrado na figura 5.17, e, logo, determinam a cinética do processo de hidrólise.

Conforme mencionado no início desta seção, há dados experimentais que sugerem que a cadeia de glucano é inserida dentro do túnel catalítico das CBHs. Estudos de microscopia eletrônica de varredura e microscopia de força atômica mostram que as CBHs degradam as extremidades da fibrila de celulose [145, 181]. Entretanto, isso não necessariamente significa que a complexação com o substrato ocorra pela entrada do túnel. As extremidades da celulose são regiões com baixo grau de cristalinidade, de forma que a probabilidade das CBHs encontrarem cadeias de glucano descristalinizadas é maior nas extremidades que em outros pontos da fibrila. As CBHs podem se ligar às extremidades das cadeias de glucano através da abertura dos loops e, então, proceder com a catálise processiva, causando os efeitos observados por microscopia. Além disso, o resíduo aromático localizado na entrada do túnel catalítico – Trp40 (TrCel7A) e Trp272 (TrCel6A) – é essencial para a degradação de celulose cristalina e para a atividade processiva [153, 194]. Entretanto, tais resíduos impactam pouco [194] ou negativamente [148] a hidrólise de celulose amorfa. Isso mostra que, para substratos mais acessíveis, a complexação não depende de interações com tais resíduos e, portanto, devem estar mais relacionados com a indução de modificações na celulose cristalina que com o processo de complexação em si. Mais estudos focados nos resíduos Trp40 (TrCel7A) ou Trp272 (TrCel6A) são necessários para revelar seus papeis na hidrólise da celulose.

Movimentos funcionais e estrutura terciária

A dinâmica intrínseca da TrCel7A em diferentes estágios do processo enzimático, como caracterizada aqui, provavelmente vale para outros membros de celobiohidrolases da família GH7. A razão disso é que os movimentos funcionais de uma enzima são deteminados pelos vínculos impostos pelo tipo de enovelamento [195]. Os mapas de correlação cruzada da *apo*-TrCel7A e da *apo*-ThCel7a (figura 5.11) mostram a semelhança da dinâmica dessas duas enzimas que compartilham o mesmo enovelamento. Além disso, modelos de redes elásticas, que não levam em conta a natureza dos aminoácidos explicitamente, capturam os modos funcionais de proteínas, indicando que tais modos de movimento são característicos do tipo de enovelamento [196]. Os detalhes da estrutura primária, apesar de ter poucos efeitos sobre os movimentos característicos da enzima, devem ter influência nas taxas de transição entre diferentes estados. Por exemplo, como mostrado na seção 5.2, por conta de diferenças de cadeiais laterais nos loops 5 e 6, a ThCel7A deve sofrer abertura do túnel catalítico com uma taxa superior à da TrCel7A. Entretanto, ambas devem exibir o mesmo tipo de movimento global.

Estudos com outras proteínas mostram que flutuações estruturais observadas em diferentes estruturas cristalográficas e ensembles de estruturas obtidas por ressonância magnética nuclear no estado apo se correlacionam com as alterações estruturais necessárias para o processo de complexação [90, 197]. Além disso, análise de modos normais – em que os movimentos são restringidos num potencial harmônico – mostra que, para várias proteínas com diferentes enovelamentos, um único modo de baixa frequência e alta amplitude é capaz de predizer a direção das mudaças conformacionais necessárias para o processo de complexação [198]. Ainda mais, flutuações harmônicas em torno de estruturas cristalográficas obtidas com análise de modos normais ou redes elásticas predizem as direções de mudanças conformacionais obtidas com simulações de vários microssegundos a milissegundos executadas no supercomputador Anton [92, 106, 199]. Todos estes estudos suportam a hipótese de que as direções das flutuações observadas na simulação da *apo*-TrCel7A correspondem ao início da transição para a conformação aberta, necessária para a ligação do substrato.

Finalmente, segundo o mecanismo ilustrado na figura 5.17, a *apo*-TrCel7A tem acesso a um *ensemble* de subestados entre os estados aberto e fechado. Dentre estes estados, há aqueles que reconhecem o substrato – originando complexos enzima(aberta)-substrato cataliticamente inativos – e aqueles incapazes de reconhecer o substrato – estado fechado. O substrato, ao se ligar à *apo*-TrCel7A, deve induzir a transição do estado aberto para o fechado para criar o complexo cataliticamente ativo (observado nas estruturas cristalográficas). Esta última etapa é explicada pelo modelo do encaixe induzido proposto por Koshland, em 1958 [197, 200, 201]. O mecanismo de associação do substrato através da entrada da cadeia de glucano pelo túnel da TrCel7A está mais próximo do antigo modelo da chave e fechadura proposto por Fischer em 1894 [202], que superenfatiza, muitas vezes de forma não-realística, a complementaridade geométrica entre a enzima e o substrato na forma ativa.

5.4 Base estrutural da inibição da TrCel7A por celobiose

5.4.1 Introdução

Celobio
hidrolases clivam ligações glicosídicas nas extremidades da celulose e geram
celobiose como produto. A celobiose, por sua vez, inibe a atividade das celulases,
tanto celobio
hidrolases quanto endoglucanases [70]. Experimentos e simulações
computacionais mostram que, dentre as celulases, as CBHs da família GH7 são as
mais sensíveis à inibição por produto, seguidas por CBHs da família GH6 e por en-
doglucanases [154, 203]. As CBHs da família GH7, então, constituem os principais
alvos de inibição pela celobiose. β -glucosidases são usadas para reduzir hidrolisar
moléculas de celobiose em glicose e, assim, reduzir a inibição das celobio
hidrola-
ses [204]. A suplementação de diferentes coquetéis celulolíticos com β -glucosidases
reduz suas diferenças de eficiência, sugerindo que a inibição por celobiose é um
limitante do processo de hidrólise [205]. Logo, entender as razões moleculares da
inibição constitui um passo importante para desenvolvimentos focados em reduzir
a inibição.

A estrutura cristalográfica da TrCel7A na presença de oligossacarídeos [134] mostra que a enzima é adaptada para efetuar a hidrólise da ligação glicosídica que une uma unidade de celobiose ligada aos sítios +1 e +2 (figura 5.18), com o restante do polímero, ligado aos sítios (-). O sítio +1 contém um resíduo de Trp e o sítio +2 contém vários resíduos polares; juntos, estes resíduos exercem a função de orientar a cadeia polissacarídica em torno do ponto de clivagem [132, 134, 206,

207], sendo, portanto, essenciais para a hidrólise. No entanto, as interações que prendem o substrato para a catálise dificultam também a dissociação do produto, o que retarda o início de um novo ciclo catalítico. Para estudar as interações que dificultam a expulsão da celobiose da TrCel7A, simulações de dinâmica molecular foram realizadas com o complexo TrCel7A-substrato-produto (TrCel7A-SP), como ilustrado na figura 5.18.



Figura 5.18. Complexo TrCel7A-substrato-produto (TrCel7A-SP). A celobiose (produto) é mostrada em alaranjado e o substrato em verde. A numeração se refere aos sítios de ligação de cada unidade de glicose da cadeia polissacarídica. Sítios (+) ficam do lado da extremidade redutora (à esquerda do ponto de hidrólise) e sítios (-) ficam do lado da extremidade não redutora (à direita do ponto de hidrólise) [208].

5.4.2 Metodologia

Para gerar o complexo TrCel7A-SP, uma configuração do sistema *holo*-TrCel7A, da seção anterior, foi tomada e a ligação glicosídica que une a unidade de celobiose da extremidade redutora ao restante da cadeia foi removida. Coordenadas internas do campo de força CHARMM36 [193] foram utilizadas para adicionar os grupos OH e H ao oligossacarídeo hidrolisado. Os mesmos procedimentos descritos nas seções anteriores foram empregados para gerar uma trajetória de 150 ns para o complexo TrCel7A-SP.

5.4.3 Resultados e discussão

A análise da trajetória da TrCel7A-SP revela uma flexibilidade do sítio de ligação do produto, que pode assumir duas conformações (figura 5.19). Além da conformação cristalográfica, a simulação mostra uma outra conformação em forma de túnel. Enquanto a conformação cristalográfica exibe o sítio do produto completamente exposto ao solvente, a conformação em forma de túnel reduz substancialmente a exposição da celobiose ao meio.



Figura 5.19. Configurações do sítio de ligação da celobiose vistas na simulação. A configuração à esquerda é semelhante à estrutura cristalográfica. A configuração à direita exibe um túnel sobre o sítio de ligação da celulose, e não é vista na estrutura cristalográfica. A TrCel7A-SP tem acesso a ambos os estados.

A formação do túnel sobre o sítio do produto é, conforme ilustrado na figura 5.20a, dependente dos resíduos Arg251 (loop 5) e Tyr381 (loop 7). Quando o sítio do produto está exposto ao solvente, o resíduo Tyr381 fica numa conformação mais voltada para o interior do loop 7 (mostrado em verde na figura 5.20a). Sempre que a cadeia lateral da Tyr381 se volta para o resíduo Arg251, assumindo a configuração representada em lilás na figura 5.20a, no interior da superfície translúcida, o túnel se forma sobre a celobiose.

A figura 5.20b mostra a distância entre os resíduos Arg251 e Tyr381 em função do tempo no complexo TrCel7A-SP e, para comparação, na *apo*-TrCel7A e na *holo*-TrCel7A. Estes dados são representados na forma de distribuição na figura 5.19c. Observa-se que a distância Arg251-Tyr381 varia entre 2 e 6 Å na TrCel7A-SP, indicando transições frequentes entre a conformação de túnel (distâncias menores que 4 Å) e a conformação aberta (distâncias maiores que 4 Å). Por outro lado,



Figura 5.20. (a) Resíduos envolvidos na formação do túnel do sítio de ligação do produto. O resíduo Tyr381, inicialmente voltado para o interior do loop 7, se estende em direção ao resíduo Arg251 do loop oposto, assim criando o túnel que envolve a celobiose. A superfície translúcida corresponde à superfície dos resíduos que formam o sítio de ligação da celobiose. (b) Evolução temporal da distância Arg251-Tyr381 na *holo*-TrCel7A (preto), *apo*-TrCel7A (verme-lho) e TrCel7A-SP (azul). (c) Distribuição das distâncias Arg251-Tyr381 na *holo*-TrCel7A, *apo*-TrCel7A e TrCel7A-SP. A dinâmica do sítio do produto exibe propriedades intermediárias entre a *apo*- e *holo*-TrCel7A.

na *holo*-TrCel7A esta distância flutua estritamente em torno de 6 Å, enquanto na *apo*-TrCel7A a distância Arg251-Tyr381 raramente atinge 6 Å. Isso mostra que a conformação de túnel, muito pouco (ou nunca) visitada pela *holo*-TrCel7A, é a conformação predominante na *apo*-TrCel7A. Logo, o sítio de ligação da celobiose possui propriedades conformacionais intermediárias entre a *holo*-TrCel7A e a *apo*-TrCel7A. A incorporação da conformação de túnel da *apo*-TrCel7A na TrCel7A-SP deve ser consequência de interações enzima-substrato-produto menos estáveis que as interações enzima-substrato na *holo*-TrCel7A. De fato, uma análise do rmsd da celobiose na *holo*-TrCel7A e na TrCel7A-SP mostra que esta fica mais móvel após

a hidrólise (figura 5.21), sugerindo uma redução do acoplamento da celobiose com a enzima após a hidrólise. Como consequência, a conformação de túnel passa a ser acessível, dificultando a dissociação da celobiose.



Figura 5.21. (a) Distribuições do rmsd da celobiose em relação a sua estrutura média na *holo*-TrCel7A e TrCel7A-SP, mostrando que nesta última, a celobiose exibe maior flexibilidade em seu sítio de ligação.

Estes resultados sugerem que a inibição da TrCel7A pela celobiose é dependente da dinâmica da enzima. Além disso, apontam três mutações que deveriam resultar num menor grau de inibição: Arg251Ala, Tyr381Ala e a dupla mutação Arg251Ala-Tyr381Ala. Estes mutantes devem exibir menores graus de inibição por produto, mas medidas de atividade enzimática devem ser feitas para saber se tais resíduos são importantes para catálise. Simulações computacionais publicadas durante a realização deste trabalho mostraram que a energia livre de ligação da celobiose, de fato, é menor nos mutantes Arg251Ala e Tyr381Ala [209].

Finalmente, de acordo com a estrutura cristalográfica da TrCel7A [134] e de outras CBHs como ThCel7A [162] e PcCel7D [175, 210], o sítio de ligação da celobiose é completamente exposto ao solvente, mas as simulações indicam que a dinâmica desta região na TrCel7A pode dificultar a dissociação do produto. O papel da dinâmica do sítio do produto é particularmente importante em CBHs que possuem um loop colapsado sobre o sítio do produto [211]. Logo, as diferenças de inibição exibidas por CBHs que possuem sítios de ligação ao produto idênticos, segundo estruturas cristalográficas, podem ser explicadas por possíveis variações na dinâmica do sítio do produto.

5.5 Considerações finais

Neste estudo, um extensivo conjunto de simulações de dinâmica molecular foi utilizado para entender aspectos físicos do ciclo catalítico de CBHs da família GH7, abordando os seguintes tópicos: (i) associação e dissociação do substrato, (ii) processividade e (iii) expulsão do produto.

Enquanto o estudo da dinâmica intrínseca da TrCel7A mostra que CBHs exibem movimentos característicos que facilitam a associação do substrato pela abertura dos loops, o estudo comparativo da ThCel7A e TrCel7A sugere que a amplitude/taxa de transição de tais movimentos são determinados por detalhes da estrutura primária. As maiores flutuações exibidas pela ThCel7A em relação à TrCel7A devem facilitar sua associação com o substrato, o que explica sua maior eficiência catalítica.

Com a ligação ao substrato, a dinâmica da TrCel7A se altera: a enzima deixa de executar movimentos abre-fecha e passa a exibir flutuações características do movimento processivo. Por esta razão, a dissociação do substrato provavelmente ocorre não pela abertura dos loops, já que a enzima não mais exibe estes movimentos, mas pela saída através do túnel catalítico.

Por fim, simulações da TrCel7A na presença da celobiose mostram que a enzima incorpora uma dinâmica local no sítio do produto que é intermediária entre a *apo*-TrCel7A – que exibe conformações na forma de túnel – e a *holo*-TrCel7A – que exibe conformações abertas –, resultado do menor acoplamento celobiose-enzima quando o substrato é hidrolisado. Como consequência, estados que dificultam estericamene a dissociação do produto são termicamente acessíveis pela enzima, o que deve ser parte da causa da potente inibição pela celobiose observada experimentalmente.

Estes resultados revelam uma pequena fração da enorme complexidade do mecanismo enzimático das celobiohidrolases. Estudos adicionais, tanto experimentais quanto computacionais, são necessários para compreender mais detalhadamente as etapas moleculares da digestão enzimática da biomassa, o que vem sendo feito com sucesso nos últimos anos [4, 144].

Capítulo 6

Base molecular da termofilicidade de laminarinases

6.1 Introdução

Laminaranos são polissacarídeos formados por resíduos de glicose unidos por ligações glicosídicas $\beta(1 \rightarrow 3)$. São encontrados em paredes celulares de algas, fungos e leveduras; logo, apresentam potencial para utilização como matéria-prima para a produção de açúcares fermentáveis [212, 213]. Isso depende da despolimerização do laminarano, que pode ser efetivamente alcançada com laminarinases. Enquanto as celulases atacam ligações glicosídicas $\beta(1 \rightarrow 4)$, laminarinases são glicosídeo hidrolases que clivam ligações $\beta(1 \rightarrow 3)$, sendo, portanto, classificadas como exoou endo- β -1,3-glucanases [212].

Nas aplicações práticas, é desejável que as glicosídeo hidrolases apresentem estabilidade em altas temperaturas. Enzimas termoestáveis e termofílicas¹ apresentam maiores atividades específicas, maiores estabilidades, menos inibição por produto e proporcionam flexibilidade para a otimização de processos industriais [154, 215]. Estratégias de engenharia de proteínas têm sido aplicadas para aumentar a estabilidade térmica de enzimas, o que muitas vezes requer o conhecimento de fatores que regulam seus comportamentos térmicos [216–218]. Em inúmeros casos, enzimas similares, com poucas diferenças estruturais, apresentam diferentes graus de estabilidade térmica [217, 219, 220, 220]. Isso sugere que aspectos dinâmicos im-

¹Termoestabilidade é a capacidade da enzima de permanecer na ativa em altas temperaturas; é medida pela meia-vida da enzima em altas temperaturas [214, 215]. Termofilicidade é capacidade da enzima de apresentar atividade em altas temperaturas; é medida em termos de atividade enzimática sobre dado substrato em altas temperaturas [214].

portantes para a termoestabilidade e termofilicidade podem emergir de diferenças estruturais sutis. Assim, o estudo das bases moleculares da termoestabilidade e termofilicidade de glicosídeo hidrolases e outras enzimas de interesse biotecnológico tem sido assunto de amplo interesse [221–224].

Por serem adaptados a sobreviver em ambientes extremos, organismos extremófilos são fontes de enzimas termofílicas, embora organismos mesofílicos muitas vezes também expressam enzimas termofílicas [215]. A bactéria gram-negativa *Rhodotermus marinus* é conhecida por sobreviver em condições inóspitas – altas temperaturas e salinidade – de habitats geotérmicos [225]. Como tal, suas enzimas apresentam atividade e estabilidade em altas temperaturas, podendo alcançar temperaturas ótimas de até 100 °C, que é o caso de algumas de suas celulases [226]. A estrutura cristalográfica da laminarinase de *Rhodothermus marinus* (RmLam) foi resolvida por colaboradores em 2011. A caracterização bioquímica desta enzima, feita em 1998 por Krah *et al.* [227], mostrou que sua temperatura ótima é de 88 °C e apresenta temperatura de desnaturação acima de 90 °C. A RmLam é, pois, hipertermoestável e hipertermofílica², possuindo características desejáveis para aplicações biotecnológicas.

Laminarinases estruturalmente similares de outros organismos, no entanto, não são hipertermofílicas. É o caso da laminarinase mesofílica de *Phanerochaete chry*sosporium (PcLam) [229, 230], e da termofílica de alcaliphilic Nocardiopsis sp. (aN-Lam) [231, 232]. A figura 6.1 mostra a estrutura cristalográfica das laminarinases PcLam [233], aNLam [231] e RmLam [213]. Como membros da família GH16, estas laminarinases possuem enovelamento do tipo β -jelly-roll, em que duas camadas de folhas β se unem formando um sanduíche, dentro do qual reside o núcleo hidrofóbico da proteína. Apesar das razões moleculares da termoestabilidade serem múltiplas e variarem de proteína para proteína [217, 220], há inúmeros casos em que enzimas mesoestáveis são estabilizadas através da introdução de elementros obtidos de homólogas (hiper)termofílicas [217, 234]. Assim, a detecção de elementos termoestabilizadores em enzimas naturalmente termoestáveis podem fornecer uma plataforma para a incorporação destes elementos em outras homólogas, visando ao

²Enzimas mesofílicas são aquelas que apresentam atividade ótima entre 11 °C e 45 °C, as termofílicas entre 46 °C e 75 °C e as hipertermofílicas acima de 75 °C [228].



Figura 6.1. Estrutura cristalográfica das laminarinases PcLam (PDB: 2CL2) [233], aNLam (PDB: 2HYK) [231] e RmLam (PDB: 3ILN) [213]. Como membros da família GH16, estas laminarinases possuem enovelamento do tipo β -*jelly-roll*, formado por um sanduíche de duas folhas β , dentro do qual reside o núcleo hidrofóbico das proteínas. As folhas β possuem uma concavidade e formam um canal de ligação ao substrato, delimitado por vários loops.

aumento de estabilidade térmica.

Neste trabalho, simulações de dinâmica molecular foram executadas visando contribuir para o entendimento das bases moleculares da hipertermoestabilidade e hipertermofilicidade da laminarinase RmLam através de estudos comparativos com as enzimas homólogas PcLam (mesofílica) e aNLam (termofílica). Este estudo foi realizado em conjunto com outros membros do grupo; os resultados reportados aqui dizem respeito às partes das quais o autor desta tese participou. Resultados e detalhes adicionais estão reportados no artigo resultante desta pesquisa [213] e nas teses de doutorado dos outros autores do trabalho [235, 236]

6.2 Metodologia

Simulações de dinâmica molecular foram executadas para as enzimas PcLam, aN-Lam e RmLam nas temperaturas de 25 °C e 90 °C, sob pressão constante de 1 bar, utilizando o programa NAMD [171]. As estruturas iniciais destas enzimas foram tomadas do banco de dados PDB (www.rcsb.org) [168], identificadas pelos seguintes códigos: 2CL2 (PcLam) [233], 2HYK (aNLam) [231] e 3ILN (RmLam) [213]. Átomos de hidrogênio foram adicionados às estruturas, respeitando os estados de protonação obtidos com o servidor H++ [169, 170] em pH neutro. Caixas de simulação cúbicas de aproximadamente 80 Å de aresta e contendo 15.000 moléculas de água foram construídas para cada sistema utilizando o programa Packmol [237]. Nestas caixas, 50 íons sódio e cloreto (resultando numa concentração de aproximadamente 0,16 M) foram adicionados, além de contra-íons adicionais para tornar os sistemas eletricamente neutros. Todas as simulações foram executadas com NAMD [171] utilizando condições de contorno periódicas, com as interações de curto alcance truncadas em 12 Å e com as interações de longo alcance computadas via particle mesh Ewald [105]. Ligações envolvendo átomos de hidrogênio foram mantidas em seus comprimentos de equilíbrio e um passo de integração de 2 fs foi empregado. O campo de força CHARMM22/CMAP foi utilizado para as proteínas [172, 173] e o modelo TIP3P/CHARMM [172, 174] para a água.

Inicialmente, 500 passos de minimização de energia foram executados para cada sistema para a remoção de maus contatos cristalográficos. Em seguida, 10 ps de dinâmica molecular foram conduzidos com as proteínas fixas, seguidos de 100 ps de dinâmica molecular com os carbonos α fixos. Estas etapas foram seguidas por 890 ps de dinâmica molecular com todo o sistema livre, para equilibração na temperatura de 25 °C. Para os sistemas em 90 °C, 500 ps de dinâmica molecular numa temperatura intermediária, 55 °C, foram conduzidos para, em seguida, equilibrálos em 90 °C por 1 ns. Para os sistemas em 25 °C, 10 simulações independentes de 12 ns foram feitas para a RmLam, 8 simulações de 12 ns para a aNLam e 8 simulações de 9 ns para a PcLam. Em 90 °C, 5 simulações independentes de 9 ns forma executadas para cada enzima.

6.3 Resultados e discussão

6.3.1 Termoestabilidade

As simulações mostraram que a estabilidade térmica das laminarinases PcLam, aNLam e RmLam se correlaciona com a estabilidade de seus núcleos hidrofóbicos - a região delimitada pelas duas folhas β (figura 6.1). Na temperatura de 25 °C, não se observa a penetração de água no núcleo hidrofóbico das laminarinases. As moléculas de água presentes (3 na PcLam, 2 na aNLam e 1 na RmLam) são oriundas da estrutura cristalográfica e lá permanecem durante o curso das simulações (figura 6.2). Por outro lado, na temperatura de 90 °C, moléculas de água adicionais penetram o núcleo hidrofóbico da mesofílica PcLam (até 6 moléculas), indicando o início de um processo de desnaturação. As (hiper)termoestáveis aNLam e Rm-Lam, no entanto, não sofrem penetração de água e permanecem com seus núcleos hidrofóbicos intactos, como na temperatura ambiente. Estes resultados indicam que a estabilidade térmica das laminarinases está associada à capacidade do núcleo hidrofóbico de resistir à penetração de água em altas temperaturas. Além disso, a estabilidade térmica das (hiper)termofílicas aNLam e RmLam se correlaciona com o número de moléculas de água presentes em seus núcleos hidrofóbicos: apenas 1 molécula de água na hipertermofílica RmLam e 2 moléculas de água na termofílica aNLam. Isso sugere uma maior suscetibilidade do núcleo hidrofóbico da aNLam em relação ao da RmLam.

A razão molecular desse comportamento é a distribuição de pontes salinas nas estruturas das laminarinases. A mesoestável PcLam possui, na entrada de seu núcleo hidrofóbico, resíduos carregados que funcionam como uma porta de entrada para as moléculas de água (figura 6.2). Nas (hiper)termoestáveis aNLam e RmLam, as pontes salinas, embora mais numerosas que na PcLam, se distribuem sobre a superfície da proteína em contato com o meio aquoso, de forma que seus núcleos hidrofóbicos abrigam também resíduos hidrofílicos [213, 235, 236]. Logo, a distribuição espacial das pontes salinas das laminarinases constitui um fator determinante de suas estabilidades térmicas.

Embora elementos que conferem termoestabilidade a proteínas sejam múltiplos,


Figura 6.2. Configurações das laminarinases PcLam, aNLam e RmLam tomadas das simulações, mostrando as moléculas de água presentes no núcleo hidrofóbico de cada enzima nas temperaturas de 25°C e 90°C. A 25°C, as moléculas de água presentes no núcleo hidrofóbico são as mesmas presentes na estrutura cristalográfica. Ao aumentar a temperatura, uma ponte salina facilita a entrada de até 6 moléculas de água no núcleo hidrofóbico da mesoestável PcLam. A ausência desta ponte salina na termoestável aNLam e na hipertermoestável RmLam impede a penetração de água em seus núcleos hidrofóbicos, permanecendo, assim, estáveis em altas temperaturas.

a existência de pontes salinas na superfície das proteínas é o fator que mais se correlaciona com a termoestabilidade [217, 220]. Outros fatores, como número de ligações de hidrogênio, composição de aminoácidos ou tipo de estrutura secundária, correlacionam-se mais fracamente [220]. Além de possuir um número elevado de pontes salinas em sua superfície [213, 235, 236], a RmLam não possui resíduos polares em seu núcleo hidrofóbico, indicando um duplo fator termoestabilizador. Estes resultados mostram ainda que pontes salinas e/ou resíduos polares na entrada dos núcleos hidrofóbicos de glicosídeo hidrolases em geral devem ser alvos de mutações visando ao aumento da estabilidade térmica.

6.3.2 Termofilicidade

A análise das flutuações estruturais locais das laminarinases mostrou que a acessibilidade do sítio ativo ao substrato pode mudar ao longo do tempo. Estas alterações estruturais são devido a flutuações de um conjunto de cadeias laterais de resíduos do canal de ligação. Três estados conformacionais majoritários foram observados e classificados conforme o grau de acessibilidade ao substrato: (i) estado obstruído ou fechado, em que resíduos do sítio de ligação impedem o acesso do substrato ao sítio ativo; (ii) estado aberto, em que o canal de ligação está completamente exposto ao solvente e ao substrato; e (iii) estado na forma de túnel, em que resíduos do canal de ligação criam uma estrutura parcialmente fechada, mas que não impede a acessibilidade do substrato. A figura 6.3a ilustra cada um destes estados e a frequência com que eles ocorrem nas simulações. O procedimento para a obtenção destas frequências está descrito em detalhes na seção seguinte.

As simulações mostram que, embora todos estes três estados sejam acessíveis pelas laminarinases PcLam, aNLam e RmLam, a população predominante de cada um dos estados depende da enzima. Mais importante, a resposta das populações predominantes com o aumento de temperatura reflete o comportamento meso-, termo- ou hipertermofílico das laminarinases. A 25 °C, a população dos três estados é similar para as laminarinases PcLam e aNLam, em que o estado conformacional predominante é o aberto. Por outro lado, a 90 °C, enquanto a população relativa dos três estados permanece essencialmente inalterada para a termofílica aNLam, o estado obstruído se torna dominante para a mesofílica PcLam. Em termos de atividade enzimática, o estado aberto, por ser acessível ao substrato, é



Figura 6.3. Estados do sítio de ligação das laminarinases e suas frequências. O estado obstruído possui resíduos colapsados ao longo do canal de ligação das laminarinases e, por isso, possui o sítio ativo inacessível ao substrato, sendo, portanto, inativo. O estado aberto, visto nas estruturas cristalográficas, possui o canal de ligação ao substrato completamente exposto, sendo uma forma ativa das enzimas. O estado túnel é acessível ao substrato e possui resíduos formando uma possível região de confinamento para o substrato. Enquanto a população dos estados da mesofílica PcLam é sensível ao aumento de temperatura, a população das (hiper)termofílicas aNLam e RmLam é insensível. As populações predominantes em cada temperatura reflete a atividade relativa de cada uma das enzimas, conforme descrito no texto.

uma forma ativa da enzima, enquanto que o estado obstruído, por possuir o sítio ativo inacessível, é uma forma inativa. Consistentemente, os resultados mostram que a mesofílica PcLam se encontra em estados predominantemente ativos somente na temperatura de 25 °C; em 90 °C a PcLam passa a visitar predominantemente os estados inativos (obstruídos), perdendo sua atividade enzimática. Por ser termofílica, a aNLam resiste ao aumento de temperatura e permanece em estados predominantemente ativos em ambas as temperaturas. As simulações mostram, portanto, a razão molecular da perda e retenção de atividade enzimática em altas temperaturas da PcLam e aNLam, respectivamente.

Semelhantemente à aNLam, a população dos estados do sítio de ligação da Rm-Lam é insensível ao aumento de temperatura. No entanto, o estado predominante da RmLam é o da forma de túnel, e não o aberto. Como o estado aberto, o túnel não impede o acesso do substrato ao sítio ativo, sendo, portanto, provavelmente ativo, consistente com a atividade enzimática da RmLam em altas temperaturas. Adicionalmente, o túnel cria um espaço nanoscopicamente confinado, que poderia: (i) reduzir a taxa de dissociação do substrato antes da ocorrência da reação química; e (ii) estabilizar o estado de transição por meio de interações enzima-substrato mais efetivas que no caso da conformação completamente aberta. Assim, o estado túnel pode ser o mais ativo, e o fato de ser o estado predominante em altas temperaturas reflete a hipertermofilicidade da RmLam. Medidas de atividade enzimática da aNLam e RmLam com os mesmos substratos são necessárias para corroborar a ideia de que o estado túnel é mais ativo que o estado aberto.

O fato das laminarinases RmLam e aNLam exibirem predominância de estados ativos tanto em 25 °C quanto em 90 °C sugere que estas enzimas não são completamente inativas em temperaturas muito abaixo de suas temperaturas ótimas. Embora enzimas termofílicas tendam a exibir baixa atividade em temperaturas próximas à ambiente, isto não constitui uma regra; há casos em que enzimas exibem atividade numa ampla faixa de temperatura [217]. Segundo as simulações, este pode ser o caso das laminarinases aNLam e RmLam. Para outras glicosídeo hidrolases termofílicas, simulações mostraram que a exposição do sítio ativo só ocorre em altas temperaturas [238–240], indicando que simulações computacionais são capazes de indicar estados com baixa acessibilidade ao substrato em baixas temperaturas, se esse for o caso. No caso das laminarinases (hiper)termofílicas, estados acessíveis ao substrato predominam também em temperatura ambiente. De fato, há experimentos que mostram que a RmLam exibe atividade em 37 °C [241] e que a aNLam exibe atividade em 30 °C [230], muito embora não haja na literatura curvas de atividade enzimática em função da temperatura. Sendo assim, não há dados para dizer o quão mais baixo da atividade máxima é a atividade próximo à temperatura ambiente destas enzimas. Experimentos são necessários para corroborar os resultados das simulações.

6.3.3 Populações dos estados do sítio de ligação

Embora classificados em três estados bem definidos, os estados do sítio de ligação são consequência de flutuações de várias cadeias laterais em conjunto. Assim, a definição de um espaço conformacional de baixa dimensionalidade que contenha os três estados não é trivial. Diante da dificuldade em definir uma coordenada conformacional simples, que defina bem os estados, um esquema baseado em superfície acessível ao solvente foi designado para obter a população dos estados.

Numa primeira etapa, a superfície acessível ao solvente foi obtida utilizando uma esfera de raio 1,4 Å para varrer a superfície de van der Waals [242, 243] do canal de ligação. Em seguida, o espaço delimitado pelo sítio de ligação das laminarinases foi preenchido por pontos de uma malha retangular de resolução 1,0 Å, de tal forma que os pontos ocupassem toda a extensão do canal de ligação. O conjunto de pontos externos à superfície acessível ao solvente foram então analisados. Estes pontos definem a região não ocupadas por átomos da enzima. A figura 6.4 ilustra este procedimento.

O conjunto tridimensional de pontos foi então projetado no plano yz, conforme o sistema de eixos definido na figura 6.5. O eixo z é paralelo à direção do canal de ligação ao substrato das laminarinases, e o eixo y aponta em direção ao solvente. Os padrões bidimensionais obtidos da projeção dos pontos no plano yzforam associados aos estados obstruído, aberto e túnel, conforme ilustrado na figura 6.6. O estado obstruído é caracterizado por um padrão contendo uma região vazia contínua (regiões isentas de pontos correspondem a regiões preenchidas pela proteína). Ao contrário, o estado aberto é caracterizado por um padrão isento de vazios, o que significa que o canal de ligação está acessível. O estado túnel, por sua vez, é caracterizado por um buraco isolado no padrão de pontos, que reflete a extremidade superior do túnel formado pela proteína.

Para discriminar os três estados computacionalmente, o seguinte algoritmo foi implementado num programa computacional:



Figura 6.4. Contrução da malha tridimensional de pontos ao longo do canal catalítico das laminarinases. Somente os pontos externos à superfície acessível ao solvente das enzimas são considerados para análise, de maneira que os estados do sítio de ligação sejam capturados.



Figura 6.5. Projeções da malha tridimensional de pontos no plano xy e yz. O eixo y aponta em direção ao solvente e o eixo z é paralelo ao canal de ligação ao substrato. A projeção no plano yz permite que os três estados das laminarinases sejam univocamente discriminados.



Figura 6.6. Padrões da projeção da malha tridimensional sobre o plano yz. O estado obstruído é caracterizado por um padrão contendo uma região vazia, o estado aberto por uma região completamente preenchida e o estado túnel por uma região vazia isolada entre os pontos da malha.

- se há uma região vazia contínua, isto é, não acessível ao carboidrato, até a base da projeção, o estado é classificado como "obstruído" (figura 6.6, esquerda);
- se não for possível encontrar nenhuma região vazia ao longo das projeções, o estado é classificado como "aberto" (figura 6.6, centro);
- se houver uma região vazia, mas isolada, o estado é contado como "túnel" (figura 6.6, direita).

O algoritmo foi aplicado às trajetórias das laminarinases e as estatísticas obtidas foram compiladas na figura 6.3.

6.4 Considerações finais

Neste trabalho, aspectos moleculares relacionados ao comportamento térmico de laminarinases meso-, termo- e hipertermofílica foram investigados. Embora as três laminarinases sejam homólogas e compartilhem estruturas terciárias semelhantes, diferenças locais em suas estruturas primárias parecem influenciar suas estabilidades térmicas. Observou-se que a facilidade com que a água penetra o núcleo hidrofóbico das laminarinases se correlaciona com suas estabilidades térmicas, indicando que elementos estruturais que isolam o núcleo hidrofóbico do meio externo contribuem para a estabilidade térmica das enzimas. A presença de uma ponte salina na entrada do núcleo hidrofóbico da mesofílica PcLam, que funciona como uma porta de entrada de água para dentro do núcleo hidrofóbico, é um exemplo claro de um núcleo hidrofóbico não completamente protegido do meio externo aquoso. Em termos práticos, estes resultados revelam que resíduos polares dentro ou próximos de núcleos hidrofóbicos devem ser alvos de mutação visando ao aumento da estabilidade térmica de glicosídeo hidrolases em geral.

As laminarinases possuem um sítio de ligação ao substrato altamente dinâmico. Além da conformação aberta revelada pelas estruturas cristalográficas, as simulações indicam dois estados adicionais: um dos estados, supostamente inativo, possui o sítio ativo completamente obstruído; outro, supostamente o mais ativo, é acessível ao substrato e possui uma configuração que permitiria o substrato ser efetivamente confinado para sofrer hidrólise. Estes estados fazem parte de um equilíbrio dinâmico, o qual é sensível a variações de temperatura apenas para a laminarinase mesofílica PcLam. Nas (hiper)termofílicas aNLam e RmLam, estados ativos dominam o equilíbrio mesmo em altas temperaturas. A população destes estados parece refletir a atividade relativa das laminarinases em altas e baixas temperaturas.

Em suma, tais resultados mostram que a estabilidade e a atividade de glicosídeo hidrolases em altas temperaturas estão intimamente relacionadas às suas estruturas, dinâmicas e interações com o solvente, e contribuem para o entendimento dos vários fatores que influenciam o comportamento térmico de enzimas em geral. Sem dúvida, tais informações são fundamentais para o desenvolvimento estratégias de engenharia de enzimas baseadas em desenho racional, a fim de aumentar a estabilidade e atividade de glicosídeo hidrolases em altas temperaturas, características essas necessárias para a eficiente conversão enzimática da biomassa em biocombustíveis.

Capítulo 7 Disrupção não-hidrolítica de biomassa por expansinas

7.1 Introdução

As paredes celulares de plantas oferecem suporte e proteção às células vegetais. Como descrito no Capítulo 2, as paredes celulares são constituídas de microfibrilas de celulose envoltas por uma matriz polimérica amorfa contendo hemiceluloses, lignina, pectina e proteínas. Esta matriz polimérica se arranja numa rede que mantém as microfibrilas unidas por ligações não-covalentes [5]. Arranjadas dessa forma, a parede celular possui a resistência mecânica necessária para suportar as pressões de turgor originadas do citoplasma das células. Durante o período de crescimento, entretanto, a parede celular deve ser flexível o bastante para acompanhar o aumento do volume das células. Em alguns casos, as células podem sofrer um aumento de volume de 30.000 vezes antes de atingir a maturidade [244, 245]. Assim, a arquitetura da parede celular, deve, de alguma forma, ser modificada para permitir o crescimento da planta.

Expansinas são proteínas que regulam a extensão da parede celular de plantas em desenvolvimento [244–246], e vários estudos genéticos e funcionais demonstram suas funções biológicas. Expansinas induzem extensão e relaxação em plantas submetidas a tensões externas [247–250], e estão envolvidas em processos que requerem o afrouxamento da arquitetura da parede celular, como amadurecimento de frutos [251, 252], abscisão (queda) de folhas e frutos [253], crescimento de raízes [254] e germinação de sementes [255]. Enquanto tecidos em fase de crescimento naturalmente expressam genes de expansina, o silenciamento desses genes impede o crescimento da planta [256–258]. Além disso, a adição de expansinas exógenas a plantas normais estimula ainda mais o crescimento, indicando que as expansinas endógenas podem limitar o desenvolvimento das plantas [245, 249, 259].

A atividade da expansina é medida em termos da taxa de extensão que ela induz em paredes celulares submetidas a forças externas. A figura 7.1 mostra o aparato (extensômetro) utilizado nos experimentos. Parte do tecido vegetal em crescimento é submetido a um tratamento térmico para inativar as expansinas. Em seguida, este tecido é submetido a uma força externa constante e seu comprimento é medido em função do tempo. Observa-se que, devido à inativação das expansinas,



Figura 7.1. Experimento de extensão de parede celular por expansinas. Paredes celulares com expansinas inativadas termicamente não respondem à força externa, mas, quando na presença de expansinas exógenas, respondem imediatamente à força externa, apresentando aumento significativo da taxa de extensão. Adaptado de [244] com permissão de Nature Publishing Group.

o tecido tratado termicamente exibe pouca ou nenhuma extensão como resposta à força externa. Quando, porém, expansinas são adicionadas ao meio, provocam um aumento da taxa de extensão sob força externa constante [244].

Duas famílias de expansinas (expansinas- α e expansinas- β) ocorrem em plantas, e proteínas similares são encontradas em outros organismos, como bactérias patogênicas que colonizam raízes de plantas [260–262] e fungos que degradam biomassa lignocelulósica [74, 75, 263]. Todas estas proteínas induzem modificações nas paredes celulares vegetais por um mecanismo não-hidrolítico e, por isso, são conhecidas como proteínas disruptivas não-hidrolíticas [78]. Além do papel biológico, estas proteínas disruptivas afetam a degradação enzimática de biomassa. A parede celular é recalcitrante à degradação enzimática em parte devido à reduzida acessibilidade da celulose [2, 4, 17, 57]. Quando misturada com celulases, as proteínas disruptivas aumentam a atividade enzimática, supostamente aumentando a acessibilidade do substrato às enzimas [71, 74–78, 261]. Embora tem sido proposto que tais proteínas agem rompendo as redes de ligações de hidrogênio dentro da parede celular [78], nenhuma evidência experimental ou computacional deste mecanismo foi reportada até o momento.

7.1.1 Expansina EXLX1 de Bacillus subtilis

A expansina EXLX1 da bactéria *Bacillus subtilis* tem recebido atenção especial ultimamente. Vários estudos relacionados à estrutura e função de expansinas vêm utilizando a EXLX1 como expansina-modelo, principalmente devido às dificuldades de expressão heteróloga de expansinas de plantas em bactérias. Devido às similaridades com as expansinas de plantas, o conhecimento obtido para a EXLX1 provavelmente se aplica a expansinas de plantas e a outras proteínas disruptivas. A EXLX1 exibe atividade de extensão de parede celular qualitativamente similar a expansinas- β [260], apresenta atividade de enfraquecimento de celulose [72, 264] e potencializa a ação de celulases sobre celulose [72]. Associado a isso, a EXLX1 possui estrutura similar à de expansinas de plantas, em que dois domínios, D1 e D2, se contatam e são conectados por um pequeno *linker*, como ilustrado na figura 7.2 [260, 265]. O domínio D1 exibe similaridade distante com glicosídeo hidrola-



Figura 7.2. Estrutura 3D da expansina EXLX1 (PDB: 3D30 [260]), mostrando os domínios D1 e D2, e os aminoácidos aromáticos e básicos, presentes em D2, e polares/carregados presentes em D1.

ses da família 45 (GH45) [266], mas nenhuma atividade hidrolítica foi detectada para a expansina até o momento. Como as enzimas GH45, o domínio D1 possui uma superfície populada por resíduos polares/carregados, os quais são importantes para a atividade da expansina [264]. O domínio D2 é um domínio do tipo imunoglobulina (*Ig-like*) e foi recentemente reconhecido como um módulo de ligação a carboidratos do tipo A – que reconhece superfícies cristalinas –, pertencente à família 63 (CBM63). Como tal, o domínio D2 exibe uma superfície plana formada por resíduos aromáticos, adaptada ao reconhecimento da celulose cristalina (figura 7.2) [135, 260, 264, 265, 267].

Georgelis *et al.* [264] mostraram que ambos os domínios D1 e D2 devem estar conectados pelo *linker* para que a EXLX1 apresente atividade de extensão de parede celular. O domínio D1 ou o D2 isolados não exibem atividade, e nem os domínios D1 e D2 presentes no meio, mas não conectados pelo *linker*. Assim, o mecanismo de ação da EXLX1 envolve a ação conjunta de ambos os domínios conectados pelo *linker*. A mutação dos resídos aromáticos do domínio D2 elimina completamente tanto a capacidade de ligação da EXLX1 como sua atividade em parede celular e papel filtro (celulose pura). Isso indica que há uma relação entre ligação da EXLX1 e sua atividade. Finalmente, a ligação da EXLX1 aos polissa-carídeos não-celulósicos da parede celular (hemicelulose e pectina) é mediado por resíduos básicos na região do domínio D2 oposta à superfície aromática (figura 7.2). Recentemente, estudos de RMN de estado sólido revelaram que domínios da celulose ricos em xiloglucano são os alvos da expansina EXLX1 na parede celular [268]. Semelhantemente, Park & Cosgrove [269] demonstraram que a atividade de extensão de expansina- α é reduzida em plantas transgênicas deficientes em xiloglucano. Estes estudos sugerem que a ação das expansinas envolve ligação à celulose e algum outro tipo de envolvimento com a fração de xiloglucanos fortemente ligada à celulose cristalina.

Apesar dos efeitos biofísicos das expansinas em paredes celulares serem bem descritos, as bases moleculares por trás desses efeitos ainda são pouco compreendidas, principalmente devido às dificuldades de estudar mecanismos biofísicos em sistemas biológicos complexos na nanoescala. Nesta parte da tese, um extenso conjunto de simulações de dinâmica molecular foram realizadas para entender as seguintes questões: (1) a adsorção e dinâmica da EXLX1 sobre celulose cristalina; (2) a interação da EXLX1 com uma cadeia de glucano; (3) as bases moleculares da inativação da EXLX1 com as mutações Asp82Asn e Tyr73Ala. O efeito de tais mutações, como descrito adiante, fornecem informações sobre o mecanismo da EXLX1.

7.2 Metodologia

7.2.1 Construção dos sistemas

EXLX1 nativa

As coordenadas iniciais da expansina EXLX1 foram tomadas da estrutura cristalográfica obtida por Kerff et al. (código PDB: 3D30) [260]. Átomos de hidrogênio foram adicionados de acordo com valores de pK_a em pH=7, estimados usando o servidor H++ [169, 170]. O pH neutro foi escolhido considerando que a EXLX1 é ativa em valores de pH variando de 5,5 a 9,5 [264]. A estrutura foi então solvatada por uma caixa retangular contendo moléculas de água e íons Na⁺ e Cl⁻ na concentração de 0,15 M, com excesso de contra-íons para neutralizar o sistema eletricamente. A caixa de simulação foi construída com o programa Packmol [237] de tal forma que a distância mínima entre a proteína e as extremidades da caixa fosse de 16 Å. A estrutura assim construída foi então submetida ao seguinte procedimento: (1) com todos os átomos da EXLX1 fixos, 1000 passos de minimização de energia com gradientes conjugados seguidos de 200 ps de dinâmica molecular; (2) com todos carbonos α da EXLX1 fixos, 1000 passos de minimização de energia com gradientes conjugados seguidos de 200 ps de dinâmica molecular; (3) 600 ps de dinâmica molecular com todos os átomos livres. Após estas etapas de preparação, uma trajetória de 150 ns foi obtida e utilizada para as análises.

Mutantes Asp82Asn e Tyr73Ala

Da simulação de 150 ns da EXLX1 nativa, uma configuração foi tomada aleatoriamente – já que as configurações ao longo dos 150-ns de simulação diferem pouco entre si – para construir os sistemas dos mutantes Asp82Asn e Tyr73Ala. As coordenadas dos resíduos mutantes foram inseridas conforme as coordenadas internas do campo de força CHARMM22 [172]. Para o mutante Asp82Asn, um cátion Na⁺ foi removido para compensar a substituição do resíduo Asp negativamente carregado pelo resíduo Asn eletricamente neutro. Ambos os sistemas foram submetidos a 100 passos de minimização de energia seguidos de 10 ns de dinâmica molecular, com todos os átomos livres. Uma trajetória de 150 ns foi então gerada para cada um dos mutantes.

EXLX1 sobre celulose cristalina

Para as simulações envolvendo a celulose cristalina, o modelo hexagonal de fibrila elementar de 36 cadeias foi empregado (figura 2.4a) [14]. Uma fibrila elementar de celulose cristalina I β , de grau de polimerização 24, foi gerada com o programa cellulose-builder [9, 270]. Este grau de polimerização foi escolhido de forma que o comprimento da fibrila fosse aproximadamente duas vezes o tamanho da EXLX1. Para estudar o processo de adsorção, a EXLX1 foi colocada a aproximadamente 10 Å da superfície hidrofóbica da celulose numa orientação em que a superfície aromática do domínio D2 está paralela à superfície. Uma camada de água de pelo menos 20 Å da EXLX1 ou da celulose foi criada usando Packmol [237]. Tal arranjo permitiu que moléculas de água ocupassem o espaço entre a EXLX1 e a superfície da celulose, de forma que o processo de adsorção pudesse ser observado. Antes de gerar uma trajetória de 50 ns – tempo suficiente para a EXLX1 se adsorver sobre a celulose – em temperatura e volume constantes, o sistema foi preparado com o seguinte procedimento: (1) 1000 passos de minimização de energia seguidos de 500 ps de dinâmica molecular com todos os átomos da EXLX1 e da celulose fixos; (2) 1000 passos de minimização de energia seguidos de 2 ns de dinâmica molecular com todo o sistema livre. O passo (1) foi executado para permitir a relaxação da camada de água entre a EXLX1 e a celulose e o passo (2) para equilibrar a temperatura e a densidade do sistema (esta etapa foi feita com temperatura e pressão constantes). Para estudar a dinâmica da EXLX1 adsovida sobre a celulose, as coordenadas do complexo EXLX1-celulose foram tomadas do final da simulação de adsorção e uma nova caixa de simulação foi construída, dessa vez com uma camada de água mais fina, mas ainda com pelo menos 20 Å de distância entre as extremidades da caixa e o complexo EXLX1-celulose. Uma simulação de 200 ns foi então executada após a repetição das mesmas etapas de preparação (1) e (2)da simulação de adsorção.

EXLX1 ligada a cadeia de glucano

A estrutura inicial da EXLX1 complexada com um oligossacarídeo foi obtida tomando as coordenadas da EXLX1 e da cadeia de glucano alinhada à superfície aromática no complexo EXLX1-celulose. Este sistema (sem solvente) foi então submetido a 10.000 passos de minimização de energia seguidos de 200 ps de dinâmica molecular com a EXLX1 fixa. Isso foi feito para permitir o oligossacarídeo se ajustar a ambos os domínios D1 e D2 da EXLX1. Após este procedimento, o sistema foi solvatado por uma camada de água de pelo menos 16 Å numa caixa retangular contendo NaCl na concentração de 0,15 M. Contra-íons em excesso foram adicionados para neutralizar o sistema. Para relaxar/equilibrar o sistema, foram adotados os mesmos passos usados para o sistema da EXLX1 nativa na ausência de substrato. Inicialmente, uma trajetória de 200 ns foi gerada. Nos primeiros 100 ns, o substrato permaneceu ligado a ambos os domínios D1 e D2. Após este período, o substrato se dissociou do domínio D1 e começou a explorar conformações no meio. Para aumentar a amostragem, três novas simulações foram executadas partindo de estruturas tomadas aleatoriamente dos primeiros 100 ns da simulação (as configurações diferem pouco entre si). Destas três, o substrato permaneceu no estado ligado durante os primeiros 60, 38 e 32 ns. A dissociação do substrato do domínio D1 é esperada, já que este domínio não possui função de ligação, que é exercida pelo domínio D2 [264]. As estruturas com substrato complexado foram então reunidas e utilizadas para análise.

7.2.2 Detalhes computacionais

Todas as simulações foram executadas com o programa NAMD [171]. O campo de força CHARMM22 com a correção CMAP foi usado para modelar as interações da EXLX1 [172, 173] e o campo de força CHARMM36 para a celulose [193]. Para as moléculas de água, o modelo TIP3P/CHARMM foi empregado [172, 174]. Condições de contorno periódicas foram empregadas; as interações de longo alcance foram computadas usando *particle mesh* Ewald [105] e as de curto alcance foram truncadas utilizando raio de corte de 12 Å. Ligações químicas envolvendo átomos de hidrogênio foram vinculadas em seus comprimentos de equilíbrio e um passo de 2.0 fs foi usado para integrar as equações de movimento numericamente. Exceto nas simulações de fibrila de celulose, em todos os sistemas a temperatura e a pressão foram mantidas constantes em 300 K e 1 atm através do termostato e pistão de Langevin, respectivamente, como implementado no NAMD. Nos sistemas envolvendo celulose, as simulações foram realizadas em volume e temperatura constantes.

7.3 Resultados

7.3.1 Adsorção da EXLX1 sobre celulose cristalina

Para estudar a dinâmica da EXLX1 interagindo com uma fibrila elementar de celulose, simulações de dinâmica molecular foram executadas para um sistema contendo uma fibrila de celulose e a EXLX1 dispersa no meio. A EXLX1 foi posicionada próximo à face hidrofóbica da celulose e o sistema foi permitido evoluir naturalmente no tempo. Dentro dos primeiros 20 ns de simulação, o domínio D2 dirige a adsorção da EXLX1 sobre a superfície da celulose. A figura 7.3a mostra as distâncias entre a superfície da celulose e os domínios D1 e D2, com as configurações do sistema nos tempos t = 0, t = 7,5 ns e t > 20 ns. A adsorção inicia com o domínio D2 se aproximando da celulose através do resíduo aromático Trp125 (t = 7, 5 ns) (figura 7.2). À medida que os outros resíduos aromáticos do domínio D2 se aproximam da celulose, o domínio D1 é forçado a se aproximar da celulose devido aos vínculos geométricos da EXLX1. Durante este processo, a camada de água presente na região entre o domínio D2 e a superfície da celulose é expulsa para o meio. Estes resultados indicam que as interações entre a superfície aromática da EXLX1 e a superfície hidrofóbica da celulose – seguida pela expulsão das águas interfaciais – constituem a força motriz da adsorção da EXLX1.

As energias de interação entre a fibrila de celulose e a EXLX1 fornecem detalhes adicionais sobre a natureza das interações EXLX1-celulose. A energia de interação total entre a EXLX1 e a celulose decresce durante o processo de adsorção devido a uma forte atração de van der Waals; ao mesmo tempo, a componente eletrostática



Figura 7.3. Adsorção da expansina EXLX1 sobre a superfície hidrofóbica da celulose cristalina. (a) Distância entre a superfície da fibrila de celulose e os domínios D1 e D2 durante o processo de adsorção, com configurações representativas da simulação. Observa-se que interações do domínio D2 com a superfície da celulose dirige o processo de adsorção da EXLX1. (b) Energias de interação entre a celulose e a EXLX1 intacta (esquerda), somente o domínio D1 (central) e somente o domínio D2 (direita) durante o processo de adsorção. Ocorre uma redução na energia de van der Waals entre a celulose e o domínio D2 durante a adsorção.

da energia flutua em torno de valores positivos (figura 7.3b, painel da esquerda). A origem principal destas interações de van der Waals são as interações do domínio D2, e não do D1, o qual interage com a celulose principalmente através de forças eletrostáticas (figura 7.3b, painel central e da direita). Além disso, a energia de interação total entre o domínio D2 e a celulose converge para valores próximos de zero nos estágios finais da simulação. Isso sugere a predominância de fatores de estabilização entrópicos – associado à expulsão de moléculas de água – em relação a fatores entálpicos na ligação da EXLX1 à celulose. Logo, estes resultados indicam que as interações EXLX1-celulose são de natureza hidrofóbica e que forças entrópicas dirigem o processo de adsorção.

7.3.2 Translação longitudinal da EXLX1 sobre a celulose

Após o processo de adsorção ter sido finalizado em 50 ns, a trajetória foi estendida por mais 200 ns para estudar o comportamento dinâmico da EXLX1 adsorvida sobre a celulose. A figura 7.4 mostra as energias de interação entre a celulose e a EXLX1 adsorvida. Observa-se que as energias flutuam em torno de valores médios aproximadamente constantes, indicando que a EXLX1 não está num estado transiente de adaptação à superfície.



Figura 7.4. Energias de interação entre a celulose e a EXLX1 intacta (esquerda), somente o domínio D1 (central) e somente o domínio D2 (direita) com a expansina adsorvida sobre a fibrila de celulose. As flutuações das energias em torno de valores médios aproximadamente constantes indicam que a EXLX1 não está num estado transiente de adaptação sobre a celulose.

A simulação mostra que a EXLX1 pode transladar longitudinalmente sobre a superfície hidrofóbica da celulose. Observou-se que, em torno de 80 ns, a EXLX1 desliza ao longo de um resíduo de glicose na superfície (figura 7.5). A cadeia de celulose sobre a qual os três resíduos aromáticos da EXLX1 (Trp125, Trp126 e Tyr157) se alinham funciona como um trilho sobre o qual a EXLX1 translada. Este resultado sugere que o domínio D2 exerce funções além de simplesmente ligar a EXLX1 à celulose, já que a translação longitudinal da EXLX1 pode ser essencial para o mecanismo das expansinas. Translação longitudinal ao longo da face hidrofóbica da celulose foi observada em vários estudos para CBMs do tipo A isolados, que possuem uma superfície aromática que reconhece superfícies cristalinas [137, 271, 272].



Figura 7.5. Movimento translacional da EXLX1 sobre a superfície da celulose. (a) Evolução temporal da distância entre o carbono α do resíduo aromático Trp125 (domínio D2) e o resíduo de glicose com o qual está em contato inicialmente. Observa-se uma transição abrupta em torno de 80 ns, indicativa da translação da EXLX1 ao longo de uma unidade de glucose sobre a celulose. (b) Posições inicial e final da EXLX1 sobre a superfície da celulose, evidenciando que ocorreu uma translação da proteína.

A EXLX1 não induziu alterações na estrutura da celulose nos 200 ns de simulação. Resultado similar foi reportado por Wang *et al.* com 40 ns de simulação da EXLX1 sobre celulose [268]. Tal observação não é completamente inesperada, já que a escala de tempo dos efeitos macroscópicos induzidos pelas expansinas em paredes celulares é da ordem de segundos a horas [260].

7.3.3 Centro de torção: interação entre EXLX1 e oligossacarídeo

A hipótese atual para o mecanismo de ação das expansinas é a de que elas rompem ligações de hidrogênio entre componentes da parede celular, assim enfraquecendo sua estrutura. Uma vez que já foi demonstrado que o domínio D2 exibe função de ligação à celulose e hemicelulose, enquanto o domínio D1 possui vários resíduos polares/carregados que são importantes para atividade – mas de função desconhecida –, é razoável supor que o domínio D1 é responsável pela função disruptiva das expansinas [264]. Para compreender o papel do domínio D1, a EXLX1 foi simulada complexada com uma cadeia oligossacarídica. Já que o oligossacarídeo não está empacotado na estrutura cristalina da celulose, este modelo permite investigar as interações EXLX1-substrato que seriam estabelecidas uma vez que tenham sido rompidas (seja por alguma ação da EXLX1 ou devido à presença de defeitos préexistentes na superfície) as interações intermoleculares entre as cadeias de celulose cristalina. O modelo representa também interações entre a EXLX1 e regiões de uma cadeia de xiloglucano isenta de ramificações, cuja cadeia principal coincide com a cadeia de glucano.

A figura 7.6 mostra a estrutura da EXLX1 complexada à cadeia oligossacarídica em sua conformação predominante. De azul a vermelho, as cores representam uma escala de flutuações conformacionais de 1,5 a 3,0 Å. A mobilidade da cadeia varia ao longo dos sítios de ligação da EXLX1 (numerados na figura 7.6) e apresenta valores mínimos em duas regiões específicas (figuras 7.6a, 7.6b): sobre o resíduo Trp126 do domínio D2 e sobre o resíduo Asp82 do domínio D1. Enquanto os resíduos aromáticos seguram a cadeia numa conformação planar no domínio D2 através de interações do tipo CH- π stacking, vários resíduos polares em torno do Asp82 induzem uma torção na parte da cadeia oligossacarídica ligada ao domínio D1. Esta torção pode ser parte do mecanismo pelo qual a EXLX1 enfraquece a estrutura da celulose e da parede celular.

A análise das energias de interação entre o oligossacarídeo e cada um dos resíduos da EXLX1 mostra quais são os resíduos importantes para interagir com o substrato (figura 7.7a). No domínio D2, os resíduos aromáticos Trp125, Trp126 e Tyr157 interagem com o substrato através de *stacking* hidrofóbico CH- π enquanto o resíduo Lys119 auxilia na estabilização do *stacking* com uma ligação de hidrogênio com um resíduo de glicose. Consistentemente, a mutação de cada um desses quatro resíduos por alanina resulta na diminuição da capacidade de ligação da EXLX1 à celulose [264]. Entretanto, as constantes de ligação dos mutantes não se correlacionam com os valores das energias de interação reportadas aqui (tabela 7.1), indicando que forças entrópicas, e não somente entálpicas, regulam as interações EXLX1-substrato. Por estarem distantes dos sítios de ligação aromáticos, os resíduos Glu191, Ser192 e Gly193 interagem fracamente com o substrato (figura 7.7b). Com exceção do resíduo Asp82, os resíduos do domínio D1 envolvidos na



Figura 7.6. Conformação da cadeia oligossacarídica ligada à EXLX1. (a) Estrutura da EXLX1 ligada ao oligossacarídeo colorido conforme suas mobilidades atômicas obtidas durante o curso das simulações. De azul a vermelho, a escala de cor representa flutuações médias de <1,5 Å a >3.0Å. Enquanto o domínio D2 segura o substrato em uma conformação planar, o domínio D1 induz uma torção de 90° na cadeia (indicada pelo retângulo pontilhado). (b) Flutuação média dos resíduos de glicose ao longo da cadeia oligossacarídica. Barras de erro indicam desvio padrão de diferentes simulações. Os números dos resíduos de glicose estão indicados no painel (a). Ambos os domínios D1 e D2 possuem resíduos capazes de estabilizar a conformação do substrato.

ligação do substrato o fazem com energia de interação média de -10 kcal/mol, principalmente via ligações de hidrogênio (somente o resíduo Tyr13 não faz ligação de hidrogênio com a cadeia). O resíduo Asp82 parece ter uma importância central na interação EXLX1-substrato. Diferentemente dos outros resíduos, o Asp82 interage com energia média de -30 kcal/mol. De fato, a substituição do Asp82 por alanina ou asparagina inativa completamente a EXLX1 [264].



Figura 7.7. Interações entre EXLX1 e oligossacarídeo. (a) Energias médias de interação entre o substrato e cada um dos resíduos da EXLX1. (b) Estrutura da EXLX1 mostrando a localização dos resíduos que interagem com o substrato.

Tabela 7.1. Comparação das constantes de dissociação (K_d) associadas à interação expansina-celulose de mutantes em que os resíduos listados foram substituídos por alanina [264], com as energias de interação expansina-glucano obtidas das simulações. Os dados mostram que não há correlação entre as energias de interação com as constantes de ligação dos mutantes, sugerindo que fatores não-entálpicos são importantes na interação EXLX1-substrato. Os valores entre parênteses indicam desvio padrão.

Resíduo	K_d (mutação por alanina)	Energia (kcal/mol)
Lys119	3,53(1,02)	-40 (6)
Trp125	5,72(1,21)	-10 (2)
Trp126	4,62 $(1,58)$	-12(2)
Tyr157	$5,\!61\ (1,\!53)$	-5(2)

Entre todos os resíduos do domínio D1, os resíduos Thr14, Ser16, Asp71 e Asp82 são os que estão envolvidos diretamente com a torção do oligossacarídeo (figura 7.8). As frequências de ligações de hidrogênio computadas das simulações indicam que, destes quatro resíduos, a Thr14 e o Asp82 assumem um papel mais importante nas interações, já que eles estão mais persistentemente envolvidos em ligações de hidrogênio que os resíduos Ser16 e Asp71 (figura 7.8b). Além disso, o resíduo



Figura 7.8. Ligações de hidrogênio entre o domínio D1 e o oligossacarídeo. (a) Visão detalhada das interações do substrato com os resíduos do domínio D1 responsáveis pela torção. A numeração dos resíduos de glicose corresponde àquelas mostradas na figura 7.6a. (b) Frequência das ligações de hidrogênio entre o substrato e a EXLX1. O resíduo Asp82 se envolve na mais persistente ligação de hidrogênio com o oligossacarídeo.

Asp82 pode interagir com o substrato através de duas ligações de hidrogênio. Estes resultados são consistentes com o perfil de mobilidade dos resíduos de glicose ligados aos sítios 8, 9 e 10 do domínio D1 (figura 7.6), uma vez que as unidades de glicose mais móveis estão associadas com ligações de hidrogênio menos persistentes com resíduos do domínio D1.

Os resultados indicam que o resíduo Asp82 é o que interage mais fortemente com o substrato e é capaz de induzir uma estabilização estrutural no estado torcionado da cadeia oligossacarídica. Assim, o resíduo Asp82 – cuja estabilidade estrutural é reforçada por uma ligação de hidrogênio com o resíduo Thr12 (figura 7.8a) – parece ser uma espécie de sítio ativo da EXLX1, o que será denotado aqui por centro de torção da EXLX1.

7.3.4 Mutante Asp82Asn

De acordo com os resultados da seção anterior, o resíduo Asp82 parece ser essencial para a função da EXLX1 por criar um centro de torção que induz alterações conformacionais na cadeia oligossacarídica. Recentemente, foi demonstrado que a substituição do Asp82 por uma alanina desabilita a atividade de extensão da EXLX1 em paredes celulares [264]. A mutação Asp82Ala na expansina Exl1 de Pectobacterium carotovorum, que possui a superfície de ligação do domínio D1 similar à da EXLX1, desabilita sua atividade de amorfogênese de celulose cristalina, sugerindo que o mesmo deve acontecer com a EXLX1 mutante [273]. Isso indica que ligações de hidrogênio mediadas pelo aspartato estão envolvidas na atividade não-hidrolítica da EXLX1. O efeito da substituição do Asp82 por outros resíduo polares capazes de fazer ligações de hidrogênio não causaria, em tese, perturbações drásticas na atividade da EXLX1. No entanto, não é isso que se observa: o mutante Asp82Asn também não exibe atividade em paredes celulares [264]. A princípio, isso parece contra-intuitivo, uma vez que o resíduo de asparagina também possui capacidade de ligação de hidrogênio através de seu grupo amida. Uma possibilidade é que o resíduo Asp82 participa da catálise de uma reação química, de forma que sua substituição por Asn impediria um ataque nucleofílico ou transferência de próton, conforme os mecanismos mostrados nas figuras 2.8 e 2.7. Entretanto, diversos estudos reportaram evidências de as expansinas serem não-hidrolíticas. A fim de entender as razões do fato do mutante Asp82Asn ser inativo com base no papel sugerido para o resíduo Asp82 de centro de torção, o sistema mutado foi simulado.

Diferentemente do resíduo Asp82 na EXLX1 nativa, o resíduo mutante Asn82 não está disponível para interagir com o substrato. Através de um mecanismo envolvendo o resíduo vizinho Tyr73, o resíduo Asn82 se move regularmente para uma região inacessível ao solvente do domínio D1, impedindo o contato com o substrato. Como ilustrado na figura 7.9, os resíduos Asn82 e Tyr73 entram num equilíbrio mútuo entre dois estados conformacionais, os quais são denotados aqui como estados ativo e inativo. No estado ativo, o resíduo Asn82 e todos os outros em sua vizinhança, assumem uma configuração similar àquela assumida pela EXLX1 nativa, ilustrada com o substrato na figura 7.8a. No estado inativo, o resíduo Asn82 interagem com o resíduo Tyr73, causando alterações conformacionais que resultam no deslocamento da Asn82 para uma posição debaixo do anel aromático da Tyr73. O resíduo Thr14 parece contribuir com este processo fazendo uma ligação de hidrogênio com a Tyr73 (figura 7.9). O resultado disso é que o resíduo Asn82 no estado inativo não está mais disponível para induzir torções no substrato. Interações entre grupos amida e anéis aromáticos, como a interação entre Asn82 e Tyr73, são comuns em proteínas e são de natureza dispersiva e eletrostática [274].

A evolução temporal do rmsd do resíduo Asn82 mostra que o tempo de vida dos estados ativo e inativo é da ordem de algumas dezenas de nanossegundos (figura 7.10a) e que a população do estado inativo (60%) é somente ligeiramente maior que a do estado inativo (40%) (figura 7.10b). Assim, ambos os estados são acessados durante o curso da simulação e numa escala de tempo muito menor que a escala de tempo de ação da EXLX1 na parede celular. Embora o estado ativo represente uma fração considerável das conformações amostradas, sugerindo a existência de uma atividade residual do mutante Asp82Asn, este estado não é persistente por tempo suficiente para a ação da EXLX1. Estes resultados corroboram o papel central do resíduo Asp82 no mecanismo da EXLX1.

7.3.5 Mutante Tyr73Ala

Estudos de mutagênese sítio-dirigida mostraram que o mutante Tyr73Ala é essencialmente inativo, enquanto o mutação Tyr73Phe resulta em nenhuma alteração na atividade da EXLX1 [264]. Isso mostra que o resíduo Tyr73 exerce um papel na atividade da EXLX1 através de seu anel aromático e que não está envolvido em ligações de hidrogênio. Diante da ideia desenvolvida nas seções anteriores, de que a atividade da EXLX1 é determinada por resíduos do centro de torção, não é claro como o resíduo Tyr73 influencia a função da EXLX1. De fato, a análise de energias de interação entre a EXLX1 e a cadeia oligossacarídica mostra que o resíduo Tyr73 não interage com o substrato (figura 7.7a). Para compreender a função do resíduo Tyr73 e sua relação com o centro de torção, o mutante Tyr73Ala foi simulado.

Os resultados sugerem que a EXLX1 mutante carece da estabilidade estrutural em torno do centro de torção necessária para estabelecer interações proteína-



Figura 7.9. Conformação do centro de torção (domínio D1) no mutante Asp82Asn, mostrando o equilíbrio conformacional entre os estados ativo e inativo, onde o par Asn82/Tyr73 assume duas conformações distintas. No estado inativo, o resíduo Asn82 torna-se indisponível para interagir com o substrato.



Figura 7.10. Ocorrência dos estados ativo e inativo no mutante Asp82Asn. (a) Rmsd do resíduo Asn82 em relação às coordenadas iniciais (após relaxação) da simulação. O rmsd flutua em torno de dois valores, correspondentes ao estado ativo (rmsd ≈ 1 Å) e inativo (rmsd ≈ 2 Å) do mutante Asp82Asn. (b) Distribuição dos valores de rmsd do resíduo Asn82, mostrando as probabilidades (área integrada) dos estados ativo (40%) e inativo (60%).

substrato efetivas. A comparação da dinâmica estrutural da EXLX1 nativa e mutante indica que a mutação perturba o domínio D1 em várias regiões, como mostrado na figura 7.11a. A mutação perturba particularmente a região de loops perto do resíduo Ala73, como ilustrado pela representação em escala de cores da mobilidade dos resíduos na estrutura da EXLX1 (figura 7.11b). Outras regiões do domínio D1, distantes do resíduo Ala73, que incluem resíduos funcionais como Thr14 e Ser16, são também afetadas pela mutação, mostrando que a mutação Tyr73Ala induz perturbações não-locais na EXLX1.



Figura 7.11. Flutuações estrutura do mutante Tyr73Ala. (a) Rmsf da EXLX1 nativa e mutante. (b) Estruturas da EXLX1 com os rmsf mapeados numa escala de cor indo de azul (rmsf < 0.3 Å) a vermelho (rmsf > 1.5 Å). A mutação aumenta as flutuações em torno do centro de torção da EXLX1.

Observa-se também que a ligação de hidrogênio entre os resíduos Thr12 e Asp82

- resíduos estritamente conservados entre proteínas estruturalmente relacionadas, incluindo expansinas de plantas, endoglucanases da família GH45 e outras [260] – é regularmente rompida na EXLX1 mutante, mas não na nativa. Como ilustrado na figura 7.12, a distribuição da distância entre os resíduos Thr12 e Asp82 é centrada no valor de 1,8 Å na EXLX1 nativa e bimodal na mutante, com picos em torno de 1,8 e 2,6 Å. Estes resultados indicam que o resíduo Tyr73 participa do mecanismo da EXLX1 indiretamente, já que seu papel parece ser essencialmente manter o centro de torção intacto em sua configuração ativa. O mutante Tyr73Ala provavelmente é deficiente em sua capacidade de interagir com cadeias de glucano devido a seu centro de torção altamente flexível.



Figura 7.12. Distâncias entre os resíduos conservados Asp82 e Thr12 (à esquerda) e suas respectivas distribuições (à direita) na EXLX1 nativa e na mutante Tyr73Ala. A mutação causa tal ligação de hidrogênio conservada ser regularmente rompida.

A razão pela qual o resíduo 73 deve ser ou uma tirosina ou uma fenilalanina é que tais resíduos aromáticos possuem cadeias laterais volumosas que estericamente impedem o domínio D1 de acessar conformações inativas. A cavidade criada ao se substituir a tirosina pela alanina dá espaço para que a proteína execute movimentos de maiores amplitudes, permitindo que estados inativos sejam regularmente visitados. Para ilustrar este ponto, as trajetórias da EXLX1 nativa e mutante foram decompostas em componentes principais (PCA). As flutuações estruturais decompostas de ambas as EXLX1, mostradas na figura 7.13a, mostram que os dois primeiros componentes capturam as maiores diferenças de flutuações. Isso indica que o efeito da mutação aparece nos movimentos funcionais e de alta amplitude da EXLX1. O perfil de energia livre qualitativo¹ ao longo destes dois modos de mais





Figura 7.13. Dinâmica essencial da EXLX1 (PCA). (a) Flutuações quadráticas ao longo dos 20 primeiros modos obtidos por PCA e perfil de energia livre qualitativo ao longo do primeiro (PC1) e segundo (PC2) componentes principais para a EXLX1 (b) nativa e (c) mutante (Tyr73Ala). As principais alterações estruturais induzidas pela mutação Tyr73Ala se manifestam nos movimentos funcionais e de alta amplitude da EXLX1. Enquanto a EXLX1 nativa visita estados conformacionais em torno de um mínimo bem definido, o mutante explora múltiplos subestados em torno de diferentes poços de energia livre.

¹O termo qualitativo significa que o perfil de energia livre apenas revela a existência de diferen-

outros mínimos correspondentes a estados inativos (figura 7.13c).

7.4 Discussão

Nesta parte da tese, reporta-se um conjunto de simulações de dinâmica molecular da expansina EXLX1 da bactéria Bacillus subtilis para entender aspectos estruturais e dinâmicos envolvidos em sua atividade não-hidrolítica. Os resultados são consistentes com o conhecimento experimental disponível até o momento e fornecem informações adicionais que permitem novas interpretações sobre o mecanismo de ação das expansinas. Observa-se que a adsorção da EXLX1 sobre a face hidrofóbica da celulose cristalina é dirigida por interações entre a celulose e resíduos aromáticos presentes no domínio D2, envolvendo a expulsão de moléculas de água das superfícies de contato EXLX1-celulose. Isso é consistente com estudos recentes de calorimetria, que mostraram que, semelhantemente aos CBMs do tipo A, a ligação da EXLX1 à celulose é entropicamente dirigida [275, 276]. Análise das componentes da energia de interação fornece evidência adicional à natureza hidrofóbica das interações EXLX1-celulose, onde somente a componente de van der Waals vindo do domínio D2 descresce durante o processo de adsorção; o papel das interações de van der Waals na interação com a celulose foi também observado para o CBM do tipo A da TrCel7A [271]. Interações entre o domínio D1 e a celulose é de origem eletrostática principalmente, o que sugere que competição com a água poderia ser a razão da incapacidade do domínio D1 de se ligar à celulose isoladamente [264].

Embora as simulações não tenham mostrado alterações estruturais na celulose induzidas pela EXLX1, experimentos de RMN de estado sólido sugerem que cadeias de celulose em contato com a EXLX1 exibem propriedades conformacionais distintas das cadeias de celulose longe da EXLX1 [268], indicando que a EXLX1 de fato induz alterações na estrutura da celulose numa escala de tempo maior que

tes mínimos de energia, e não os valores absolutos da profundidades desses mínimos. Simulações muito mais longas deveriam ser executadas para obter o perfil de energia livre quantitativo.

a que é tipicamente acessada via simulações. Para explorar as interações EXLX1celulose – e contornar a questão da escala de tempo –, foram realizadas simulações da EXLX1 ligada a uma cadeia livre de oligossacarídeo. As simulações mostraram que enquanto o domínio D2 segura o substrato numa conformação planar, o domínio D1 induz uma torção de 90° ao longo da cadeia. Propõe-se aqui que esta torção seja essencial para o mecanismo de ação da EXLX1 na superfície da celulose. Ao induzir torções em cadeias individuais da superfície da celulose cristalina, a expansina seria capaz de romper ligações de hidrogênio intermoleculares da celulose, o que explica seus efeitos de potencialização de celulases [72]. Adicionalmente, a EXLX1 poderia, através de torções, romper ligações de hidrogênio entre a celulose e cadeias de xiloglucano fortemente adsorvidas sobre a superfície da celulose, já que regiões da celulose contendo xiloglucanos adsorvidos constituem o alvo da EXLX1 e de expansinas de plantas nas paredes celulares [248, 268]. Tal capacidade de quebrar ligações de hidrogênio já havia sido proposta há vários anos atras [247], mas nenhuma evidência experimental ou computacional disso havia sido reportada até o momento.

A observação de que a EXLX1 pode transladar sobre a superfície da celulose através de contatos hidrofóbicos do domínio D2, tomada em conjunto com a capacidade do domínio D1 de induzir torções nas cadeias de glucano, fornece um possível modo de atuação da EXLX1, no qual a expansina poderia transladar ao longo do cristal de celulose causando torções locais em vários pontos da superfície da celulose e em cadeias de xiloglucano adsorvidas, assim enfraquecendo da estrutura da parede celular. Nesta perspectiva, o domínio D2 parece ter não somente função de ligação ao substrato, como previamente estabelecido [264], mas também um papel ativo na atividade disruptiva da expansina. De fato, Georgelis *et al.* substituíram os três resíduos aromáticos do domínio D2 (Trp125, Trp126 e Tyr157) por alanina e não detectaram atividade de extensão da EXLX1 em paredes celulares, mesmo com altas concentrações da proteína usadas para compensar a redução de afinidade pelo substrato [264].

A função de torção do domínio D1 é mediada por quatro resíduos: Thr14, Ser16, Asp71 e Asp82 (figura 7.8a). A substituição destes resíduos por alanina causa a redução da atividade da EXLX1 em paredes celulares [264]. Entre estes resíduos, o Asp82 possui a maior capacidade de fazer ligações de hidrogênio com o substrato, o que é associado com sua capacidade de segurar o substrato torcionado mais efetivamente. Assim, o resíduo Asp82 funciona como o centro de torção do substrato e há evidências de que seja essencial para a atividade da EXLX1. Curiosamente, a substituição dos resíduos Thr14, Ser16 e Asp71 por alanina resulta na redução da atividade, mas não na inativação completa da EXLX1. Por outro lado, a substituição do Asp82 por alanina ou asparagina inativa a EXLX1 completamente [264]. Estas observações se correlacionam com a estabilidade das ligações de hidrogênio que estes resíduos fazem com o substrato: Asp82 se envolve em uma ou duas ligações de hidrogênio em aproximadamente 90% do tempo, enquanto os outros três resíduos não são tão efetivos em fazer ligações de hidrogênio com o substrato (com frequência, moléculas de água são vistas mediando as ligações de hidrogênio). Assim, os resultados suportam a conclusão de que os resíduos Thr14, Ser16 e Asp81 são importantes para a atividade da EXLX1, mas somente o resíduo Asp82 é essencial.

Se o resíduo Asp82 tem a função de fazer ligações de hidrogênio efetivas com o substrato, como as simulações sugerem, então é trivial entender por que o mutante Asp82Ala é inativo. Menos óbvio é o fato de que a substituição Asp82Asn também inativa a EXLX1, já que ambos os resíduos Asp e Asn podem se envolver em ligações de hidrogênio. O tipo de resultado observado para o mutante Asp82Asn é típico de enzimas glicosídeo hidrolases, uma vez que o resíduo de Asp participaria de um mecanismo de catálise como ácido/base geral ou nucleófilo [277]. Entretanto, vários estudos mostraram que expansinas não clivam nenhum dos componentes da parede celular. Durante o curso da simulação, observou-se que o anel aromático da Tyr73 se volta para o grupo amida planar da Asn82, sugerindo que a inatividade do mutante Asp82Asn é devido a alterações conformacionais que inserem o resíduo Asp82 debaixo da cadeia lateral da Tyr73, tornando-o indisponível ao substrato. Esta observação poderia ser testada experimentalmente acoplando a mutação Asp82Asn com uma substituição da Tyr73 por outro resíduo volumoso, mas não aromático, como Asp82Asn/Tyr73Leu ou Asp82Asn/Tyr73Met. Resultados preliminares de uma simulação do duplo mutante Asp82Asn/Tyr73Leu mostram que, embora a Asn82 seja mais móvel no mutante, a cadeia lateral da Leu73 não possui tamanho suficiente para recobri-lo. Assim, este duplo mutante deve exibir alguma atividade residual pelo fato da Asn82 estar disponível para interagir com o substrato. Experimentos com estes mutantes poderiam corroborar o papel não-hidrolítico do resíduo Asp82 e sua importância em romper ligações de hidrogênio dentro da parede celular.

De acordo com as simulações da EXLX1 complexada com uma cadeia oligossacarídica, o resíduo Tyr73 não participa diretamente das interações entre o domínio D1 e o substrato (figura 7.7a). Entretanto, a substituição de tal resíduo por alanina dramaticamente reduz a atividade da EXLX1 [264]. Diante das ideias sugeridas previamente pelas simulações, não é óbvio por que a Tyr73 é necessária para a atividade da EXLX1, já que este resíduo não está diretamente associado à função de torção do domínio D1. A dinâmica do mutante Tyr73Ala sugere que a falta de atividade está relacionada a perturbações da dinâmica estrutural de resíduos funcionais do domínio D1. A mobilidade da região em torno do centro de torção aumenta na EXLX1 mutante, sugerindo que a estabilidade das interações entre resíduos do domínio D1 e o substrato seria significativamente afetada. A razão dessa perturbação é o espaço extra liberado quando se insere a alanina no lugar do resíduo volumoso Tyr73, o que permite a proteína visitar conformações inativas com frequência, desestabilizando a estrutura biologicamente ativa. O papel puramente estérico da Tyr73 é suportado pelo fato do mutante Tyr73Phe não apresentar alterações de atividade quando comparado à EXLX1 nativa. Portanto, a integridade estrutural dos loops do domínio D1 parece ser essencial para a atividade da EXLX1.

Finalmente, a figura 7.14 mostra um modelo do mecanismo de ação da EXLX1 na parede celular. Este modelo é baseado nas seguintes informações: (1) o alvo da EXLX1 são domínios celulósicos enriquecidos com xiloglucano [248, 268]; (2) a superfície aromática do domínio D2 possui afinidade pela celulose cristalina [264, 275] e desliza sobre ela; (3) a hemicelulose (xiloglucano) se liga à superfície básica do domínio D2 (figura 7.2) [264]; e (4) o domínio D1 induz torções em cadeias de celulose e de xiloglucano sobre a celulose. Propõe-se que a atividade da EXLX1 na parede celular envolva a substituição de ligações de hidrogênio hemicelulose-celulose por ligações hemicelulose-EXLX1-celulose. Para isso, o domínio D1 promoveria o enfraquecimento de ligações de hidrogênio celulose-celulose, ao mesmo tempo que romperia os contatos hemicelulose-celulose, liberando assim as cadeias de hemicelulose para interagir com a superfície básica do domínio D2. Nessa arquitetura mediada pela EXLX1, as tensões da parede celular seriam aliviadas pelos contatos aromáticos entre o domínio D2 e a celulose, de forma que a EXLX1 funcionaria como uma graxa molecular. Estudos adicionais serão necessários para refinar este modelo.

7.5 Considerações finais

Expansinas estimulam a extensão da parede celular e enfraquecem a estrutura da celulose por um mecanismo ainda não completamente entendido. Usando simulações de dinâmica molecular, as interações entre a expansina EXLX1 de Ba*cillus subtilis* e celulose foram investigadas em detalhe. Enquanto observou-se que a EXLX1 pode transladar sobre a superfície da celulose cristalina através da superfície aromática do domínio D2, interações entre a proteína com uma cadeia oligossacarídica individual indicaram que a EXLX1 induz uma torção no substrato através do domínio D1. Tais resultados sugerem que os efeitos de enfraquecimento da celulose e da parede celular emergem de perturbações conformacionais induzidas pela EXLX1 ao longo da superfície da celulose, o que resulta no rompimento de ligações de hidrogênio celulose-celulose e celulose-hemicelulose. Os resultados também apontam que a função de torção é primariamente mediada pelo resíduo Asp82 do domínio D1. A observação experimental de que os mutantes Asp82Asn e Tyr73Ala são inativos foram associadas a perturbações estruturais em torno do centro de torção da EXLX1. Em resumo, estes resultados sugerem uma função, até então desconhecida, para o domínio D1 e indica que o domínio D2 participa de eventos além de simplesmente se ligar à celulose. As informações moleculares obtidas aqui possuem implicações para a compreensão dos princípios envolvidos na extensão da parede celular de plantas e dos efeitos sinergísticos observados


Figura 7.14. Mecanismo proposto para a ação da EXLX1 na parede celular primária. (a) Modelo de parede celular na ausência de expansina, onde a hemicelulose se liga às fibrilas de celulose através de ligações de hidrogênio. Sob estresse externo, as ligações de hidrogênio entre a celulose e hemicelulose causam um aumento de tensão na parede celular. (b) Modelo de parede celular na presença de expansina, onde a expansina se liga aos domínios de celulose ricos em hemicelulose (xiloglucano) e media os contatos hemicelulose-celulose. Uma vez que a expansina pode transladar sobre a superfície da celulose, as forças externas aplicadas à parede celular seriam aliviadas pelos contatos livre de atrito entre a superfície aromática da expansina e a celulose.

quando expansinas são misturadas com celulases para a degradação de biomassa lignocelulósica. Do ponto de vista tecnológico, o entendimento dos mecanismos das expansinas se faz necessário para a incorporação de funções disruptivas nas celulases através de engenharia de enzimas.

Parte IV

Aspectos moleculares de paredes celulares de plantas

Capítulo 8

Nanoarquitetura da parede celular primária

8.1 Introdução

Como mencionado anteriormente nesta tese, a conversão de biomassa em açúcares fermentáveis requer a desconstrução da arquitetura nativa da parede celular das plantas [4]. Durante o crescimento das células vegetais, as paredes celulares são formadas por fibrilas de celulose imersas numa matriz composta por hemicelulose e pectina. Esta parede celular, isenta de lignina, é conhecida como parede celular primária [4, 5]. Quando o crescimento da célula cessa, ocorre a deposição de lignina na matriz hemicelulósica, tornando-a mecanicamente resistente, que passa então a se denominar parede celular secundária [26]. Esta matriz de hemicelulose e lignina é responsável pelas forças de coesão internas que mantêm a arquitetura parede celular e, portanto, constitui parte da origem da recalcitrância [2]. Entender os aspectos moleculares das interações entre celulose, hemicelulose e lignina é a base para o entendimento da recalcitrância lignocelulósica.

E um fato conhecido que a recalcitrância da parede celular varia entre espécies de plantas e mesmo entre diferentes fenótipos da mesma espécie [278]. Esta variação é devido a diferenças associadas à organização nanoestrutural e à composição química da parede celular [22]. A arquitetura da parede celular determina suas propriedades mecânicas, necessárias para suportar pressões de turgor originadas no interior das células, e regula a extensão da parede celular durante o crescimento da célula [269, 279–281]. Ao mesmo tempo, o processo de degradação de biomassa lignocelulósica (pré-tratamentos e hidrólise enzimática) é sensível à composição química e à organização de seus componentes, a qual determina a acessibilidade da celulose às enzimas [7, 36, 282–284]. Isso mostra que modulações genéticas poderiam ser utilizadas racionalmente para ajustar a composição química da parede celular e, assim, ganhar controle sobre a recalcitrância. Embora análises bioquímicas e sequenciamento genético tenham permitido o estudo detalhado da composição química e rotas de síntese de componentes da parede celular [13, 20, 26], o entendimento de como estes componentes e variações em suas composições influenciam as propriedades da parede celular ainda é limitado [279, 285]. Obviamente, intervenções externas na maquinaria biossintética da parede celular podem acarretar consequências prejudiciais para o crescimento e desenvolvimento da planta [21, 52]. Logo, entender a origem das forças estruturais atuantes na parede celular constitui um primeiro passo para estabelecer a relação entre a composição química e a nanoestrutura da biomassa.

Neste capítulo, a teoria 3D-RISM/KH foi aplicada a um modelo de parede celular primária para obter as forças termodinâmicas que estabelecem a arquitetura das fibrilas de celulose [128]. Em particular, foi investigado o efeito da hemicelulose e variações em sua composição química nas forças efetivas entre nanofibrilas de celulose. A escolha da teoria 3D-RISM/KH, em vez de simulações de dinâmica molecular, deve-se ao fato de que a primeira metodologia fornece, através de expressões analíticas, as médias de *ensemble* necessárias para a obtenção das forças moleculares atuando na parede celular. Tais forças, por possuírem natureza estatística, necessitariam de simulações demasiadamente longas para serem obtidas.

8.2 Metodologia

8.2.1 Modelos moleculares e procedimentos gerais

Por ser uma teoria molecular de solvatação, a aplicação da teoria 3D-RISM requer que o sistema seja subdividido em soluto – entidade de estrutura fixa que cria um campo de potencial no espaço – e solvente – tratado por meio de funções de distribuição, obtidas através de médias sobre um *ensemble* de todas as suas possíveis configurações espaciais e orientacionais. Assim, a primeira etapa da aplicação deste método é a definição do soluto e do solvente. O objetivo deste trabalho foi a obtenção das forças efetivas entre diferentes faces (hidrofóbica/hidrofílica) de fibrilas elementares de celulose na presença de hemiceluloses de diferentes composições químicas. Tais forças efetivas incluem interações diretas (calculadas usando um campo de força) e efeitos do meio (forças de solvatação obtidas com 3D-RISM). Dessa forma, a celulose foi tratada como soluto e a hemicelulose/água como solvente, de maneira que os efeitos da matriz hidratada de hemicelulose fossem computados em termos de suas capacidades de solvatar a celulose.

Embora questões ligadas à falta de amostragem não apareçam na teoria 3D-RISM – fato que a qualifica como um método adequado para o tipo de problema abordado aqui –, o método, como formulado no Capítulo 4, permite lidar apenas com estruturas rígidas, o que impede sua aplicação a sistemas muito flexíveis, que é o caso da hemicelulose. Tratar um polímero de hemicelulose como solvente em 3D-RISM significa tratá-lo como um objeto inteiramente rígido e obter suas propriedades termodinâmicas através de médias sobre o *ensemble* de todas as suas posições e orientações no espaço. Além de ser fisicamente não-realístico, é fácil perceber que este sistema não formaria uma solução aquosa homogênea. Por ser uma teoria de estado líquido, as equações da teoria DRISM não convergiriam para fornecer um conjunto de funções de correlação para a hemicelulose. Assim, para aplicar 3D-RISM, a estrutura da hemicelulose foi simplificada de forma a transformar a matriz hemicelulósica da parede celular numa solução aquosa hemicelulósica. Com isso, os efeitos da hemicelulose nas forças efetivas entre fibrilas de celulose puderam ser estudadas.

Considerando que as alterações químicas da hemicelulose ocorrem em suas ramificações, contruiu-se um modelo de hemicelulose que consiste de monômeros dessas ramificações em solução. Tal modelo captura as especificidades químicas das interações das ramificações de hemicelulose com a celulose. O modelo foi baseado no xilano, que é a hemicelulose mais abundante em monocotiledôneas [7, 20], grupo que inclui plantas de interesse biotecnológico, como cana-de-açúcar e milho. O xilano (por vezes denominado glucuroarabinoxilano) consiste de uma cadeia principal de xilose com ramificações de ácido glucurônico/glucuronato e arabinose, conforme ilustrado na figura 2.5. Tanto a quantidade dessas ramificações quanto a proporção relativa dos diferentes tipos de ramificações variam com o fenótipo da planta [21, 51]. Diante disso, o modelo de hemicelulose consistiu de monômeros de arabinose e ácido glucurônico/gluronato em solução aquosa. A figura 8.1 mostra as estruturas desses monômeros. A arabinose foi considerada como furanose (anel de cinco membros) por ser esta a sua configuração mais frequente em hemiceluloses. Íons hidrônio (H_3O^+) foram utilizados para neutralizar as soluções de glucuronato. Íons hidrônio, e não, por exemplo, íons Ca⁺⁺, presentes na parede celular, foram empregados para que a densidade das soluções de glucuronato fosse igual à densidade das soluções de ácido glucurônico, para facilitar os estudos comparativos.



Figura 8.1. Ramificações de xilano: arabinose, ácido glucurônico e glururonato.

A ideia original deste trabalho era utilizar as fibrilas elementares de celulose de 36 cadeias, como ilustrado na figura 2.4. Embora, para solventes simples (como água pura), a teoria 3D-RISM/KH permita a obtenção da termodinâmica e estrutura de solvatação em torno de solutos grandes em tempos relativamente curtos quando comparada com simulações de dinâmica molecular [286], o custo computacional (memória RAM e tempo de CPU) de 3D-RISM/KH aumenta muito rapidamente com a complexidade do solvente e com o número de átomos do soluto. (Por complexidade, entende-se o número de espécies químicas presentes no solvente e o número de átomos de cada uma dessas espécies.) As soluções aquosas de monômeros de hemicelulose apresentaram um custo computacional proibitivo ao se tentar realizar os cálculos com as fibrilas elementares de celulose de 36 cadeias. Diante disso, como não seria conveniente simplificar as soluções de hemicelulose, um modelo simplificado de fibrila de celulose foi utilizado, denotado aqui por nanofibrila de celulose (figura 8.2). O modelo consistiu em 4 cadeias de glucano de grau de polimerização 8 e empacotadas conforme a estrutura cristalográfica da celulose I β [9]. A nanofibrila foi construída de forma tal que possuísse tanto faces hidrofóbicas quanto faces hidrofílicas expostas, conforme ilustrado na figura 8.2.



Figura 8.2. Estrutura da nanofibrila de celulose formada por 4 cadeias de glucano de grau de polimerização 8 e com ambas as faces hidrofílicas e hidrofóbicas expostas. As diferentes cores discriminam as diferentes cadeias de glucano.

Assim, o modelo de parede celular primária utilizado consistiu de duas nanofibrilas de celulose imersas numa solução aquosa de monômeros de ramificações de glucuroarabinoxilano (figura 8.3). As forças efetivas entre as nanofibrilas foram computadas como o potencial de força média (PMF), ou perfil de energia livre, ao longo de um caminho de reação que corresponde ao rompimento de duas arquiteturas distintas das nanofibrilas: (a) rompimento do contato das faces hidrofílicas (figura 8.4a) e (b) das faces hidrofóbicas (figura 8.4b). O PMF foi obtido através da expressão [114]

$$PMF(d) = [\mu(d) + U(d)] - [\mu(\infty) + U(\infty)], \qquad (8.1)$$



Figura 8.3. Modelo de parede celular primária. O modelo consiste de duas nanofibrilas de celulose imersas numa solução aquosa de monômeros de hemicelulose.



Figura 8.4. Caminhos de desagregação ao longo do qual o PMF foi computado. Em (a) contatos hidrofílicos das fibrilas são rompidos e, em (b), contatos hidrofóbicos.

onde $\mu(d)$ é a energia livre de solvatação das nanofibrilas de celulose separadas pela distância d na solução de hemicelulose, e U(d) é a energia potencial de interação entre as duas nanofibrilas de celulose, que inclui termos de interação eletrostática e de van der Waals. Os termos $\mu(\infty)$ e $U(\infty)$ correspondem, respectivamente, à energia livre de solvatação e energia de interação das nanofibrilas infinitamente separadas; isso significa que $\mu(\infty) = 2\mu_0$ e $U(\infty) = 0$, onde μ_0 a energia livre de solvatação de uma única nanofibrila. O PMF foi computado com as nanofibrilas imersas em água pura e em soluções de hemicelulose em concentrações crescentes. Em síntese, o procedimento empregado para o cálculo dos PMFs foi:

- Obtenção das funções de correlação radiais solvente-solvente (suscetibilidade) para todos os pares de sítios das soluções aquosas de hemicelulose através da teoria DRISM/KH; estes cálculos são realizados sem a celulose, para a obtenção de propriedades intrínsecas do *bulk* das soluções;
- Para cada configuração relativa (distância d) das nanofibrilas de celulose (cálculos single point), obtenção das funções de correlação tridimensionais solutosolvente via 3D-RISM/KH e das respectivas energias livres de solvatação μ(d); isso foi feito separadamente para cada uma das soluções aquosas de hemicelulose;
- 3. Obtenção dos PMFs (eq. 8.1) em função da separação das nanofibrilas;

As interações específicas de cada tipo de hemicelulose com a celulose foram analisadas com base na função de distribuição tridimensional da hemicelulose em torno da celulose (densidade local da hemicelulose) e com base na densidade de energia livre de solvatação (SFED: *solvation free energy density*), que mostra as regiões espaciais em torno do soluto que concentram interações (incluindo componentes entálpica e entrópica) que contribuem mais ou menos para a solvatação.

8.2.2 Detalhes computacionais

As equações integrais da teoria DRISM/KH foram resolvidas para obter as funções de correlação radiais solvente-solvente para as soluções aquosas de arabinose, ácido glucurônico, glucuronato e acetato nas concentrações (expressas em fração molar do soluto) de 0,02, 0,04, 0,06, 0,08 e 0,10. (Soluções de acetato foram utilizadas para responder a questões adicionais, conforme descrito na seção de resultados.) As equações foram discretizadas numa malha radial uniforme com 8192 pontos

separados entre si pela distância de 0,1 Å, e convergidas usando o algoritmo MDIIS [114, 287] até o desvio médio quadrático relativo de 10^{-10} . Íons hidrônio foram adicionados às soluções de glucuronato e acetato para manter a eletroneutralidade. A constante dielétrica das soluções aquosas foram consideradas como igual da água pura ($\epsilon = 78, 497$). As densidades foram obtidas por simulações de dinâmica molecular em pressão constante (descrito abaixo).

Para computar o PMF entre as nanofibrilas de celulose, as equações integrais da teoria 3D-RISM/KH foram resolvidas para distâncias crescentes entre as nanofibrilas imersas nas diferentes soluções, com intervalo de 0,2 Å. As nanofibrilas (figura 8.2) foram construídas utilizando o programa Cellulose-Builder [270]. As equações integrais 3D-RISM/KH foram discretizadas em uma malha uniforme de dimensões 64 Å × 64 Å × 64 Å e contendo 128 × 256 × 128 pontos. Uma malha refinada por um fator de 2 não alterou os resultados. O algoritmo MDIIS [114, 287] foi utilizado para convergir as equações até o desvio médio quadrático relativo de 10^{-4} .

Para obter a densidade das soluções aquosas (em T = 298 K e p = 1 bar), foram executadas simulações de dinâmica molecular em pressão e temperatura constantes para as diferentes soluções nas concentrações correspondentes. As caixas de simulação foram construídas com o programa Packmol [237], contendo aproximadamente 20.000 átomos. As simulações foram conduzidas utilizando condições de contorno periódicas com potenciais de curto alcance truncados num raio de corde de 12 Å, e interações eletrostáticas foram computadas utilizando *particle mesh* Ewald [105]. Com ligações químicas envolvendo átomos de hidrogênio vinculadas em seus comprimentos de equilíbrio, as simulações foram propagadas numericamente utilizando um passo de integração de 2 fs durante 1 ns, que foi suficiente para que as densidades alcançassem um valor médio constante. O programa AM-BER11 [127] foi usado para realizar as simulações. A densidade das soluções de glucuronato e acetato foram consideradas como sendo as mesmas das soluções de ácido glucurônico e ácido acético, respectivamente. As densidades obtidas estão compiladas na tabela 8.1.

As estruturas dos monômeros foram obtidas da biblioteca de oligossacarídeos

	Fração molar				
monômero	$0,\!02$	0,04	0,06	0,08	$0,\!10$
arabinose ácido glucurônico ácido acético	$1,06 \\ 1,09 \\ 1,02$	$1,09 \\ 1,14 \\ 1,02$	$1,12 \\ 1,19 \\ 1,02$	$1,15 \\ 1,23 \\ 1,02$	$1,17 \\ 1,26 \\ 1,02$

Tabela 8.1. Densidade (em g cm^{-3}) das soluções aquosas de arabinose, ácido glucurônico e ácido acético em diferentes concentrações.

Glycam [288], e os parâmetros do potencial de interação foram tomados do campo de força CHARMM36. O modelo TIP3P/CHARMM foi usado para a água e os íons hidrônio foram modelados com os parâmetros disponíveis nas referências [121, 289]. Os arquivos no formato CHARMM foram convertidos para o formato AMBER, programa utilizado para as simulações de dinâmica molecular, utilizando a ferramente CHAMBER [290]. Todos os cálculos de equações integrais foram executados com programas desenvolvidos no grupo do Dr. Kovalenko.

8.3 Resultados

8.3.1 Auto-agregação de nanofibrilas de celulose

Os PMFs – computados ao longo dos caminhos de desagregação das faces hidrofílica (figura 8.4a) e hidrofóbica (figura 8.4b) das nanofibrilas de celulose – indicam que, em água pura, as nanofibrilas se agregam preferencialmente pelas suas faces hidrofílicas. Para ambos os casos, os PMFs exibem dois mínimos bem definidos (figura 8.5). Configurações das nanofibrilas correspondentes a estes dois mínimos e ao máximo da barreira de energia livre que os separa são mostradas nas figuras 8.5b e 8.5c, onde também são mostradas superfícies de contorno das funções de distribuição $g(\mathbf{r})$ dos átomos de oxigênio da água. O primeiro mínimo, em 0 Å, corresponde às nanofibrilas agregadas em contato direto, e o segundo, em ≈ 3 Å, corresponde às nanofibrilas separadas por uma camada de água. O primeiro mínimo é ≈ 60 kcal/mol mais estável para a agregação das faces hidrofílicas em relação à agregação hidrofóbica. Este resultado está associado à formação



Figura 8.5. (a) Potencial de força média de desagregação das nanofibrilas de celulose ao longo do caminho de desagregação das faces hidrofílicas e hidrofóbicas. (b) e (c): superfície de contorno da função de distribuição $g(\mathbf{r})$ dos átomos de oxigênio da água ao longo da desagregação (b) hidrofílica e (c) hidrofóbica das nanofibrilas. As superfícies de contorno correspondem a $g(\mathbf{r}) = 2, 5$.

de ligações de hidrogênio interfibrilas na agregação hidrofílica. Por outro lado, a configuração em que uma camada de solvente separa as nanofibrilas fornece o mínimo global do arranjo via faces hidrofóbicas, mostrando que o contato direto das faces hidrofóbicas é uma configuração metaestável. Estes resultados sugerem que, em processos não-enzimáticos que alteram a arquitetura da parede celular (pré-tratamentos e crescimento da planta) contatos hidrofóbicos entre fibrilas de celulose são mais facilmente rompidos que os hidrofílicos. Ou vice-versa: em processos que causam agregação, esta ocorrerá preferencialmente através das faces hidrofílicas.

O fato da agregação hidrofóbica ser metaestável indica que a barreira de energia livre que separa os dois mínimos é importante para manter a arquitetura das macrofibrilas (controle cinético). Embora o PMF acuse a existência de uma alta barreira de energia livre separando os mínimos, esta deve ser interpretata apenas qualitativamente. O caminho de desagregação (figura 8.4) ao qual corresponde a barreira computada não é necessariamente o mais favorável. Caminhos alternativos, como o rompimento progressivo de contatos celulose-celulose, semelhantemente à abertura de um zíper, forneceria menores penalidades energéticas para a desagregação.

Embora o senso comum induza interpretar a barreira de energia livre como sendo devido à formação de um vácuo no espaço interfibrilar, a análise da função de distribuição dos átomos de oxigênio mostra que há uma camada de água separando as fibrilas no máximo do PMF (figuras 8.5b e 8.5c). Esta camada de água, no entanto, é confinada em pequenas cavidades existentes no espaço interfibrilar, sem acesso ao meio externo às interfaces, sugerindo que penalidades entrópicas associadas a estas moléculas de água contribuem para a barreira de energia livre. Isso implica que processos que induzem a agregação e a desagregação de fibrilas de celulose não necessariamente induzem uma interconversão entre dois estados idênticos. No processo de desagregação, partindo de fibrilas em contato cristalino ou quase-cristalino, a água não teria acesso ao espaço interfibrilar e o máximo corresponderia a um vácuo perfeito. No processo de agregação, partindo das fibrilas completamente solvatadas, a água tenderia a permanecer solvatando as fibrilas à medida em que elas fossem aproximadas, até alcançar o máximo. Como a expulsão dessas moléculas de água seria impedida estericamente, estados de agregação imperfeitos seriam atingidos.

8.3.2 Papel da hemicelulose nas forças de agregação

Para estudar o papel da hemicelulose nas forças moleculares que determinam a arquitetura da parede celular, o PMF de desagregação de nanofibrilas de celulose foi computado em soluções aquosas de arabinose, ácido glucurônico e glucuronato, em concentrações crescentes. Os PMFs são mostrados na figura 8.6. Os resultados mostram que a estabilidade das arquiteturas descritas acima, para nanofibrilas em água, é fortemente afetada pela presença de hemicelulose. Em relação aos PMFs em água pura, as formas dos PMFs nas soluções de hemicelulose são mantidas à medida que os monômeros são adicionados ao meio, independente de suas naturezas químicas. À medida que a concentração (fração molar x) de hemicelulose aumenta de x = 0,0 a x = 0,1, os PMFs associados ao primeiro mínimo e ao primeiro máximo diminuem, na mesma proporção. Por outro lado, a sensibilidade do segundo mínimo com a concentração de hemicelulose é mais fraca. Como consequência, o efeito dominante da hemicelulose é a estabilização do estado agregado das nanofibrilas.

As figuras 8.7a e 8.7b mostram a dependência da energia livre de agregação $[\Delta G_{agr} = \text{PMF}(d = 0)]$ com a composição química e concentração da hemicelulose. O íon glucuronato apresenta o maior efeito na estabilização do estado agregado, aproximadamente duas vezes maior que os efeitos da arabinose e ácido glucurônico. No caso da agregação das faces hidrofóbicas, que é metaestável em água pura, observa-se que ΔG_{agr} atinge valores próximos de zero à medida que a concentração de glucuronato aumenta. Assim, embora as nanofibrilas de celulose preferencialmente se agreguem através de suas faces hidrofílicas, a presenca de hemicelulose, contendo glucuronato, pode facilitar, do ponto de vista termodinâmico, a agregação através das faces hidrofóbicas. Além disso, observa-se que para concentrações de glucuronato maiores que $x = 0, 1, \Delta G_{agr}$ mudaria de sinal, indicando que a estabilidade relativa dos dois mínimos do PMF ao longo da desagregação hidrofóbica pode alterar em função da presença de hemicelulose.

É possível que o efeito estabilizador pronunciado das soluções de glucuronato tenha duas origens: interações das nanofibrilas de celulose com os grupos carboxilato ou com os contra-íons hidrônio. Para verificar estas possibilidades, cálculos adicionais foram feitos considerando as nanofibrilas imersas em soluções de acetato e hidrônio. Os resultados mostram que, para soluções contendo a mesma concentração de hidrônio, o abaixamento de ΔG_{agr} é muito maior na presença de glucuronato que na presença de acetato (figura 8.7). Assim, se o abaixamento de ΔG_{agr} fosse devido ao íon hidrônio, ambas as soluções de glucuronato e acetato resultariam no mesmo efeito. Logo, o ânion glucuronato é de fato responsável pela estabilização das nanofibrilas agregadas. Comparando os valores de ΔG_{agr} das soluções de acetato com os valores das soluções de arabinose, observa-se que a soma de seus



Figura 8.6. Potencial de força média de desagregação hidrofílica e hidrofóbica das nanofibrilas de celulose em soluções de (a) arabinose, (b) ácido glucurônico e (c) glucuronato. As curvas de diferentes cores correspondem a concentrações crescentes de hemicelulose, variando da fração molar x = 0 a x = 0, 1. Em todos os casos, o efeito da hemicelulose é a estabilização do estado agregado das nanofibrilas de celulose.



Figura 8.7. Efeito da natureza química e da concentração das ramificações de hemicelulose na energia livre de agregação (ΔG_{agr}) das faces (a) hidrofóbicas e (b) hidrofílicas da nanofibrilas de celulose. A linha pontilhada cinza corresponde à soma dos ΔG_{agr} para arabinose e acetato, conforme detalhado no texto.

respectivos ΔG_{agr}^{1} é muito próxima do ΔG_{agr} do glucuronato (também, aproximadamente, considerando ácido glucurônico em vez de arabinose). Isso indica que os efeitos devido ao anel glucosídico e ao grupo carboxilato são aproximadamente aditivos no que concerne à estabilização da parede celular. Logo, os resultados sugerem que hemiceluloses ricas em ramificações de ácido glucurônico/glucuronato devem ter efeitos mais pronunciados na arquitetura da parede celular que as hemiceluloses ricas em ramificações de arabinose.

8.3.3 Especificidade das interações celulose-hemicelulose

Os modos de interação das ramificações de hemicelulose com as nanofibrilas de celulose foram obtidos em termos de mapas espaciais tridimensionais (superfície de contorno) das funções de distribuição dos átomos de carbono da hemicelulose (figura 8.8). Estes mapas indicam regiões em torno das nanofibrilas onde a densidade local da hemicelulose é maior que no seio da solução; logo, mostram como a hemicelulose se estrutura em torno da celulose.

Os monômeros eletricamente neutros (arabinose e ácido glucurônico) interagem com a face hidrofílica da celulose principalmente como doadores de ligação de hi-

¹Nesta soma, o valor do ΔG_{agr} para x = 0 foi descontado, para não ser contado duas vezes.



Figura 8.8. Mapas da função de distribuição tridimensional $g(\mathbf{r})$ para os átomos de carbono da (a) arabinose, (b) ácido glucurônico, (c) glucuronato e (d) acetato em torno das faces hidrofílicas e hidrofóbicas das nanofibrilas de celulose. O esquema de cores é indicado pelas esferas coloridas das estruturas à direita. Os valores de $g(\mathbf{r})$ foram escolhidos para clara visualização dos modos de interação hemicelulose-celulose e correspondem a $g(\mathbf{r}) = 1, 4$ para arabinose, ácido glucurônico e acetato, e $g(\mathbf{r}) = 2, 0$ para glucuronato.

drogênio, coordenando os átomos de oxigênio disponíveis na superfície da celulose através de suas hidroxilas exocíclicas. Os ânions de glucuronato, por outro lado, funcionam como aceptores de ligação de hidrogênio que coordenam as hidroxilas exocíclicas da superfície da celulose. Os ânions acetato se distribuem sobre a face hidrofílica da celulose de forma similar ao glucuronato. Tais resultados indicam que as ramificações da hemicelulose interagem de uma forma complementar com a celulose, em que cada um dos monômeros age como doador ou aceptor de ligação de hidrogênio sobre a celulose. A orientação preferencial dos monômeros de hemicelulose sobre a superfície hidrofóbica da celulose é tambem determinada pelas interações polares entre os grupos exocíclicos da hemicelulose com a superfície da celulose. A interação dos grupos endocíclicos é menos específica, o que torna sua distribuição local pouco evidente através das funções de distribuição.

As funções de distribuição são úteis para fornecer a densidade local de cada sítio do solvente separadamente e, assim, dar informações sobre modos de ligação. Sítios atômicos que não se estruturam em torno do soluto, como é o caso de átomos de carbonos endocíclicos da hemicelulose sobre as nanofibrilas, são menos evidentes nas análises da estrutura de solvatação. Isso significa que, enquanto interações polares fornecem picos de $g(\mathbf{r})$ localizados e altos, as interações hidrofóbicas nãodirecionais estão associadas com picos de $g(\mathbf{r})$ baixos (próximos de 1) e largos. O número de coordenação, obtido com a integral de $g(\mathbf{r})$, deveria fornecer valores similares para sítios associados a interações hidrofóbicas e hidrofóbicas.

Alternativamente, para obter informações sobre as intensidades relativas das interações hidrofílicas e hidrofóbicas entre a hemicelulose e as nanofibrilas, superfícies de contorno de densidade de energia livre de solvatação (SFED) foram construídas. Diferentemente das funções de distribuição (e dos números de coordenação associados a elas), que são atribuídas a cada um dos sítios do solvente individualmente, a SFED pode ser somada sobre todos os sítios de espécies do solvente, conforme a equação 4.27. Assim, a SFED fornece as regiões espaciais associadas a interações de uma determinada espécie química que contribuem em maior ou menor grau para a energia livre de solvatação, independentemente de suas naturezas (hidrofóbica/hidrofílica).

Na figura 8.9 são mostradas superfícies de contorno da SFED associada ao glucuronato sobre a superfície hidrofóbica da celulose. A superfície de contorno da figura 8.9a corresponde a valores negativos relativamente altos da SFED e mostra regiões extremamente localizadas em torno de sítios da celulose capazes de fazer ligações de hidrogênio. Aumentanto-se o valor da SFED para valores negativos mais altos, observa-se uma segunda camada sobre a celulose (figura 8.9b) que se mostra difusa, sugerindo interações carboidrato-carboidrato do tipo *stacking*. Assim, estes mapas mostram que as ligações de hidrogênio são as principais interações entre



Figura 8.9. Mapas da densidade de energia livre de solvatação (SFED) para glucuronato. Os valores mais negativos em (a) correspondem a distribuições localizadas e indicam ligações de hidrogênio; os valores menos negativos em (b) correspondem a distribuições difusas, características de interações hidrofóbicas do tipo *stacking* entre glucuronato e a superfície da celulose.

hemicelulose e celulose, e que as interações hidrofóbicas, embora presentes, são mais fracas. Os mapas de SFED para arabinose e ácido glucurônico são qualitativamente similares, diferindo apenas nos valores da SFED e seguindo a mesma tendência dos valores de ΔG_{agr} .

8.4 Discussão

Neste trabalho, empregou-se a teoria 3D-RISM/KH para estimar as interações efetivas (PMF) entre nanofibrilas de celulose em diferentes arquiteturas e o papel da hemicelulose e sua composição química em tais interações. Os resultados revelam a existência de uma força termodinâmica que leva as nanofibrilas a se agregaram preferencialmente através de suas faces hidrofílicas, o que explica e suporta modelos derivados de microscopia [17, 291], os quais sugerem a existência de agregados macrofibrilares estabelecidos via contatos hidrofílicos em paredes celulares primárias. Além disso, espalhamento de nêutros a baixo ângulo revela um padrão de espalhamento característico de empacotamentos lado a lado de fibrilas elementares de celulose [18], o que também é evidenciado por medidas espectroscópicas [292].

Embora a configuração de contato direto das faces hidrofóbicas seja bem definida por um mínimo no PMF, a configuração em que as fibrilas são separadas por uma camada de solvente constitui mínimo global do PMF. Assim contatos fibrila-fibrila mediados por uma camada de solvente podem ocorrer dentro da parede celular. Este resultado possui implicações para o entendimento das alterações estruturais que acontecem na parede celular primária durante o crescimento da célula. A expansão da parede celular requer que as fibrilas de celulose sejam espaçadas paralelamente entre si para que possam deslizar longitudinalmente sob pressões de turgor [245, 293]. Os mínimos exibidos nos PMF mostram que estes estados estáveis/metaestáveis separados por uma camada de água podem facilitar o deslizamento paralelo das fibrilas de celulose. Experimentos de espalhamento de nêutrons a baixo ângulo sugerem, de fato, a existência tanto de agregados fibrilares em contato direto quanto separados por camadas de hidratação [16, 294].

A hemicelulose estabiliza o estado agregado das nanofibrilas e a intensidade desta estabilização é dependente do tipo de hemicelulose. Em primeiro lugar, este resultado reflete o papel da matriz hemicelulósica em gerar forças de coesão dentro da parede celular, o que é consistente com experimentos em que o bloqueio da síntese de glucuroxilanos em plantas causa uma redução da recalcitrância à hidrólise [295]. Em segundo lugar, os resultados mostram o efeito de variações de composição química da matriz hemicelulósica nas forças internas da parede celular. A importância das ramificações de glucuronato é consistente com experimentos: paredes celulares com hemiceluloses deficientes em glucuronatos são mais suscetíveis à desconstrução que aquelas com alto teor de glucuronato [23].

Em relação a outros tipos de hemicelulose, os resultados permitem fazer predições sobre seus efeitos na parede celular. Como a presença de glucuronatos na hemicelulose é o fator que mais afeta a intensidade de seu efeito, outras hemiceluloses, como galactoglucomananos ou xiloglucanos – compostos de monossacaídeos eletricamente neutros –, fariam interações semelhantes às observadas para arabinose. A presença de glucuronato, como um ânion, parece ser um fator determinante das forças que a matriz hemicelulósica exerce sobre a organização das fibrilas de celulose.

As análises de SFED mostraram que os diferentes monômeros podem se associar à celulose de duas maneiras: através de ligações de hidrogênio, que é o modo mais energeticamente favorável, e através de interações hidrofóbicas do tipo *stacking*, que são mais fracas. Como mostrado pelos mapas das funções de distribuição, as ligações de hidrogênio são específicas para cada tipo de ramificação, e são estabelecidas entre a celulose e as hidroxilas exocíclicas dos monômeros. Experimentos de calorimetria mostraram que a adsorção de xiloglucano sobre a celulose apresenta duas fases em função de sua concentração: uma fase exotérmica, em baixas concentrações, e uma fase endotérmica, em altas concentrações. Consistente com os resultados de 3D-RISM/KH, a fase exotérmica corresponde a formação de ligações de hidrogênio entre xiloglucano e a celulose enquanto que a fase endotérmica corresponde a interações hidrofóbicas termodinamicamente dirigidas por expulsão de moléculas de águas das interfaces de interação entre xiloglucano e celulose [296].

Conforme descrito na seção de métodos, os monômeros de ramificações de hemicelulose foram utilizados como um modelo de hemicelulose que leva em conta sua especificidade química. Adicionalmente, este modelo permite interpretações sobre o estado pré-tratado da parede celular, em que a hemicelulose é hidrolisada a monômeros. Nesta interpretação, os resultados obtidos são também consistentes com experimentos: um aumento da largura das fibras de celulose resulta de prétratamentos ácidos, hidrotérmicos e de explosão de vapor [18, 33, 39, 40, 297]. Isso sugere que a expulsão de água entre fibras de celulose é um processo dirigido não somente pelo aumento da entropia rotacional e translacional da água expelida em altas temperaturas, como mostrado por Langan *et al.* [18], mas também pelos efeitos estabilizadores dos monômeros de hemicelulose, conforme mostrado na figura 8.6.

8.5 Considerações finais

Neste capítulo, a teoria de equações integrais 3D-RISM/KH foi usada para investigar, semiquantitativamente, as interações efetivas entre nanofibrilas de celulose e o impacto da matriz hemicelulósica e sua composição química nas forças de interação. Os resultados mostram que a hemicelulose estabiliza estados agregados da celulose e que a presença de grupos carboxilatos, presentes nas ramificações de glucuronato, potencializa os efeitos da matriz hemicelulósica na nanoarquitetura da parede celular primária.

Estes estudos se inserem num contexto em que faz-se necessário o entendimento de relações entre composição química e nanoarquitetura da parede celular, e entre nanoarquitetura e recalcitrância à degradação. Modificações genéticas de plantas têm mostrado ser possível modular a recalcitrância [298], mas tais estudos são essencialmente empíricos e carecem de uma base fundamentada em aspectos moleculares da parede celular [2]. Os resultados obtidos aqui constituem um dos primeiros esforços para entender como a recalcitrância lignoceluósica é afetada pela composição química da parede celular. No futuro, à medida em que a complexidade da parede celular é compreendida em maiores detalhes, informações moleculares deverão ser utilizada para guiar o avanção da tecnologia de conversão de biomassa.

Capítulo 9

Forças termodinâmicas em paredes celulares secundárias

9.1 Introdução

A lignina é o segundo biopolímero mais abundante do planeta e um componente majoritário da parede celular secundária de plantas [25, 26]. Nas plantas, a lignina está envolvida em regulação da condução de água, suporte estrutural e resistência a patógenos [3, 299]. Tais efeitos protetores e estruturais criam desafios para o processamento industrial da biomassa lignocelulósica, por ser a lignina recalcitrante à degradação química e biológica. Na produção de biocombustíveis, o impacto negativo da lignina é reconhecido como um dos maiores problemas [4], demandando a utilização de pré-tratamentos severos que compromentem a viabilidade econômica de todo o processo [2, 28]. Assim, solucionar o problema da recalcitrância da lignina é uma das necessidades para a produção de biocombustíveis em larga escala para suprir a demanda mundial [29, 57].

Como descrito no Capítulo 2, a lignina é um polímero formado por unidades de fenilpropanóides que se associam formando uma complexa rede tridimensional na parede celular secundária. Juntamente com a hemicelulose, a lignina preenche os espaços entre as microfibrilas de celulose e cria uma barreira resistente que dificulta a digestibilidade da celulose por reduzir a acessibilidade de glicosídeo hidrolases. Além disso, devido à possibilidade de adsorção nos microporos da matriz de lignina, as enzimas se difundem lentamente até o substrato celulósico, prejudicando a cinética do processo [4, 300–302]. De fato, pré-tratamentos que reduzem o conteúdo de lignina – enquanto mantêm a arquitetura nativa da parede celular – aumentam a digestibilidade da biomassa [17, 27]. Pré-tratamentos demasiadamente severos, contudo, podem levar à formação de pseudo-lignina a partir de produtos da degradação de carboidratos e lignina e ser ainda mais prejudicial à hidrólise enzimática que a própria lignina [49, 50]. Assim, pré-tratamentos ideais devem ser moderados e ao mesmo tempo capazes de romper a barreira da lignina para expor a celulose ao ataque enzimático.

Plantas geneticamente modificadas que possuem menor conteúdo e/ou diferentes composições químicas de lignina exibem uma maior digestibilidade e requerem pré-tratamentos menos severos para fornecer o mesmo rendimento que as plantas nativas [29, 53, 303–305]. Assim, além de estar correlacionada com o conteúdo de lignina, a recalcitrância lignocelulósica é também influenciada pela composição química da lignina. A razão entre os conteúdos de siringil e guaiacil (S/G) na lignina (vide figura 9.1) é considerada um dos parâmetros chave que governam a recalcitrância. Não obstante, resultados contraditórios têm sido obtidos em estudos com plantas geneticamente modificadas: alguns estudos mostram que uma alta razão S/G levam ao aumento da digestibilidade enquanto outros mostram que uma baixa razão S/G aumenta a digestibilidade [3]. Studer *et al.* encontraram que, dependendo da razão S/G, a digestibilidade depende mais ou menos do conteúdo total de lignina, sugerindo relações complexas entre a digestibilidade da biomassa e atributos ligados à lignina [306]. No conjunto, estes estudos mostram que o efeito da composição química da lignina na recalcitrância lignocelulósica é ainda pouco compreendido e requer investigações mais profundas.

Somente nos últimos anos é que aspectos estruturais e físico-químicos das interações entre lignina e outros componentes da parede celular passaram a ser explorados, o que é reflexo da complexidade da biomassa. Esforços multidisciplinares têm sido necessários para a obtenção de uma visão da recalcitrância em múltiplas escalas. Microscopias e tomografias eletrônicas e espalhamento de nêutros a baixo ângulo mostraram que pequenos agregados de lignina se formam na biomassa prétratada com ácido diluído e que tais agregados se coalescem em gotas e se redistribuem sobre a superfície da parede celular [32, 34], revelando alterações estruturais induzidas pelo pré-tratamento. Em outro estudo, um conjunto multimodal de microscopias – reunindo microscopia de força atômica, microscopia de espalhamento Raman estimulado, microscopia confocal de varredura a laser e microscopia óptica de campo claro – foi usado para visualizar a ação de celulases em paredes celulares nativas e deslignificadas em tempo real e em condições próximas à fisiológica, fornecendo imagens de alta resolução do impacto negativo da lignina na degradação enzimática da biomassa [17]. Uma combinação de simulações de dinâmica molecular e espalhamento de nêutrons a baixo ângulo mostraram que os agregados de lignina possuem estrutura autosimilar em múltiplas escalas (fractal) com poros de dimensões suficientes para permitir a adsorção de celulases [307]. Simulações também mostraram que o colapso da lignina em temperatura ambiente, bem como a associação da lignina com domínios cristalinos da celulose, são processos entropicamente dirigidos [308, 309]. Simulações de dinâmica molecular de um modelo de parede celular secundária mostraram que existe uma intepenetração entre a lignina e hemicelulose ao longo de uma interface lignina-hemicelulose [310], sugerindo a existência de uma apreciável afinidade entre estes dois componentes da parede celular. Por outro lado, simulações sugerem que ocorre uma separação de fases entre a lignina e hemicelulose quando estes polímeros são submetidos às altas temperaturas dos pré-tratamentos termoquímicos [18], mostrando que as interações efetivas entre os componentes da parede celular são moduladas pelas condições termodinâmicas do meio.

Embora estes estudos começaram a lançar luz sobre aspectos moleculares e microscópicos da lignina, o mecanismo pelo qual a composição química da lignina afeta as forças termodinâmicas dentro da parede celular ainda é desconhecido. Dada a complexidade e a variabilidade da biomassa lignocelulósica, o estudo isolado de fatores que afetam a recalcitrância – por exemplo, a composição química da lignina – não é trivial. Nesta parte da tese, o efeito da composição química da lignina nas forças termodinâmicas de interação lignina-lignina e lignina-hemicelulose foi investigado usando a teoria 3D-RISM/KH, com o intuito de obter os princípios moleculares que governam a modulação de tais forças nas paredes celulares secundárias de plantas.

9.2 Metodologia

9.2.1 Modelos moleculares e procedimentos gerais

Para estudar as interações lignina-lignina e lignina-hemicelulose, a termodinâmica de solvatação de oligômeros de lignina e hemicelulose em soluções de monômeros de lignina foi obtida por 3D-RISM/KH. Os cálculos realizados simulam experimentos de calorimetria em que os oligômeros de lignina e hemicelulose são adicionados à água e, em seguida, é medida a variação de energia livre à proporção em que monômeros de lignina são progressivamente adicionados à solução.

Os monômeros de lignina *p*-hidroxifenil (H), guaiacil (G) e siringil (G), ilustrados na figura 9.1a, foram considerados neste estudo. Por conveniência, os monômeros de lignina serão denotados aqui por monolignóis, embora, a rigor, os monolignóis sejam os precursores das unidades H, G e S (vide figura 2.6). Uma vez que a ligação monolignol-monolignol mais predominante na lignina é a do tipo β -O-4' [3, 25], em que o carbono β é ligado diretamente ao átomo de oxigênio do grupo fenólico (figura 9.1b), os grupos hidroxila envolvidos na formação deste tipo de ligação foram substituídos por grupos metóxi para melhor representar os resíduos do polímero, como mostrado na figura 9.1c. Estas estruturas foram otimizadas usando mecânica molecular em vácuo e então usadas para definir soluções aquosas de monolignóis em diferentes concentrações. As funções de correlação radiais solvente-solvente destas soluções foram obtidas com a teoria DRISM/KH.

Os oligômeros xilohexaose – cadeia linear formada por 6 resíduos de xilose unidos por ligações glicosídicas $\beta(1 \rightarrow 4)$ – e hexaguaiacil – cadeia linear formada por 6 unidades de guaiacil unidas por ligações β -O-4' – foram usados como modelos para a hemicelulose (xilano) e lignina, respectivamente (figura 9.1d). Cinco conformações diferentes destes oligômeros foram tratadas nos cálculos de 3D-RISM/KH. Os confôrmeros foram obtidos usando simulações de dinâmica molecular em água explícita. Em seguida, a teoria 3D-RISM/KH foi empregada para computar a termodinâmica de solvatação dos oligômeros em soluções aquosas de monolignóis em concentrações variadas.

Em todos os cálculos, o campo de força CHARMM36 [193] foi usado para a



Figura 9.1. (a) Unidades monoméricas da lignina, como aparecem no polímero, em que os substituintes R1 e R2 determinam as subunidades H, G ou S. (b) Dímero de lignina destacando a ligação β -O-4' em azul. (c) Monômero de lignina usado nos cálculos, aqui denotados de monolignóis. (d) Oligômeros de lignina e hemicelulose.

hemicelulose, o modelo TIP3P/CHARMM [172, 174] para a água e os parâmetros CHARMM obtidos por Petridis & Smith [311] para a lignina.

9.2.2 Detalhes computacionais

Para obter as funções de correlação radiais solvente-solvente, cálculos de DRISM/KH foram executados para as soluções aquosas de monolignóis em concentrações (fração

molar) variando de 0 a 10^{-2} . A densidade e a constante dielétrica das soluções foram consideradas como sendo as mesmas da água pura (1.0 g cm⁻³ e 78,5, respectivamente) na temperatura de 300 K. Para computar a entropia, que envolve a derivada da energia livre de solvatação em relação à temperatura, cálculos de DRISM/KH semelhantes foram realizados para as temperaturas de 295 e 305 K. As equações integrais foram resolvidas numa malha radial contendo 8192 pontos separados por 0,1 Å. Malhas mais finas foram consideradas em testes, mas os resultados não se alteraram. As equações foram resolvidas iterativamente até o desvio médio quadrático relativo de 10^{-10} usando o algoritmo modificado de inversão direta no subespaço iterativo (MDIIS), que é um procedimento para acelerar a convergência de equações integrais de teorias do estado líquido [114, 287]

As equações integrais da teoria 3D-RISM/KH foram resolvidas numa malha tridimensional contendo $128 \times 128 \times 128$ pontos numa caixa cúbica de dimensões 64 Å × 64 Å × 64 Å. Estas dimensões foram suficientes para que as funções de correlação decaíssem dentro dos limites da caixa. Uma malha mais fina contendo $256 \times 256 \times 256$ pontos não alterou os resultados significativamente. O algoritmo MDIIS [114, 287] foi usado para iterar as equações integrais até o desvio médio quadrático relativo de 10^{-4} . As funções de correlação soluto-solvente obtidas por 3D-RISM/KH foram então utilizadas para obter a termodinâmica de solvatação do sistema. Cálculos da teoria DRISM/KH e 3D-RISM/KH foram realizados com programas desenvolvidos no grupo do Dr. Kovalenko.

As simulações de dinâmica molecular empregadas para amostrar conformações dos oligômeros foram executadas usando condições de contorno periódicas, *particle mesh* Ewald para computar interações eletrostáticas [105] e um raio de corte de 12 Å para truncar as interações de curto alcance. Os oligômeros foram solvatados com moléculas de água numa caixa cúbica usando Packmol [237]. Ligações envolvendo átomos de hidrogênio foram vinculadas em seus comprimentos de equilíbrio e um passo de integração de 2 fs foi usado para resolver as equações de movimento. A temperatura e a pressão foram mantidas constantes em 300 K e 1 atm, respectivemente, utilizando o termostato e pistão de Langevin [171]. As simulações foram executadas com o programa NAMD [171] por 10 ns. Cinco configurações de lignina

e hemicelulose foram extraídas a cada 2 ns das trajetórias, para posteriormente serem utilizadas nos cálculos da teoria 3D-RISM/KH.

9.3 Resultados e discussão

9.3.1 Interações de monolignóis com lignina e hemicelulose

Para investigar as interações mediadas pela lignina, a energia livre de solvatação dos oligômeros de hemicelulose e lignina em soluções aquosas dos monolignóis H, G e S em concentrações crescentes foram computadas. A presença de diferentes monolignóis no ambiente aquoso tem o efeito de abaixar a energia livre de solvatação tanto da hemicelulose quanto da lignina em diferentes intensidades, indicando que as interações lignina-lignina e hemicelulose-lignina são termodinamicamente favoráveis e dependentes da composição química da lignina.

A figura 9.2a mostra as energias livres de solvatação da hemicelulose em função da concentração dos monolignóis, com relação à energia livre de solvatação em água pura. Os dados mostram que o aumento da concentração de monolignol resulta no abaixamento da energia livre de solvatação da hemicelulose. À medida que a concentração de monolignol varia de 0 a 0,01 (expressa em fração molar), a energia livre de solvatação decresce em 10-15 kcal/mol, dependendo da natureza do monolignol. Este decréscimo na energia livre indica que os monolignóis são capazes de estabilizar a hemicelulose em meio aquoso. A figura 9.2a mostra também que a variação da energia livre de solvatação da hemicelulose cresce com a natureza dos monolignóis na seguinte ordem: H < G < S. Comparando as inclinações de um ajuste linear dos pontos da figura 9.2a, que reflete uma medida da afinidade dos monolignóis pela hemicelulose, nota-se que as interações da hemicelulose com os monolignóis G e S são, respectivamente, em torno de 20% e 50% mais fortes que as interações com o monolignol H. Estes resultados indicam que as substituições metóxi no anel aromático da lignina (figura 9.1) intensificam as interações efetivas entre hemicelulose e lignina em meio aquoso.

Os resultados mostrados na figura 9.2b mostram que o aumento da concentração



Figura 9.2. Energia livre de solvatação da (a) hemicelulose e (b) lignina em soluções aquosas dos monolignóis H, G e S em diferentes concentrações. A afinidade dos monolignóis pela hemicelulose e lignina é evidente pela redução da energia livre de solvatação à medida que a concentração dos monolignóis aumenta. A natureza dos monolignóis influencia a estabilização da hemicelulose e lignina na ordem H < G < S.

de monolignóis até a fração molar de 0,01 causa um decréscimo monotônico na energia livre de solvatação da lignina de \approx 17-27 kcal/mol, indicando que a lignina possui afinidade por si mesma e, portanto, a tendência de auto-agregação. Além disso, como nas interações hemicelulose-monolignol, (i) a dependência da energia livre de solvatação com o tipo de monolignol se intensifica na ordem H < G < S, e (ii) as interações mediadas pelos monolignóis G e S são, respectivamente, em torno de 20% e 50% mais intensas que as interações mediadas pelo monolignol H.

A redução da energia livre de solvação da lignina e hemicelulose à medida que a concentração dos monolignóis aumenta indica que a estabilidade termodinâmica da parede celular secundária – e, portanto, sua recalcitrâcia a desconstrução – é afetada tanto por interações entre lignina e hemicelulose quanto pelas interações entre a própria lignina. Isso é consistente com o fato de que a redução do conteúdo de lignina pode aumentar a digestibilidade da biomassa [303]. Microgotas de lignina que aparecem em paredes celulares após tratamentos termoquímicos [32] fornecem evidência da auto-agregação da lignina e os resultados apresentados aqui indicam que as forças internas que mantêm estes agregados são função da composição da lignina em termos dos monolignóis H, G e S. Além disso, a composição da lignina afeta sua interação com a hemicelulose; tal fato é importante em biomassa em que a remoção de hemicelulose é um passo chave para a digestão enzimática, como em *Panicum virgatum*, conhecida como *switchgrass* em inglês [278]. Nestes casos, a composição da lignina pode afetar a severidade dos pré-tratamentos necessários para a remoção da hemicelulose.

As interações efetivas mediadas pelo monolignol S são cerca de 25% mais fortes que as mediadas pelo monolignol G, um efeito que poderia ser significante na modulação da recalcitrância lignocelulósica. Assim, no que diz respeito às interações físicas da parede celular secundária, uma reduzida razão S/G resultaria em estruturas mais suscetíveis à degradação. Esta tendência tem sido, de fato, observada experimentalmente por diferentes grupos. Hung et al. demonstraram que um reduzido conteúdo de lignina e reduzida razão S/G melhoram a digestibilidade da cana-de-açúcar [312]. Fu et al. reportaram que uma razão S/G reduzida em switchgrass leva a um fenótipo de crescimento normal, maior eficiência na digestão enzimática e necessita de pré-tratamentos menos severos para obter o mesmo rendimento de acúcares fermentáveis quando comparado com plantas nativas [53]. Recentemente, Vanholme et al. mostraram que o aumento do conteúdo do monolignol H na lignina de plantas transgênicas causa um aumento de quatro vezes na eficiência enzimática [305]. Entretanto, estas tendências não se verificam em todos os casos: há vários relatos de que quanto maior a razão S/G, maior a digestibilidade da biomassa [313, 314]. Esta discrepância é, em parte, atribuída a fatores genéticos, bioquímicos e ambientais envolvidos nos experimentos [313], tornando difícies as comparações. É importante salientar que a recalcitrância lignocelulósica emerge de vários fatores interdependentes, os quais não estão incorporados no modelo. Além do conteúdo de lignina e sua composição, influenciam a recalcitrância: porosidade da parede celular, presença de complexos carboidrato-lignina, inibição enzimática pela lignina, e a acessibilidade e cristalinidade das fibras de celulose [2–4, 7]. Além disso, a suscetibilidade química da lignina tem um papel importante: durante pré-tratamento ácido, a lignina pode sofrer fragmentação através da clivagem de ligações β -O-4' [3]. Deste ponto de vista, a presença de monolignóis S tornam a lignina mais suscetível por haver menores possibilidades dos monolignóis S se acoplarem via ligações diferentes da β -O-4', como a ligação β -5,

já que o carbono 5 é metoxilado no monolignol S. Assim, a lignina rica em monolignóis G tende a ser mais ramificada e exibir ligações químicas carbono-carbono mais recalcitrantes [7, 29].

9.3.2 Efeito da conformação da hemicelulose e lignina

Os resultados da seção anterior indicam que o efeito de alterar a concentração e o tipo de monolignol na energia livre de solvatação da hemicelulose e lignina é similar, ou seja, para ambos os oligômeros ocorre um aumento das interações com os monolignóis, na ordem H < G < S, à medida que suas concentrações aumentam no meio. Isso sugere que, em nível molecular, as interações ligninalignina e hemicelulose-lignina são de natureza similar. A comparação das energias livres de solvatação da hemicelulose e da lignina em diferentes conformações revela que suas interações com os monolignóis são não-específicas e insensíveis ao estado conformacional dos oligômeros. A figura 9.3 mostra as energias livres de solvatação para cada uma das cinco conformações consideradas. As alterações de energia livre com a concentração de monolignol – obtidas como as inclinações de um ajuste linear – são essencialmente as mesmas para quaisquer das cinco conformações (as curvas são paralelas).

Embora as conformações da hemicelulose e lignina não afetem as interações com os monolignóis, os valores absolutos das energias livres de solvatação são sensíveis à conformação. Considerando que a água é o principal co-solvente nas soluções de monolignóis, estas energias livres refletem majoritariamente a hidratação dos oligômeros. A dependência das energias livres com a conformação dos oligômeros resulta de interações específicas com a água através de ligações de hidrogênio.

A não-especificidade das interações lignina-monolignol e hemicelulose-monolignol sugere que tais interações são predominantemente hidrofóbicas, mesmo as que envolvem a hemicelulose hidrofílica. No contexto da parede celular secundária, esta não-especificidade pode ser de importância fundamental para o estabelecimento de sua estrutura amorfa e resistente, já que requerimentos estéricos mínimos seriam necessários para a formação de estruturas estáveis na matriz não-celulósica da parede celular. Além disso, a similaridade das interações dos monolignóis com a



Figura 9.3. Energia livre de solvatação da hemicelulose (à esquerda) e da lignina (à direita) em função da concentração dos monolignóis H, G e S, para os cinco diferentes confôrmeros de hemicelulose e lignina. As diferentes cores representam as diferentes conformações. As energias livres de solvatação são mostradas com relação ao valor mais alto entre os confôrmeros. É evidente que o efeito da concentração dos monolignóis é independente da conformação dos oligômeros, como pode ser inferido pelas inclinações essencialmente iguais das curvas. Entretanto, os valores das energias de solvatação dependem da conformação e refletem a hidratação da hemicelulose e lignina.

hemicelulose e lignina implica que estes polímeros podem ser misturados na parede celular com baixa penalidade termodinâmica decorrente de diferenças da natureza das interações. De fato, há evidências de que, na arquitetura da parede celular nativa, a hemicelulose e a lignina estão empacotadas mutuamente de tal forma que o colapso da lignina é impedido. Estudos demonstram que, para a biomassa imersa em ácido diluído na temperatura ambiente, a lignina se mantém associada com a hemicelulose. Sob altas temperaturas, a hemicelulose sofre hidrólise e então a lignina se auto-agrega [34]. Dessa forma, a auto-agregação da lignina é dependente da remoção da hemicelulose, indicando que esta última possui um papel em manter a distribuição da lignina na parede celular nativa. Inclusive, interações lignina-hemicelulose têm sido apontadas como fatores-chave da recalcitrância, e não a própria lignina. Em *Populus*, a associação da lignina com xilanos é o principal fator ligado à digestibilidade [315]. O mesmo foi mostrado para cana-de-açúcar [316] e switchgrass [278]. Além disso, a inibição da síntese de xilanos resultou em paredes celulares secundárias finas e deformadas, demonstrando a importância do entrelaçamento lignina-hemicelulose nas propriedades da parede celular secundária [295, 317]. Em suma, a lignina pode estar tanto associada com hemicelulose (em paredes celulares nativas) ou auto-agregada (em paredes pré-tratadas com ácido). Em ambos os casos, a natureza das interações intermoleculares parecem ser similares de acordo com os resultados aqui apresentados, os quais fornecem uma interpretação molecular para tais observações experimentais.

9.3.3 Forças termodinâmicas de solvatação

Para entender a força termodinâmica que dirige as interações hemicelulose-lignina e lignina-lignina, as contribuições parciais da água e dos monolignóis para a energia livre de solvatação da hemicelulose e lignina foram computadas. Os resultados mostram que a contribuição parcial da água para a energia livre de solvatação da hemicelulose decresce (no sentido de abaixamento da energia livre) à medida que a concentração de monolignol aumenta (figura 9.4a), enquanto que a contribuição parcial dos monolignóis aumenta ligeiramente sob tais condições. A contribuição termodinamicamente desfavorável dos monolignóis é, em parte, reflexo da perda de liberdade translacional e rotacional ao interagirem com a hemicelulose. Logo, o efeito termodinâmico favorável é dirigido pela contribuição da água, que torna maior à medida que a concentração de monolignol aumenta.



Figura 9.4. Contribuição parcial da água e dos monolignóis para a energia livre de solvatação da (a) hemicelulose e da (b) lignina. A contribuição da água é termodinamicamente favorável enquanto que a dos monolignóis é desfavorável.

Adicionalmente, o efeito da composição química dos monolignóis é significativamente mais pronunciado na contribuição da água que dos próprios monolignóis (figura 9.4a). A contribuição da água para a energia livre de solvatação decresce na mesma ordem que a energia livre total, isto é, na ordem H < G < S. De forma similar, o efeito da conformação da hemicelulose é negligenciável da parte dos monolignóis, enquanto que a contribuição da água é fortemente afetada (figura 9.5).

Semelhantemente, as interações dos monolignóis com a lignina estão também associadas com uma redução da contribuição da água para a energia livre de solvatação, e o efeito do tipo de monolignol se manifesta principalmente na contribuição da água (figura 9.3b). O efeito da conformação da lignina, mostrado na figura 9.5, evidencia a baixa contribuição dos monolignóis e a forte contribuição de efeitos dependentes da água.

A tabela 9.1 mostra as energias livres de solvatação e as entropias de solvatação da hemicelulose e lignina nas soluções dos monolignóis H, G e S na fração molar de 0,01. Os dados indicam que a entropia de solvatação aumenta à medida que o número de substituições metoxi aumenta nos monolignóis (H < G < S) e que


Figura 9.5. Efeito do estado conformacional da hemicelulose (à esquerda) e da lignina (à direita) nas contribuições parciais da água e dos monolignóis para a energia livre de solvatação. A contribuição dos monolignóis é insensível à conformação tanto da hemicelulose quanto da lignina, enquanto contribuição da água depende do confôrmero.

corresponde à maior parte da energia livre. Isso mostra que a estabilização das interações lignina-hemicelulose e lignina-lignina devido a mudanças na composição química da lignina é um processo entropicamente dirigido.

Tabela 9.1. Energia livre de solvatação $(\Delta \mu)$ e entropia de solvatação multiplicada pela temperatura $(T\Delta S)$ da hemicelulose e lignina em soluções aquosas dos monolignóis H, G e S na fração molar de 0,01. Os valores são relativos ao monolignol H e expressos em kcal/mol.

	Hemicelulose		Lignina	
Monolignol	$\Delta \mu$	$T\Delta S$	$\Delta \mu$	$T\Delta S$
Н	0,0	0,0	0,0	0,0
G	-2,2	2,2	-4,1	5,0
\mathbf{S}	-4,5	3,8	-8,3	6,8

Estes resultados indicam que as interações hemicelulose-lignina e lignina-lignina são de natureza hidrofóbica e sugerem que a contribuição favorável da água para a energia livre de solvatação é devido a uma desidratação das interfaces hemicelulosemonolignol e lignina-monolignol. Além disso, o aumento do grau de metoxilação aromática dos diferentes monolignóis causa um aumento do efeito hidrofóbico nas interações, o que é acompanhado de um aumento na entropia de solvatação. A hidrofobicidade relativa dos monolignóis é evidenciada pelas compressibilidades isotérmicas das soluções aquosas de monolignóis. Tem sido estabelecido que a compressibilidade isotérmica da água próxima a superfícies hidrofóbicas é maior que a da água bulk [309, 318, 319] e, consistentemente, a compressibilidade das soluções de monolignóis aumenta com a concentração, com a dependência com o tipo de monolignol aumentando na ordem H < G < S. Estes resultados são consistentes com extensivas simulações de dinâmica molecular que preveem que a lignina colapsa por um mecanismo dirigido entropicamente [308]. Associação da lignina com celulose cristalina foi também mostrada ser dirigida por desidratação das interfaces mútuas de interação [309]. De acordo com os resultados reportados aqui, a composição química da lignina provavelmente afeta a recalcitrância lignocelulósica através das interações lignina-celulose pelos mesmos mecanismos entrópicos das interações lignina-hemicelulose.



Figura 9.6. Compressibilidades isotérmicas das soluções de monolignóis obtidas pela teoria DRISM/KH. À medida que a concentração de monolignol aumenta, a compressibilidade também aumenta monotonicamente. Além disso, as compressibilidades aumentam na ordem H < G < S, indicando que esta é a ordem relativa das hidrofobicidades dos monolignóis.

9.3.4 Aspectos estruturais das interações

Para obter uma visão molecular das interações que levam ao comportamento termodinâmico descrito nas seções anteriores, mapas tridimensionais de densidade de energia livre de solvatação (SFED) [128] foram construídos. Na figura 9.7, as SFED em torno da hemicelulose são mostradas. Os mapas representam superfícies de contorno com os mesmos valores de SFED. Tais valores foram definidos com base da primeira camada de solvatação da água, o que é equivalente a dizer que um valor ligeiramente maior de SFED estaria associado com uma segunda camada de solvatação em torno da hemicelulose. Definidos assim os valores de SFED, as superfícies de contorno das SFED dos monolignóis foram plotadas. Para a água, a SFED revela regiões altamente localizadas em torno dos grupos polares da hemicelulose, indicativas de ligações de hidrogênio (figura 9.7a). As superfícies de SFED em torno das faces mais hidrofóbicas da hemicelulose são negligenciáveis, demonstrando que as interações água-hemicelulose são estabelecidas essencialmente através de ligações de hidrogênio. Em contraste, as SFED associadas com as interações da hemicelulose com os monolignóis (figura 9.7b-9.7d) são representadas



Figura 9.7. Distribuição tridimensional da densidade de energia livre de solvatação em torno da hemicelulose (em alaranjado), mostrando regiões espaciais que concentram as interações favoráveis com a (a) água e com os monolignóis (b) H, (c), G e (d) S.

como superfícies difusas que cobrem, principalmente, as superfícies hidrofóbicas da hemicelulose, indicando que interações hidrofóbicas do tipo *stacking* entre os grupos C–H axiais da hemicelulose e a nuvens π do anel aromático dos monolignóis são predominantes. A figura 9.7 também mostra diferenças entre as SFED associadas com os diferentes monolignóis. Suas interações com a hemicelulose se tornam mais efetivas na ordem H < G < S, como indicado pela cobertura mais efetiva da SFED do monolignol S (figura 9.7d) na superfície da hemicelulose em relação a distribuições menos efetivas do monolignóis G (figura 9.7c) e H (figura 9.7b), para mesmos valores de SFED. Estes resultados mostram claramente que as substituições metóxi aromáticas da lignina tem o efeito de aumentar a habilidade dos monolignóis de se envolverem em interações do tipo *stacking* com as faces hidrofóbicas da hemicelulose.

A figura 9.8 mostra as SFED em torno da lignina. Para a superfície de contorno associada à água (figura 9.8a), os sítios de ligação de hidrogênio da lignina são as

regiões mais favoráveis de interação com a água, enquanto os anéis aromáticos da lignina repelem a água. Por outro lado, as SFED dos monolignóis (figuras 9.8b-9.8d) são espacialmente concentradas especialmente sobre os anéis aromáticos da lignina, mostrando que as interações monolignol-lignina são predominantemente do tipo *stacking* hidrofóbico $\pi - \pi$. Tais interações aumentam em intensidade na ordem H < G < S, refletindo o papel dos grupos metóxi em aumentar a hidrofobicidade da lignina.

A figura 9.9 fornece uma visão ampliada das SFED associadas ao monolignol S e à água superpostas em torno da hemicelulose e lignina, mostrando que a água coordena preferencialmente sítios de ligação de hidrogênio, enquanto que os monolignóis se espalham pelas superfícies hidrofóbica da hemicelulose e aromática da lignina. Assim, os monolignóis expulsam a água das regiões hidrofóbicas para os sítios hidrofílicos dos oligômeros, reduzindo assim seu contato desfavorável com regiões não-hidrofílicas. Tais visões moleculares corroboram a ideia de que a lignina tende a minimizar sua área superficial em contato com a água – causando sua coalescência e a formação de esferas em biomassa pré-tratada [32, 34] – e, além disso, permite extrapolar tal mecanismo para as interações hemicelulose-lignina.

9.3.5 Ligações de hidrogênio

Embora as interações do tipo *stacking* hidrofóbicas constituam o tipo mais importante de interações entre a lignina e a hemicelulose e determinem a termodinâmica das interações, algumas ligações de hidrogênio são reveladas pela análise das SFED (figuras 9.7 e 9.8). Os monolignóis, como definidos na figura 9.1c, possuem dois conjuntos de sítios de ligação de hidrogênio: (1) os átomos de oxigênio dos grupos metóxi ligados ao anel aromático e (2) os grupos hidroxila da cadeia alifática anexada ao anel aromático. A fim de avaliar a contribuição relativa destes dois grupos para as possíveis ligações de hidrogênio envolvendo a lignina e a hemicelulose, a intensidade das ligações de hidrogênio foram calculadas como

$$W = -k_B T \ln g^{max},\tag{9.1}$$



Figura 9.8. Distribuição tridimensional da densidade de energia livre de solvatação em torno da lignina (em alaranjado), mostrando regiões espaciais que concentram as interações favoráveis com a (a) água e com os monolignóis (b) H, (c), G e (d) S.

onde g^{max} é o valor máximo da função de distribuição tridimensional $g(\mathbf{r})$, que corresponde ao primeiro pico de $g(\mathbf{r})$ próximo ao sítio de ligação de hidrogênio do soluto (hemicelulose e lignina), k_B é a constante de Boltzmann e T é a temperatura do sistema. Esta definição de W é baseada na definição do potencial de força média entre o soluto e um sítio do solvente como [114]

$$PMF(\mathbf{r}) = -k_B T \ln g(\mathbf{r}). \tag{9.2}$$

Assim, o valor de W corresponde ao primeiro mínimo do PMF ao longo de um caminho arbitrário para levar um átomo do solvente do contato com o soluto até o seio da solução. O cálculo de W está ilustrado na figura 9.10 para o caso da água pura. A função de distribuição radial entre o oxigênio e hidrogênio da água mostra um pico na separação próxima de 1,8 Å, correspondente à ligação de hidrogênio entre duas moléculas de água. Este pico dá origem ao mínimo do PMF, que é utilizado aqui como medida da intensidade de uma ligação de hidrogênio.



Figura 9.9. Visão ampliada da SFED da água (em azul) e do monolignol S (em amarelo) em torno da (a) hemicelulose e (b) lignina. Enquanto a água faz ligações de hidrogênio com a hemicelulose, os monolignóis se distribuem sobre as faces hidrofóbicas disponíveis, reduzindo assim o contato da água com tais faces e, ao mesmo tempo, permitindo a formação de ligações de hidrogênio entre a água e os sítios hidrofílicos da hemicelulose e da lignina.



Figura 9.10. Análogo unidimensional do cálculo da força da ligação de hidrogênio como definida pela equação 9.1. Em (a), é mostrada a função de distribuição radial g(r) entre átomos de hidrogênio e oxigênio da água pura. O PMF obtido por PMF= $-k_BT \ln g(r)$ é mostrado em (b). Observa-se que o pico g^{max} de g(r) corresponde ao mínimo do PMF, que fornece a medida da força da ligação de hidrogênio W.

O valor de W foi computado considerando todos os sítios polares da hemicelulose e da lignina nas cinco diferentes conformações. Deste conjunto de valores de W, o valor mínimo (mais negativo) obtido para as ligações de hidrogênio entre o soluto (hemicelulose e lignina) e água e cada um dos sítios (1) e (2) descrito acima foi utilizado para comparações.

As ligações de hidrogênio envolvendo os átomos de oxigênio dos grupos metóxi aromáticos dos monolignóis e a hemicelulose alcançam, no máximo, W = -0.6 kcal/mol, enquanto que as ligações de hidrogênio envolvendo os grupos hidrolixa dos monolignóis podem chegar a W = -1,0 kcal/mol. Em contrapartida, a água faz ligações de hidrogênio com a hemicelulose com intensidade de até W = -1,2 kcal/mol. Semelhantemente, as ligações de hidrogênio envolvendo os grupos metóxi dos monolignóis e a lignina apresentam W = -0.7 kcal/mol enquanto as ligações de hidrogênio envolvendo os grupos hidroxi apresentam W =-1,1 kcal/mol. As ligações de hidrogênio água-lignina são mais intensas, apresentando W = -1,3 kcal/mol. Logo, os principais sítios de ligação de hidrogênio dos monolignóis são as hidroxilas presentes na cauda hidroxipropílica, enquanto que os substituintes metóxi aromáticos participam muito fracamente de ligações de hidrogênio com a hemicelulose e lignina. Estes resultados corroboram a observação de que os grupos metóxi aromáticos causam um aumento do caráter hidrofóbico da lignina, e não um aumento de sua hidrofilicidade.

9.4 Considerações finais

Neste estudo, foram investigados os detalhes das interações entre os dois polímeros majoritários da matriz não-celulósica da parede celular secundária de plantas, lignina e hemicelulose, e os efeitos da natureza química da lignina na intensidade de tais interações. As interações lignina-hemicelulose e lignina-lignina possuem caráter similares: são ambas hidrofóbicas e não-específicas, o que deve ser um requisito essencial para a agregação da parte amorfa da parede celular secundária. Como tal, estas interações são entropicamente dirigidas, em que moléculas de água são afastadas das faces hidrofóbicas. Além disso, os efeitos da composição química da lignina nas forças de coesão da parede celular são manifestações entrópicas decorrentes da modulação da hidrofobicidade das unidades monoméricas da lignina.

Estes resultados poderiam ser testados experimentalmente com compostos modelo (como os oligômeros de hemicelulose e lignina empregados neste estudo) usando técnicas calorimétricas, como titulação calorimétrica isotérmica (ITC) [320]. Tais experimentos poderiam fornecer as energias livres relativas associadas com as interações lignina-lignina e hemicelulose-lignina. Além disso, compostos modelo mais sofisticados poderiam ser utilizados para cuidadosamente investigar as interações de paredes celulares. A utilização de compostos modelo, em vez das complexas paredes celulares nativas, poderia fornecer informações importantes sobre os aspectos moleculares que governam a recalcitrância lignocelulósica.

Finalmente, nota-se que interações do tipo *stacking* estão presentes em toda a parede celular: nos empacotamentos de cadeias de glucano na celulose por interações stacking C-H/C-H [9], nas interações hemicelulose-lignina e ligninalignina por meio de stacking C-H/ π e π - π , respectivamente, e nas interações celulose-hemicelulose (stacking C-H/C-H) [128] e celulose-lignina (stacking C- H/π) [309]. Da mesma forma, enzimas e proteínas auxiliares (como expansinas) envolvidas na desconstrução da parede celular possuem resíduos aromáticos, principalmente triptofanos e tirosinas, e argininas que reconhecem componentes da parede celular através de interações do tipo stacking C-H/ π , π - π e cátion π - π [132, 134, 163, 275, 302]. Esta similaridade na natureza das interações permite às proteínas competirem eficientemente com as forças de coesão internas das paredes celulares. Assim, interações do tipo *stacking* hidrofóbicas parecem ser as interações fundamentais utilizadas pela natureza tanto para formar e desconstruir a arquitetura das paredes celulares. De acordo com estas observações, o desenvolvimento inteligente de pré-tratamentos e novas enzimas deveriam ser baseados em vencer as interações hidrofóbicas presentes na parede celular.

Parte V

Conclusões

Capítulo 10 Conclusões e comentários finais

O avanço da tecnologia do bioetanol de segunda geração depende da compreensão tanto da natureza das forças internas presentes no substrato quanto dos mecanismos moleculares adotados pelas enzimas e proteínas auxiliares em sua despolimerização. Nesta tese, estas duas questões ligadas à degradação de biomassa lignocelulósica foram estudadas sob o ponto de vista atômico-molecular utilizando, para isso, duas diferentes técnicas computacionais: simulações de dinâmica molecular e a teoria de equações integrais 3D-RISM/KH.

A. Simulações de dinâmica molecular

Simulações de dinâmica molecular foram empregadas para estudar aspectos dinâmicos envolvidos nos mecanismos de celobiohidrolases, laminarinases e expansinas.

Celobiohidrolases Três estudos independentes foram realizados para entender como a dinâmica estrutural de celobiohidrolases afeta diferentes etapas de seus mecanismos, tais como ligação ao substrato, processividade e expulsão do produto. As simulações mostraram que o fato da celobiohidrolase Cel7A de *Trichoderma harzianum* apresentar um túnel catalítico com maiores flutuações que o da Cel7A de *Trichoderma reesei* se correlaciona com maiores parâmetros cinéticos e menor grau de inibição por produto, sugerindo que a acessibilidade ao substrato é um aspecto chave do mecanismo de celobiohidrolases. A inibição por produto, por sua vez, parece estar relacionada a aspectos dinâmicos do sítio de ligação da celobiose, que exibe flutuações que dificultam estericamente sua dissociação. Por fim, a dinâmica intrínseca das celobiohidrolases é modulada pelo substrato, o que permite que a enzima altere sua função conforme a etapa de seu ciclo catalítico. Em particular, na ausência do substrato, a celobiohidrolase Cel7A de *T. reesei* exibe flutuações que favorecem o processo de complexação, enquanto que na presença do substrato as flutuações se alteram e passam a ter relação com o movimento processivo do substrato dentro do túnel catalítico.

Laminarinases Neste estudo, a dinâmica das laminarinases PcLam, aNLam e RmLam foi investigada em diferentes temperaturas para obter as bases moleculares da meso-, termo- e hipertermofilicidade/estabilidade destas enzimas estruturalmente similares. A estabilidade térmica se correlaciona com a capacidade do núcleo hidrofóbico das laminarinases de resistir à penetração de água, o que ocorre mais facilmente na mesofílica através de uma porta hidrofílica formada por uma ponte salina numa das entradas do núcleo hidrofóbico. Nas laminarinases (hiper)termoestáveis, não há resíduos no núcleo hidrofóbico, e a entrada de água é dificultada, resultando em maiores estabilidades térmicas. A termofilicidade das laminarinases está associada à capacidade de manter a acessibilidade do sítio ativo em altas temperaturas. Enquanto a mesofílica PcLam exibe o sítio ativo exposto somente em temperatura ambiente, as (hiper)termofílicas aNLam e RmLam mantêm a acessibilidade do sítio ativo em altas temperaturas.

Expansinas Um conjunto de simulações de dinâmica molecular mostrou que a expansina EXLX1 interage com a face hidrofóbica da celulose através de interações de van der Waals mediadas pela superfície aromática do domínio D2 que, por sua vez, permite à EXLX1 deslizar sobre a celulose. Quando em contato com uma cadeia isolada de celulose, o domíno D1 da EXLX1 induz uma torção em sua conformação, processo no qual o resíduo Asp82 participa através de ligações de hidrogênio. Os mutantes Asp82Asn e Tyr73Ala, que causam a inativação da EXLX1, desestabilizam estruturalmente o centro de torção, compromentendo sua a habilidade de induzir torções em cadeias de glucano. O fato de deslizar sobre a celulose e de induzir torções em cadeias individuais de glucano sugere que a

EXLX1 possa caminhar sobre a celulose rompendo interações xiloglucano-celulose e celulose-celulose, assim enfraquecendo a arquitetura da parede celular e gerando substratos mais suscetíveis ao ataque enzimático das celulases.

B. Teoria de equações integrais

A teoria de equações integrais 3D-RISM/KH foi empregada para entender como aspectos termodinâmicos intrínsecos de paredes celulares primárias e secundárias – que possuem ligação com a recalcitrância à desconstrução – emergem diretamente dos detalhes moleculares de sua composição química e nanoarquitetura.

Paredes celulares primárias Em água pura, nanofibrilas de celulose se agregam preferencialmente através de suas faces hidrofílicas e o contato de nanofibrilas através de suas faces hidrofóbicas constitui um estado metaestável. Quando monômeros de hemicelulose são adicionados ao meio, estes induzem a estabilização do estado agregado das nanofibrilas. A intensidade desta estabilização é dependente da natureza química dos monômeros: unidades de glucuronato negativamente carregadas afetam a agregação das nanofibrilas mais intensamente que unidades não carregadas, como arabinose e açúcares similares. Estes estudos mostram o papel da hemicelulose em modular termodinamicamente o arranjo preferencial das fibrilas de celulose na parede celular.

Paredes celulares secundárias Com um procedimento que simula experimentos de titulação calorimétrica, a teoria 3D-RISM/KH foi utilizada para estudar a natureza das interações de lignina e hemicelulose com as três unidades monoméricas de lignina – H, G e S – as quais variam no grau de metoxilação aromática. Os resultados mostraram que metoxilações aumentam a área superficial aromática dos monômeros de lignina e intensificam interações hidrofóbicas com oligômeros de lignina e hemicelulose. A desidratação das superfícies de contato lignina-lignina e hemicelulose-lignina origina a força termodinâmica que causa a agregação destes componentes da parede celular sedundária, que varia com a natureza dos monômeros, na ordem H < G < S.

Perspectivas futuras

Estes cinco estudos constituem contribuições científicas básicas no contexto biotecnológico de conversão de biomassa lignocelulósica. Embora uma explosão científica no campo de mecanismos das enzimas e estrutura de biomassa venha acontecendo nos últimos 5 anos, devido, principalmente, a pesquisas realizadas no National Renewable Energy Laboratory (NREL) e no Oak Ridge National Laboratory (ORNL), nos Estados Unidos, há ainda muito a ser compreendido a respeito desses processos em escala molecular. De acordo com a tendência da literatura, os próximos anos devem testemunhar a revelação de mais detalhes dos mecanismos de celobiohidrolases, que constituem enzimas-chave na despolimerização da celulose, e de proteínas auxiliares, englobando expansinas e enzimas com atividades auxiliares. Além disso, técnicas avançadas de microscopias devem possibilitar a visualização de detalhes cada vez maiores da arquitetura da parede celular e os efeitos de pré-tratamentos. Nestes desenvolvimentos, simulações computacionais cada vez mais realísticas e em escala de tempo cada vez maiores terão um papel essencial, tanto na verificação de hipóteses geradas por experimentos, quanto na formulação de novas hipóteses a serem verificadas experimentalmente, num processo iterativo.

Apêndice A Nomenclatura e estrutura dos aminoácidos

A tabela abaixo lista a estrutura e a abreviatura dos aminoácidos.



Nome,	Fórmula	Nome,	Fórmula
Abreviações	estrutural	Abreviações	estrutural
Aminoácidos de Glicina Gly G	e cadeia lateral apolar COO HCH NH ₃ +	Metionina Met M	COO HCCH ₂ CH ₂ SCH ₃ NH ₃ +
Alanina Ala A	CCO HCH3 NH3+	Prolina Pro P	C H H H H
Valina Val V	$H - C - CH_{3}$ $H - C - CH_{3}$ $H - CH_{3}$	Fenilalanina Phe F	$H - C - CH_2 - O$ NH_3^+
Leucina Leu L	$H - C - CH_2 - CH_3$ $H - H_3^+ - CH_3$	Triptofano Trp W	H-C-CH ₂ NH ₃ ⁺
Isoleucina Ile I	COO CH ₃ HC C*CH ₂ CH ₃ NH ₃ + H		
Aminoácidos de Serina Ser S	e cadeia lateral polar sem carga COO HCH2OH NH3 ⁺	Glutamina Gln Q	$H - C - CH_2 -$
Treonina Thr T	COO H HC [*] CH ₃ NH ₃ ⁺ OH	Tirosina Tyr Y	
Asparagina Asn N	COO H-C-CH2-C NH3+ NH2	Cisteína Cys C	COO HCH2 SH NH3+

=

Referências Bibliográficas

- J. M. Otero, G. Panagiotou, and L. Olsson. Fueling industrial biotechnology growth with bioethanol. Adv. Biochem. Eng. Biotechnol., 108:1–40, 2007.
- [2] M. E. Himmel, S-.Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, and T. D. Foust. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315:804–807, 2007.
- [3] Y. Pu, F. Hu, F. Huang, B. H. Davison, and A. J. Ragauskas. Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments. *Biotechnol. Biofuels*, 6:15, 2013.
- [4] S. P. S. Chundawat, G. T. Beckham, M. E. Himmel, and B. E. Dale. Decontruction of lignocellulosic biomass to fuels and chemicals. Annu. Rev. Chem. Biomol. Eng., 2:121–145, 2011.
- [5] K. Keegstra. Plant cell walls. *Plant Physiol.*, 154:483–486, 2010.
- [6] Z. A. Popper, G. Michel, C. Hervé, D. S. Domozych, W. G. T. Willats, M. G. Tuohy, B. Kloareg, and D. B. Stengel. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.*, 62:567–590, 2011.
- [7] N. Sorek, T. H. Yeats, H. Szemenyei, H. Youngs, and C. R. Someville. The implications of lignocellulosic biomass chemical composition for the production of advanced biofuels. *Bioscience*, 64:192–201, 2014.
- [8] A. C. O'Sullivan. Cellulose: the structure slowly unravels. *Cellulose*, 4: 173–207, 1997.
- [9] Y. Nishiyama, P. Langan, and H. Chanzy. Crystal structure and hydrogenbonding system in cellulose Iβ from synchrotron x-ray and neutron fiber diffraction. J. Am. Chem. Soc., 124:9024–9082, 2002.

- [10] C. M. Payne, M. E. Himmel, M. F. Crowley, and G. T. Beckham. Decrystalization of oligosaccharides from the cellulose $I\beta$ surface with molecular simulation. J. Phys. Chem. Lett., 2:1546–1550, 2011.
- [11] D. Gao, S. P. S. Chundawat, A. Sethi, V. Balan, S. Gnanakaran, and B. E. Dale. Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. *Proc. Natl. Acad. Sci. USA*, 110:10922–10927, 2013.
- [12] G. T. Beckham, J. F. Matthews, B. Peters, Y. J Bomble, M. E. Himmel, and M. F. Crowley. Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. J. Phys. Chem. B, 115:4118–4127, 2011.
- [13] C. Somerville. Cellulose synthesis in higher plants. Annu. Rev. Cell Dev. Biol., 22:53–78, 2006.
- [14] S. Y. Ding and M. E. Himmel. The maize primary cell wall microfibril: a new model derived from direct visualization. J. Agric. Food Chem., 54:597–606, 2006.
- [15] Y. Nishiyama. Structure and properties of the cellulose microfibril. J. Wood. Sci., 55:241–249, 2009.
- [16] L. H. Thomas, V. T. Forsyth, A. Sturcová, C. J. Kennedy, R. P. May, C. M. Altaner, D. C. Appeley, T. J. Wess, and M. C. Jarvis. Structure of cellulose microfibrils in primary cell walls from collenchyma. *Plant Physiol.*, 161:465–476, 2013.
- [17] S. Y. Ding, Y. S. Liu, Y. Zeng, M. E. Himmel and J. O. Baker, and E. A. Bayer. How does plant cell wall nanoscale architecture correlates with enzymatic digestibility? *Science*, 338:1055–1060, 2012.
- [18] P. Langan, L. Petridis, H. M. O'Neill, S. V. Pingali, M. Fouston, Y. Nishiyama, R. Shultz, B. Lindner, B. L. Hanson, S. Harton, W. T. Heller, V. Urban, B. R. Evans, S. Gnanakaran, A. J. Ragauskas, J. C. Smith, and B. H. Davison. Common processes drive the thermochemical pretreatment of lignocellulosic biomass. *Green Chem.*, 16:63–68, 2014.
- [19] R. P. Chandra, R. Bura, W. E. Mabee, A. Berlin, X. Pan, and J. N. Saddler. Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? *Adv. Biochem. Engin. Biotechnol.*, 108:67–93, 2007.

- [20] H. V. Scheller and P. Ulvskov. Hemicelluloses. Annu. Rev. Plant Biol., 61: 263–289, 2010.
- [21] M. Pauly and K. Keegstra. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.*, 54:559–568, 2008.
- [22] M. Pauly and K. Keegstra. Plant cell wall as precursors for biofuels. Curr. Opin. Plant Biol., 13:305–312, 2010.
- [23] J. C. Mortimer, G. P. Miles, D. M. Brown, Z. Zhang, M. P. Segura, T. Weimar, X. Yu, K. A. Seffen, E. Stephens, S. R. Turner, and P. Dupree. Absence of branches from xylan in arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proc. Natl. Acad. Sci. USA, 107: 17409–17414, 2010.
- [24] M. Galbe and G. Zacchi. Pretreatment of lignocellulosic materials for efficient bioethanol production. Adv. Biochem. Engin. Biotechnol., 108:41–65, 2007.
- [25] W. Boerjan, J. Ralph, and M. Baucher. Lignin biosynthesis. Annu. Rev. Plant Biol., 54:519–546, 2003.
- [26] N. D. Bonawits and C. Chapple. The genetics of lignin biosynthesis: connecting genotype to phenotype. Annu. Rev. Genet., 44:337–363, 2010.
- [27] C. I. Lacayo, M. S. Hwang, S. Y. Ding, and M. P. Thelen. Lignin depletion enhances the digestibility of cellulose in cultured xylem cells. *Plos One*, 8: e68266, 2013.
- [28] V. B. Agbor, N. Cicek, R. Sparling, A. Berlin, and D. B. Levin. Biomass pretreatment: fundamentals toward application. *Biotechnol. Adv.*, 29:675– 685, 2011.
- [29] R. Vanholme, K. Morreel, C. Darrah, P. Oyarce, J. H. Grabber, J. Ralph, and W. Boerjan. Metabolic engineering of novel lignin in biomass crops. *New Phytologist*, 196:978–1000, 2012.
- [30] R. Kumar, G. Mago, V. Balan, and C. E. Wyman. Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatments technologies. *Biores. Technol.*, 100:3948–3962, 2009.
- [31] S. P. S. Chundawat, R. Vismeh, L. N. Sharma, J. F. Humpula, L. C. Sousa, C. K. Chambliss, A. D. Jones, V. Balan, and B. E. Dale. Multifaceted characterization of cell wall decomposition products formed during ammonia fiber expansion (AFEX) and dilute acid based pretreatments. *Biores. Technol.*, 101:8429–8438, 2010.

- [32] B. Donohoe, S. Decker, M. Tucker, M. E. Himmel, and T. Vinzant. Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment. *Biotechnol. Bioeng.*, 101:913–925, 2008.
- [33] Y. Nishiyama, P. Langan, H. M. O'Neill, S. Pingali, and S. Harton. Structural coarsening of aspen wood by hydrothermal pretreatment monitored by small- and wide-angle scattering of X-rays and neutrons on oriented specimens. *Cellulose*, 21:1015–1024, 2014.
- [34] S. V. Pingali, V. S. Urban, W. T. Heller, J. McGaughey, H. O'Neill, M. Foston, D. A. Myles, A. Ragauskas, and B. R. Evans. Breakdown of cell wall nanostructure in dilute acid pretreated biomass. *Biomacromolecules*, 11: 2329–2335, 2010.
- [35] J. L. Wen, T. Q. Yuan, S. L. Sun, F. Xu, and R. C. Sun. Understanding the chemical transformations of lignin during ionic liquid pretreatment. *Green Chem.*, 16:181–190, 2014.
- [36] V. C. Coletta, C. A. Rezende, F. R. Conceição, I. Polikarpov, and F. E. G. Guimarães. Mapping the lignin distribution in pretreated sugarcane bagasse by confocal and fluorescence lifetime imaging microscopy. *Biotechnol. Bio-fuels*, 6:43, 2013.
- [37] K. Igarashi, M. Wada, and M. Samejima. Activation of crystalline cellulose to cellulose III_{I} results in efficient hydrolysis by cellobiohydrolase. *FEBS J.*, 274:1785–1792, 2007.
- [38] M. Foston and A. J. Ragauskas. Changes in lignocellulosic supramolecular and ultrastructure during dilute acid pretreatment of *Populus* and switchgrass. *Biomass and Bioenergy*, 34:1885–1895, 2010.
- [39] C. Driemeier, M. T. B. Pimenta, G. J. M. Rocha, M. M. Oliveira, D. B. Mello, P. Mazieiro, and A. R. Gonçalves. Evolution of cellulose crystals during prehydrolysis and soda delignification of sugarcane lignocellulose. *Cellulose*, 18:1509–1519, 2011.
- [40] S. Pingali, H. M. O'Neill, Y. Nishiyama, L. He, Y. B. Melnichenko, V. Urban, L. Petridis, B. Davison, and P. Langan. Morphological changes in the cellulose and lignin components of biomass occur at different stages during steam pretreatment. *Cellulose*, pages 1–6, 2014.
- [41] X. Meng, M. Foston, J. Leisen, J. DeMartini, C. E. Wyman, and A. J. Ragauskas. Determination of porosity of lignocellulosic biomass before and

after pretreatment by using Simons' stain and NMR techniques. *Biores. Technol.*, 144:467–476, 2013.

- [42] S. P. S. Chundawat, G. Bellesia, N. Uppugundla, L. C. Sousa, D. Gao, A. M. Cheh, U. P. Agarwal, C. M. Bianchetti, G. N. Phillips, P. Langan, V. Balan, S. Gnanakaran, and B. E. Dale. Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. J. Am. Chem. Soc., 133:11163–11174, 2011.
- [43] R. P. Swatloski, S. K. Spear, J. D. Holbrey, and R. D. Rogers. Dissolution of cellulose with ionic liquids. J. Am. Chem. Soc., 124:4974–4975, 2002.
- [44] S. S. Y. Tan, D. R. MacFarlane, J. Upfal, L. A. Edye, W. O. S. Doherty, A. F. Patti, J. M. Pringle, and J. L. Scott. Extraction of lignin from lignocellulose at atmospheric pressure using alkylbenzenesulfonate ionic liquid. *Green Chem.*, 11:339–345, 2009.
- [45] N. Sun, M. Rahman, Y. Qin, M. L. Maxim, H. Rodriguez, and R. D. Rogers. Complete dissolution and partial delignification of wood in the ionic liquid 1-ethyl-3-methylimidazolium acetate. *Green Chem.*, 11:646–655, 2009.
- [46] H. Tadesse and R. Luque. Advances on biomass pretreatment using ionic liquids: an overview. *Energy Environ. Sci.*, 4:3913–3929, 2011.
- [47] H. M. Cho, A. S. Gross, and J. W. Chu. Dissecting force interactions in cellulose deconstruction reveals the required solvent versatility for overcoming biomass recalcitrance. J. Am. Chem. Soc., 133:14033–14041, 2011.
- [48] R. Rinaldi. Instantaneous dissolution of cellulose in organic electrolyte solutions. Chem. Comm., 47:511–513, 2011.
- [49] F. Hu, S. Jung, and A. Ragauskas. Pseudo-lignin formation and its impact on enzymatic hydrolysis. *Biores. Technol.*, 117:7–12, 2012.
- [50] P. Sannigrahi, D. H. Kim, S. Jung, and A. Ragauskas. Pseudo-lignin and pretreatment chemistry. *Energy Environ. Sci.*, 4:1306–1310, 2011.
- [51] F. M. Gírio, C. Fonseca, F. Carvalheiro, L. C. Duarte, S. Marques, and R. Bogel-Lukasik. Hemicelluloses for fuel ethanol: a review. *Biores. Technol.*, 101:4775–4800, 2010.
- [52] M. C. McCann and N. C. Carpita. Designing the deconstruction of plant cell walls. *Curr. Opin. Plant Biol.*, 11:314–320, 2008.

- [53] C. Fu, J. R. Mielenz, X. Xiao, Y. Ge, C. Y. Hamilton, M. R. Rodriguez, F. Chen, M. Foston, A. Ragauskas, J. Bouton, R. A. Dixon, and Z. Y. Wang. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. USA*, 108:3803–3808, 2011.
- [54] C. G. Wilkerson, S. D. Mansfield, F. Lu, S. Withers, J. -Y. Park, S. D. Karlen, E. Gonzales-Vigil, D. Padmakshan, F. Unda, J. Rencoret, and J. Ralph. Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science*, 344:90–93, 2014.
- [55] D. M. Harris, K. Corbin, T. Wang, R. Gutierrez, A. L. Bertolo, C. A. Petti, D. M. Smilgies, J. M. Estevez, D. Bonetta, B. R. Urbanowicz, D. W. Ehrhardt, C. R. Somerville, J. K. C. Rose, M. Hong, and S. DeBolt. Cellulose microfibril crystallinity is reduced by mutating c-terminal transmembrane region residues CESA1^{a903v} and CESA3^{t942i} of cellulose synthase. *Proc. Natl. Acad. Sci. USA*, 109:4098–4103, 2012.
- [56] J. L. W. Morgan, J. Strumillo, and J. Zimmer. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature*, 493:181–187, 2013.
- [57] B. Yang, Z. Dai, S. Y. Ding, and C. E. Wyman. Enzymatic hydrolysis of cellulosic biomass. *Biofuels*, 2:412–450, 2011.
- [58] D. B. Wilson. Cellulases and biofuels. Curr. Opin. Biotechnol., 20:295–299, 2009.
- [59] D. Martinez, R. M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S. E. Baker, J. Chapman, O. Chertkov, P. M. Coutinho, D. Cullen, E. G. J. Danchin, I. V. Grigoriev, P. Harris, M. Jackson, C. P. Kubicek, C. S. Han, I. Ho, L. F. Larrondo, A. L. Leon, J. K. Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A. A. Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C. L. Schoch, J. Yao, R. Barabote, M. A. Nelson, C. Detter, D. Bruce, C. R. Kuske, G. Xie, P. Richardson, D. S. Rokhsar, S. M. Lucas, E. M. Rubin, N. Dunn-Coleman, M. Ward, and T. S. Brettin. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat. Biotechnol., 26:553–560, 2008.
- [60] G. Davies and B. Henrissat. Structures and mechanisms of glycosyl hydrolases. *Structure*, 3:853–859, 1995.
- [61] B. Henrissat. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, 280:309–316, 1991.

- [62] V. Lombard, H. G. Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat. The carbohydrate-active enzymes database (CAZy) in 2013. Nucl. Acids Res., 42: D490–D495, 2013.
- [63] B. Henrissat and A. Bairoch. Updating the sequence-based classification of glycosil hydrolases. *Biochem. J.*, 316:695–696, 1996.
- [64] M. L. Sinnott. Catalytic mechanisms of enzymic glycosyl transfer. Chem. Rev., 90:1171–1202, 1990.
- [65] J. Gebler, N. R. Gilkes, M. Claeyssens, D. B. Wilson, P. Béguin, W. W. Wakarchuk, D. G. Kilburn, R. C. Miller, R. A. J. Warren, and S. G. Withers. Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-xylanases. J. Biol. Chem., 267:12559–12561, 1992.
- [66] T. M. Gloster, J. P. Turkenburg, J. R. Potts, B. Henrissat, and G. J. Davies. Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chem. Biol.*, 15:1058–1067, 2008.
- [67] S. A. K. Jongkees and S. G. Withers. Unusual enzymatic glycoside cleavage mechanisms. Acc. Chem. Res., 47:226–235, 2014.
- [68] A. V. Gusakov. Alternatives to *Trichoderma reesei* in biofuel production. *Trends Biotechnol.*, 29:419–425, 2011.
- [69] T. T. Teeri. Crystalline cellulose degradation: new insights into the function of cellobiohydrolases. *Trends Biotechnol.*, 15:160–167, 1997.
- [70] M. Gruno, P. Valjamae, G. Petterson, and G. Johansson. Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnol. Bioeng.*, 86:503–511, 2004.
- [71] V. Arantes and J. N. Saddler. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol. Biofuels*, 3:4, 2010.
- [72] E. S. Kim, H. J. Lee, W. G. Bang, I. G. Choi, and K. H. Kim. Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.*, 102:1342–1353, 2009.
- [73] J. O. Baker, M. R. King, W. S. Adney, S. R. Decker, T. B. Vinzant, S. E. Lantz, R. E. Nieves, S. R. Thomas, L. Li, D. J. Cosgrove, and M. E. Himmel. Investigation of the cell-wall loosening protein expansin as a possible

additive in the enzymatic saccharification of lignocellulosic biomass. *Applied Biochemistry and Biotechnology*, 84-86:217–223, 2000.

- [74] G. Jäger, M. Girfoglio, F. Dollo, R. Rinaldi, H. Bongard, U. Commandeur, R. Fischer, A. C. Spiess, and J. Buchs. How recombinant swollenin from *Kluyveromyces lactis* affects cellulosic substrates and accelerates their hydrolysis. *Biotechnol. Biofuels*, 4:33, 2011.
- [75] R. E. Quiroz-Castañeda, C. Martínez-Anaya, L. I. Cuervo-Soto, L. Segovia, and J. L. Folch-Mallol. Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*. *Microb. Cell Fact.*, 10:8, 2011.
- [76] K. Gourlay, J. Hu, V. Arantes, M. Andberg, M. Saloheimo, M. Penttila, and J. Saddler. Swollenin aids in the amorphogenesis step during the enzymatic hydrolysis of pretreated biomass. *Biores. Technol.*, 142:498–503, 2013.
- [77] B. Bunterngsook, W. Mhuantong, V. Champreda, A. Thamchaiphenet, and L. Eurwilaichitr. Identification of novel bacterial expansins and their synergistic actions on cellulose degradation. *Biores. Technol.*, 159:64–71, 2014.
- [78] K. Gourlay, V. Arantes, and J. N. Saddler. Use of substructure-specific carbohydrate binding modules to track changes in cellulose accessibility and surface morphology during the amorphogenesis step of enzymatic hydrolysis. *Biotechnol. Biofuels*, 5:51, 2012.
- [79] S. J. Horn, G. Vaaje-Kolstad, B. Westereng, and V. G. H. Eijsink. Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels*, 5:45, 2012.
- [80] G. R. Hemsworth, G. J. Davies, and P. H. Walton. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Curr. Opin. Struct. Biol.*, 23:660–668, 2013.
- [81] S. Kim, J. Ståhlberg, M. Sandgren, R. S. Paton, and G. T. Beckham. Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc. Natl. Acad. Sci.* USA, 111:149–154, 2014.
- [82] P. V. Harris, D. Welner, K. C. McFarland, E. Re, J. C. N. Poulsen, K. Brown, R. Salbo, H. Ding, E. Vlasenko, S. Merino, F. Xu, J. Cherry, S. Larsen, and L. Leggio. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family. *Biochemistry*, 49:3305–3316, 2010.

- [83] G. Vaage-Kolstad, B. Westereng, S. J. Horn, Z. Liu, H. Zhai, M. SØrlie, and V. G. H. Eijsink. An oxydative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*, 330:219–222, 2010.
- [84] R. J. Quinlan, M. D. Sweeney, L. Leggio, H. Otten, J. C. N. Poulsen, K. Johansen, K. B. R. M. Krogh, C. I. Jørgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, and P. H. Walton. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. USA*, 108:15079–15084, 2011.
- [85] W. T. Beeson, C. M. Phillips, J. H. D. Cate, and M. A. Marletta. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. J. Am. Chem. Soc., 134:890–892, 2012.
- [86] D. Frenkel and B. Smit. Understanding molecular simulations. Academic Press, San Diego, 2001.
- [87] M. P. Allen and D. Tildesley. Computer simulations of liquids. Clarendon Press, Oxford, 1987.
- [88] A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education. Prentice Hall, 2001.
- [89] O. M. Becker, A. D. MacKerrel Jr., B. Roux, and M. Watanabe. Computational biochemistry and biophysics. Marcel Dekker, New York, 2001.
- [90] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. Rev.*, 110:1463–1497, 2010.
- [91] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.
- [92] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.*, 19:120–127, 2009.
- [93] M. W. van der Kamp and A. J. Mulholland. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, 52:2708–2728, 2013.
- [94] H. M. Senn and W. Thiel. QM/MM methods for biomolecular systems. Angewandte Chem., 48:1198–1229, 2009.

- [95] R. L. Silveira. Simulações de dinâmica molecular do receptor ativado de proliferadores de peroxissomos isoforma γ. Dissertação de Mestrado em Química, Instituto de Química – Universidade Estadual de Campinas, Campinas, São Paulo, 2010.
- [96] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000.
- [97] C. Chipot and A. Pohorille. Free Energy Calculations: Theory and Applications in Chemistry and Biology. Springer Series in Chemical Physics. Physica-Verlag, 2007.
- [98] M. Griebel, S. Knapek, and G. Zumbusch. Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications. Texts in Computational Science and Engineering. Springer, 2007.
- [99] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys., 23:327–341, 1977.
- [100] H. C. Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. J. Comput. Phys., 52:24–34, 1983.
- [101] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem., 13:952– 962, 1992.
- [102] J. W. Ponder and D. A. Case. Force fields for protein simulation. Adv. Prot. Chem., 66:27–85, 2003.
- [103] P. Gibbon and G. Sutmann. Long-range interactions in many-particle simulation. In J Grotendorst, D Marx, and A Muramatsu, editors, *Quantum* simulations of complex many-body systems: from theory to algorithms, volume 10 of NIC series. John von Neumann Institute for Computing, Jülich, 2002.
- [104] J. R. Reitz, F. J. Milford, and R. W. Christy. Foundations of Electromagnetic Theory. Addison-Wesley, 4th edition, 2009.
- [105] P. Darden, D. York, and L. Pedersen. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J. Chem. Phys., 98:10089–10092, 1993.

- [106] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330:341–346, 2010.
- [107] H. Lei and Y. Duan. Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.*, 17:187–191, 2007.
- [108] W. Humphrey, A. Dalke, and K. Schulten. VMD Visual Molecular Dynamics. J. Mol. Graph., 14:33–38, 1996.
- [109] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.
- [110] S. Hayward and B. L. de Groot. Normal modes and essential dynamics. In A Kukol, editor, *Molecular modeling of proteins*, volume 443 of *Methods in Molecular Biology*. Humana Press, Totowa, New Jersey, USA, 2008.
- [111] R. C. Tolman. The principles of statistical mechanics. Dover, New York, 1979.
- [112] J. P. Hansen and I. R. McDonald. Theory of simple liquids. Elsevier, 2nd edition, 1990.
- [113] A. Kovalenko. Multiscale modeling of solvation in chemical and biological nanosystems and in nanoporous materials. *Pure Appl. Chem.*, 85:159–199, 2013.
- [114] A. Kovalenko. Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces. In F. Hirata, editor, *Molecular Theory of Solvation*. Kluwer Academid Publishers, Dordrecht, The Netherlands, 2003.
- [115] T. L. Hill. Statistical Mechanics: Principles and Selected Applications. Dover Books on Physics Series. McGraw-Hill, 1956.
- [116] T. L. Hill. An Introduction to Statistical Thermodynamics. Addison-Wesley series in chemistry. Dover Publications, 1960.
- [117] F. Hirata. Theory of molecular liquids. In F. Hirata, editor, Molecular Theory of Solvation. Kluwer Academid Publishers, Dordrecht, The Netherlands, 2003.
- [118] D. S. Palmer, J. SØrensen, B. SchiØtt, and M. V. Fedorov. Solvent binding analysis and computational alanine scanning of the bovine chymosin-bovine κ-casein complex using molecular integral equation theory. J. Chem. Theory Comput., 9:5706–5717, 2013.

- [119] D. S. Palmer, A. L. Frolov, E. L. Ratkova, and M. V. Fedorov. Towards a universal method for calculating hydration free energies: a 3d reference interaction site model with partial molar volume correction. J. Phys.: Condens. Matter, 22:492101, 2010.
- [120] D. Chandler and H. C. Andersen. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. J. Chem. Phys., 57:1930–1937, 1972.
- [121] S. Phongphanphanee, N. Yoshida, and F. Hirata. On the proton exclusion of aquaporins: a statistical mechanical study (Material Suplementar). J. Am. Chem. Soc., 130:1540–1541, 2008.
- [122] J. Perkyns and B. M. Pettitt. A site-site theory for finite concentration saline solutions. J. Chem. Phys., 97:7656–7666, 1992.
- [123] A. Kovalenko and F. Hirata. Self-consisten description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. J. Chem. Phys., 110:10095, 1999.
- [124] S. Phongphanphanee, T. Rungrotmongkol, N. Yoshida, S. Hannongbua, and F. Hirata. Proton transport through the influenza a M2 channel: Threedimensional reference interaction site model study. J. Am. Chem. Soc., 132 (28):9782–9788, 2010.
- [125] S. Gusarov, B. S. Pujari, and A. Kovalenko. Efficient treatment of solvation shells in 3d molecular theory of solvation. J. Comput. Chem., 33:1478–1494, 2012.
- [126] A. Kovalenko F. Hirata. Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. J. Chem. Phys., 112:10391, 2000.
- [127] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. Amber11. University of California, San Francisco, 2011.
- [128] R. L. Silveira, S. R. Stoyanov, S. Gusarov, M. S. Skaf, and A. Kovalenko. Plant biomass recalcitrance: effect of hemicellulose composition on nanoscale

forces that control cell wall strength. J. Am. Chem. Soc., 135:19048–19051, 2013.

- [129] T. T. Teeri, A. Koivula, M. Linder, T. Reinikainen, L. Ruohonen, M. Srisodsuk, M. Claeyssens, and T. A. Jones. Modes of action of two *Trichoderma reesei* cellobiohydrolases. In Birte Svensson Steffen B. Petersen and Sven Pedersen, editors, *Carbohydrate Bioengineering Proceedings of an International Conference*, volume 10, pages 211–224. Elsevier, 1995.
- [130] M. H. Momeni, C. M. Payne, H. Hansson, N. E. Mikkelsen, J. Svedberg, A. Engström, M. Sandgren, G. T. Beckham, and J. Ståhberg. Structural, biochemical and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. J. Biol. Chem., 288:5861–5872, 2013.
- [131] C. M. Payne, M. G. Resch, L. Chen, M. F. Crowley, M. E. Himmel, L. E. Taylor II, M. Sandgren, J. Ståhlberg, I. Stals, Z. Tan, and G. T. Beckham. Glycosilated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proc. Natl. Acad. Sci. USA*, 2013:14646–14651, 2013.
- [132] C. Divne, J. Ståhlberg, T. Reinikainen, L. Ruohonen, G. Petterson, J. K. Knowles, T. T. Teeri, and T. A. Jones. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science*, 265:524–528, 1994.
- [133] G. J. Kleywegt, J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Ståhlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri, and T. A. Jones. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol., 272:383–397, 1997.
- [134] C. Divne, J. Ståhlberg, T. T. Teeri, and T. A. Jones. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstron long tunnel of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol., 275:309–325, 1998.
- [135] A. B. Boraston, D. N. Bolam, H. J. Gilbert, and G. J. Davies. Carbohydratebinding modules. fine-tuning polysaccharide recognition. *Biochem. J.*, 382: 769–781, 2004.
- [136] P. J. Kraulis, G. M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles, and A. M. Gronenborn. Determination of the three-dimensional solu-

tion structure of the c-terminal domain of cellobiohydrolase I from *Tricho*derma reesei. a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry*, 28:7241–7257, 1989.

- [137] M. R. Nimlos, G. T. Beckham, J. F. Matthews, L. Bu, M. E. Himmel, and M. F. Crowley. Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. J. Biol. Chem., 287:20603–20612, 2012.
- [138] G. T. Beckham, J. F. Matthews, Y. J. Bomble, L. Bu, W. S. Adney, M. E. Himmel, M. R. Nimlos, and M. F. Crowley. Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. J. Phys. Chem. B, 114:1447–1453, 2010.
- [139] G. T. Beckham, Y. J. Bomble, J. F. Matthews, C. B. Taylor, M. G. Resch, J. M. Yarbrough, S. R. Decker, L. Bu, X. Zhao, C. McCabe, J. Wohlert, M. Bergenstråhle, J. W. Brady, W. S. Adney, M. E. Himmel, and M. F. Crowley. The O-glycosylated linker from the *Trichoderma reesei* family 7 cellulase is a flexible, disordered protein. *Biophys. J.*, 99:3773–3781, 2010.
- [140] L. H. F. Lima, V. I. Serpa, F. R. Rosseto, M. O. Neto, G. R. Sartori, L. Martínez, and I. Polikarpov. Small-angle X-ray scattering and structural modeling of full-length: cellobiohydrolase I from *Trichoderma harzianum*. *Cellulose*, 20:1573–1585, 2013.
- [141] D. W. Sammond, C. M. Payne, R. Brunecky, M. E. Himmel, M. F. Crowley, and G. T. Beckham. Cellulase linkers are optimized based on domain type and function: insights from sequence analysis, biophysical measurements, and molecular simulation. *Plos One*, 7:e48615, 2012.
- [142] X. Zhao, T. R. Rignall, C. McCabe, W. S. Adney, and M. E. Himmel. Molecular simulation evidence for processive motion of *Trichoderma reesei* Cel7A during cellulose depolymerization. *Chem. Phys. Lett.*, 460:284–288, 2008.
- [143] I. Stanković. Dinâmica molecular de celulases: estudos de reconhecimento de substrato e propriedades conformacionais. Tese de Doutorado em Química, Instituto de Química – Universidade Estadual de Campinas, Campinas, São Paulo, 2014.
- [144] G. T. Beckham, J. Ståhlberg, B. C. Knott, M. E. Himmel, M. F. Crowley, M. Sandgren, M. Sørlie, and C. M. Payne. Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. *Curr. Opin. Biotechnol.*, 27:96–106, 2014.

- [145] C. Boisset, C. Fraschini, M. Schülein, B. Henrissat, and H. Chanzy. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Appl. Environ. Microb.*, 66:1444–1452, 2000.
- [146] M. Kurasin and P. Valjamae. Processivity of cellobiohydrolases is limited by the substrate. J. Biol. Chem., 286:169–177, 2011.
- [147] P. K. GhattyVenkataKrishna, E. M. Alekozai, G. T. Beckham, R. Shulz, M. F. Crowley, E. C. Uberbacher, and X. Cheng. Initial recognition of a cellodextrin chain in the cellulose-binding tunnel may affect cellobiohydrolase directional specificity. *Biophys. J.*, 104:904–912, 2013.
- [148] A. Nakamura, T. Tsukada, S. Auer, T. Furuta, M. Wada, A. Koivula, K. Igarashi, and M. Samejima. The tryptophan residue at the active site tunnel entrance of *Trichoderma reesei* cellobiohydrolase Cel7A is important for initiation of degradation of crystalline cellulose. J. Biol. Chem., 288:13503–13510, 2013.
- [149] B. C. Knott, M. H. Momeni, M. F. Crowley, L. F. Mackenzie, A. W. Götz, M. Sandgren, S. G. Withers, J. Ståhlberg, and G. T. Beckham. The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. J. Am. Chem. Soc., 136:321–329, 2014.
- [150] J. Li, L. Du, and L. Wang. Glycosidic-bond hydrolysis mechanism catalyzed by cellulase Cel7A from *Trichoderma reesei*: A comprehensive theoretical study by performing MD, QM, and QM/MM calculations. J. Phys. Chem. B, 114:15261–15268, 2010.
- [151] C. B. Barnett, K. A. Wilkinson, and K. J. Naidoo. Molecular details from computational reaction dynamics for the cellobiohydrolase I glycosylation reaction. J. Am. Chem. Soc., 133:19474–19482, 2011.
- [152] S. Yan, T. Li, and L. Yao. Mutational effects on the catalytic mechanism of cellobiohydrolase I from *Trichoderma reesei*. J. Phys. Chem. B, 115:4982– 4989, 2011.
- [153] K. Igarashi, A. Koivula, M. Wada, S. Kimura, M. Pentilla, and M. Samejima. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. J. Biol. Chem., 284:36186–36190, 2009.

- [154] H. Teugjas and P. Valjamae. Product inhibition of cellulases studied by ¹⁴C-labeled cellulose substrates. *Biotechnol. Biofuels*, 6:104, 2013.
- [155] I. von Ossowski, J. Ståhlberg, A. Koivula, K. Piens, D. Becker, H. Boer, R. Harle, M. Harris, C. Divne, S. Mahdi, Y. Zhao, H. Driguez, M. Claeyssens, M. L. Sinnott, and T. T. Teeri. Engineerging the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. a comparison with *Phanerochaete chrysosporium* Cel7D. J. Mol. Biol., 333:817–829, 2003.
- [156] A. M. Castro, K. C. N. R. Pedro, J. C. Cruz, M. C. Ferreira, S. G. F. Leite, and N. Pereira Jr. *Trichoderma harzianum* IOC-4038: a promising strain for the production of a cellulolytic complex with significant β-glucosidase activity from sugarcane bagasse cellulignin. *Appl. Biochem. Biotechnol.*, 162: 2111–2122, 2010.
- [157] B. Benoliel, F. Torres, and L. Moraes. A novel promising *Trichoderma har-zianum* strain for the production of a cellulolytic complex using sugarcane bagasse in natura. *SpringerPlus*, 2:656, 2013.
- [158] M. A. C. Horta, R. Vicentini, P. S. Delabona, P. Laborda, A. Crucello, S. Freitas, R. M. Kuroshu, I. Polikarpov, J. G. C. Pradella, and A. P. Souza. Transcriptome profile of *Trichoderma harzianum* IOC-3844 sugarcane bagasse. *Plos One*, 9:e88689, 2014.
- [159] F. Colussi, V. Serpa, P. S. Delabona, L. R. Manzine, M. L. Voltatodio, R. Alves, B. L. Mello, N. Pereira Jr., C. S. Farinas, A. M. Golubev, M. A. M. Santos, and I. Polikarpov. Purification, and biochemical and biophysical characterization of cellobiohydrolase i from *Trichoderma harzianum* IOC 3844. J. Microb. Biotechnol., 21:808–817, 2011.
- [160] F. Colussi, W. Garcia, F. R. Rosseto, B. L. S. Mello, M. O. Neto, and I. Polikarpov. Effect of pH and temperature on the global compactness, structure, and activity of cellobiohydrolase Cel7A from *Trichoderma harzianum*. *Eur. Biophys. J.*, 41:89–98, 2012.
- [161] R. N. Maeda, V. I. Serpa, V. A. L. Rocha, R. A. A. Mesquita, L. M. M. S. Anna, A. M. Castro, C. E. Driemeier, N. Pereira, and I. Polikarpov. Enzymatic hydrolysis of pretreated sugar cane bagasse using *Penicillium funiculosum* and *Trichoderma harzianum* cellulases. *Proc. Biochem.*, 46:1196–1201, 2011.
- [162] L. C. Textor, F. Colussi, R. L. Silveira, V. Serpa, B. L. Mello, J. R. C. Muniz, F. M. Squina, N. Pereira, M. S. Skaf, and I. Polikarpov. Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from

Trichoderma harzianum: deciphering the structural features of cellobiohydrolase catalytic activity. FEBS J., 280:56–69, 2013.

- [163] E. T. Prates, I. Stankovic, R. L. Silveira, M. V. Liberato, F. Henrique-Silva, N. Pereira Jr., I. Polikarpov, and M. S. Skaf. X-ray structure and molecular dynamics simulations of endogluanase 3 from *Trichoderma harzianum*: structural organization and substrate recognition by endoglucanases that lack cellulose binding module. *Plos One*, 8:e59069, 2013.
- [164] J. Ståhlberg, C. Divne, A. Koivula, K. Piens, M. Claeyssens, T. T. Teeri, and T. A. Jones. Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol., 264: 337–349, 1996.
- [165] M. Claeyssens, H. van Tilbeurgh, P. Tomme, T. M. Wood, and S. I. McRae. Fungal cellulase systems: comparison of the specificities of the cellobiohydrolases isolated from *Penicillium pinophilum* and *Trichoderma reesei*. *Biochem.* J., 261:819–825, 1989.
- [166] S. P. Voutilainen, T. Puranen, M. Siika-aho, A. Lappalainen, M. Alapuranen, J. Kallio, S. Hooman, L. Viikari, J. Vehmaanpera, and A. Koivula. Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnol. Bioeng.*, 101:515–528, 2008.
- [167] S. Vonhoff, K. Piens, M. Pipelier, C. Braet, M. Claeyssens, and A. Vasella. Inhibition of cellobiohydrolases from *Trichoderma reesei*. Synthesis and evaluation of some glucose-, cellobiose-, and cellotriose-derived hydroximolactams and imidazoles. *Helv. Chimi. Acta*, 82:963–980, 1999.
- [168] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, 2000.
- [169] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids. Res.*, 33:W368–W371, 2005.
- [170] J. Myers, G. Grothaus, S. Narayanan, and A. Onufriev. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins*, 63:928–938, 2006.
- [171] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. J. Comput. Chem., 26:1781–1802, 2005.

- [172] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B, 102:3586–3616, 1998.
- [173] A. D. Mackerell, M. Feig, and C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comput. Chem., 25:1400–1415, 2004.
- [174] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys., 79:926–935, 1983.
- [175] I. G. Muñoz, W. Ubhayasekera, H. Henriksson, I. Szabó, G. Petterson, G. Johansson, S. L. Mowbray, and J. Ståhlberg. Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. J. Mol. Biol., 314:1097–1111, 2001.
- [176] A. Grassick, P. G. Murray, R. Thompson, C. M. Collins, L. Byrnes, G. Birrane, T. M. Higgins, and M. G. Tuohy. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the *cel7* gene from the filamentous fungus, *Talaromyces emersonii*. *Eur. J. Biochem.*, 271:4495–4506, 2004.
- [177] M. G. Tuohy, D. J. Walsh, P. G. Murray, M. Claeyssens, M. M. Cuffe, A. V. Savage, and M. P. Coughlan. Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. Biochim. Biophys. Acta, 1596:366–380, 2002.
- [178] C. M. Payne, W. Jiang, M. R. Shirts, M. E. Himmel, M. F. Crowley, and G. T. Beckham. Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. J. Am. Chem. Soc., 135:18831–18839, 2013.
- [179] A. Nakamura, H. Watanabe, T. Ishida, T. Uchihashi, M. Wada, T. Ando, K. Igarashi, and M. Samejima. Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose. J. Am. Chem. Soc., 136:4584–4592, 2014.

- [180] J. Jalak and P. Väljamäe. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnol. Bioeng.*, 106: 871–883, 2010.
- [181] K. Igarashi, T. Uchihashi, A. Koivula, M. Wada, S. Kimura, T. Okamoto, M. Penttila, T. Ando, and M. Samejima. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science*, 333:1279–1282, 2011.
- [182] S. P. Voutilainen, H. Boer, M. Alapuranen, J. Jänis, J. Vehmaanperä, and A. Koivula. Improving the thermostability and activity of *Melanocarpus albomyces* cellobiohydrolase Cel7B. *Appl. Microbiol. Biotechnol.*, 83:261–272, 2009.
- [183] T. Parkkinen, A. Koivula, J. Vehmaanpera, and J. Rouvinen. Crystal structures of *Melanocarpus albomyces* cellobiohidrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding. *Protein Sci.*, 17:1383–1394, 2008.
- [184] J. Ståhlberg, G. Johansson, and G. Pettersson. *Trichoderma reesei* has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. *Biochim. Biophys. Acta*, 1157:107–113, 1993.
- [185] Y. Amano, M. Shiroishi, K. Nisizawa, E. Hoshino, and T. Kanda. Fine substrate specificities of four exo-type cellulases produced by *Aspergillus niger*, *Trichoderma reesei*, and *Irpex lacteus* on $(1 \rightarrow 3), (1 \rightarrow 4)$ - β -D-glucans and xyloglucan. J. Biochem., 120:1123–1129, 1996.
- [186] S. Armand, S. Drouillard, M. Schülein, B. Henrissat, and H. Driguez. A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. J. Biol. Chem., 272:2709–2713, 1997.
- [187] A. Varrot, M. Schülein, and G. J. Davies. Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry*, 38:8884–8891, 1999.
- [188] A. Varrot, T. P. Frandsen, I. von Ossowski, V. Boyer, S. Cottaz, H. Driguez, M. Schülein, and G. J. Davies. Structural basis for ligand binding and processivity in cellobiohydrolase Cel6A from *Humicola insolens*. Structure, 11: 855–864, 2003.
- [189] M. Kern, J. E. McGeehan, S. D. Streeter, R. N. A. Martin, K. Besser, L. Elias, W. Eborall, G. P. Malyon, C. M. Payne, M. E. Himmel, K. Schnorr, G. T. Beckham, S. M. Cragg, N. C. Bruce, and S. J. McQueen-Mason. Structural
characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proc. Natl. Acad. Sci. USA*, 110:10189–10194, 2013.

- [190] L. Bu, M. F. Crowley, M. E. Himmel, and G. T. Beckham. Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. J. Biol. Chem., 288:12175–12186, 2013.
- [191] S. V. Pingali, H. M. O'Neill, J. McGaughey, V. S. Urban, C. S. Rempe, L. Petridis, J. C. Smith, B. R. Evans, and W. T. Heller. Small angle neutron scattering reveals pH-dependent conformational changes in *Trichoderma reesei* cellobiohidrolase I. J. Biol. Chem., 286:32801–32809, 2011.
- [192] J. M. Fox, S. E. Levine, D. S. Clark, and H. W. Blanch. Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry*, 51:422–452, 2012.
- [193] O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor, and A. D. MacKerell. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. J. Chem. Theory Comput., 5:2353–2370, 2009.
- [194] A. Koivula, T. Kinnari, V. Harjunpää, L. Ruohonen, A. Teleman, T. Drakenberg, J. Rouvinen, T. A. Jones, and T. T. Teeri. Tryptophan 272: an essential determinant of crystalline cellulose degradation by *Trichoderma re*esei cellobiohydrolase Cel6A. *FEBS Lett.*, 429:341–346, 1998.
- [195] A. Leo-Macias, P. Lopez-Romero, D. Lupyan, D. Zerbino, and A. R. Ortiz. An analysis of core deformations in protein superfamilies. *Biophys. J.*, 88: 1291–1299, 2005.
- [196] A. R. Altigan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, 80:505–515, 2001.
- [197] D. Tobi and I. Bahar. Structural changes involved in protein binding correlate with intrinsic motions of protein in the unbound state. *Proc. Natl. Acad. Sci.* USA, 102:18908–18913, 2005.
- [198] S. E. Dobbins, V. I. Lesk, and M. J. E. Sternberg. Insights into protein flexibility: the relatioship between normal modes and conformational changes upon protein-protein docking. *Proc. Natl. Acad. Sci. USA*, 105:10390–10395, 2008.

- [199] M. Gur, E. Zomot, and I. Bahar. Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions. J. Chem. Phys., 139:121912, 2013.
- [200] D. E. Koshland. Application of a theory of enzyme specificity to protein synthesis. Proc. Natl. Acad. Sci. USA, 44:98–104, 1958.
- [201] A. Bakan and I. Bahar. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci. USA*, 106:14349–14354, 2009.
- [202] E. Fisher. Einfluss der configuration auf die wirkung der enzyme. Ber. Disch. Chem. Ges., 27:2984–2993, 1894.
- [203] L. Bu, M. R. Nimlos, M. R. Shirts, J. Ståhlberg, M. E. Himmel, M. F. Crowley, and G. T. Beckham. Product binding varies dramatically between processive and nonprocessive cellulase enzymes. J. Biol. Chem., 287:24807– 24813, 2012.
- [204] H. Teugjas and P. Valjamae. Selecting β -glucosidases to support cellulases in cellulose saccharification. *Biotechnol. Biofuels*, 6:105, 2013.
- [205] A. Berlin, N. Gilkes, D. Kilburn, R. Bura, A. Markov, A. Skomarovsky, O. Okunev, A. Gusakov, V. Maximenko, D. Gregg, A. Sinitsyn, and J. Saddler. Evaluation of novel fungal cellulase preparations for ability to hydrolyze softwood substrates – evidence for the role of accessory enzymes. *Enzyme* and Microb. Technol., 37:175–184, 2005.
- [206] C. B. Taylor, C. M. Payne, M. E. Himmel, M. F. Crowley, C. McCabe, and G. T. Beckham. Binding site dynamics and aromatic-carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases. J. Phys. Chem. B, 117:4924–4933, 2013.
- [207] C. M. Payne, Y. J. Bomble, C. B. Taylor, C. McCabe, M. E. Himmel, M. F. Crowley, and G. T. Beckham. Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. J. Biol. Chem., 286:41028–41035, 2011.
- [208] G. J. Davies, K. S. Wilson, and B. Henrissat. Nomenclature of sugar-binding subsites in glycosyl hydrolases. *Biochem. J.*, 321:557–559, 1997.
- [209] L. Bu, G. T. Beckham, M. R. Shirts, M. R. Nimlos, W. S. Adney, M. E. Himmel, and M. F. Crowley. Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. J. Biol. Chem., 286:18161–18169, 2011.

- [210] W. Ubhayasekera, I. G. Muñoz, A. Vasella, J. Ståhlberg, and S. L. Mowbray. Structures of *Phanerochaete chrysosporium* Cel7D in complex with product and inhibitors. *FEBS J.*, 272:1952–1964, 2005.
- [211] M. Wu, L. Bu, T. V. Vuong, D. B. Wilson, M. F. Crowley, M. Sandgren, J. Ståhlberg, G. T. Beckham, and H. Hansson. Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. J. Biol. Chem., 288:33107–33117, 2013.
- [212] V. V. Zverlov, I. Y. Volkov, T. V. Velikodvorskaya, and W. H. Schwarz. Highly thermostable endo-1,3-β-glucanase (laminarinase) LamA from *Thermotoga neapolitana*: nucleotide sequence of the gene and characterization of the recombinant gene product. *Microbiology*, 143:1701–1708, 1997.
- [213] L. Bleicher, E. T. Prates, T. C. F. Gomes, R. L. Silveira, A. S. Nascimento, A. L. Rojas, A. Golubev, L. Martínez, M. S. Skaf, and I. Polikarpov. Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. J. Phys. Chem. B, 115:7940–7949, 2011.
- [214] J. Georis, F. L. Esteves, J. Lamotte-Brasseur, V. Bougnet, B. Devreese, F. Giannotta, B. Granier, and J.-M. Frère. An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study. *Protein Sci.*, 9:466–475, 2000.
- [215] L. Viikari, M. Alapuranen, T. Puranen, J. Vehmaanperä, and M. Siika-aho. Thermostable enzymes in lignocellulose hydrolysis. Adv. Biochem. Engin. Biotechnol., 108:121–145, 2007.
- [216] I. Asial, Y. X. Cheng, H. Engman, M. Dollhopf, B. Wu, P. Nordlund, and T. Cornvik. Engineering protein thermostability using a generic activityindependent biophysical screen inside the cell. *Nat. Comm.*, 4:2901, 2013.
- [217] B. van den Burg and V. G. H. Eijsink. Selection of mutations for increased protein stability. *Curr. Opin. Biotechnol.*, 13:333–337, 2002.
- [218] M. Lehmann and M. Wyss. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.*, 12:371–375, 2001.

- [219] C. Dumon, A. Varvak, M. A. Wall, J. E. Flint, R. J. Lewis, J. H. Lakey, C. Morland, P. Luginbühl, S. Healey, T. Todaro, G. DeSantis, M. Sun, L. Parra-Gessert, X. Tan, D. P. Weiner, and H. J. Gilbert. Engineering hyperthermostability into GH11 xylanase is mediated by subtle changes to protein structure. J. Biol. Chem., 283:22557–22564, 2008.
- [220] A. Szilágyi and P. Závodszky. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, 8:493 – 504, 2000.
- [221] R. Fonseca-Maldonado, D. S. Vieira, J. S. Alponti, E. Bonneil, P. Thibault, and R. J. Ward. Engineering the pattern of protein glycosylation modulates the thermostability of a GH11 xylanase. J. Biol. Chem., 288:25522–25534, 2013.
- [222] Y. S. Cheng, C. C. Chen, C. H. Huang, T. P. Ko, W. Luo, J. W. Huang, J. R. Liu, and R. T. Guo. Structural analysis of a glycoside hydrolase family 11 xylanase from *Neocallimastix patriciarun*: insights into the molecular basis of a thermophilic enzyme. J. Biol. Chem., early:early, 2014.
- [223] C. R. Santos, J. H. Paiva, M. L. Sforça, J. L. Neves, R. Z. Navarro, J. Cota, P. K. Akao, Z. B. Hoffmam, A. N. Meza, J. H. Smetana, M. L. Nogueira, I. Polikarpov, J. Xavier-Neto, F. M. Squina, R. J. Ward, R. Ruller, A. C. Zeri, and M. T. Murakami. Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochem. J.*, 441:95–104, 2012.
- [224] K. Manjunath and K. Sekar. Molecular dynamics perspective on the protein thermal stability: A case study using SAICAR synthetase. J. Chem. Inf. Model., 53(9):2448–2461, 2013.
- [225] A. G. Alfredsson, J. K. Kristjansson, S. Hjorleifsdottir, and K. O. Stetter. *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in iceland. *J. Gen. Microbiol.*, 134:299–306, 1988.
- [226] S. H. Bjornsdottir, T. Blondal, G. O. Hreggvidsson, G. Eggertsson, S. Petursdottir, S. Hjorleifsdottir, S. H. Thorbjarnardottir, and J. Kristjansson. *Rhodothermus marinus*: physiology and molecular biology. *Extremophiles*, 10:1–16, 2006.
- [227] M. Krah, R. Misselwitz, O. Politz, K. K. Thomsen, H. Welfle, and R. Borriss. The laminarinase from the thermophilic eubacterium *Rhodothermus*

marinus: conformation, stability, and identification of active site carboxylic residues by site-directed mutagenesis. *Eur. J. Biochem.*, 257:101–111, 1998.

- [228] H. Zheng and H. Wu. Gene-centric association analysis for the correlation between the guanine-cytosine content levels and temperature range conditions of prokaryotic species. *BMC Bioinf.*, 11(Suppl 11):S7, 2010.
- [229] J. Vasur, R. Kawai, K. H. M. Jonsson, G. Widmalm, A. Engström, M. Frank, E. Andersson, H. Hansson, Z. Forsberg, K. Igarashi, M. Samejima, M. Sandgren, and J. Ståhlberg. Synthesis of cyclic β-glucan using laminarinase 16A glycosynthase mutant from the basidiomycete *Phanerochaete chrysosporium*. J. Am. Chem. Soc., 132:1724–1730, 2010.
- [230] R. Kawai, K. Igarashi, M. Yoshida, M. Kitaoka, and M. Samejima. Hydrolysis of β-1,3/1,6-glucan by glycoside hydrolase family 16 endo-1,3(4)-β-glucanase from basidiomicete *Phanerochaete chrysosporium*. Appl. Microbiol. Biotechnol., 71:898–906, 2006.
- [231] G. Fibriansah, S. Masuda, N. Koizumi, S. Nakamura, and T. Kumasaka. The 1.3 Å crystal structure of a novel endo-β-1,3-glucanase of glycoside hydrolase family 16 from alkaliphilic *Nocardiopsis sp.* strain f96. *Proteins*, 69:683–690, 2007.
- [232] S. Masuda, K. Endo, N. Koizumi, T. Hayami, T. Fukazawa, R. Yatsunami, T. Fukui, and S. Nakamura. Molecular identification of a novel β-1,3-glucanase from alkaliphilic Nocardiopsis sp. strain F96. Extremophiles, 10:251–255, 2006.
- [233] J. Vasur, R. Kawai, A. M. Larsson, K. Igarashi, M. Sandgren, M. Samejima, and J. Ståhlberg. X-ray crystallographic native sulfur SAD structure determination of laminarinase Lam16A from *Phanerochaete chrysosporium*. Acta Cryst. D, 62:1422–1429, 2006.
- [234] D. Perl, U. Mueller, U. Heinemann, and F. X. Schmid. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.*, 7:380–383, 2000.
- [235] E. T. Prates. Dinâmica molecular de hidrolases para sacarificação de celulose e proteínas correlatas. Tese de Doutorado em Química, Instituto de Química – Universidade Estadual de Campinas, Campinas, São Paulo, 2013.
- [236] T. C. F. Gomes. Simulações por dinâmica molecular de sistemas biológicos relacionados à hidrólise de biomassa lignocelulósica. Tese de Doutorado em

Química, Instituto de Química – Universidade Estadual de Campinas, Campinas, São Paulo, 2013.

- [237] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. Packmol: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.*, 30:2157–2164, 2009.
- [238] M. T. Murakami, R. K. Arni, D. S. Vieira, L. Degrève, R. Ruller, and R. J. Ward. Correlation of temperature induced conformation change with optimum catalytic activity in the recombinant G/11 xylanase A from bacillus subtilis strain 168 (1a1). FEBS Lett., 579:6505–6510, 2005.
- [239] D. S. Vieira, L. Degrève, and R. J. Ward. Characterization of temperature dependent and substrate-binding cleft movements in *Bacillus circulans* family 11 xylanase: a molecular dynamics investigation. *Biochem. Biophys. Acta*, 1790:1301–1306, 2009.
- [240] D. S. Vieira and R. J. Ward. Conformation analysis of a surface loop that controls active site access in the GH11 xylanase A from *Bacillus subtilis*. J. Mol. Model., 18:1473–1479, 2012.
- [241] K. N. Neustroev, A. M. Golubev, M. L. Sinnott, R. Borriss, M. Krah, H. Brumer III, E. V. Eneyskaya, S. Shishlyannikov, K. A. Shabalin, V. T. Peshechonov, V. G. Korolev, and A. A. Kulminskaya. Transferase and hydrolytic activities of the laminarinase from *Rhodothermus marinus* and its M133A, M133C and M133W mutants. *Glycoconj. J*, 23:501–511, 2006.
- [242] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol., 55:379–400, 1971.
- [243] F. M. Richards. Areas, volumes, packing and protein structure. Annu. Rev. Biophys. Bioeng., 6:151–176, 1977.
- [244] D. J. Cosgrove. Loosening of plant cell walls by expansins. *Nature*, 407: 321–326, 2000.
- [245] D. J. Cosgrove. Growth of plant cell walls. Nat. Rev. Mol. Cell Biol., 6: 850–861, 2005.
- [246] D. Choi, H. T. Cho, and Y. Lee. Expansins: expanding importance in plant growth and development. *Physiol. Plant*, 126:511–518, 2006.
- [247] S. J. McQueen-Mason and D. J. Cosgrove. Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. *Proc. Natl. Acad. Sci. USA*, 91:6574–6578, 1994.

- [248] S. J. McQueen-Mason and D. J. Cosgrove. Expansin mode of action on cell walls. analysis of wall hydrolysis, stress relaxation, and binding. *Plant Physiol.*, 107:87–100, 1995.
- [249] D. J. Cosgrove. Relaxation in a high-stress environment: the molecular bases of extendable cell walls and cell enlargement. *Plant Cell*, 9:1031–1041, 1997.
- [250] S. E. C. Whitney, M. J. Gildey, and S. J. McQueen-Mason. Probing expansin action using cellulose/hemicellulose composites. *Plant J.*, 22:327–334, 2000.
- [251] D. A. Brummell, M. H. Harpster, P. M. Civello, J. M. Palys, A. B. Bennett, and P. Dunsmuir. Modification of expansin protein abundance in tomato fruit alters softening and cell wall polymer metabolism during ripening. *Plant Cell*, 11:2203–2216, 1999.
- [252] P. M. Civello, A. L. T. Powell, A. Sabehat, and A. B. Bennett. An expansin gene expressed in ripening strawberry fruit. *Plant Physiol.*, 121:1273–1279, 1999.
- [253] E. J. Belfield, B. Ruperti, J. A. Roberts, and S. J. McQueen-Mason. Changes in expansin activity and gene expression during ethylene-promoted leaflet abscission in *Sambucus nigra. J. Exp. Bot.*, 56:817–823, 2005.
- [254] H. -T. Cho and D. J. Cosgrove. Regulation of root hair initiation and expansin gene expression in arabidopsis. *Plant Cell*, 14:3237–3253, 2002.
- [255] F. Chen, P. Dahal, and K. J. Bradford. Two tomato expansin genes show divergent expression and localization in embryos during seed development and germination. *Plant Physiol.*, 127:928–936, 2001.
- [256] H. T. Cho and D. J. Cosgrove. Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA, 97:9783–9788, 2000.
- [257] D. Choi, Y. Lee, H. T. Cho, and H. Kende. Regulation of expansin gene expression affects growth and development in transgenic rice plants. *Plant Cell*, 15:1383–1398, 2003.
- [258] S. Pien, J. Wyrzykowska, S. J. McQueen-Mason, C. Smart, and A. Fleming. Local expression of expansin induces the entire process of leaf development and modifies leaf shape. *Proc. Natl. Acad. Sci. USA*, 98:11812–11817, 2011.
- [259] A. J. Fleming, S. McQueen-Mason, T. Mandel, and C. Kuhlemeier. Induction of leaf primordia by the cell wall protein expansin. *Science*, 276:1415–1418, 1997.

- [260] F. Kerff, A. Amoroso, R. Herman, E. Sauvage, S. Petrella, P. Filée, P. Charlier, B. Joris, A. Tabuchi, N. Nikolaidis, and D. J. Cosgrove. Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proc. Natl. Acad. Sci. USA*, 105:16876–16881, 2008.
- [261] H. J. Lee, S. Lee, H. Ko, K. H. Kim, and I. Choi. An expansin-like protein from *Hahella chejuensis* binds cellulose and enhances cellulase activity. *Mol. Cells*, 29:379–385, 2010.
- [262] N. Georgelis, N. Nikolaidis, and D. J. Cosgrove. Biochemical analysis of expansin-like proteins from microbes. *Carb. Polymers*, 100:17–23, 2014.
- [263] M. Saloheimo, M. Paloheimo, S. Hakola, J. Pere, B. Swanson, E. Nyyssonen, A. Bhatia, M. Ward, and M. Penttila. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *Eur. J. Biochem.*, 269:4202–4211, 2002.
- [264] N. Georgelis, A. Tabuchi, N. Nikolaidis, and D. J. Cosgrove. Structurefunction analysis of the bacterial expansin EXLX1. J. Biol. Chem., 286: 16814–16823, 2011.
- [265] N. H. Yennawar, L. C. Li, D. M. Dudzinski, A. Tabuchi, and D. J. Cosgrove. Crystal structure and activities of EXPB1 (Zea m 1), a β-expansin and group-1 pollen allergen from maize. *Proc. Natl. Acad. Sci. USA*, 103:14664–14671, 2006.
- [266] G. J. Davies, G. G. Dodson, R. E. Hubbard, S. P. Tolley, Z. Dauter, K. S. Wilson, C. Hjort, J. M. Mikkelsen, G. Rasmussen, and M. Schulein. Structure and function of endoglucanase V. *Nature*, 365:362–364, 1993.
- [267] I. J. Kim, H. J. Ko, T. W. Kim, I. G. Choi, and K. H. Kim. Characteristics of the binding of a bacterial expansin (BsEXLX1) to microcrystalline cellulose. *Biotechnol. Bioeng.*, 110:401–407, 2013.
- [268] T. Wang, Y. B. Park, M. A. Caporini, M. Rosay, L. Zhong, D. J. Cosgrove, and M. Hong. Sensitivity-enhanced solid-state NMR detection of expansin's target in plant cell walls. *Proc. Natl. Acad. Sci. USA*, 110:16444–16449, 2013.
- [269] Y. B. Park and D. J. Cosgrove. Changes in cell wall biomechanical properties in the xyloglucan-deficient xxt1/xxt2 mutant of arabidopsis. *Plant Physiol.*, 158:465–475, 2012.

- [270] T. C. F. Gomes and M. S. Skaf. Cellulose-Builder: a toolkit for building crystalline structures of cellulose. J. Comput. Chem., 33:1338–1346, 2012.
- [271] E. M. Alekozai, P. K. GhattyVenkataKrishna, E. C. Uberbacher, M. F. Crowley, J. C. Smith, and X. Cheng. Simulation analysis of the cellulase Cel7A carbohydrate binding module on the surface of the cellulose Iβ. *Cellulose*, 21:951–971, 2014.
- [272] J. Wohlert and L. A. Berglund. A coarse-grained model for molecular dynamics simulations of native cellulose. J. Chem. Theory Comput., 7:753–560, 2011.
- [273] M. Olarte-Lozano, M. A. Mendonza-Nuñez, N. Pastor, L. Segovia, J. Folch-Mallol, and C. Martínez-Anaya. PcExl1 a novel acid expansin-like protein from the plant pathogen *Pectobacterium carotovorum*, binds cell walls differently to BsEXLX1. *Plos One*, 9:e95638, 2014.
- [274] G. Duan, V. H. Smith, and D. F. Weaver. Characterization of aromatic-amide (side-chain) interactions in proteins through systematic ab initio calculations and data mining analyses. J. Phys. Chem. A, 104:4521–4532, 2000.
- [275] N. Georgelis, N. Yennawar, and D. J. Cosgrove. Structural basis for entropydriven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc. Natl. Acad. Sci. USA*, 109:14830–14935, 2012.
- [276] A. L. Creagh, E. Ong, E. Jervis, D. G. Kilburn, and C. A. Haynes. Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl. Acad. Sci. USA*, 93:12229–12234, 1996.
- [277] H. D. Ly and S. G. Withers. Mutagenesis of glycosidades. Annu. Rev. Biochem., 68:487–522, 1999.
- [278] J. D. DeMartini, S. Pattathil, J. S. Miller, H. Li, M. G. Hahn, and C. E. Wyman. Investigating plant cell wall componentes that affect biomass recalcitrance in poplar and switchgrass. *Energy Environ. Sci*, 6:898–909, 2013.
- [279] J. P. Knox. Revealing the structural and functional divesity of plant cell walls. *Curr. Opin. Plant Biol.*, 11:308–313, 2008.
- [280] Y. B. Park and D. J. Cosgrove. A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. *Plant Physiol.*, 158:1933–1943, 2012.

- [281] I. Burgert and T. Keplinger. Plant micro- and nanomechanics: experimental techniques for plant cell-wall analysis. J. Exp. Bot., 64:4617–4633, 2013.
- [282] J. A. Rollin, Z. Zhu, N. Sathitsuksanoh, and Y. H. P. Zhang. Increasing cellulose accessibility is more important than removing lignin: a comparison of cellulose solvent-based lignocellulosic fractionation and soaking in aqueous ammonia. *Biotechnol. Bioeng.*, 108:22–30, 2011.
- [283] N. Xu, W. Zhang, S. Ren, F. Liu, C. Zhao, H. Liao, Z. Xu, J. Huang, Q. Li, Y. Tu, B. Yu, Y. Wang, J. Jiang, J. Qin, and L. Peng. Hemicelluloses negatively affect lignocellulose crystallinity for high biomass digestibility under NaOH and H₂SO₄ pretreatments in *Miscanthus. Biotechnol. Biofuels*, 5:58, 2012.
- [284] T. Jeoh, C. I. Ishizawa, M. F. Davis, M. E. Himmel, W. S. Adney, and D. K. Johnson. Cellulase digestibility of pretreated biomass is limited by cellulose accessibility. *Biotechnol. Bioeng.*, 98:112–122, 2007.
- [285] T. Wang, O. Zabotina, and M. Hong. Pectinâcellulose interactions in the Arabidopsis primary cell wall from two-dimensional magic-angle-spinning solidstate nuclear magnetic resonance. *Biochemistry*, 51:9846–9856, 2012.
- [286] M. C. Stumpe, N. Blinov, D. Wishart, A. Kovalenko, and V. S. Pande. Calculation of local water densities in biological systems: a comparison of molecular dynamics simulations and the 3D-RISM-KH molecular theory of solvation. J. Phys. Chem. B, 115:319–328, 2011.
- [287] A. Kovalenko, S. Ten-no, and F. Hirata. Solution of the three-dimensional RISM/HNC equations for spc water by the modified method of direct inversion in the iterative subspace. J. Comput. Chem., 20:928–936, 1999.
- [288] Woods Group. (2005-2014) GLYCAM Web. Complex carbohydrate research center. University of Georgia, Athens, GA. (http://www.glycam.com).
- [289] U. W. Schmitt and G. A. Voth. The computer simulation of proton transport in water. J. Chem. Phys., 111:9361–9381, 1999.
- [290] M. F. Crowley, M. J. Williamson, and R. C. Walker. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. *Int. J. Quantum Chem.*, 109:3767–3772, 2009.
- [291] C. T. Anderson, A. Carroll, L. Akhmetova, and C. Somerville. Real-time imaging of cellulose reorientation during cell wall expansion in Arabidopsis roots. *Plant Physiol.*, 152:787–796, 2010.

- [292] Y. B. Park, C. M. Lee, B.-W. Koo, S. Park, D. J. Cosgrove, and S. H. Kim. Monitoring meso-scale ordering of cellulose in intact plant cell walls using sum frequency generation spectroscopy. *Plant Physiol.*, 163:907–913, 2013.
- [293] F. Marga, M. Grandbois, D. J. Cosgrove, and T. I. Baskin. Cell wall extension results in the coordinate separation of parallel microfibrils: evidence from scanning electron microscopy and atomif force microscopy. *Plant J.*, 43:181– 190, 2005.
- [294] A. N. Fernandes, L. H. Thomas, C. M. Altaner, P. Callow, V. T. Forsyth, D. C. Apperley, C. J. Kennedy, and M. C. Jarvis. Nanostructure of cellulose microfibrils in spruce wood. *Proc. Natl. Acad. Sci. USA*, 108:E1198–E1203, 2011.
- [295] C. Lee, Q. Teng, W. Huang, R. Zhong, and Z.-H. Ye. Down-regulation of PoGT47C expression in poplar results in a reduced glucuroxylan content and an increased wood digestibility by cellulase. *Plant Cell Physiol.*, 50: 1075–1089, 2009.
- [296] M. Lopez, H. Bizot, G. Chambat, M. F. Marais, A. Zykwinska, M.-C. Ralet, H. Driguez, and A. Buléon. Enthalpic studies of xyloglucan-cellulose interactions. *Biomacromolecules*, 11:1417–1428, 2010.
- [297] K. Leppanen, S. Andersson, M. Torkkeli, M. Knaapila, N. Kotelnikova, and R. Serimaa. Structure of cellulose and microcrystalline cellulose from various wood species, cotton and flax studied by X-ray scattering. *Cellulose*, 16:999– 1015, 2009.
- [298] S. Lu, L. Li, and G. Zhou. Genetic modification of wood quality for secondgeneration biofuel production. *GM Crops*, 1:230–236, 2010.
- [299] F. J. Ruiz-Dueñas and A. T. Martínez. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.*, 2:164–177, 2009.
- [300] A. Berlin, M. Balakshin, N. Gilkes, J. Kadla, V. Maximenko, S. Kubo, and J. Saddler. Inhibition of cellulose, xylanase and β-glucosidase activities by softwood lignin preparations. J. Biotechnol., 125:198–209, 2006.
- [301] R. Martín-Sampedro, J. L. Rahikainen, L. S. Johansson, K. Marjamaa, J. Laine, K. Kruus, and O. J. Rojas. Preferential adsorption and activity of monocomponent cellulases on lignocellulose thin films with varying lignin content. *Biomacromolecules*, 14:1231–1239, 2013.

- [302] M. A. Lima, M. Oliveira-Neto, M. A. Kadowaki, F. R. Rosseto, E. T. Prates, F. M. Squina, A. F. Leme, M. S. Skaf, and I. Polikarpov. Aspergillus niger β-glucosidase has a cellulase-like tadpole molecular shape: insights into GH3 beta-glucosidases structure and function. J. Biol. Chem., 288:32991–33005, 2013.
- [303] F. Chen and R. A. Dixon. Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.*, 25:759–761, 2007.
- [304] L. A. Jackson, G. L. Shadle, R. Zhou, J. Nakashima, F. Chen, and R. A. Dixon. Improving saccharification efficiency of alfafa stems through modification of the terminal stages of monolignol biosynthesis. *Bioenerg. Res.*, 1: 180–192, 2008.
- [305] R. Vanholme, I. Cesarino, K. Rataj, Y. Xiao, L. Sundin, G. Goeminne, H. Kim, J. Cross, K. Morreel, P. Araujo, L. Welsh, J. Hustraete, C. Mc-Clellan, B. Vanholme, J. Ralph, G. G. Simpson, C. Halpin, and W. Boerjan. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway. *Science*, 341:1103–1106, 2013.
- [306] M. H. Studer, J. D. DeMartíni, M. F. Davis, R. W. Sykes, B. Davison, M. Keller, G. A. Tuskan, and C. E. Wyman. Lignin content in natural *Populus* variants affests sugar release. *Proc. Natl. Acad. Sci. USA*, 108: 6300–6305, 2011.
- [307] L. Petridis, S. V. Pingali, V. Urban, W. T. Heller, H. M. O'Neill, M. Foston, A. Ragauskas, and J. C. Smith. Self-similar multiscale structure of lignin revealed by neutron scattering and molecular dynamics simulations. *Phys. Rev. E*, 83:061911, 2011.
- [308] L. Petridis, R. Schulz, and J. C. Smith. Simulation analysis of the temperature dependence of lignin structure and dynamics. J. Am. Chem. Soc., 133: 20277–20287, 2011.
- [309] B. Lindner, L. Petridis, R. Schulz, and J. C. Smith. Solvent-driven preferential association of lignin with regions of crystalline cellulose in molecular dynamics simulation. *Biomacromolecules*, 14:3390–3398, 2013.
- [310] L. Charlier and K. Mazeau. Molecular modeling of the structural and dynamical properties of secondary plant cell walls: influence of lignin chemistry. J. Phys. Chem. B, 116:4163–4174, 2012.
- [311] L. Petridis and J. C. Smith. A molecular mechanics force field for lignin. J. Comput. Chem., 30:457–467, 2009.

- [312] J. H. Hung, W. M. Fouad, W. Vermerris, M. Gallo, and F. Altpeter. RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. *Plant Biotechnol. J.*, 10:1067– 1076, 2012.
- [313] X. Li, J. K. Weng, and C. Chapple. Improvement of biomass through lignin modification. *Plant J.*, 54:569–581, 2008.
- [314] S. D. Mansfield. Solutions for dissolution engineering cell walls for deconstruction. Curr. Opin. Biotechnol., 20:286–294, 2009.
- [315] J. D. DeMartini, S. Pattathil, U. Avci, K. Szekalski, K. Mazumder, M. G. Hahn, and C. E. Wyman. Application of monoclonal antibodies to investigate plant cell wall deconstruction for biofuels production. *Energy Environ. Sci.*, 4:4332–4339, 2011.
- [316] S. Lv, Q. Yu, X. Zhuang, Z. Yuan, W. Wang, Q. Wang, W. Qi, and X. Tan. The influence of hemicellulose and lignin removal on the enzymatic digestibility from sugarcane bagasse. *Bioenerg. Res.*, 6:1128–1134, 2013.
- [317] S. Persson, K. H. Caffall, G. Freshour, M. T. Hilley, S. Bauer, P. Poindexter, M. G. Hahn, D. Mohnen, and C. Somerville. The Arabidopsis irregular xylem8 mutant is deficient in glucuroxylan and homogalacturonan, which are essential for secondary cell wall integrity. *Plant Cell*, 19:237–255, 2007.
- [318] S. Sarupria and S. Garde. Quantifying water density fluctuations and compressibility of hydration shells of hydrophobic solutes and proteins. *Phys. Rev. Lett.*, 103:037803, 2009.
- [319] S. N. Jamadagni, R. Godawat, and S. Garde. Hydrophobicity of proteins and interfaces: insights from density fluctuations. Annu. Rev. Chem. Biomol. Eng., 2:147–171, 2011.
- [320] E. Freire, O. L. Mayorga, and M. Straume. Isothermal titration. Anal. Chem., 62:950A-959A, 1990.