



THIAGO COSTA FERREIRA GOMES

SIMULAÇÕES POR DINÂMICA MOLECULAR DE
SISTEMAS BIOLÓGICOS RELACIONADOS À
HIDRÓLISE DE BIOMASSA LIGNOCELULÓSICA

CAMPINAS

2013



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA

THIAGO COSTA FERREIRA GOMES

SIMULAÇÕES POR DINÂMICA MOLECULAR DE
SISTEMAS BIOLÓGICOS RELACIONADOS À
HIDRÓLISE DE BIOMASSA LIGNOCELULÓSICA

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 THIAGO
COSTA FERREIRA GOMES, E ORIENTADA PELO PROF. DR. MUNIR SALOMÃO SKAF.

Assinatura do Orientador

CAMPINAS

2013

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Química
Danielle Dantas de Sousa - CRB 8/6490

G585s Gomes, Thiago Costa Ferreira, 1983-
Simulações por dinâmica molecular de sistemas biológicos relacionados à hidrólise de biomassa lignocelulósica / Thiago Costa Ferreira Gomes. – Campinas, SP : [s.n.], 2013.

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

1. Dinâmica molecular. 2. Glicosideo hidrolases. 3. Bioetanol. 4. Celulose. 5. Hemicelulose. I. Skaf, Munir Salomão. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Biological systems related to lignocellulosic biomass hydrolysis studied by means of molecular dynamics simulations

Palavras-chave em inglês:

Molecular dynamics

Glycoside hydrolases

Bioethanol

Cellulose

Hemicellulose

Área de concentração: Físico-Química

Titulação: Doutor em Ciências

Banca examinadora:

Munir Salomão Skaf [Orientador]

Leandro Martínez

Pedro Antônio Muniz Vazquez

Gustavo de Miranda Seabra

Charles Rubber de Almeida Abreu

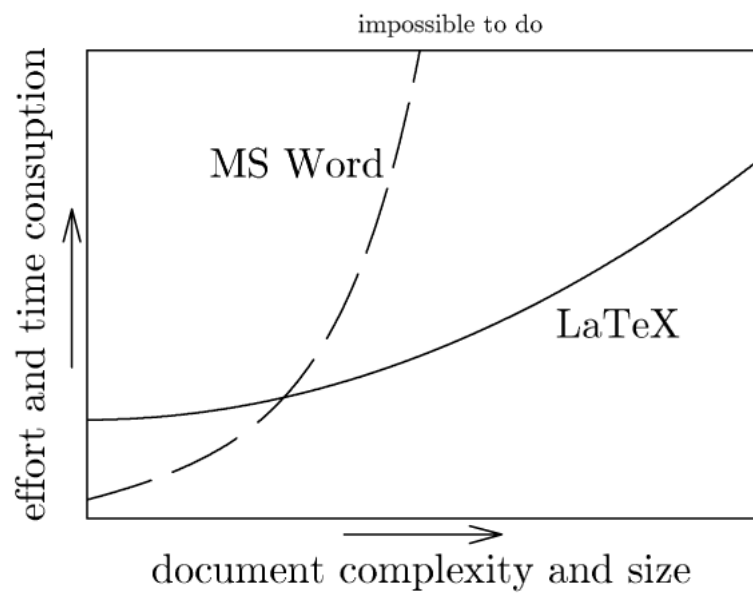
Data de defesa: 17-12-2013

Programa de Pós-Graduação: Química

Epígrafe

*“Fools and fanatics are always certain of themselves,
but wiser people are full of doubts.”*

Frase atribuída a Bertrand Arthur William Russell



Agradecimentos

- à Clarisse,
- à toda minha família, em especial aos meus pais e aos meus irmãos,
- ao Professor Munir,
- aos Professores membros da banca examinadora,
- a todos os outros Professores do IQ, em especial aos Professores Pedro Vazquez, Roy Bruns, Adalberto Bassi, Yoshiyuki Hase, Rogério Custódio, Leandro Martínez, Cláudia Longo, Marco-Aurélio de Paoli, Marcelo Ganzarolli, Pedro Volpe, Regina Buffon, Lúcia Baptistella, Paulo Imamura, Heloise Pastore, Inés Joeques (*in memoriam*), Watson Loh, Edvaldo Sabadini, Teresa Atvars, Celso Davanzo, Celso Bertran, Paulo Moran, José Alberto Fracassi, Paulo Rosa, Isabel Felisberti e Francisco Pessine,
- aos Professores externos ao IQ Alex Antonelli (IFGW), Frederico Tavares (EQ-UFRJ), Roberto Haiduke (USP São Carlos), Marcos Buckeridge (IB-USP), Weber Amaral (ESALQ-USP), Sergio Pantano (Institut Pasteur de Montevideo), Richard Lavery (CNRS), Luca Monticelli (CNRS), Philippe Derreumaux (Institut Universitaire de France), Sophie Sacquin-Mora (CNRS), Stephane Redon (INRIA), Marc Baaden (CNRS), Edward Maginn (University of Notre Dame, EUA) e Paul Dupree (University of Cambridge, Reino Unido),
- a todos os funcionários do IQ, em especial aos da CPG-IQ, Bel, Miguel e Gabriela,
- ao coordenador da CPG-IQ, Professor Cláudio Tormena,
- ao IQ, e à UNICAMP,
- ao CNPq pela bolsa de doutorado (processo 140978/2009-7),

-
- à FAPESP, pela infraestrutura do laboratório,
 - à comissão Fulbright, em especial à Thais Coser e ao Mike,
 - a todos os colegas do grupo de pesquisas, em especial ao Dr. Paulo César Telles de Souza e ao Rodrigo Leandro Silveira,
 - aos colegas, amigos e mestres Prof. Dr. Luciano Vidal, Dr. Marcelo (P2) de Souza, Dr. Cleber Borges, João Viçoso, Sérgio Faria, Dr. Maurício Chagas, José Carlos de Lima, Dra. Sílvia Moya, Dra. Déborah D'Avila, Eduardo Meneghetti, Pedro Villalobos, Saulo Yamaki, Lara Tavares, Janaína Gianfelice, Dra. Cassiana Montagner, Dra. Amedea Seabra, Ana Lúcia Malagodi, Ana Paula Rogieri, Diego Robert, David Britto, Leonardo Chang, Renato Ramponi, Francisco Ide Sales, Dr. Arthur Zampieri, Dr. André Zuin, Dr. Mathias Strauss, Cláudia Jovanini, Vanessa Morales, Prof. Carlos Rafael Longo, Natalie Machado, Kaiser e Camila, Jorge Haddad, Paulo Henrique e Giz, Lívia Streit, Cristiane Schwalm e Lutiano, Sr. Noriel, Rafael Caceres, Prof. Homero, Prof. Abílio, Prof. Sávio, Prof. Beto, Prof. Heron, i Profa. Regina, Prof. Hélio, Prof. Henrique, Prof. Mauro, Profa. Elzira, Prof. Sam, Prof. Sidney, Profa. Bete, Profa. Shirley, Profa. Rita Carè, Thiago Campos, Fernando Eto, André Turu, Bruno Cardoso, Chico, Inácio, Dudu, Karen, Gabi Strafacci, Marta Lopes, Petri Lintinen, Leandro Alvarez, Cristiano, Eduardo, Taís Vendrame, Sarah, Amanda, Tani Lemos, João Rafael, Ivelise, Miriam Almeida, (Sebas)Tião da livraria e família,
 - a todos os cientistas, pesquisadores e pensadores, cuja lista completa de nomes seria impossível de elencar, que desenvolveram pesquisas ou produziram obras que de alguma maneira contribuíram para a execução desta tese, para minha formação e para o progresso da ciência e da humanidade em geral, desde a antiguidade até os tempos modernos,
 - a todas as pessoas que desenvolvem e disponibilizam *software* livre.
 - Peço desculpas às pessoas cujo nome eu tenha esquecido de mencionar nestes agradecimentos.

Curriculum Vitae

Currículo na Plataforma Lattes: <http://lattes.cnpq.br/9623357218158305>

Formação acadêmica

- Doutorado em Química, UNICAMP (mar/2009 - em conclusão)
- Mestrado em Química, UNICAMP (mar/2006 - nov/2008. Concluído)
- Bacharelado em Química, UNICAMP (mar/2001 - dez/2005. Concluído)

Experiência em pesquisa

- Iniciação Científica (fev/2002 - dez/2003), sob orientação do Prof. Dr. Marco-Aurélio de Paoli (IQ-UNICAMP): desenvolvimento de eletrólitos poliméricos para a utilização em dispositivos (foto)eletro-químicos, com interesse especial em células solares flexíveis de estado sólido.
- Iniciação Científica (fev/2004 - jun/2005), sob orientação do Prof. Dr. Marcelo Ganzarolli de Oliveira (IQ-UNICAMP): investigação da química do óxido nítrico e de nitroso-tióis em suas interações com hemoglobina, utilizando espectroscopia UV-visível e ressonância paramagnética de elétrons (EPR).
- Mestrado em Química (mar/2006 - nov/2008. Concluído), sob orientação do Prof. Dr. Pedro A. M. Vazquez e co-orientação do Prof. Dr. Roy E. Bruns (IQ-UNICAMP): cálculo e interpretação de intensidades de bandas de transição fundamentais no infravermelho de clorofluormetanos utilizando o modelo CCFDF e cargas e dipolos atômicos ChelpG.

Experiência em ensino

- Estágio docente em disciplina experimental, laboratório de Físico-Química I (sigla QF-632), pelo Programa de Estágio Docente (PED B) em vigor na UNICAMP, ministrada no segundo semestre letivo de 2010.

-
- Estágio docente em disciplina teórica, Físico-Química I (sigla QF-431), pelo Programa de Estágio Docente (PED C) em vigor na UNICAMP, ministrada no segundo semestre letivo de 2006.
 - Assistente em disciplina experimental de Química geral experimental (sigla QG-102), através do Programa de Apoio Didático (PAD) em vigor na UNICAMP, ministrada no primeiro semestre letivo de 2005.

Artigos publicados

- *Cellulose-Builder: A toolkit for building crystalline structures of cellulose*, Thiago C. F. Gomes, Munir S. Skaf. *Journal of Computational Chemistry* (2012) Vol. 33, pp. 1338-1346. DOI 10.1002/jcc.22959.
- *Molecular Basis of the Thermostability and Thermophilicity of Laminarinases: X-ray Structure of the Hyperthermostable Laminarinase from *Rhodothermus marinus* and Molecular Dynamics Simulations*, Lucas Bleicher, Érica T. Prates, Thiago C. F. Gomes, Rodrigo L. Silveira, Alessandro Nascimento, Adraina L. Rojas, Alexander Golubev, Leandro Martinez, Munir S. Skaf, Igor Polikarpov. *Journal of Physical Chemistry B* (2011) Vol. 115, pp. 7940-7949. DOI 10.1021/jp200330z.
- *Chelpg and QTAIM atomic charge and dipole models for the infrared fundamental intensities of the fluorochloromethanes*, Thiago C. F. Gomes, João V. da Silva Jr., Luciano N. Vidal, Pedro A. M. Vazquez, Roy E. Burns. *Theoretical Chemistry Accounts* (2008) Vol. 121, pp. 173-179. DOI 10.1007/s00214-008-0461-4.
- *Implementação computacional do modelo Carga-Fluxo de Carga-Fluxo de Dipolo para cálculo e interpretação das intensidades do espectro infravermelho*, Thiago C. F. Gomes, João V. da Silva Jr., Luciano N. Vidal, Pedro A. M. Vazques, Roy E. Bruns. *Química Nova* (2008) Vol. 31, pp. 1750-1754.

Aprovação em concurso público

- Aprovado, em primeiro lugar, em concurso público para provimento de vaga para o cargo de Professor Auxiliar, no Instituto Tecnológico de Aeronáutica (ITA), no campo de conhecimento de “Físico-Química de materiais energéticos”, conforme homologação divulgada no Diário Oficial da União aos 25 de outubro de 2013, seção 3, página 22.

Resumo

SIMULAÇÕES POR DINÂMICA MOLECULAR DE SISTEMAS BIOLÓGICOS RELACIONADOS À HIDRÓLISE DE BIOMASSA LIGNOCELULÓSICA. Nesta Tese foram investigados sistemas biológicos relacionados à hidrólise da biomassa lignocelulósica. O trabalho de tese está subdividido em três partes. Na primeira parte foram investigadas, mediante simulações por dinâmica molecular, as características estruturais e dinâmicas que conferem termoestabilidade a laminarinases, enzimas da família das glicosídeo hidrolases de grande interesse no desenvolvimento de tecnologias viáveis para a produção industrial de etanol celulósico. Na segunda parte do trabalho, desenvolvemos e tornamos publicamente disponível um programa de computador, denominado cellulose-builder, que visa automatizar a construção de arquivos em formato PDB que representem cristalitos de celulose. Os arquivos em formato PDB gerados pelo nosso programa podem ser utilizados como configuração inicial em simulações por dinâmica molecular de sistemas que contenham fases cristalinas da celulose. O programa permite gerar cristalitos de forma e tamanho arbitrários, permitindo expor qualquer face cristalográfica de quatro alomorfos da celulose cristalina. Visando contribuir para o melhor entendimento em nível molecular da estrutura da parede celular vegetal e elucidar os modos de interação entre a celulose e a hemicelulose (dois componentes importantes da parede celular vegetal), na terceira parte deste trabalho empregamos o programa cellulose-builder para construir cristais de celulose com xilanos (hemiceluloses) adsorvidos em diversas faces cristalográficas e utilizamos simulações por dinâmica molecular para estudar em nível atômico a interação entre xilanos e diversas faces cristalográficas do alomorfo I-beta da celulose. Também investigamos o efeito de substituintes comumente presentes em hemiceluloses (acetil e glucuronil) nos parâmetros geométricos e energéticos da interação entre celulose e hemicelulose.

Abstract

BIOLOGICAL SYSTEMS RELATED TO LIGNOCELLULOSIC BIOMASS HYDROLYSIS STUDIED BY MEANS OF MOLECULAR DYNAMICS SIMULATIONS. This Thesis comprises the study of biological systems related to the hydrolysis of lignocellulosic biomass. The work is divided in three parts. In the first part we investigated, by means of molecular dynamics simulations, the structural and dynamical features that confer thermostability to laminarinases, enzymes belonging to the glycoside hydrolase family, which are of great interest in the development of viable technologies aiming at the industrial production of cellulosic ethanol. In the second part of the Thesis work we developed and made publicly available a computer program, named cellulose-builder, whose purpose is to automatize the construction of PDB-format files representing cellulose crystallites. The PDB-format files generated by our program can be used as starting configurations in molecular dynamics simulations of systems containing crystalline phases of cellulose. The program enables one to generate crystallites of arbitrary size and shape, making it possible to expose any crystallographic face of four cellulose allomorphs. Aiming to contribute to a better understanding of the plant cell wall structure at the molecular level and to elucidate the modes of interaction between cellulose and hemicellulose (two major components of the plant cell wall), in the third part of this work we employed the computer program cellulose-builder to set up cellulose crystals with xylans (hemicelluloses) adsorbed onto several crystallographic faces and applied molecular dynamics simulations to study, at the atomic level, the interactions between xylans and several crystallographic faces of cellulose's allomorph I-beta. We also investigated the effect of substituents commonly occurring in hemicelluloses (namely acetyl and glucuronyl) in geometric and energetic parameters resulting from the interaction between cellulose and hemicellulose.

Índice

Lista de Tabelas	xix
Lista de Figuras	xxi
1 Introdução	1
1.1 A parede celular vegetal	3
1.2 Polissacarídeos	6
1.2.1 Carboidratos: nomenclatura, classificação e estrutura . . .	7
1.2.2 Polissacarídeos: diversidade e complexidade estrutural . .	20
1.3 Enzimas Glicosídeo Hidrolases	26
1.3.1 Termoestabilidade em laminarinases	32
1.4 Estrutura da celulose cristalina	33
1.5 Interações entre celulose e xilanos	36
2 Metodologia	41
2.1 Simulações por dinâmica molecular	41
2.1.1 Campos de força para simulações de glicanos	44
2.2 Cellulose-builder	44
3 Termoestabilidade em laminarinases	53
4 Cellulose-Builder	63
5 Interações entre celulose e xilanos	73
5.1 Xilanos adsorvidos sobre faces hidrofílicas	74
5.2 Xilanos adsorvidos sobre faces hidrofóbicas	84
5.3 O cristal de celulose	88
5.4 Xilanos substituídos	92

6 Conclusões	103
6.1 Perspectivas	104
Bibliografia	105

Lista de Tabelas

1.1	Dissacarídeos formados a partir da condensação de duas unidades (idênticas) de D-glicose. ¹⁹	20
1.2	Estrutura primária ^a de alguns glicanos ¹⁹	21
1.3	Primeiro nível do sistema EC de classificação de enzimas, de acordo com a IUPAC e IUBMB ^a	27
3.1	Número de pontes salinas na RmLam e em outras estruturas de enzimas GH da família 16 depositadas no <i>Protein Data Bank</i> . . .	57
3.2	Dados computados a partir das simulações por DM das três laminarinases homólogas, RmLam, aNLam, and PcLam a 25°C e 90°C.	58
4.1	Tempos de execução do <i>cellulose-builder</i> para produzir arquivos PDB e PSF para as fibrilas elementares de celulose representadas na Figura 4.6, em função do número de unidades monoméricas de celobiose em cada fibrila. Tempo real decorrido (<i>wall clock time</i>), tempo de CPU que o processo utiliza diretamente (em <i>user mode</i>), tempo de utilização da CPU pelo sistema operacional por causa do processo (em <i>kernel mode</i>), e porcentagem de utilização da CPU pelo <i>cellulose-builder</i> , conforme fornecidos pelo <i>/usr/bin/time</i> . Todos os tempos estão em segundos. A porcentagem de utilização de CPU equivale aos tempos (<i>user + system</i>) divididos pelo tempo real decorrido. A quantidade de memória RAM utilizada para construir o maior sistema foi apenas 360MB.	70

5.1	<p>Porcentagem do tempo de simulação em que ligações de hidrogênio ocorrem entre a molécula de xilano M0 e a molécula de celulose M4, na face (010) (vide Figuras 5.1 e 5.3). As ligações de hidrogênio estão divididas por resíduo de xilose. Apenas os resíduos pares em M0 participam de ligações de hidrogênio com M4. Em negrito: para a maioria dos resíduos de xilose, o tipo de ligação de hidrogênio predominante é aquela em que os oxigênios O2 das xiloses atuam como doadores de prótons para os oxigênios O6 das glicoses, que atuam como aceptores de próton. Para o resíduo 0, que é a extremidade não redutora do xilano M0, o padrão é diferente devido à maior mobilidade do resíduo. Os oxigênios O3 das xiloses estabelecem ligações de hidrogênio menos freqüentemente com M4, atuando principalmente como aceptores de próton para os oxigênios O6 das glicoses, que atuam como doadores de próton neste caso.</p>	77
5.2	<p>Porcentagem do tempo de simulação em que ligações de hidrogênio ocorrem entre a molécula de xilano M25 e a molécula de celulose M21, na face (020) (vide Figura 5.1). As ligações de hidrogênio estão divididas por resíduo de xilose. Apenas os resíduos ímpares em M25 participam de ligações de hidrogênio com M21. Em negrito: o tipo de ligação de hidrogênio predominante é aquela em que os oxigênios O2 das xiloses atuam como doadores de prótons para os oxigênios O6 das glicoses, que atuam como aceptores de próton. Ligações de hidrogênio envolvendo o oxigênio O3 das xiloses ocorrem menos freqüentemente que aquelas envolvendo o O2 no mesmo resíduo de xilose.</p>	79

Lista de Figuras

1.1	Modelo para a estrutura microscópica da parede celular vegetal. . .	3
1.2	Representação esquemática da hidrólise da celulose por um coquetel enzimático, com enzimas agindo sinergisticamente.	5
1.3	a)Estruturas do gliceraldeído, o menor monossacarídeo, e da di-hidroxiacetona em projeção de Fischer. b)D e L-gliceraldeído. . .	9
1.4	Projeções de Fischer para as formas acíclicas da série D das aldoses, obtida a partir do D-gliceraldeído, indo da triose às hexoses. . . .	10
1.5	Formação de hemiacetal cíclico, fórmula de Haworth e conformação cadeira 4C_1 da β -D-glicopiranosose.	11
1.6	Formas anoméricas da D-glicopiranosose em projeções de Haworth. À esquerda anômero α e à direita anômero β	12
1.7	Mapeamento esférico das conformações em piranoses.	14
1.8	Conformações <i>gt</i> , <i>gg</i> e <i>tg</i> para o grupo exocíclico hidroximetil, em hexopiranoses.	16
1.9	Conformações do anel das furanoses.	17
1.10	Estrutura do metil- β -D-celobiosídeo, derivado metilado da β -D-celobiose.	19
1.11	Representação do monômero de um exopolissacarídeo produzido pela linhagem LBB.B26 de <i>Lactobacillus delbrueckii</i>	24
1.12	Estrutura em tripla hélice do escleroglucano.	25
1.13	Classificação das enzimas GH segundo sua capacidade de hidrolisar glicanos a partir das extremidades.	28
1.14	Classificação das enzimas GH segundo seu mecanismo de reação. .	29
1.15	Representação esquemática das enzimas envolvidas na hidrólise da celulose (A) e hemicelulose (B).	31
1.16	Cristalito de celulose $I\beta$ gerado utilizando o programa <i>cellulose-builder</i>	34

1.17	Outros alomorfos da cellulose suportados pelo programa <i>cellulose-builder</i> , vistos a partir das extremidades das cadeias de cellulose. a) I α b) II c) III _I . As escalas diferem. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf ³ , Cellulose-Builder: A toolkit for building crystalline structures of cellulose. Journal of Computational Chemistry, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).	35
1.18	Representação da superfície de Van der Waals de cadeias de xilano (em vermelho) adsorvidas sobre diferentes superfícies da cellulose cristalina I β (em cinza). (A) Xilanos adsorvidos sobre superfícies hidrofílicas da cellulose I β . (B) Xilanos adsorvidos sobre superfícies hidrofóbicas da mesma fibrila de cellulose.	37
1.19	Xilano simples e xilanos ramificados. (A) Xilano. (B) Glucoronoxilano com substituintes nas posições 3 e 7, padrão par de substituição (neste caso em i e em $i + 4$, com $i = 3$). (C) Glucoronoxilano com substituintes nas posições 3 e 8 (padrão ímpar, substituição em i e em $i + 5$).	38
1.20	Acetilxilano com padrão de substituição par, adsorvido nas faces (A) hidrofílicas e (B) hidrofóbicas da cellulose cristalina I β	39
2.1	Critérios geométricos utilizados para contabilizar ligações de hidrogênio nas simulações por DM de xilanos adsorvidos à cellulose. O ângulo α deve ser igual ou menor que 60° e a distância d deve ser menor ou igual a 3.5\AA	43
2.2	Representação da primeira etapa executada pelo programa <i>cellulose-builder</i>	48
2.3	Representação da etapa de replicação executada pelo programa <i>cellulose-builder</i> após ter construído a cela unitária a partir de sua unidade assimétrica.	49
2.4	Fluxograma que representa simplificadaamente o fluxo de trabalho do programa <i>cellulose-builder</i>	51
3.1	Alinhamento, baseado em estrutura terciária, das seqüências primárias das enzimas RmLam, aNLam e PcLam.	54
3.2	Alinhamento das estruturas terciárias da RmLam (azul), aNLam (vermelho) e PcLam (verde). As fitas- β , numeradas de S1 a S15, são as porções de maior sobreposição nestas três estruturas.	55

3.3	As duas regiões mais importantes na estrutura das laminarinases são a fenda catalítica e o núcleo hidrofóbico. O núcleo hidrofóbico compreende uma extensa região entre as duas folhas- β , na qual estão presentes apenas cadeias laterais de resíduos hidrofóbicos.	56
3.4	Características topológicas das pontes salinas na laminarinases estudadas.	59
3.5	Topologia das pontes salinas em laminarinases. (a) RmLam, hipertermofílica. (b) aNLam, termofílica. (c) PcLam, mesofílica.	60
3.6	Detalhes da penetração de água no núcleo de resíduos hidrofóbicos encerrado entre as folhas β interna e externa das laminarinases.	61
3.7	Número de moléculas de água detectadas dentro dos núcleos hidrofóbicos da RmLam (azul/ciano), aNLam (vermelho, laranja), e PcLam (verde claro e verde escuro) em função do tempo. Cores claras correspondem a resultados de simulações individuais enquanto cores escuras correspondem à média.	62
4.1	Formato necessário para que um arquivo PDB seja utilizável por programas de simulação por DM.	64
4.2	Dois cristalitos expondo diferentes superfícies da celulose cristalina $I\beta$. À esquerda: o cristalito expõe predominantemente a face cristalográfica (010), hidrofílica. À direita: o cristalito expõe predominantemente a face cristalográfica (100), hidrofóbica. As imagens A e B representam o mesmo cristalito observado de pontos de vista diferentes. Idem para C e D.	65
4.3	Cristalitos de celulose $I\beta$ aptos para a aplicação de condições periódicas de contorno ao longo das direções cristalográficas a (esquerda), b (meio) e ambas a e b (direita).	66
4.4	O método de esculpir um cristalito a partir de outro cristalito maior é utilizado pelo <i>cellulose-builder</i> para construir automaticamente fibrilas com a seção transversal representada na Figura 4.5. Este método também pode ser utilizado para produzir manualmente cristalitos com formato totalmente geral, como por exemplo os cristalitos nesta figura, que expõem as faces cristalográficas (110) e (1-10) da celulose $I\beta$. À direita, as faces opostas entre si têm os mesmos índices de Miller.	67

4.5	Modelo proposto por Ding e Himmel ⁹⁰ para fibrila elementar de celulose em parede celular primária de milho, vista a partir de suas extremidades não redutoras. A fibrila elementar é composta de 36 cadeias de celulose.	67
4.6	Fibrilas elementares de celulose I β com diferentes graus de polimerização (número de unidades de celobiose, k , igual a 5, 25, 50, 100, 500), geradas com o programa <i>cellulose-builder</i> . A fibrila com $k = 5$ foi ampliada. Todas as fibrilas tem a seção transversal mostrada na Figura 4.5.	69
4.7	Representação do número de acessos ao sítio http://code.google.com/p/cellulose-builder/ em escala de cores, por país.	71
5.1	Conjunto de rótulos para os sistemas estudados, no qual cada cadeia polimérica recebe um rótulo.	73
5.2	Representação estatística da dinâmica de moléculas de xilano adsorvidas sobre as faces hidrofílicas da celulose I β , utilizando superfícies de nível, ou <i>isosuperfícies</i> , de ocupação.	75
5.3	<i>Frame</i> estatisticamente representativo das ligações de hidrogênio entre a cadeia de xilano M0 e a cadeia de celulose M4, posicionada imediatamente acima (vide Figura 5.1), na face (010).	76
5.4	Distribuição do número de ligações de hidrogênio estabelecidas entre cadeias de xilano e as faces hidrofílicas da celulose.	78
5.5	Cadeias de xilano assumindo conformações em hélices. A) Hélice 2 ₁ . B) Hélice 3 ₁	80
5.6	Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M0, sobre a face (010).	81
5.7	Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M25, sobre a face (020).	82
5.8	Distância entre o oxigênio glicosídico que une os resíduos 3 e 4 do xilano M25 e o oxigênio glicosídico logo abaixo dele na cadeia de celulose M21, em função da soma ($\phi + \psi$) para o par de resíduos 3 e 4.	83
5.9	Distribuição da soma dos diedros glicosídicos ϕ e ψ dos nove pares de diedros glicosídicos em uma molécula de $\beta(1 \rightarrow 4)$ -xilano composta por dez unidades de D-xilose, em água pura.	84

5.10	Representação estatística da dinâmica de moléculas de xilano adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$, utilizando superfícies de nível, ou <i>isosuperfícies</i> , de ocupação.	85
5.11	À esquerda: distribuição do número de ligações de hidrogênio entre a cadeia de xilano M0 e o cristal de celulose $I\beta$. À direita: distribuição do número de ligações de hidrogênio estabelecidas entre a cadeia de xilano M25 e a celulose.	86
5.12	Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M0, sobre a face (100).	87
5.13	Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M25, sobre a face (200).	88
5.14	Representação da distorção na geometria dos cristais de celulose $I\beta$, que ocorre durante as simulações por DM.	89
5.15	Painel que ilustra histogramas que representam a distribuição das conformações do grupo exocíclico hidroximetil, correspondente ao carbono C6 nos anéis de glicose (vide Figura 1.8), em cada uma das 24 cadeias de celulose presentes no cristal.	91
5.16	Representação da dinâmica de moléculas de acetilxilano e glucuronoxilano adsorvidas sobre as faces hidrofílicas da celulose $I\beta$, utilizando isosuperfícies de ocupação.	94
5.17	Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas nas moléculas de acetilxilano adsorvidas sobre as faces hidrofílicas da celulose $I\beta$	96
5.18	Representação da dinâmica de moléculas de acetilxilano e glucuronoxilano adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$, utilizando isosuperfícies de ocupação.	97
5.19	Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas nas moléculas de acetilxilano adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$	99

Capítulo 1

Introdução

O recente interesse pelo desenvolvimento de novas tecnologias que permitam a utilização de fontes de energia renováveis e não fósseis tem atraído grande atenção dos pesquisadores para o estudo da biomassa lignocelulósica *, hoje considerada uma importante e promissora fonte alternativa para a geração de etanol, entre outros compostos químicos. De fato, a celulose compõe grande parte da parede celular dos vegetais (“parede celulósica”), sendo assim considerada¹ a mais abundante substância orgânica presente na crosta terrestre. A celulose é um polímero de glicose que se for hidrolisado formará açúcares que podem ser convertidos a etanol por processos microbiológicos como a fermentação alcoólica. As técnicas atuais de produção de etanol combustível, seja a partir da sacarose de cana-de-açúcar, seja a partir do hidrolisado de amido de milho, competem com as lavouras de alimentos pela utilização de extensões de terra agricultável e portanto tem seu potencial de expansão seriamente comprometido. Esta competição por terras agricultáveis constitui uma importante limitação para a expansão da produção de qualquer *biocombustível* obtido a partir de produtos agrícolas. Entretanto, a possibilidade de utilizar os componentes químicos da biomassa vegetal - principalmente a celulose, mas também outros glicanos † - como matéria prima para obtenção de açúcares fermentáveis a etanol pode significar uma imensa expansão da quantidade de matéria prima disponível para tal processo, uma vez que não apenas o bagaço da cana-de-açúcar poderia servir de insumo, como potencialmente qualquer sobra ou resíduo agrícola, industrial ou urbano constituído de biomassa vegetal *in natura* ou processada (palha, folhas, madeira, grama, aparas de papel, papelão *etc*).

*A biomassa obtida de vegetais superiores, *i.e.* Gimnospermas e Angiospermas.

†O nome moderno, atualmente recomendado pela IUPAC para oligo e polissacarídeos

Contudo, a grande dificuldade de hidrolisar a celulose e outros glicanos contidos na biomassa vegetal a seus monômeros básicos é o principal obstáculo à idéia de utilizar a biomassa lignocelulósica como matéria prima para obtenção de etanol. A estrutura microscópica e molecular da parede celular vegetal ainda é objeto de intensa pesquisa e, embora a estrutura da parede celular possa variar dependendo da espécie ou do tecido, algumas características gerais parecem ser conservadas de forma a tornar a biomassa vegetal extremamente resistente (ou *recalcitrante*) a ataques químicos ou enzimáticos que visem à hidrólise da celulose contida na biomassa (vide Seção 1.1 a seguir). Portanto, o problema da resistência (ou *recalcitrância*) da lignocelulose à hidrólise é o problema central que boa parte das pesquisas na área do etanol celulósico tentam de alguma forma entender e resolver. O foco da pesquisa desenvolvida pelo grupo de dinâmica molecular do IQ-UNICAMP na área do bioetanol consiste em estudar aspectos relacionados à hidrólise enzimática dos principais componentes da biomassa produzida por organismos fotossintetizadores²⁻⁵: a celulose, polímeros da matriz lignocelulósica, e outros polissacarídeos. O entendimento do processo de deconstrução da biomassa para fins de geração de combustíveis e outros produtos químicos de alto valor agregado exige o estudo aprofundado das enzimas *glicosídeo hidrolases* (GH) - enzimas que catalisam a hidrólise de glicanos - , o estudo do substrato lignocelulósico, de seus aspectos estruturais e dinâmicos e das interações deste substrato com as enzimas de interesse.

Neste contexto, a contribuição que esta Tese de Doutorado pretende fornecer pode ser dividida em três partes, que serão descritas nas seções a seguir. Esta divisão é proposta para melhor descrever o trabalho e sistematizar os resultados. No título deste trabalho optou-se por designar o objeto de estudo por “*sistemas biológicos*” uma vez que os sistemas estudados são:

- Três enzimas GH da família 16, Seção 1.3,
- Vários polimorfos da celulose, Seção 1.4,
- Xilanos adsorvidos em celulose, Seção 1.5,

e que estes sistemas estão presentes em seres vivos, com exceção de alguns polimorfos da celulose que não ocorrem naturalmente. De forma análoga ao que será feito no capítulo introdutório deste texto, os capítulos nos quais serão descritos Metodologia, Resultados e Discussões e Conclusões serão divididos em capítulos e seções referentes a cada uma das três partes do trabalho de tese, conforme os itens elencados acima.

Para melhor contextualizar o leitor e introduzir os detalhes dos problemas que foram tratados nesta tese, na Seção 1.1 a seguir será brevemente abordado o modelo atualmente aceito para a estrutura microscópica da parede celular vegetal.

1.1 A parede celular vegetal

Paredes celulares são estruturas que revestem a parte exterior da membrana celular de células vegetais. A sua composição química é bastante complexa e varia conforme a espécie e o tecido considerado, mas em geral pode ser representada pelo modelo ilustrado na Figura 1.1. A parede celular é uma matriz composta de vários tipos de polissacarídeos, reticulados ou não, lignina e proteínas glicosiladas. Os componentes majoritários da parede celular são a celulose, a hemicelulose e a lignina. Como pode-se observar na Figura 1.1, a celulose está presente em paredes celulares na forma de feixes de moléculas de celulose que usualmente apresentam vastos domínios cristalinos. A estes feixes de moléculas de celulose dá-se o nome de *microfibrilas*. Na parede celular as microfibrilas de celulose são recobertas por outros polissacarídeos genericamente denominados *hemicelulose*. A

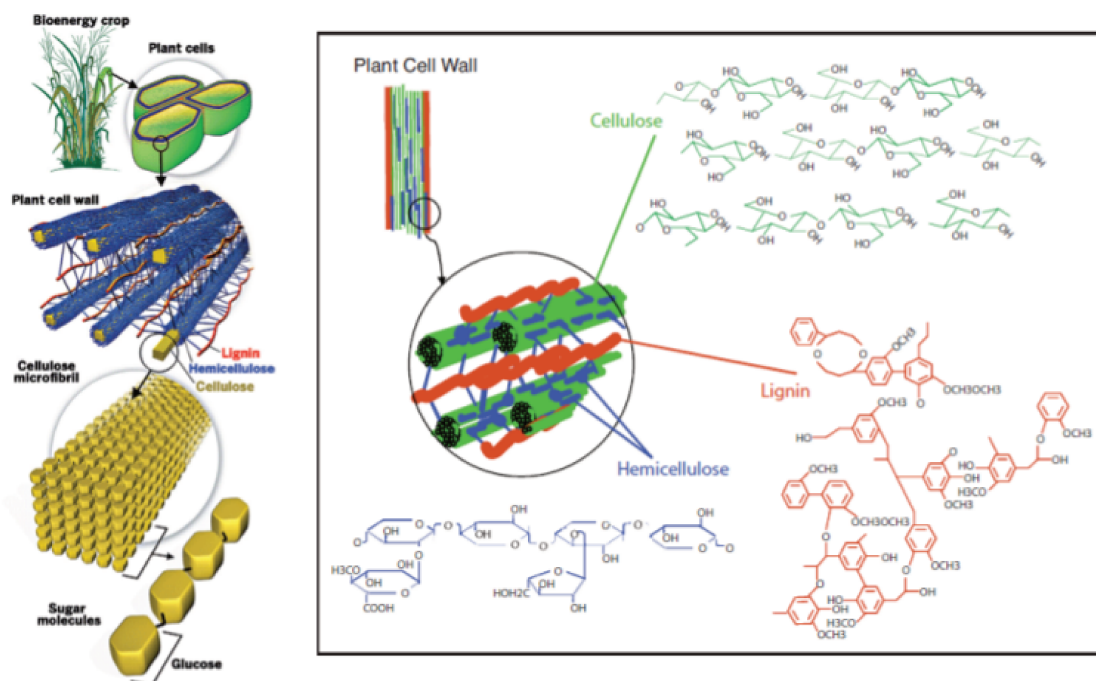


Figura 1.1: Modelo para a estrutura microscópica da parede celular vegetal. Adaptado de Stephen K. Ritter. Lignocellulose: a complex biomaterial. Chemical & Engineering News, 86(49):15, 2008 e reproduzido com permissão do U.S. Department of Energy Genomic Science Program.⁶

estrutura e composição de polissacarídeos será considerada mais adiante, na Seção 1.2. Por ora, basta afirmar que a celulose é um polímero de glicose que, em plantas, pode ter centenas ou milhares de unidades monoméricas; e *hemicelulose* é um nome genérico dado a diversos tipos de polissacarídeos que em geral são polímeros de xilose e que podem ter vários tipos de ramificações (grupos substituintes nos anéis de xilose da cadeia principal), incluindo acetil, arabinose, galactose, glucose, manose e outros açúcares. As cadeias de hemicelulose podem ter de 500 a 3000 unidades monoméricas. A lignina é um polímero reticulado constituído de três tipos diferentes de fenóis substituídos. Este componente age como um “cimento”, ocupando o volume livre entre os outros componentes da parede celular e mantendo unida a matriz lignocelulósica. A lignina pode estabelecer ligações covalentes com a hemicelulose (vide Figura 1.15) e contribui para a resistência mecânica da parede celular e para a sua resistência a ataques de parasitas[‡].

A estrutura microscópica da parede celular vegetal é o que determina as propriedades macroscópicas da biomassa lignocelulósica, como sua resistência mecânica e estabilidade química. São estas propriedades que propiciaram o uso da biomassa lignocelulósica na sua forma mais bruta, a madeira, desde tempos imemoriais para construção de móveis, barcos ou edifícios que chegaram a durar séculos. São estas mesmas propriedades, determinadas pela estrutura microscópica da parede celular vegetal, que impõem os maiores obstáculos ao uso da biomassa lignocelulósica como matéria prima industrial para obtenção de compostos químicos de alto valor agregado derivados de açúcares. Em paredes celulares, as fibrilas de celulose estão recobertas por hemicelulose e imersas na matriz de lignina, tornando inviável o acesso à superfície da celulose a enzimas que catalisem a sua hidrólise. Por este motivo é necessário pré-tratar a biomassa lignocelulósica antes de processar a celulose em si. Diversos tipos de pré-tratamento foram e estão sendo investigados pela comunidade científica e alguns deles envolvem tratamento com ácidos, álcalis, altas temperaturas, explosão a vapor *etc.* O objetivo desta primeira etapa de pré-tratamento é deixar a superfície da celulose exposta e disponível ao ataque enzimático. A Figura 1.2 ilustra o fato de que, mesmo com a superfície

[‡]Quando pensamos em parasitas de plantas tendemos a imaginar herbívoros ou insetos. Fungos, entretanto, podem ser parasitas ou decompositores de plantas superiores e são o exemplo de maior interesse no nosso caso. Fungos são organismos que tipicamente fazem digestão extracelular, o que significa que excretam enzimas para fora de suas células para que estas enzimas digiram matéria extracelular, que após a digestão poderá incorporada pelas células do fungo como alimento. As enzimas expressas por organismos como bactérias e fungos que naturalmente parasitam ou decompõem plantas são de especial interesse para a área do *bioetanol*, pois tais enzimas são especializadas na deconstrução da parede celular vegetal. Algumas destas enzimas serão discutidas na Seção 1.3.

da celulose totalmente exposta, diversas enzimas são necessárias para hidrolisar a celulose a monômeros de glicose. A composição ideal desta mistura enzimática para a hidrólise da celulose é objeto de intensa pesquisa, e o que se busca obter é um *coquetel enzimático* que seja quimicamente eficiente e economicamente viável para aplicação em escala industrial.

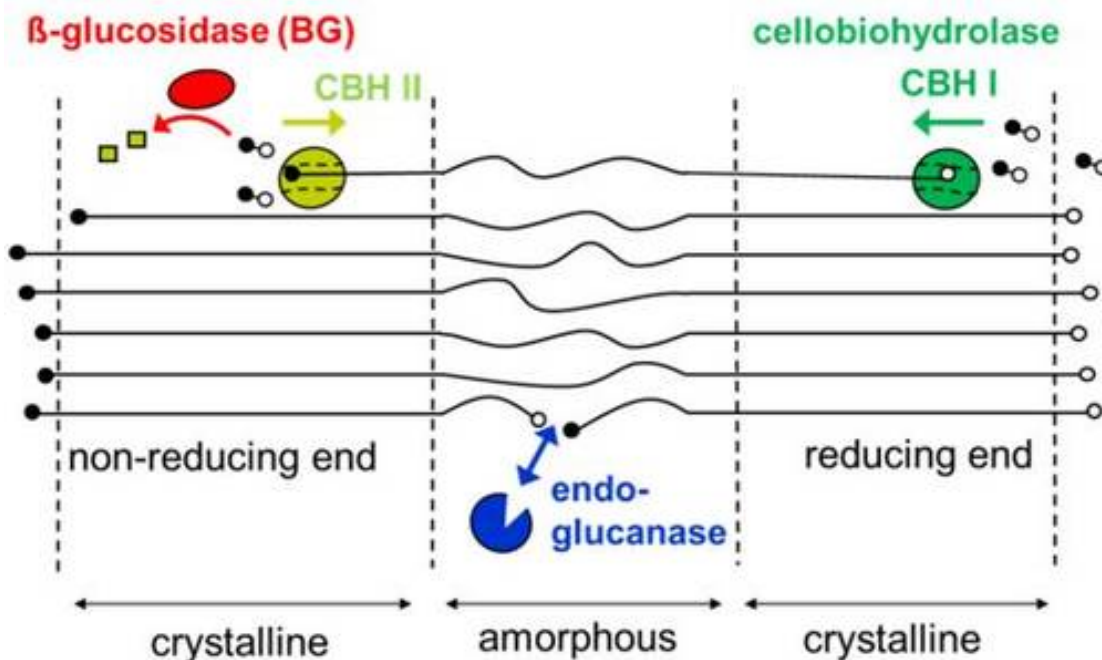


Figura 1.2: Representação esquemática da hidrólise da celulose por um coquetel enzimático, com enzimas agindo sinergisticamente. *Endoglucanases* (EG) são enzimas capazes de catalisar a hidrólise de uma cadeia de celulose, em uma região amorfa, clivando uma ligação glicosídica no meio de uma cadeia de celulose e produzindo duas novas extremidades. *Celobiohidrolases* (CBH) catalisam a hidrólise a partir das extremidades das cadeias de celulose, podendo agir nas regiões cristalinas, e a reação de hidrólise catalisada por elas forma celobiose (dímeros de glicose solúveis em água). As enzimas denominadas *β-glicosidases* (BG) só hidrolisam moléculas de celobiose livres em solução, levando à formação de duas moléculas de glicose. Outras proteínas, denominadas *swolleninas* ou *expansinas* (não mostradas), promovem a transição de fase entre a celulose cristalina e a celulose amorfa, e agem na superfície da celulose cristalina. Figura extraída de <http://www.avt.rwth-aachen.de/AVT/index.php?id=888&L=1> e reproduzida com a permissão de Prof. Dr.-Ing. Antje Spieß. Esta figura, por sua vez, é baseada na Figura 1 do trabalho de Tuula T. Teeri⁷.

Estas considerações deixam claro que, para que um dia exista uma indústria do etanol celulósico ou uma indústria de produtos químicos obtidos a partir de biomassa lignocelulósica (que alguns chegam a apostar que um dia poderá substituir a indústria petroquímica), é necessário aprofundar nosso entendimento sobre a

estrutura molecular da parede celular vegetal e sobre a físico-química das enzimas que catalisam a reação de hidrólise de polissacarídeos. Esta tese representa uma pequena contribuição a este entendimento, investigando três destas enzimas (Capítulo 3), alguns alomorfos da celulose cristalina (Capítulo 4), bem como a interação entre dois dos principais componentes da parede celular vegetal: a hemicelulose e a celulose (Capítulo 5). Antes de iniciar qualquer discussão mais aprofundada sobre os polissacarídeos que constituem a parede celular vegetal, ou sobre as enzimas que catalisam a sua hidrólise, é útil introduzir alguns conceitos fundamentais. Na próxima seção é discutida a importância e a ocorrência dos polissacarídeos e uma introdução aos aspectos fundamentais da química dos carboidratos, como nomenclatura, classificação e estrutura, é apresentada.

1.2 Polissacarídeos: importância científica e tecnológica

Os polissacarídeos têm imensa importância científica e tecnológica. Em primeiro lugar, os polissacarídeos compõem boa parte da matéria viva (biomassa) sobre a crosta terrestre. Como componentes dos seres vivos eles desempenham inúmeras funções primordiais como estoque de energia, construção de tecidos, sinalização, codificação e reconhecimento molecular. Os polissacarídeos constituem boa parte da comida que ingerimos todos os dias e nos fornecem a maior parte da energia que nos mantém vivos. Enquanto polímeros de função estrutural, estão presentes no exoesqueleto de crustáceos, aracnídeos e insetos⁸ e na parede celular tanto de células vegetais (Seção 1.1) como de microorganismos⁹. As cápsulas de polissacarídeos que revestem a membrana celular de determinados microorganismos⁹⁻¹¹ (polissacarídeos denominados *exopolissacarídeos* justamente por estarem no exterior da membrana celular) estão diretamente envolvidas em mecanismos de reconhecimento e defesa, e os polissacarídeos que as compõem encontram aplicação em imunologia e na produção de vacinas.^{12;13}

Recentemente muito estudo tem sido dedicado à obtenção de energia ou de combustíveis a partir de polissacarídeos^{14;15}, como o amido de milho que é transformado em etanol nos Estados Unidos ou como a celulose, cuja hidrólise com viabilidade econômica e industrial poderia transformar biomassa vegetal em uma nova fonte de combustíveis. Em última instância, combustíveis ou energia provenientes da biomassa vegetal vêm da energia solar, e os átomos de carbono que

compõem os polissacarídeos da biomassa vêm do dióxido de carbono retirado da atmosfera¹⁶, logo estes polissacarídeos poderão servir não apenas como fonte de energia, mas também como fonte de material sintético para uma infinidade de compostos como plásticos e fármacos, e potencialmente poderão tornar-se matérias primas alternativas ou substitutas para a síntese dos produtos da indústria química, farmacêutica e petroquímica. Algumas rotas para obtenção de matérias primas básicas são a desidratação do etanol a etileno¹ e a desidratação de açúcares a 5-hidroximetilfurfural (HMF)^{17;18}, ambos produtos de partida extremamente versáteis para sínteses. Polissacarídeos encontram aplicação na indústria têxtil: fibras e tecidos de algodão, linho e cânhamo são quase que inteiramente constituídos de polissacarídeos. Rayon e celofane são obtidos a partir da celulose, um polissacarídeo. O celofane encontra larga aplicação na indústria de embalagens, na qual também são utilizados outros produtos da celulose como o papelão e o papel, este último tendo sido utilizado para a quase totalidade da comunicação escrita desde os antigos egípcios até o surgimento dos meios eletrônicos.

Portanto, ainda que mais aplicações pudessem ser citadas, é evidente que os polissacarídeos (e carboidratos em geral) constituem uma classe de compostos importantes. Seu estudo mais aprofundado exige uma revisão dos aspectos mais fundamentais da química destes compostos como sua nomenclatura, critérios de classificação e estrutura, a começar pela descrição de suas unidades básicas de construção, os monossacarídeos, passando pelos dissacarídeos com os quais é possível estudar as características das ligações entre estas unidades de construção, até chegar na descrição da estrutura de polissacarídeos.

1.2.1 Carboidratos: nomenclatura, classificação e estrutura

Para descrever e estudar polissacarídeos é necessário primeiro compreender alguns aspectos relacionados à estrutura química de carboidratos mais simples, para então desenvolver a análise de carboidratos mais complexos.

Os compostos que fazem parte da classe dos carboidratos foram um dia conhecidos como “hidratos de carbono”, pois suas fórmulas moleculares em muitos casos (mas não todos) correspondem a $C_n(H_2O)_m$. Entretanto, é mais realístico definir carboidratos como compostos carbonílicos polihidroxilados, um ponto de vista mais próximo de sua realidade estrutural e mais sugestivo em termos das propriedades e reatividade química destes compostos.¹ A palavra sacarídeo, que

deriva do latim *saccharum* (açúcar) que por vez sua deriva do sânscrito *sarkhara*, é a base de um sistema de classificação de carboidratos. Um monossacarídeo é definido como a unidade estrutural básica que não pode ser dividida em carboidratos mais simples, ou seja, é um carboidrato simples que não sofre hidrólise e não forma carboidratos com massa molecular menor. Um dissacarídeo é um carboidrato que ao ser hidrolisado forma dois monossacarídeos, idênticos ou não. Da mesma forma tri, tetra, e penta sacarídeos formam respectivamente três, quatro e cinco monossacarídeos ao sofrer hidrólise. Oligossacarídeos formam de três a dez monossacarídeos ao sofrer hidrólise completa (embora haja divergências sobre o tamanho máximo e mínimo dos oligossacarídeos^{1;19}) enquanto um polissacarídeo pode formar de poucos a milhares de monossacarídeos ao sofrer hidrólise completa.

Monossacarídeos

Os monossacarídeos, geralmente representados por alguns autores pela fórmula molecular $C_n(H_2O)_n$ ¹⁹, por outros pela fórmula $C_x(H_2O)_n$ ²⁰, são constituídos de uma cadeia carbônica em que cada carbono está ligado a um grupo hidroxila, com exceção de um carbono que possui uma carbonila, como pode-se observar na Figura 1.3. De acordo com a posição desta carbonila, os monossacarídeos podem ser classificados em polihidroxialdeídos ou polihidroxicetonas: os primeiros são denominados *aldoses* e têm a carbonila em uma extremidade da cadeia carbônica, enquanto as segundas, chamadas *cetoses*, podem ter a carbonila em qualquer posição exceto nas extremidades. As cetoses são ainda classificadas em 2-cetoses, 3-cetoses, e assim por diante, indicando a posição do grupo carbonílico. Como as 2-cetoses são as cetoses mais abundantemente encontradas na natureza, convencionou-se que se um texto ou autor não especifica a cetose, ele está se referindo a uma 2-cetose. Os monossacarídeos podem ainda ser classificados de acordo com o número de átomos de carbono que possuem: trioses, tetroses, pentoses, hexoses são monossacarídeos com 3, 4, 5, 6 carbonos respectivamente. É comum ver os dois critérios de classificação utilizados em conjunto, por exemplo: uma aldohexose é uma aldose com seis carbonos e uma cetopentose é uma cetose com cinco carbonos. O menor dos monossacarídeos é o gliceraldeído que, como pode-se ver pela Figura 1.3, possui um centro estereogênico e portanto dois enantiômeros são possíveis, o D-gliceraldeído e o L-gliceraldeído, em que os prefixos D e L referem-se à configuração do carbono assimétrico. Pode-se imaginar todas as aldoses maiores como derivadas do gliceraldeído: as que derivam do D-gliceraldeído pertencem à série D (Figura 1.4), enquanto as derivadas do L-gliceraldeído pertencem à série L das

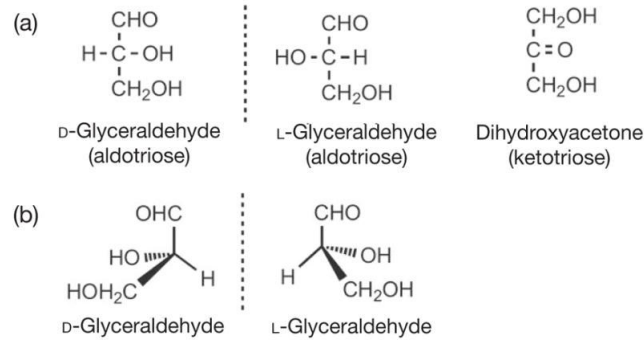


Figura 1.3: a) Estruturas do gliceraldeído, o menor monossacarídeo, e da dihidroxiacetona em projeção de Fischer. b) D e L-gliceraldeído. Figura extraída de Carolyn R. Bertozzi and David Rabuka²⁰, *Essentials of Glycobiology*, chapter Structural Basis of Glycan Diversity. Cold Spring Harbor Laboratory Press, 2nd edition, 1999 (copyright 1999).

aldoses. A dihidroxiacetona, a cetose mais simples, não possui um centro estereogênico como o gliceraldeído e por isso às vezes não é considerada um carboidrato. Assim sendo, as séries das D- e L-cetoses são obtidas a partir da eritrulose¹⁹ (a 2-cetotetose), que contém um centro estereogênico, como composto de referência.

Monossacarídeos com mais de seis carbonos não são abundantes na natureza, mas este fato não impede que estes sejam sintetizados, em todas as suas formas estereoisoméricas possíveis, resultando em uma grande diversidade de estruturas. Para se ter uma idéia de como a diversidade estrutural dos monossacarídeos cresce com seu tamanho, basta observar que o número de centros estereogênicos num monossacarídeo de n carbonos é $n - 2$ para aldoses e $n - 3$ para cetoses, e o número de estereoisômeros é 2^k , onde k é o número de centros estereogênicos. São conhecidos monossacarídeos de até nove carbonos, logo para aldoses o número de estereoisômeros é $2^7 = 128$, 64 D-nonoses e 64 L-nonoses. Como pode-se ver na Figura 1.4, o prefixo D na série das D-aldoses é referente à configuração do carbono assimétrico com maior numeração, ou seja, o centro estereogênico mais distante do grupo carbonílico. Na série D, a configuração do carbono assimétrico com mais alta numeração é sempre análoga àquela do carbono assimétrico do D-gliceraldeído, na série das L-aldoses, o carbono assimétrico com maior numeração terá sempre configuração análoga à do carbono assimétrico do L-gliceraldeído. Para as cetoses, da mesma forma, a configuração do carbono assimétrico com maior numeração é análoga à do carbono assimétrico da eritrulose (L ou D dependendo da série). Das séries das cetoses, cujas estruturas não serão mostradas aqui, temos, além da eritrulose, (até cetohexoses): ribulose e xilulose (pentoses);

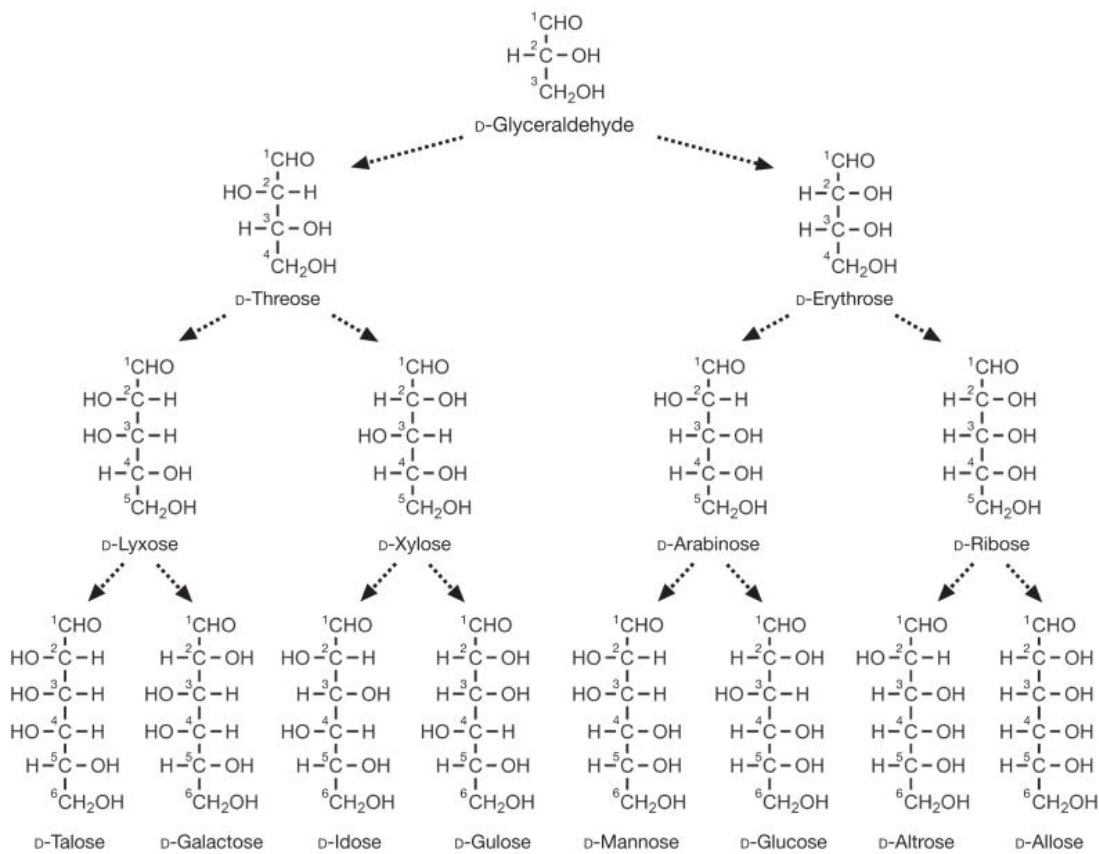


Figura 1.4: Projeções de Fischer para as formas acíclicas da série D das aldoses, obtida a partir do D-gliceraldeído, indo da triose às hexoses. Figura extraída de Carolyn R. Bertozzi and David Rabuka²⁰, *Essentials of Glycobiology*, chapter Structural Basis of Glycan Diversity. Cold Spring Harbor Laboratory Press, 2nd edition, 1999 (copyright 1999).

psicose, frutose, sorbose e tagatose (hexoses).¹⁹ Para cada composto em uma série D existe sua imagem especular na série L correspondente e a cada nome como os que constam na Figura 1.4 (alose, altrose, etc.) está associado um par enantiomérico. Dois monossacarídeos que diferem na configuração de um único carbono assimétrico são denominados *epímeros*. Por exemplo, pode-se observar na Figura 1.4 que a D-glicose § difere da D-manose apenas na configuração do carbono 2, portanto estes dois monossacarídeos são epímeros em relação ao carbono 2. De maneira similar D-glicose e D-alose são epímeros em relação ao carbono 3 e D-glicose e D-galactose são epímeros em relação ao carbono 4. Como a configuração no carbono 5 determina se a aldohexose pertence à série D ou à L, o epímero da D-glicose com relação ao carbono 5 é a L-idose.

§ A glicose é às vezes chamada de dextrose, ou então glucose.

A utilização destas séries ajuda a sistematizar os monossacarídeos e todas as suas possíveis formas estereoisoméricas, já que as séries percorrem sistematicamente todas as combinações de configurações possíveis para os carbonos assimétricos. Após discutir a sistemática de classificação dos monossacarídeos, os aspectos básicos de sua estrutura química e as suas possíveis configurações, temos agora que descrever as formas cíclicas dos monossacarídeos, que em solução ocorrem em equilíbrio com as formas acíclicas já apresentadas, para posteriormente entrar na análise conformacional das formas cíclicas que são as formas de ocorrência predominante dos monossacarídeos. Essas discussões terão implicação mais adiante, no estudo de estrutura e análise conformacional de polissacarídeos.

Formas cíclicas dos monossacarídeos

As aldoses e cetoses são compostos carbonílicos polihidroxilados e, portanto, possuem grupos funcionais capazes de reagirem entre si: a parte alcóolica (hidroxilada) pode reagir com a carbonila. A adição nucleofílica de um álcool a um grupo carbonílico forma um hemiacetal.¹ Quando a hidroxila e a carbonila fazem parte

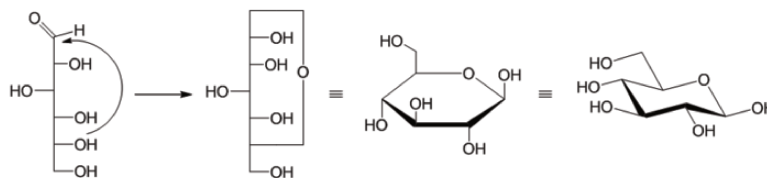


Figura 1.5: Formação de hemiacetal cíclico, fórmula de Haworth e conformação cadeira 4C_1 da β -D-glicopiranosose.

da mesma molécula, o produto formado é um hemiacetal cíclico, Figura 1.5. A formação de hemiacetais cíclicos é mais comum quando os anéis formados são compostos por cinco ou seis átomos, pois esses têm maior estabilidade e hemiacetais cíclicos de monossacarídeos com ciclos de cinco e seis membros são denominados *furanoses* e *piranoses*, respectivamente. A formação de monossacarídeos cíclicos origina um novo centro estereogênico no carbono 1 no caso de aldoses, ou no carbono 2 no caso de cetoses. Este carbono, que nas formas cíclicas é ligado a dois átomos de oxigênio, é chamado carbono *anomérico*. Os pares de formas isoméricas de monossacarídeos que diferem entre si apenas pela configuração do carbono anomérico são chamados *anômeros*: furanoses e piranoses podem existir em duas formas diastereoisoméricas denominadas α e β . Na forma α a hidroxila do carbono anomérico está do mesmo lado que a hidroxila do centro estereogênico

com mais alta numeração, quando visualizadas em uma projeção de Haworth. Na forma β a hidroxila do carbono anomérico está do lado oposto. A Figura 1.6 ilustra as formas anoméricas α e β da D-glicopiranosose em projeção de Haworth, que nada mais é do que uma variação da projeção de Fischer adaptada para sistemas cíclicos. A nomenclatura das formas cíclicas consiste em utilizar o nome do monos-

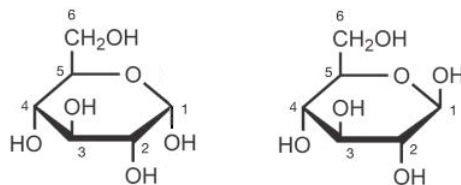


Figura 1.6: Formas anoméricas da D-glicopiranosose em projeções de Haworth. À esquerda anômero α e à direita anômero β . Figura adaptada de Carolyn R. Bertozzi and David Rabuka²⁰, *Essentials of Glycobiology*, chapter Structural Basis of Glycan Diversity. Cold Spring Harbor Laboratory Press, 2nd edition, 1999 (copyright 1999).

sacarídeo em conjunto com a denominação do tipo de anel formado, por exemplo: α -D-ribofuranose é um hemiacetal cíclico de D-ribose com anel de cinco membros e carbono anomérico na configuração α , enquanto β -D-glicopiranosose é a forma cíclica da D-glicose, cujo anel possui seis membros e o carbono anomérico tem configuração β . Em uma solução de um monossacarídeo em equilíbrio, as formas cíclicas coexistem com a forma acíclica e, embora a quantidade desta última seja ínfima na maioria dos casos, pelo menos duas formas anoméricas (α e β) costumam estar presentes em quantidades significativas, entretanto. Ao se dissolver cristais de uma forma anomérica pura, parte das moléculas em solução se transformarão na outra forma anomérica, tendo a forma aberta (acíclica) como intermediário reacional, até que a composição de equilíbrio seja atingida. Este processo é denominado *mutarotação*. Muitos monossacarídeos em solução apresentam até cinco espécies em equilíbrio tautomérico, presentes em quantidades detectáveis, como por exemplo a D-ribose para a qual experimentos de espectroscopia de RMN de 1H e ^{13}C de uma solução aquosa deste monossacarídeo mostram que cinco formas tautoméricas estão presentes: α -D-ribofuranose (6%), α -D-ribopiranosose (20%), β -D-ribofuranose (18%), β -D-ribopiranosose (56%) e a D-ribose acíclica (menos de 1%).¹

É comum utilizar as fórmulas de Haworth (Figuras 1.5 e 1.6) para representar as configurações das formas cíclicas dos monossacarídeos, entretanto estas fórmulas não fornecem informações a respeito da *conformação* destes anéis. Conformações são diferentes arranjos espaciais de uma mesma molécula que são gerados

por torções em torno de suas ligações químicas. Diferentes conformações de uma mesma molécula são às vezes chamadas *confôrmeros* ou *rotâmeros*. A princípio existem infinitas conformações possíveis para um monossacarídeo com uma dada configuração. Na prática, entretanto, as conformações mais estáveis predominam sobre as outras. O estudo detalhado das conformações dos compostos químicos é chamado *análise conformacional*. Para representar conformações adequadamente é necessário utilizar desenhos em perspectiva destes rotâmeros, como os que aparecem na Figura 1.5 (último desenho) e na Figura 1.7. Para o estudo das conformações dos polissacarídeos é necessário estudar as conformações de piranoses e furanoses (suas unidades, cíclicas, básicas) e posteriormente os graus de liberdade conformacional advindos de torções em torno das ligações que unem estes anéis, ligações estas presentes em dissacarídeos, oligo e polissacarídeos.

Análise conformacional em piranoses

As inúmeras conformações assumidas por piranoses são classificadas em diferentes famílias denominadas cadeira, bote, bote torcido, meia cadeira e envelope, representadas por símbolos que consistem em uma letra maiúscula acompanhada de índices e superíndices. A cada família corresponde uma letra maiúscula: C, B, S, H, E, respectivamente, do inglês *chair*, *boat*, *skew-boat*, *half-chair* e *envelope*. Os índices e superíndices podem ser os inteiros 1, 2, 3, 4, 5 ou o caractere "O". Os números correspondem à numeração dos carbonos na piranose, a letra O representa o oxigênio do ciclo. As famílias de conformações C, B, S e H são esquematicamente representadas¹⁹ por quatro átomos coplanares e dois átomos fora deste plano, acima ou abaixo. O número (ou a letra O) correspondente a um átomo acima do plano constará no símbolo que representa a conformação como superíndice, e o número (ou a letra O) correspondente a um átomo abaixo do plano constará como índice. Por exemplo, a conformação representada pelo símbolo 1C_4 é uma conformação da família cadeira (*chair*) com o carbono 1 da piranose acima do plano em que estão os quatro átomos C2, C3, C5 e O, e o carbono 4 abaixo deste plano. ¶ O que é acima ou abaixo do plano dos quatro átomos é definido da seguinte maneira: vendo este plano de cima, a numeração (crescente) dos átomos no anel deve estar em sentido horário. Para a família de conformações B, os dois átomos fora do referido plano estão ambos acima ou abaixo deste, e com isso as

¹⁹Neste texto os símbolos de elementos químicos estão grafados na mesma fonte tipográfica que o texto, como em C, H ou O, e os símbolos que representam conformações são grafados em fonte sem serifas, como em C, B, S, H, E.

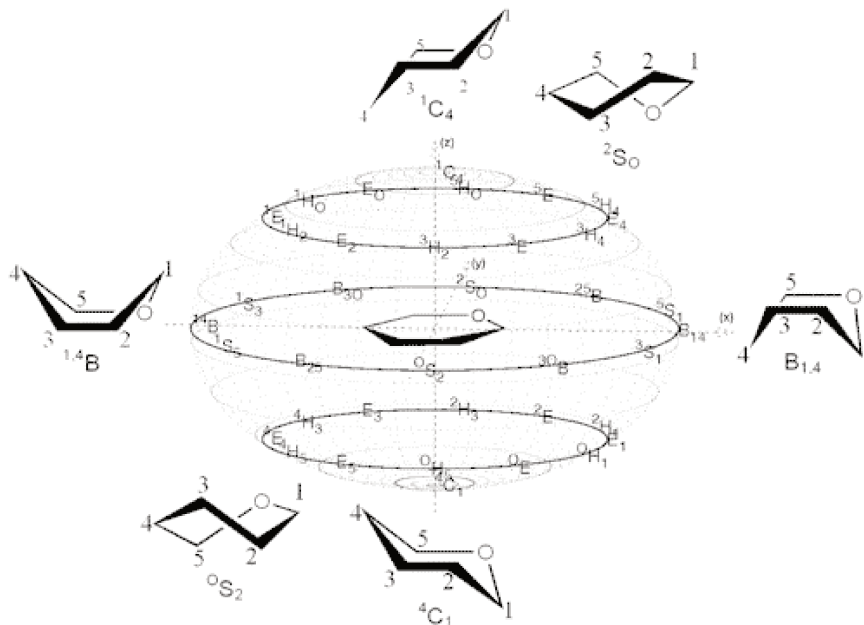


Figura 1.7: Mapeamento esférico das conformações em piranoses. Figura extraída de Ionescu *et al.*²¹, *Journal of Physical Chemistry A*, 109(36):8096–8105, 2005 e reproduzida com permissão da American Chemical Society (Copyright 2005).

conformações são denotadas com dois índices (como em $B_{1,4}$) ou dois superíndices (como em ${}^{1,4}B$). Nas famílias C e B, os dois átomos do ciclo que estão fora do plano são o i e o $i + 3$, enquanto que para a família de conformações S são os átomos i e $i + 2$ que estão fora do plano definido pelos outros quatro. Além disso, há mais de uma maneira de definir o “melhor plano” contendo estes quatro átomos (ou seja, qual conjunto de quatro átomos está mais próximo de ser coplanar) e portanto há duas notações equivalentes (os símbolos 0S_2 e 3S_5 , por exemplo, correspondem à mesma conformação). Na família H, os átomos i e $i + 1$ é que estão fora do plano definido pelos outros quatro, e na família E há apenas um átomo fora de um plano contendo os outros cinco e, portanto, os símbolos que representam conformações desta última família têm apenas um indicador, seja ele índice ou superíndice.

Todas as conformações possíveis para uma piranose podem ser representadas por uma combinação linear de uma conformação da família cadeira, uma da família bote e uma da família bote torcido, sendo mais conveniente usar as conformações 1C_4 , ${}^{1,4}B$ e 0S_2 , chamadas conformações *canônicas*, para este propósito.²² O universo de conformações de uma piranose pode também ser representado por uma esfera, uma idéia proposta por Hendrickson²³, sendo que a representação mais famosa é aquela proposta por Cremer e Pople²⁴. Entretanto,

a definição de coordenadas internas naturais, por Pulay e colaboradores^{25;26}, é a base para a uma representação mais moderna proposta por Bérces, Whitfield e Nukada²², e Nukada e colaboradores²¹, onde definem-se seis coordenadas internas de torção, $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6$, que correspondem, respectivamente, aos diedros C1-C2-C3-C4, C2-C3-C4-C5, C3-C4-C5-O, C4-C5-O-C1, C5-O-C1-C2 e O-C1-C2-C3 do anel de piranose. Com as seis coordenadas de torção assim definidas, e considerando a conformação planar correspondendo à origem de um sistema de coordenadas como na Figura 1.7, a coordenada interna

$$q_1 = \tau_1 - \tau_2 + \tau_3 - \tau_4 + \tau_5 - \tau_6$$

descreve o deslocamento em direção à conformação 1C_4 , a coordenada

$$q_2 = \tau_2 - \tau_3 + \tau_5 - \tau_6$$

descreve o deslocamento em direção à conformação ${}^{1,4}B$ e a coordenada

$$q_3 = \tau_1 - (\tau_2/2) - (\tau_3/2) + \tau_4 - (\tau_5/2) - (\tau_6/2)$$

descreve o deslocamento em direção à conformação 0S_2 . Portanto, o conjunto destas coordenadas internas (q_1, q_2, q_3) é um conjunto de base para o espaço de conformações originadas por quaisquer torções dos diedros do ciclo. Entretanto, ainda restam os diedros *exocíclicos*, que compreendem torções em torno de todas as ligações entre hidroxilas e os carbonos do ciclo e também as torções do grupo hidroximetil, se este estiver presente (vide por exemplo a Figura 1.5, onde é fácil visualizar as torções exocíclicas).

No esquema mais simples para tais torções, são consideradas 3 conformações discretas mais estáveis para cada hidroxila exocíclica, então para cada conformação do anel da piranose (Figura 1.7), haveriam 3^m diferentes conformações possíveis para as m hidroxilas exocíclicas. Além dessas, se o anel possuir um grupo hidroximetil, consideram-se ainda três conformações discretas mais estáveis, chamadas *gg*, *gt* e *tg*, para a torção da ligação entre o carbono do grupo hidroximetil e o carbono do ciclo ao qual ele está ligado, como mostrado na Figura 1.8. Para o caso das aldohexopiranoses esta ligação é a C5-C6. As duas letras dos símbolos *gg*, *gt* e *tg* (*g* para *gauche*, *t* para *trans*), no caso de aldohexopiranoses, se referem à posição relativa entre o O6 e o O5, e entre o O6 e o C4 respectivamente, em uma projeção de Newman¹ em torno da ligação C5-C6. Este ângulo de torção é

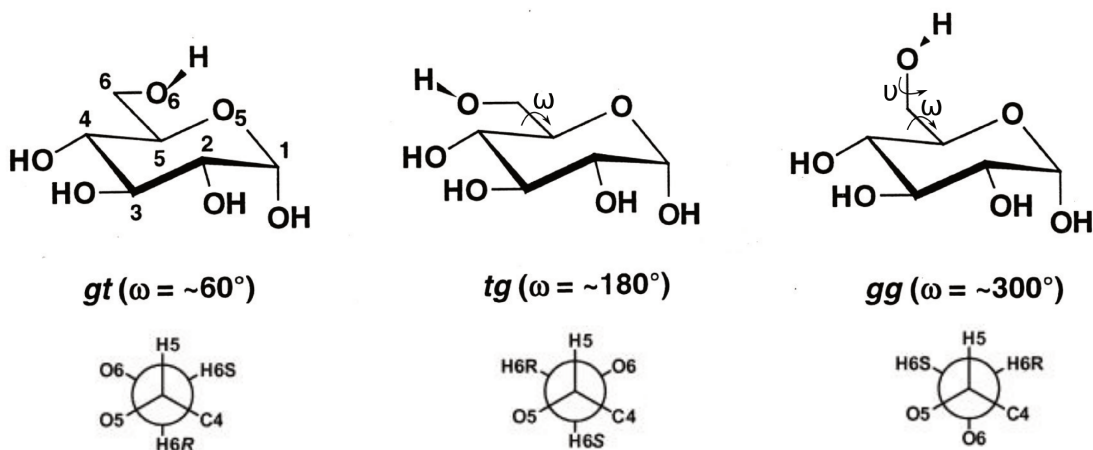


Figura 1.8: Conformações mais estáveis para o grupo exocíclico hidroximetil, em hexopiranoses. O ângulo ω corresponde à rotação em torno da ligação C5-C6 ou, mais especificamente, à coordenada de torção do diedro O5-C5-C6-O6. Esta coordenada tem valor igual a 60° para o confôrmero *gt*, 180° para o confôrmero *tg* e 300° para o confôrmero *gg*. As duas letras dos símbolos *gt*, *gg* e *tg* indicam, respectivamente, se as posições relativas entre O6 e O5 e entre O6 e C4 são *gauche* ou *trans*. Estas posições relativas são mais facilmente visualizadas em projeções de Newman em torno do eixo da ligação C5-C6: no confôrmero *gt*, por exemplo, o átomo O6 está *gauche* ao O5 e *trans* ao C4. Figura adaptada de Karl N. Kirschner and Robert J. Woods²⁷, Solvent interactions determine carbohydrate conformation. Proceedings of the National Academy of Sciences, 98(19):10541–10545, 2001 (copyright 2001) .

geralmente denominado ω e está representado nas Figuras 1.8 e 1.10, juntamente com o ângulo de torção ν em torno da ligação entre o oxigênio e o carbono do grupo hidroximetil.

Análise conformacional em furanoses

Um formalismo análogo ao que foi apresentado para o mapeamento das conformações do anel em piranoses existe para furanoses. Nesse formalismo, são definidas coordenadas internas de torção ϕ_0 , ϕ_1 , ϕ_2 , ϕ_3 , ϕ_4 correspondentes às torções em torno das ligações químicas do anel de furanose (Figura 1.9). Cada conformação é representada por duas coordenadas P e ϕ_m :

$$\tan P = \frac{(\phi_2 + \phi_4) - (\phi_1 + \phi_3)}{3.077\phi_0} ,$$

$$\phi_m = \phi_0 / \cos P .$$

P é chamada ângulo de fase da pseudorotação, e representa o tipo de conformação, e ϕ_m é uma medida da magnitude do deslocamento daquela conformação em relação à forma planar.^{28;29} As conformações do anel de furanose são classificadas

em *envelope* e *twist*. A forma envelope é deformada em relação à conformação planar por ter um átomo fora do plano em que estão os outros quatro, enquanto a forma twist tem dois átomos fora do plano definido pelos outros três. Uma permutação cíclica e ordenada dos arranjos possíveis com um e dois átomos fora do plano descreve um itinerário de pseudorotação, conforme representado na Figura 1.9. Assim como para as conformações das piranoses, os símbolos usados para representar as conformações das furanoses consistem em uma letra maiúscula (E ou T neste caso) que leva índices ou superíndices de acordo com quantos e quais átomos estão fora do plano em que os demais átomos do anel estão contidos. As furanoses também possuem liberdade conformacional referente às torções exocíclicas.

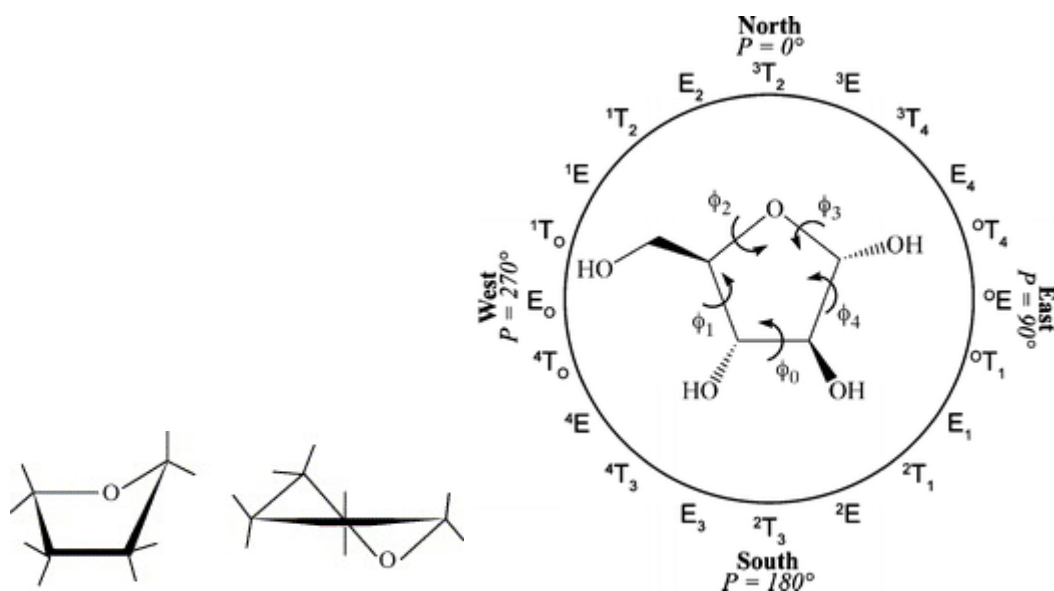


Figura 1.9: Conformações do anel das furanoses. À esquerda: conformação envelope, com apenas um átomo fora do plano que contém os outros quatro. No centro: conformação twist, com um átomo acima e outro abaixo do plano definido pelos três demais. À direita: itinerário pseudorotacional em furanoses. Adaptada de Hashem *et al.*²⁹, Conformational analysis of arabinofuranosides: Prediction of (3)J(H,H) using MD Simulations with DFT-derived spin-spin coupling profiles. *Journal of Chemical Theory and Computation*, 6(1): 212–222, 2010, e reproduzida com permissão da American Chemical Society (copyright 2010).

Em resumo, para a análise conformacional de monossacarídeos é necessário considerar as torções em torno das ligações que compõem o ciclo e as torções exocíclicas. Embora a análise conformacional de monossacarídeos não seja simples, mais torções devem ser levadas em conta quando monossacarídeos se combinam em dissacarídeos ou em carboidratos mais complexos. As ligações formadas entre os monossacarídeos implicam em coordenadas diedrais adicionais para descrever as conformações dos dissacarídeos formados.

Dissacarídeos

Dissacarídeos, por definição, são compostos que formam dois monossacarídeos ao sofrer hidrólise. O processo inverso, a *condensação* ^{||} de dois monossacarídeos leva à formação de um dissacarídeo e de uma molécula de água. Esse processo envolve, mais precisamente, a condensação da hidroxila do carbono anomérico de um monossacarídeo com qualquer hidroxila do outro monossacarídeo. Por este motivo, ainda que os monossacarídeos envolvidos na reação de condensação sejam idênticos, vários dissacarídeos diferentes podem ser formados. Duas moléculas de D-glicose, por exemplo, podem formar onze dissacarídeos diferentes (Tabela 1.1).

As ligações entre os monossacarídeos são denominadas ligações *glicosídicas* (Figura 1.10). O átomo de oxigênio que conecta os dois monossacarídeos é o oxigênio glicosídico. O monossacarídeo cujo carbono anomérico está envolvido na ligação glicosídica é chamado de *unidade não redutora*, o outro monossacarídeo é a *unidade redutora*. A unidade redutora de um dissacarídeo sofre mutarotação, logo, assim como os monossacarídeos, os dissacarídeos podem existir em duas formas anoméricas distintas que se interconvertem em solução.

Na unidade não redutora, a configuração do carbono anomérico é fixa devido à formação do dissacarídeo, e a configuração neste carbono é outro fator que contribui, além das múltiplas hidroxilas, para que sejam muitos os dissacarídeos que podem ser formados a partir de dois monossacarídeos idênticos: há duas possibilidades para a configuração do carbono anomérico da unidade não redutora, α e β . Se for a α , a ligação glicosídica é denotada α -(1 \rightarrow i), sendo i a numeração do carbono da unidade redutora envolvido na ligação glicosídica. Se a configuração do carbono anomérico da unidade não redutora for a β , a ligação glicosídica é denotada β -(1 \rightarrow i).

Duas aldoses também podem estar unidas por uma ligação glicosídica do tipo (1 \leftrightarrow 1) e nesse caso, ambas as unidades são não redutoras e suas configurações anoméricas podem ser as mesmas ou não. Portanto, é possível a formação de três dissacarídeos diferentes com ligações do tipo (1 \leftrightarrow 1), ou quatro se os monossacarídeos não forem idênticos. A sacarose ^{††}, um dissacarídeo formado

^{||}A palavra condensação não está sendo usada para denominar uma mudança de estado físico da matéria (transição de fase), mas sim para designar uma reação química na qual duas moléculas são unidas para formar uma molécula maior e também uma pequena molécula, neste caso água. A palavra condensação é bastante usada com este significado na área de polímeros, para fazer referência a um tipo de processo de polimerização em que os monômeros ligam-se quimicamente e, simultaneamente, formam uma pequena molécula que pode ser água, *HCl* etc.

^{††}Açúcar comum de mesa.

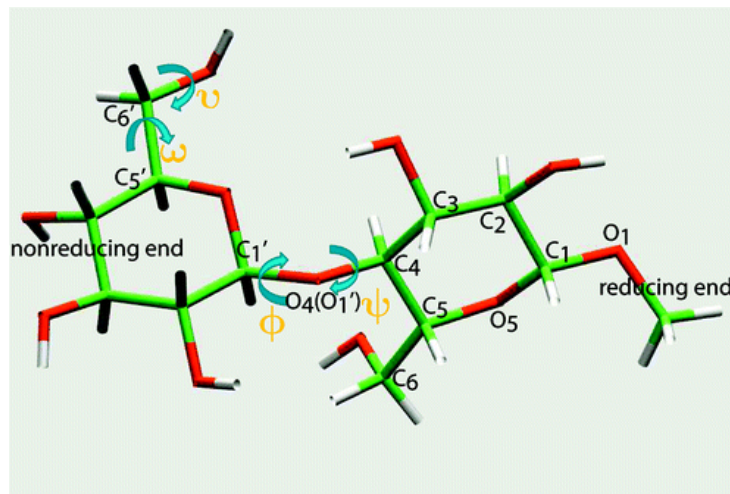


Figura 1.10: Estrutura do metil- β -D-celobiosídeo, derivado metilado da β -D-celobiose, um dissacarídeo (o sufixo *osídeo* indica substituição no oxigênio 1 da unidade redutora). As ligações glicosídicas são aquelas em que participa o oxigênio glicosídico, e torções em torno destas ligações são simbolizadas pelas coordenadas de torção ϕ e ψ . Estas duas coordenadas estão sendo apresentadas agora, pois elas advém da formação de ligações glicosídicas entre monossacarídeos. As coordenadas ω e ν , entretanto, já foram apresentadas na Figura 1.8. Figura extraída de Tongye *et al.*³⁰, Conformational flexibility of soluble cellulose oligomers: chain length and temperature dependence. Journal of the American Chemical Society, 131(41):14786–14794, 2009, e reproduzida com permissão da American Chemical Society (copyright 2009) .

pela condensação de D-glicose e D-frutose, é um exemplo de dissacarídeo não redutor, ou seja, no qual as duas unidades são não redutoras. Entretanto, a ligação glicosídica é do tipo (2 \leftrightarrow 1) já que a frutose é uma cetose (e não uma aldose como a glicose) e para cetoses o carbono anomérico tem numeração 2 (vide definição de carbono anomérico, Seção 1.2.1). Segundo a sistemática que vem sendo exposta, a sacarose é denominada β -D-frutofuranosil-(2 \leftrightarrow 1)- α -D-glicopiranosídeo. De maneira geral, portanto, as ligações glicosídicas são do tipo (1 \rightarrow i) para aldoses e (2 \rightarrow i) para cetoses.

O exemplo da Tabela 1.1 ilustra como duas unidades de um único tipo de monossacarídeo pode formar vários dissacarídeos distintos. Imagine-se a diversidade de dissacarídeos possíveis considerando agora combinações com um monossacarídeo e sua imagem especular (no caso do exemplo dado, D-glicose e L-glicose). Ou então, imagine-se todas as combinações possíveis incluindo os diferentes monossacarídeos. Ou seja, não bastasse a grande diversidade estrutural dos monossacarídeos, o número de dissacarídeos possíveis na natureza é ainda imensamente amplificada pelo fato de serem inúmeras as possibilidades de formação de

Tabela 1.1: Dissacarídeos formados a partir da condensação de duas unidades (idênticas) de D-glicose.¹⁹

Ligação glicosídica	Nome comum	Classificação
α -(1 \leftrightarrow 1)- α	α,α -Trealose	α -D-glicopiranosil-(1 \leftrightarrow 1)- α -D-glicopiranosídeo
α -(1 \leftrightarrow 1)- β	α,β -Trealose	α -D-glicopiranosil-(1 \leftrightarrow 1)- β -D-glicopiranosídeo
β -(1 \leftrightarrow 1)- β	β,β -Trealose	β -D-glicopiranosil-(1 \leftrightarrow 1)- β -D-glicopiranosídeo
α -(1 \rightarrow 2)	Kojibiose	α -D-glicopiranosil-(1 \rightarrow 2)-D-glicose
α -(1 \rightarrow 3)	Nigerose	α -D-glicopiranosil-(1 \rightarrow 3)-D-glicose
α -(1 \rightarrow 4)	Maltose	α -D-glicopiranosil-(1 \rightarrow 4)-D-glicose
α -(1 \rightarrow 6)	Isomaltose	α -D-glicopiranosil-(1 \rightarrow 6)-D-glicose
β -(1 \rightarrow 2)	Soforose	β -D-glicopiranosil-(1 \rightarrow 2)-D-glicose
β -(1 \rightarrow 3)	Laminarabiose	β -D-glicopiranosil-(1 \rightarrow 3)-D-glicose
β -(1 \rightarrow 4)	Celobiose	β -D-glicopiranosil-(1 \rightarrow 4)-D-glicose
β -(1 \rightarrow 6)	Gentiobiose	β -D-glicopiranosil-(1 \rightarrow 6)-D-glicose

ligações de natureza estrutural diferente entre dois monossacarídeos, uma vez que estes são compostos polihidroxilados, quirais e altamente assimétricos, possuindo duas configurações possíveis para o carbono anomérico. Como veremos a seguir, do ponto de vista estrutural, em alguns casos é mais correto definir não os monossacarídeos, mas os dissacarídeos como unidades monoméricas dos polissacarídeos. Se a diversidade de dissacarídeos é grande, a de óligo e polissacarídeos é ainda maior.

1.2.2 Polissacarídeos: diversidade e complexidade estrutural

Embora também sejam carboidratos, é conveniente, no presente contexto, introduzir uma seção exclusiva para os polissacarídeos. Oligo e polissacarídeos são chamados, em uma denominação mais moderna, de *glicanos* ou *glicanas*. Os glicanos apresentam diferenças estruturais peculiares se comparados a outros biopolímeros, como proteínas e ácidos nucleicos, no que diz respeito a sua estrutura primária. Pelo que foi exposto nas seções anteriores fica claro que a diversidade de monômeros que podem constituir um glicano (mais de 200 tipos) é muito maior do que em proteínas (20 tipos de aminoácidos) ou ácidos nucleicos (4 ou 5 tipos de nucleotídeos), o que, por si só, torna muito grande a diversidade de glicanos possíveis na natureza. Ocorre que, devido à estrutura química de seus monômeros, os glicanos tem ainda outra peculiaridade, em relação aos outros biopolímeros, que aumenta ainda mais a sua diversidade: a possibilidade de ramificação da cadeia

Tabela 1.2: Estrutura primária^a de alguns glicanos¹⁹.

Nome	Unidades de monossacarídeos	Ligações
Ácido algínico	ácidos β -D-manurônico e α -L-gulurônico	(1 \rightarrow 4)
Amilopectina	α -D-glicose	(1 \rightarrow 4) com ramificações (1 \rightarrow 6)
Amilose	α -D-glicose	(1 \rightarrow 4)
Carragenano	β -D-galactose	(1 \rightarrow 3) e (1 \rightarrow 4) alternantes
Celulose	β -D-glicose	(1 \rightarrow 4)
“Crown-gall”	β -D-glicose	(1 \rightarrow 2)
Dextrano	α -D-glicose	Maioria (1 \rightarrow 6) com ramificações (1 \rightarrow 3)
Glicogênio	α -D-glicose	(1 \rightarrow 4) com ramificações (1 \rightarrow 6)
Laminarano ^b	β -D-glicose	(1 \rightarrow 3)
Manano	β -D-manose	(1 \rightarrow 4)
Micodextrano	α -D-glicose	(1 \rightarrow 3) e (1 \rightarrow 4) alternantes
Quitina	N-acetil- β -D-glicosamina	(1 \rightarrow 4)
β -(1 \rightarrow 3) Xilano	β -D-xilose	(1 \rightarrow 3)
β -(1 \rightarrow 4) Xilano	β -D-xilose	(1 \rightarrow 4)

^a Todos os monossacarídeos estão na forma de piranoses. ^b Também denominada laminarina, abordada na Seção 1.3.1.

polimérica. Como foi visto na seção anterior, dois monossacarídeos podem ligar-se de várias maneiras distintas, enquanto existe apenas uma única maneira de ligar dois aminoácidos ou dois nucleotídeos. Essa característica estrutural permite não só a existência de diferentes polímeros a partir do mesmo monômero (Tabelas 1.1 e 1.2), mas também a existência de glicanos ramificados (Tabela 1.2).

Diante dessa diversidade estrutural, os glicanos são classificados em *homoglicanos*, quando são compostos apenas de um tipo de monossacarídeo, ou *heteroglicanos* se são compostos de mais de um tipo de monossacarídeo. Homo e heteroglicanos podem ser classificados ainda em lineares (não ramificados) ou ramificados. Todos os glicanos envolvem necessariamente as formas cíclicas do monossacarídeos, na sua maioria piranoses, mas também furanoses.¹⁰ O primeiro glicano na Tabela 1.2, o ácido algínico, é composto de monossacarídeos que tiveram suas hidroxilas terminais oxidadas à ácidos carboxílicos. Esse glicano, às vezes chamado de algina, também pode ocorrer sob a forma de um sal, caso em que é denominado alginato (de sódio ou de cálcio por exemplo). Glicanos podem, portanto, ser polieletrólitos. Outro glicano que tem seus monossacarídeos modificados é a quitina, principal constituinte do exoesqueleto dos artrópodes, e seus monossacarídeos tiveram uma hidroxila exocíclica substituída por um grupo acetilamina. Esses são apenas dois exemplos mostrando que glicanos podem ser constituídos de monossacarídeos derivatizados (ou funcionalizados), e que é grande a diversidade de possibilidades de derivatização. Outros derivados importantes dos monossacarídeos são os *deoxi* açúcares e os *anidro* açúcares. Em sistemas biológicos, glicanos são encontrados

quimicamente ligados a proteínas, aminoácidos e lipídeos, em sistemas denominados glicoproteínas, proteoglicanos, peptidoglicanos e glicolipídeos, desempenhando inúmeras funções.

Como já foi definido, o termo glicano se refere a qualquer polissacarídeo, entretanto existem denominações mais específicas para determinadas famílias de homoglicanos. Por exemplo, os homoglicanos constituídos apenas de D-glucose (aqui a palavra *apenas* é redundante, pois o prefixo *homo* já denomina constituição por um único tipo de monossacarídeo) são chamados D-glucanos. Como a maioria dos monossacarídeos encontrados em organismos vivos são isômeros D, muitas vezes o prefixo D é omitido. Assim homoglicanos constituídos por D-manose, por exemplo, podem ser denominados simplesmente *mananos*. Analogamente, homoglicanos de D-xilose são *xilanos*, de D-arabinose são *arabinanos*, de D-galactose são *galactanos*, de D-frutose são *frutanos*, e assim por diante. Também usam-se as flexões de gênero feminino destes substantivos (como em *glucanas* ou *xilanas*).

De maneira ainda mais específica, dentro da família dos glucanos, por exemplo, pode-se falar em α -glucanos ou em β -glucanos, se as ligações glicosídicas forem todas do tipo α ou todas do tipo β . Amilose, amilopectina e glicogênio são exemplos de α -glucanos, enquanto celulose e laminarina são exemplos de β -glucanos. Para α -glucanos e β -glucanos lineares, é estruturalmente mais correto definir os correspondentes dissacarídeos como seus monômeros constituintes, pois a identidade do glucano é dada não só pelo tipo de monossacarídeo envolvido (D-glucose em todos esses casos), mas também pelo tipo de ligação glicosídica formada entre os monossacarídeos. A ligação glicosídica é tão importante que ela determina não apenas a identidade, mas todas as propriedades estruturais e físico-químicas do glucano. Basta dizer que amilose e celulose, embora sejam glucanos lineares com ligações (1 \rightarrow 4), têm propriedades totalmente diferentes pois na amilose as ligações glicosídicas são α -(1 \rightarrow 4) e na celulose elas são β -(1 \rightarrow 4). As cadeias de amilose têm estrutura secundária helicoidal, enquanto as cadeias de celulose são fitas planares. A amilose (um dos componetes do amido, junto à amilopectina) pode ser digerida pelos animais superiores pois estes possuem em seu trato digestivo enzimas específicas para a hidrólise de α -glucanos, genericamente classificadas como α -glicosidases (um exemplo é a α -amilase presente em nossa saliva). Por outro lado, apenas microorganismos (como, por exemplo, bactérias simbiontes presentes no sistema digestivo de ruminantes, ou protozoários presentes no sistema digestivo de cupins) e alguns fungos parasitas ou decompositores de plantas são capazes de hidrolisar a celulose, pois expressam β -glicosidases. Portanto, no

caso desses importantes glucanos lineares, define-se a maltose (vide Tabela 1.1) como unidade monomérica da amilose, e a celobiose (Tabela 1.1) como unidade monomérica da celulose. Definidos dessa maneira, as estruturas desses polímeros podem ser matematicamente geradas a partir de repetidas operações de simetria (de grupo espacial) aplicadas a um único monômero. Assim sendo, poderia-se dizer que polímeros regulares (não randômicos) são cristais unidimensionais, e as cadeias poliméricas destes glucanos de fato assumem estruturas regulares em suas diversas fases cristalinas. Em solução, entretanto, as cadeias poliméricas podem assumir uma infinidade de conformações randômicas.

A definição da unidade estrutural básica de um glicano é variada. Ao tentar estudar a estrutura de amilose ou celulose por hidrólise completa, chegaria-se à conclusão de que são homoglicanos cujo monômero é a D-glucose. Entretanto, como foi discutido, a unidade estrutural quimicamente correta de cada um desses glucanos é um dissacarídeo (maltose num caso, celobiose no outro). Também ocorre que a unidade estrutural biológica de um glicano pode ser uma seqüência de alguns monossacarídeos a qual apresente um tamanho mínimo a partir do qual o glicano possa ser reconhecido por mecanismos bioquímicos de reconhecimento molecular.³¹

Os heteroglicanos, ao contrário dos homoglicanos, são formados por mais de um tipo de monossacarídeo, mas também podem ter estrutura primária linear ou ramificada. Algumas denominações comuns dadas a heteroglicanos presentes em tecidos vegetais são, por exemplo, arabinoxilanos, arabinogalactanos e galactomananos. Entretanto, esses nomes não informam nada sobre a estrutura primária desses glicanos, exceto pelos tipos de monossacarídeos que os compõem. Para representar adequadamente a estrutura primária de glicanos é necessário recorrer a diagramas como aquele da Figura 1.11, em que os nomes dos monossacarídeos constituintes aparecem abreviados (como em *Glc* para glicose, *Gal* para galactose, etc.) e em que consta toda a informação estrutural primária, como o tipo de ligação glicosídica formada entre os monossacarídeos, e qual das formas cíclicas está sendo adotada por cada monossacarídeo (a letra *p*, após as abreviações *Glc*, *Gal* etc., está para indicar piranoses e a letra *f* indica furanoses). A Figura 1.11 representa a estrutura de um monômero daquilo que alguns autores¹⁹ poderiam chamar de *poli-oligossacarídeos*. Essa classificação não deve ser confundida com heteroglicanos, uma vez que um homoglicano com ramificações regulares ao longo de sua cadeia principal também poderia ser chamado de poli-oligossacarídeo.

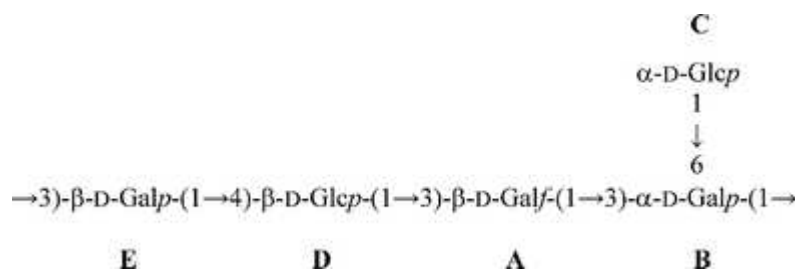


Figura 1.11: Representação do monômero de um exopolissacarídeo produzido pela linhagem LBB.B26 de *Lactobacillus delbrueckii*. Figura extraída de Sanchez-Medina *et al.*¹⁰, Conformational analysis of the neutral exopolysaccharide produced by *Lactobacillus delbrueckii* ssp *bulgaricus* LBB.B26. *Organic & Biomolecular Chemistry*, 7(2):280–287, 2009, e reproduzida com permissão da Royal Society of Chemistry (copyright 2009).

Como em outros biopolímeros, uma determinada estrutura^{‡‡} primária pode levar à formação de uma estrutura secundária. A estrutura secundária é a disposição relativa dos átomos em alguns segmentos localizados da cadeia polimérica, ou seja, é a forma adotada por alguns segmentos do biopolímero (Figura 1.12), enquanto a estrutura terciária é a que descreve as posições dos átomos no espaço tridimensional. Os segmentos de um biopolímero que apresentam estrutura secundária o fazem porque assumem conformações que são estabilizadas por ligações de hidrogênio entre monossacarídeos que não necessariamente são adjacentes em termos de estrutura primária. Uma vez que as estruturas secundárias na realidade representam regiões do espaço total de conformações do glicano, o estudo de suas estruturas secundárias (e de mais alta ordem) exige a sua análise conformacional.

O problema da análise conformacional em glicanos

As conformações de um glicano serão determinadas pela conformação individual dos anéis que compõem a cadeia polimérica (Seção 1.2.1), e também pelo valor das coordenadas de torção diedral ϕ e ψ (vide Figura 1.10). Um glicano linear composto de N monossacarídeos pode ser descrito especificando-se $(N - 1)$ pares (ou triplas em alguns casos, vide parágrafo seguinte) de ângulos diedros, como em $(\phi_1, \psi_1; \phi_2, \psi_2; \dots \phi_i, \psi_i; \dots \phi_{N-1}, \psi_{N-1})$. Se esses ângulos diedros forem iguais em cada oxigênio glicosídico (ou seja, $\phi_1 = \phi_2 = \dots = \phi_{N-1}$ e $\psi_1 = \psi_2 = \dots = \psi_{N-1}$), o glicano assume uma forma helicoidal a qual, geralmente, é encontrada no estado sólido e pode ser determinada por técnicas de difração de raios-X. Entretanto,

^{‡‡}Para proteínas e ácido nucleicos bastaria dizer *seqüência* primária, mas para glicanos o termo *estrutura* primária é mais adequado.

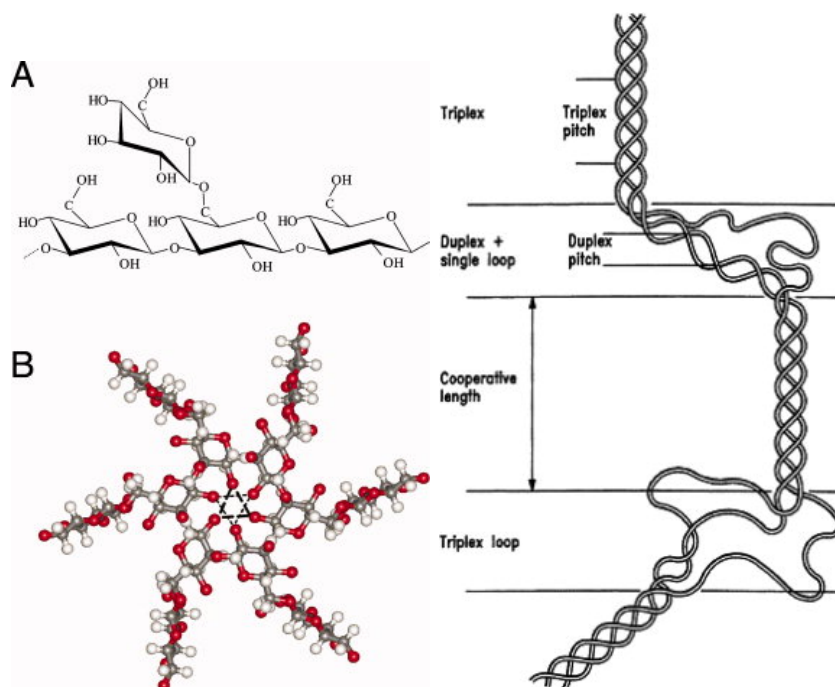


Figura 1.12: À esquerda: A) Unidade monomérica do escleroglucano, um β -(1 \rightarrow 3)-glucano com ramificações β -(1 \rightarrow 6). B) Seção transversal da estrutura em tripla hélice do escleroglucano. Figura extraída de Marit Sletmoen e Bjorn T. Stokke³², Review : Higher order structure of (1,3)-beta-D-glucans and its influence on their biological activities and complexation abilities. *Biopolymers*, 89(4):310-321, 2008, reproduzida com permissão da John Wiley & Sons (copyright 2008). À direita: ilustração esquemática das possíveis imperfeições em uma tripla hélice renaturada. Figura extraída de B. H. Falch and B. T. Stokke³³, Structural stability of (1-3)- β -D-glucan macrocycles. *Carbohydrate Polymers*, 44(2):113-121, 2001, reproduzida com permissão da Elsevier (copyright 2001).

se o valor desses ângulos diedros não forem correlacionados, os glicanos assumem formas irregulares ou randômicas.

Em relação aos movimentos conformacionais dos anéis individuais, muitos trabalhos sobre glicanos constituídos por piranoses consideram estes anéis conformacionalmente rígidos, adotando a sua conformação mais estável (por exemplo, a conformação mais estável de glucopiranoses é a cadeira 4C_1), embora existam trabalhos^{30;34-36} em que a liberdade conformacional das piranoses é considerada, pelo menos em certa extensão, admitindo-se transições para outras conformações (por exemplo para a cadeira 1C_4 no caso de glucopiranoses). Em glicanos cuja cadeia principal envolva ligações glicosídicas (1 \rightarrow 6), três coordenadas diedrais são importantes: ϕ , ψ e ω . Em glicanos com ramificações (1 \rightarrow 6), o fato de haver torção em torno dessas três coordenadas diedrais confere grande liberdade de movimento

aos resíduos ** laterais. As furanoses, ainda que ocorram com menor frequência, são geralmente consideradas bem mais flexíveis que as piranoses, e seus movimentos conformacionais são considerados importantes para o movimento do glicano como um todo.^{9;10;29}

Antes da ampla disponibilidade de técnicas experimentais instrumentais, usava-se uma bateria de testes químicos para obter informações sobre a estrutura dos glicanos, principalmente sobre a sua estrutura primária. Hoje em dia, técnicas como polarimetria funcionam bem para mono e dissacarídeos. Para oligossacarídeos, técnicas poderosas com a ressonância magnética nuclear podem ser de grande valia, mas tornam-se menos úteis à medida que aumentam o tamanho e a flexibilidade do glicano. A difração de raios-X também é uma técnica valiosa na determinação estrutural de glicanos em fase cristalina. Para glicanos de alta massa molecular, dissolvidos, intumescidos ou suspensos em solventes, entretanto, o número de conformações possíveis pode ser tão grande que a sua determinação experimental é muito difícil, ou até mesmo impossível. Portanto, para o estudo das conformações e da dinâmica de grandes glicanos é imprescindível fazer uso de abordagens computacionais. Neste contexto, o presente trabalho de tese pretende trazer contribuições para o entendimento da dinâmica de grandes glicanos, cujos resultados são apresentados nos Capítulos 4 e 5, mas também para o estudo das enzimas que catalisam a hidrólise de tais glicanos, que são introduzidas na próxima seção.

1.3 Enzimas Glicosídeo Hidrolases

As enzimas que catalisam a hidrólise de ligações glicosídicas são genericamente denominadas *Glicosídeo Hidrolases* ou GH. Enzimas GH são proteínas abundantemente encontradas em vários organismos pelo seu papel essencial na degradação de glicanos. Dada a enorme diversidade de glicanos existentes, há uma enorme diversidade de enzimas GH e há também vários sistemas de classificação destas enzimas. Há dois destes sistemas de classificação que provavelmente são os mais úteis para os químicos: o primeiro é o sistema de classificação “EC” (sigla para a expressão em inglês *Enzyme Commission*)³⁷ e o segundo é um sistema de classificação baseado na seqüência dos aminoácidos da enzima.^{38;39} Estes dois sistemas

**Resíduo é o nome genérico empregado para fazer referência a monossacarídeos, aminoácidos ou nucleotídeos, após a sua condensação, quando eles são parte de um biopolímero qualquer. O nome resíduo não é usado para os monômeros livres (não condensados).

Tabela 1.3: Primeiro nível do sistema EC de classificação de enzimas, de acordo com a IUPAC e IUBMB^a.

Classe EC	Exemplos
EC 1: Oxidoreduases	Dehidrogenase, oxidase, monooxigenase
EC 2: Transferases	Transaminase, quinase
EC 3: Hidrolases	Lipase, amilase, peptidase
EC 4: Liases	Decarboxilase
EC 5: Isomerases	Isomerase, mutase
EC 6: Ligases	Sintase, Sintetase

^a Acrônimo para o inglês *International Union of Biochemistry and Molecular Biology*.

de classificação de enzimas serão brevemente expostos a seguir, como maneira de contextualizar a primeira parte deste trabalho de doutorado.

No sistema de classificação EC uma enzima recebe um determinado número, denominado número EC, que identifica o tipo de reação química que a enzima catalisa. A cada número EC corresponde um nome recomendado para a respectiva enzima, entretanto os números EC não especificam enzimas mas sim reações químicas catalisadas por enzimas: se duas enzimas distintas, provenientes de organismos diferentes, catalisam a mesma reação química, elas recebem o mesmo número EC. Há enzimas que catalisam mais de um tipo de reação química e, portanto, recebem mais de um número EC. Uma desvantagem do sistema de classificação EC é que ele só permite classificar enzimas para as quais uma função bioquímica tenha sido identificada. Como mostrado na Tabela 1.3, o sistema de classificação EC abrange todas as enzimas cuja atividade bioquímica já tenha sido caracterizada, não só as hidrolases. No sistema de classificação EC, as reações catalisadas por enzimas são divididas em seis grandes classes e a cada uma destas classes é associado um número de 1 a 6. Este número corresponde ao primeiro algarismo do número EC para uma enzima pertencente a uma determinada classe. Por exemplo, o primeiro algarismo do número EC para todas as hidrolases é 3. Os algarismos seguintes do número EC representam subdivisões em outras subclasses de reações catalisadas por enzimas. Por exemplo, todas as glicosídeo hidrolases (GH) têm número EC começando em 3.2, e todas as GH que catalisam a hidrólise de compostos O-glicosil ou S-glicosil[§] têm número EC iniciado em 3.2.1. Uma GH que catalise, por exemplo, a hidrólise de uma ligação glicosídica no meio de uma cadeia polimérica de β -(1-3)-D-glucano tem número EC 3.2.1.39, e seu nome recomendado segundo a sistemática EC é β -(1-3)-D-glucano glucanohidrolase (o

[§]Glicanos cujas ligações glicosídicas sejam formadas por átomos de oxigênio ou de enxofre.

nome sistemático consiste no nome do substrato seguido de um nome que especifica a classe de ação da enzima, neste caso uma hidrolase). Além do nome sistemático recomendado, o sistema EC também informa outros nomes aceitos: para enzimas com número EC 3.2.1.39 também são aceitos os nomes endo- β -(1-3)-D-glucosidase, endo- β -(1-3)-D-glucanase ou laminarinase, pois o glucano hidrolisado neste caso é genericamente denominado laminarina (vide Tabela 1.2). Enzimas GH cujo número EC é 3.2.1.6, por sua vez, catalisam uma reação de hidrólise com uma “regra de seleção” diferente: todas as vezes que em um β -D-glucano linear houver uma ligação glicosídica β -(1-3), a próxima ligação glicosídica em tal glucano poderá ser hidrolisada pela enzima se ela for do tipo β -(1-3) ou β -(1-4). Em outras palavras, a enzima exige que uma ligação β -(1-3) anteceda a ligação glicosídica a ser hidrolisada, que só será de fato hidrolisada se for do tipo β -(1-3) ou β -(1-4). Ou seja, estas enzimas só catalisam a hidrólise ao reconhecerem um determinado padrão na estrutura primária do substrato, e estes padrões de reconhecimento molecular são muito comuns em biomoléculas. Como esta regra permite às enzimas EC 3.2.1.6 hidrolisarem também laminarina, entre outros substratos, os nomes aceitos para estas enzimas incluem endo- β -(1-3),(1-4)-D-glucanase e *laminarinase*. Portanto, percebe-se que o nome genérico laminarinase é utilizado para se referir a enzimas com diferentes números EC. É comum enzimas que tenham regras ligeiramente diferentes para as reações químicas que catalisam serem capazes de atuar em um mesmo substrato, como no caso exemplificado acima.

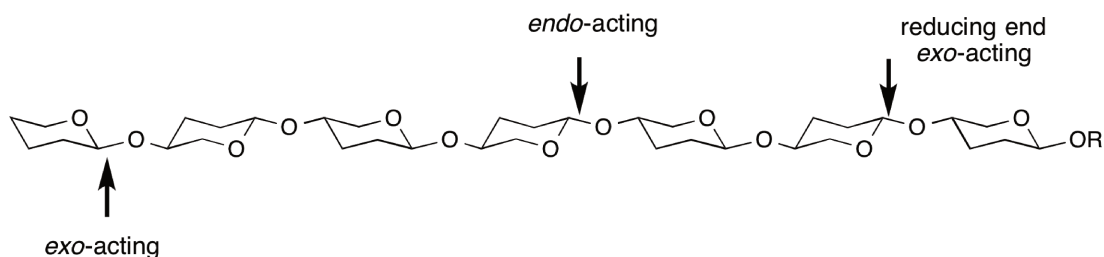
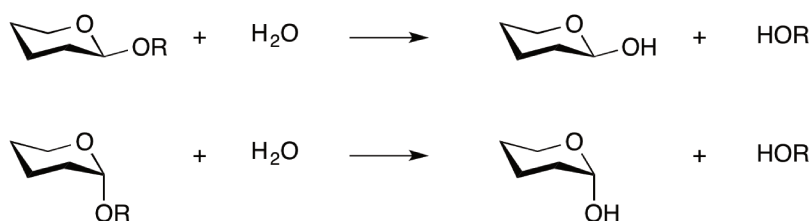


Figura 1.13: Classificação das enzimas GH segundo sua capacidade de hidrolisar glicanos a partir das extremidades. Se a enzima não é capaz de hidrolisar a partir das pontas, ela é denominada *endo*. Caso contrário ela é denominada *exo*. As GH *exo* podem ter atividade hidrolítica específica para uma das extremidades. A extremidade do glicano na qual o carbono anomérico não está envolvido em ligação glicosídica é denominada extremidade *reduzora*. No caso específico das GH que catalisam a hidrólise da celulose, existem as *endo*-glucanases e as *exo*-glucanases, estas últimas também denominadas *celobiohidrolases* (vide Figuras 1.2 e 1.15). Figura extraída de Steve Withers, Spencer Williams, “Glycoside Hydrolases” em CAZypedia <http://www.cazypedia.org/i>, acessado em dezembro de 2013, e reproduzida com permissão.

Laminarinases de três organismos diferentes foram objeto de estudo desta tese, e serão introduzidas na Seção 1.3.1. Em um outro sistema de classificação de enzimas, baseado nas seqüências de aminoácidos e que será tratado mais adiante nesta seção, as laminarinases estudadas nesta tese pertencem à família 16 das GH. Antes de falar deste outro sistema de classificação, entretanto, é útil deixar claro que as GH também são classificadas em *endo* e *exo*, de acordo com sua capacidade de hidrolisar o substrato em uma de suas extremidades, e também podem ser classificadas de acordo com seu mecanismo de reação. O critério de classificação mais simples para o mecanismo da reação de hidrólise consiste em especificar se a configuração no carbono anomérico (que é assimétrico) é invertida ou é mantida após a reação de hidrólise. Este dois tipos mais simples de classificação de GH são ilustrados nas Figuras 1.13 e 1.14.

Retaining glycoside hydrolases:



Inverting glycoside hydrolases:

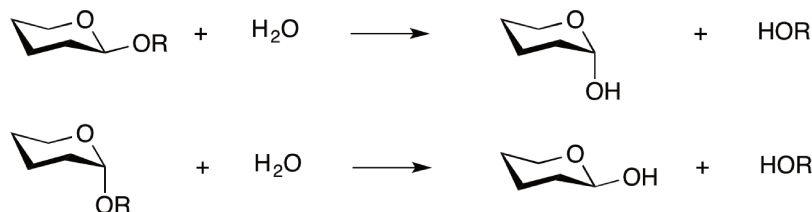


Figura 1.14: Classificação das enzimas GH segundo seu mecanismo de reação. O carbono anomérico da ligação glicosídica hidrolisada pode reter ou inverter a sua configuração estereoquímica após a reação de hidrólise. Figura extraída de Steve Withers, Spencer Williams, “Glycoside Hydrolases” em CAZypedia <http://www.cazypedia.org/i>, acessado em dezembro de 2013, e reproduzida com permissão.

Já no sistema de classificação baseado na seqüência de aminoácidos, diversos algoritmos são utilizados para analisar a seqüência primária de aminoácidos das proteínas e classificá-las em famílias. Os membros das diferentes famílias possuem grande similaridade em determinados trechos de sua seqüência primária e, por corolário, no seu padrão tridimensional de enovelamento. Este sistema de

classificação em famílias é muito útil para fazer previsões acerca dos membros de uma determinada família, uma vez que a maquinária catalítica, a geometria em torno da ligação glicosídica e os detalhes moleculares do mecanismo reacional são conservados dentro de uma mesma família, para a vasta maioria das famílias GH⁴⁰. Uma desvantagem desta sistemática é que ela só pode ser aplicada a proteínas para as quais informações acerca da seqüência primária estejam disponíveis. Por outro lado, ela permite a classificação de inúmeras proteínas para as quais não há evidência experimental da função bioquímica, por exemplo as milhares de seqüências peptídicas não caracterizadas que são originadas através dos esforços de seqüenciamento genético no mundo inteiro. Métodos de classificação baseados na seqüência primária são, portanto, bastante diferente da sistemática de classificação EC, todavia complementares em muitos aspectos. As enzimas estudadas nesta tese são membros da família 16 das GH, que abrange enzimas com diferentes números EC, entre eles EC 3.2.1.6 e EC 3.2.1.39.

As enzimas GH serão importantes para uma eventual indústria de etanol celulósico, pois elas contribuirão para viabilizar economicamente o processo de hidrólise da biomassa lignocelulósica (ilustrado nas Figuras 1.2 e 1.15) feito em escala industrial. Como já mencionado em outras seções, a escolha criteriosa destas enzimas será decisiva e, para que sejam utilizáveis industrialmente, terão que apresentar características que assegurem a reutilização destas enzimas no processo industrial pelo maior número de vezes possível, uma vez que produzir enzimas específicas em quantidade é algo muito caro. Para serem utilizáveis as GH deverão ser resistentes a variações de pH e salinidade, entre outros fatores, e deverão apresentar alta atividade em uma ampla faixa de temperaturas, de forma que enzimas termoestáveis e termofílicas serão largamente preferíveis. A busca por enzimas resistentes a condições mais drásticas geralmente leva os pesquisadores interessados em obter este tipo de enzima a realizar experimentos de evolução dirigida de proteínas em culturas de microorganismos através de técnicas de mutagênese⁴¹, ou a prospectar enzimas naturalmente resistentes expressas por determinados microorganismos. Enzimas resistentes a condições extremas são naturalmente expressas por microorganismos denominados *extremófilos*, que suportam viver em ambientes térmica ou quimicamente hostis à maioria das outras formas de vidas. Em particular o *Rhodothermus Marinus*, um destes microorganismos extremófilos, vive em temperaturas entre 54 e 77 °C, com condições ótimas de crescimento em 65 °C, e a maioria das enzimas deste microorganismo já caracterizadas tem atividade ótima entre 55 e 100 °C. O fato de um grupo de colaboradores ter conseguido

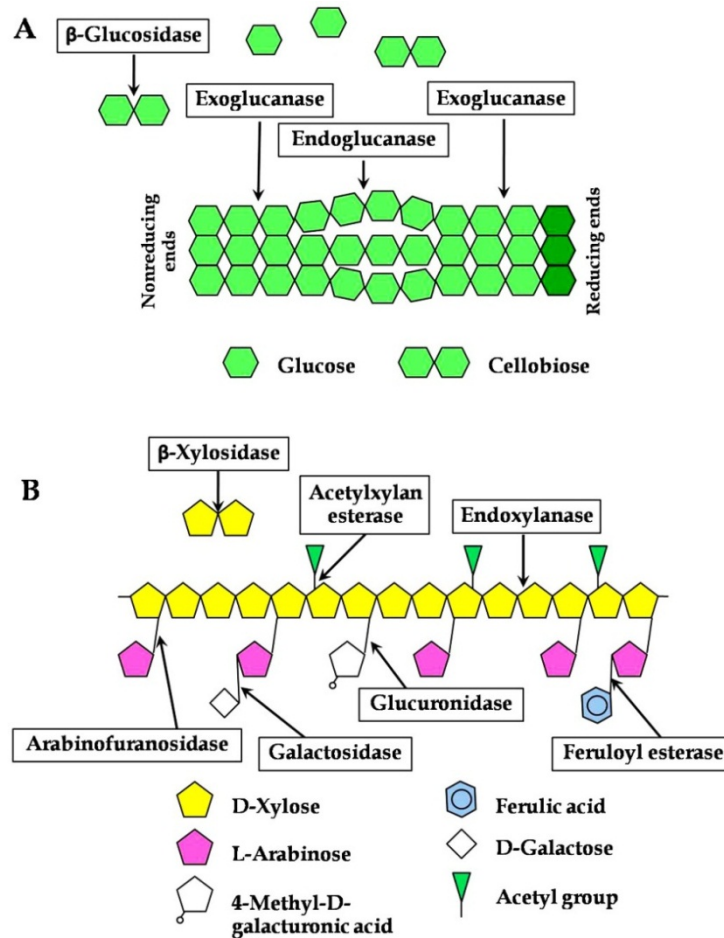


Figura 1.15: Representação esquemática das enzimas envolvidas na hidrólise da celulose (A) e hemicelulose (B). Os nomes das enzimas aparecem dentro dos retângulos e as ligações hidrolisadas por elas são indicadas por flechas. A hemicelulose é um polímero de xiloses, um xilano, que pode ter diversos tipos de ramificações, cada um exigindo um tipo diferente de enzima para ser hidrolisado. A lignina pode formar ligações éster com a hemicelulose, nas quais o feruloil da lignina liga-se a uma das hidroxilas de um resíduo lateral da hemicelulose. Figura extraída de Ratanakhanokchai *et al.*⁴², Biomass Now - Cultivation and Utilization, chapter Paenibacillus curdlandolyticus strain B-6 multienzyme complex: a novel system for biomass utilization. InTech, 2013, sob licença Creative Commons Attribution 3.0 Unported License (CC BY 3.0). Esta figura por sua vez foi adaptada de Aro, Pakula e Penttilä⁴³, Transcriptional regulation of plant cell wall degradation by filamentous fungi, FEMS Microbiology Reviews, 29(4):719–739, 2005, e portanto foi obtida permissão para reprodução também da John Wiley & Sons (copyright 2005).

expressar, purificar, cristalizar e caracterizar a estrutura tridimensional de uma laminarinase⁴⁴ expressa por esta bactéria extremófila nos permitiu e motivou a realização de investigações, mediante simulações de dinâmica molecular, sobre os

fatores que conferem termoestabilidade à laminarinase do *Rhodothermus Marinus*. Duas outras laminarinases, provenientes de organismos diferentes, também foram estudadas para permitir determinadas comparações. Estas investigações consistem na primeira parte deste trabalho de tese, e são introduzidas na seção a seguir.

1.3.1 Termoestabilidade em laminarinases

Uma enzima GH (da família 16, *laminarinase*, EC 3.2.1.6, código PDB 3ILN) expressa pela bactéria *Rhodothermus Marinus*⁴⁵, cuja estrutura cristalográfica foi determinada por colaboradores⁴⁴, foi investigada por nós mediante simulações de dinâmica molecular na primeira parte deste trabalho de doutorado, parte cuja introdução será brevemente apresentada nesta seção. Nesta parte do trabalho, procuramos entender as bases moleculares (estruturais e dinâmicas) que conferem termoestabilidade e termofilicidade a esta enzima, denominada *RmLam*. Esta enzima tem atividade ótima em pH 5.5 e temperatura de 88 °C sendo classificada como *hipertermoestável*. A *RmLam* catalisa a reação de hidrólise de um glicano denominado *laminarina* (ou *laminarano*) e, portanto, é genericamente chamada de *laminarinase*. A laminarina é um polímero de armazenamento energético encontrado em algas marrons, composto por unidades de D-glucose unidas por ligações glicosídicas $\beta(1 \rightarrow 3)$. A laminarina compõe a parte rica em nutrientes de algas que alimentam diretamente centenas de espécies de animais marinhos. As algas marrons crescem muito rapidamente, chegando a atingir 30 centímetros por dia, e sintetizam laminarina diretamente a partir da fotossíntese. Portanto, ainda que não estejamos tratando de biomassa lignocelulósica neste caso, algas também são geralmente consideradas fontes muito promissoras de biomassa renovável. Além disso, é de nosso interesse entender as características que tornam enzimas GH termoestáveis ou hipertermoestáveis.

Para investigar as bases moleculares da termoestabilidade da *RmLam*, realizamos simulações por dinâmica molecular desta enzima e de outras duas laminarinases, expressas por outros dois organismos e cuja atividade ótima se dá em temperaturas mais baixas. Estas outras duas laminarinases, denominadas *aNLam* e *PcLam*, tinham suas estruturas cristalográficas já determinadas e depositadas no *Protein Data Bank* (PDB). A enzima *aNLam* (GH da família 16, EC 3.2.1.39, PDB 2HYK) tem atividade ótima a 77 °C e a *PcLam* (GH da família 16, EC 3.2.1.6, PDB 2CL2) é mesofílica, com atividade ótima abaixo de 50 °C. As siglas adotadas para as laminarinases são compostas de duas letras que referem-se ao organismo

de origem da enzima, seguidas do sufixo *Lam*. A enzima aNLam é expressa por bactérias alcalifílicas do gênero *Nocardiopsis* (*Nocardiopsis* sp. linhagem F96), já a PcLam é expressa por fungos da espécie *Phanerochaete chrysosporium*. As simulações por dinâmica molecular destas três enzimas foram feitas a duas temperaturas, a 25 e 90 °C, para permitir comparações entre as dinâmicas das diferentes laminarinases nestas duas condições. Os resultados deste trabalho sobre laminarinases já estão publicados em artigo² do qual o aluno, autor desta tese, é co-autor, e serão expostos no Capítulo 3. Além desta primeira parte sobre laminarinases, este trabalho de tese é composto por outras duas partes, que são introduzidas nas duas seções a seguir.

1.4 Estrutura da celulose cristalina

Como vimos na Seção 1.2.2 a celulose é homopolímero não ramificado, composto de unidades repetitivas de celobiose que, por sua vez, são constituídas de dois resíduos de D-glucose unidos por uma ligação glicosídica $\beta(1 \rightarrow 4)$. As cadeias poliméricas da celulose podem adotar diferentes arranjos conformacionais e espaciais em fase sólida, dando origem portanto a vários *alomorfos* (ou *polimorfos*) [✱] da celulose em fase cristalina. Em fase sólida, portanto, a celulose pode apresentar-se na forma amorfa, semi-cristalina ou em diferentes fases cristalinas. Quatro diferentes alomorfos da celulose já foram caracterizados mediante experimentos de difração, com resolução alta o suficiente para permitir a determinação das posições atômicas no cristal.⁴⁶⁻⁴⁹ Estes quatro alomorfos são denominados celulose I α , celulose I β , celulose II e celulose III_I e estão representados nas Figuras 1.16 e 1.17. Nos vegetais superiores, além das formas amorfa e semi-cristalina, está presente a fase cristalina I β da celulose, sendo este portanto o alomorfo de maior interesse para a área do etanol celulósico.

Para lograr a obtenção de etanol a partir da celulose é necessário hidrolisar tal composto a açúcares fermentáveis. Entretanto vimos que vários são os fatores que dificultam a hidrólise da celulose, seja na sua forma *in natura* ou na forma purificada. Muitos estudos têm sido realizados para entender as causas desta recalcitrância (resistência à hidrólise), bem como para desenvolver métodos que possibilitem contornar estes problemas, e uma das metodologias que vêm

[✱]A existência de polimorfos, ou seja, de diferentes fases cristalinas para a mesma substância pura em fase sólida é bastante comum. Um dos exemplos mais comuns é a água, com os vários tipos de gelo em seu diagrama de fases. Calcita e aragonita, dois alomorfos do carbonato de cálcio, também são freqüentemente citados.

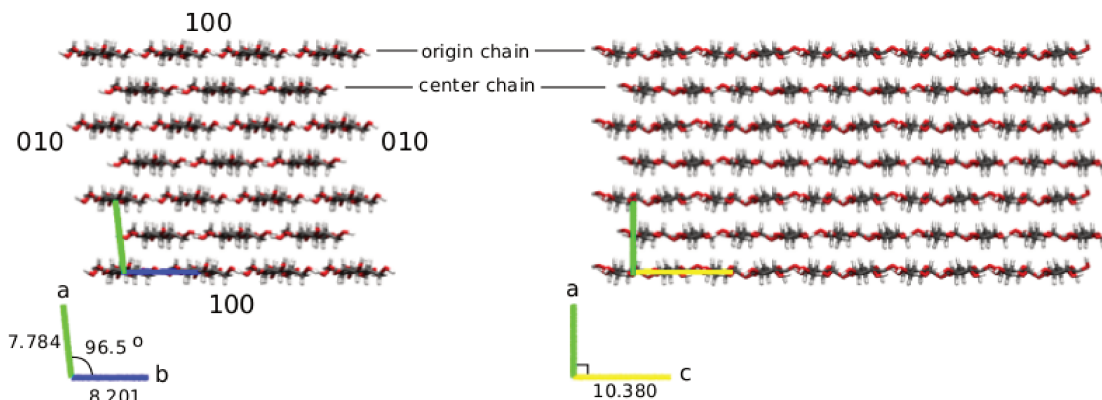


Figura 1.16: Cristalito de celulose $I\beta$ gerado utilizando o programa *cellulose-builder*, desenvolvido pelo aluno, visto a partir de suas extremidades não redutoras (esquerda), e girado 90° (direita). Faces cristalográficas estão indicadas pelos seus correspondentes índices de Miller. Camadas de cadeias de celulose *origin* e *center*⁴⁷, e eixos da cela unitária também estão indicados. As cadeias de celulose são paralelas à direção cristalográfica c . Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, *Cellulose-Builders: A toolkit for building crystalline structures of cellulose*. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

sendo utilizadas para entender as propriedades da lignocelulose é a simulação por dinâmica molecular (DM) seja da celulose cristalina *per se*^{50–52}, seja de sistemas mais complexos que envolvem proteínas^{53;54}, lignina⁵⁵ ou glicanos^{56;57} adsorvidos a superfícies de celulose cristalina. Entretanto, para realizar simulações por DM de qualquer sistema faz-se necessário obter coordenadas para os átomos que compõem o sistema. Para esta finalidade, no caso de proteínas é muito comum realizar experimentos de difração (de raios-X, de luz síncrotron, de neutrons *etc.*) que, muitas vezes, permitem a localização das posições atômicas em estruturas cristalinas com alta resolução. As estruturas de macromoléculas biológicas assim determinadas são geralmente depositadas no *Protein Data Bank* (PDB – banco de dados de proteínas) na forma de arquivos com formato específico, tornando simples a tarefa de alimentar qualquer programa de DM com as posições atômicas do sistema proteico que se deseja simular. Para a celulose, entretanto, embora a estrutura cristalina de vários alomorfos da celulose já tenha sido determinada em alta resolução^{46–49}, não estão disponíveis arquivos em formato PDB que permitam prontamente a simulação por DM de sistemas contendo este composto.

Visando facilitar a simulação de sistemas que contenham domínios de celulose cristalina mediante DM, nesta segunda parte do trabalho de tese desenvolvemos um programa capaz de produzir arquivos em formato PDB contendo coordenadas Cartesianas para os átomos de cristalitos de quatro alomorfos da

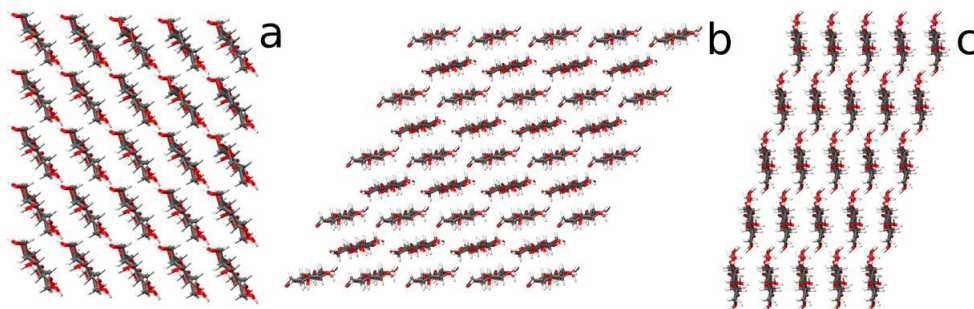


Figura 1.17: Outros alomorfos da celulose suportados pelo programa *cellulose-builder*, vistos a partir das extremidades das cadeias de celulose. a) I α b) II c) III_I. As escalas diferem. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, *Cellulose-Builder: A toolkit for building crystalline structures of cellulose*. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

celulose. As coordenadas Cartesianas dos átomos são geradas pelo programa a partir das coordenadas fracionárias relatadas por Nishiyama e colaboradores^{46–49} para a unidade assimétrica de diferentes fases cristalinas da celulose. O programa desenvolvido pelo aluno gera coordenadas para todos os átomos do cristalito, inclusive os hidrogênios, uma vez que os referidos autores^{46–49} foram capazes de determinar as posições destes átomos leves no cristal mediante difração de neutrons por fibras de celulose deuteras. Este fato é muito importante pois a correta disposição da rede de ligações de hidrogênio na celulose cristalina é determinante para a sua estrutura e propriedades. Este programa foi publicado³ com o nome de *cellulose-builder* e disponibilizado aberta e gratuitamente na rede mundial (<ftp://ftp.iqm.unicamp.br/pub/cellulose-builder/> ou <http://code.google.com/p/cellulose-builder/>). Arquivos em formato PDB, tais como os produzidos pelo programa *cellulose-builder*, são fundamentais para o desenvolvimento de qualquer trabalho na área de simulações de dinâmica molecular e a preparação manual de tais arquivos pode tornar-se extremamente laboriosa e propensa a erros dependendo do tamanho do sistema a ser estudado. Ao permitir a automatização da preparação dos arquivos PDB para cristalitos de celulose de tamanho arbitrário, este nosso trabalho³ consiste em uma importante contribuição para a pesquisa em materiais celulósicos. O programa *cellulose-builder*, suas opções e funcionalidades são discutidas na Seção 2.2 e no Capítulo 4. Nós utilizamos o programa *cellulose-builder* para construir arquivos PDB contendo as coordenadas Cartesianas iniciais das simulações por DM dos sistemas que são descritos na próxima seção.

1.5 Interações entre celulose e xilanos

Além das duas partes do trabalho já introduzidas nas duas seções anteriores, a cada uma delas correspondendo um artigo já publicado em periódico internacional^{2;3}, há ainda uma terceira parte neste trabalho de doutorado. Esta terceira parte do trabalho consiste em estudar cadeias de xilano adsorvidas nas várias faces cristalográficas da celulose I β , mediante simulações por DM. Nós utilizamos os arquivos produzidos pelo nosso programa *cellulose-builder* para viabilizar a construção dos *inputs* necessários para as simulações por DM desta terceira etapa do trabalho de tese, que trata das interações entre celulose e hemicelulose.

Como visto na Seção 1.1, a parede celular vegetal inclui as assim denominadas *hemiceluloses* em grau variado em sua composição. As hemiceluloses são polímeros responsáveis por recobrir as fibrilas elementares de celulose presentes na parede celular vegetal, protegendo-as de ataque enzimático e reforçando a matriz polimérica lignocelulósica. O nome genérico “hemiceluloses” é utilizado para se referir a diferentes tipos de glicanos, com estruturas químicas diversas. Dentre estes diversos glicanos, destacam-se o xilano e seus derivados substituídos.^{56;57} O xilano é um polímero linear (*i.e.* não ramificado) de unidades monoméricas de D-xilose. A xilose é um açúcar simples de cinco carbonos, comumente associado a dificuldades no processo de deconstrução da biomassa lignocelulósica, seja porque quanto maior o seu conteúdo de hemiceluloses maior a sua recalcitrância, como porque pentoses (açúcares de cinco carbonos) são inibidores do metabolismo de microorganismos fermentadores como *Saccharomyces cerevisiae* (levedo utilizado na fermentação alcoólica de hexoses como a glucose e a frutose). No xilano presente em parede celular vegetal, as unidades monoméricas de xilose estão na forma cíclica (xilopirranose) e ligadas umas à outras por ligações glicosídicas do tipo $\beta(1\rightarrow4)$. Em cadeias de xilano simples não há ramificações e as hidroxilas não são funcionalizadas. Há vários tipos de funcionalização possíveis para xilanos em plantas, acetilação, metilação e glicosilação sendo apenas as mais comuns. Em particular, a glicosilação envolve a substituição de hidroxilas da cadeia principal do xilano por açúcares monoméricos ou então por cadeias laterais de dois, três ou mais açúcares. Substituintes comuns em xilanos ramificados são açúcares como a arabinose (outra pentose) ou o ácido glucurônico (um derivado da glucose, oxidado no carbono 6).^{56;57} Cadeias de xilanos contendo unidades de arabinose como ramificação são denominados *arabinoxilanos*, enquanto cadeias de xilanos contendo unidades de ácido glucurônico como ramificação são denominados *glu-*

uronoxilanos e cadeias de xilanos substituídas com o grupo acetil são chamadas *acetilxilanos*.

Dado que xilanos são componentes importantes da parede celular, que interagem diretamente com as fibrilas elementares de celulose cristalina, é importante entender e descrever estas interações. Nesta terceira etapa do trabalho de doutorado foram estudados, mediante simulações por DM, xilanos simples e substituídos adsorvidos sobre diversas faces cristalográficas da celulose $I\beta$, com intuito de descrever em nível molecular as interações estabelecidas entre celulose e xilano. As configurações iniciais de algumas destas simulações são ilustradas na Figura 1.18. Como não há evidências experimentais de que existem ligações covalentes entre celulose e xilanos em plantas, é razoável supor que a interação entre estes dois componentes da parede celular ocorra por adsorção.

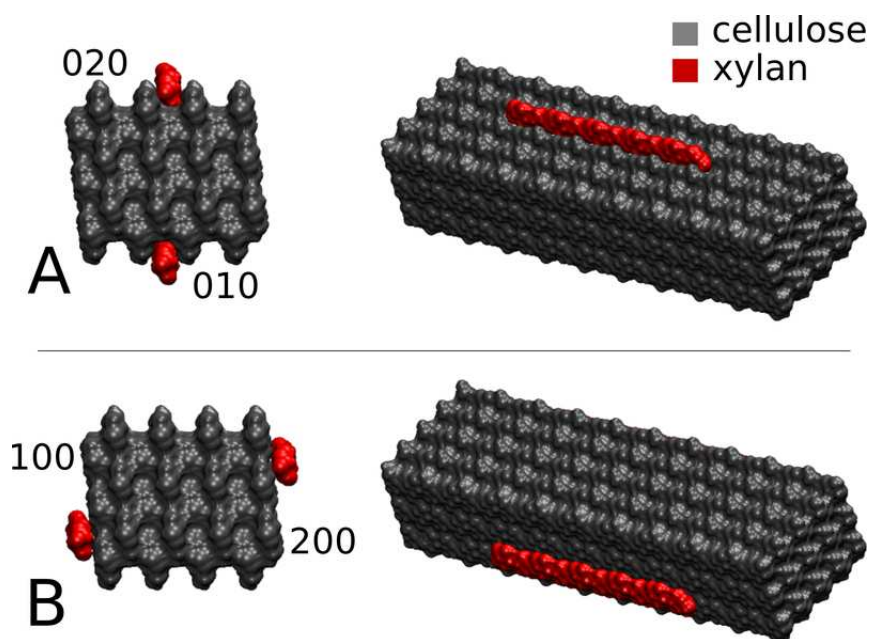


Figura 1.18: Representação da superfície de Van der Waals de cadeias de xilano (em vermelho) adsorvidas sobre diferentes superfícies da celulose cristalina $I\beta$ (em cinza). (A) Xilanos adsorvidos sobre superfícies hidrofílicas da celulose $I\beta$. (B) Xilanos adsorvidos sobre superfícies hidrofóbicas da mesma fibrila de celulose.

Ao caracterizar experimentalmente hemiceluloses provenientes de plantas, Dupree e colaboradores obtiveram evidências de que há um padrão de substituição nos xilanos destas plantas em que os grupos substituintes estão presentes em padrão par⁵⁸⁻⁶⁰, conforme ilustrado na Figura 1.19. Ou seja, considerando que em uma cadeia de xilano linear haja uma ramificação na xilose i , poderá haver outra ramificação na xilose $i + 2$, $i + 4$ e assim por diante. Em colabo-

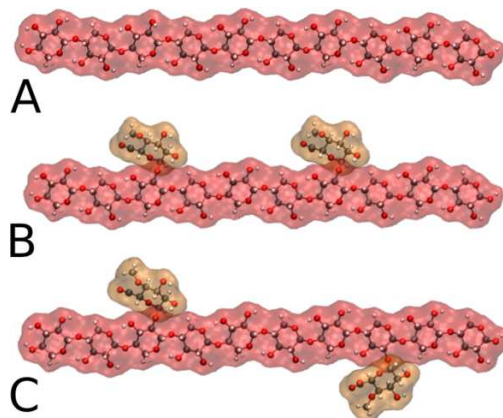


Figura 1.19: Xilano simples e xilanos ramificados. (A) Xilano. (B) Glucuronoxilano com substituintes nas posições 3 e 7, padrão par de substituição (neste caso em i e em $i + 4$, com $i = 3$). (C) Glucuronoxilano com substituintes nas posições 3 e 8 (padrão ímpar, substituição em i e em $i + 5$).

ração com Dupree, nós investigamos, mediante simulações por MD, se as cadeias de xilano com este padrão par de substituição podem formar complexos de adsorção estáveis com diferentes faces cristalográficas da celulose. A hipótese que consideramos mais plausível, em um primeiro momento, é que as cadeias de xilano possam adsorver a superfícies da celulose exatamente nas posições cristalográficas em que outra cadeia de celulose ocuparia se o cristalito fosse mais extenso, porque: i) a D-xilose é idêntica à D-glucose exceto pela inexistência do grupo exocíclico hidroximetil em moléculas de xilose (a D-xilose é igual à D-glucose sem o carbono 6, e sem a hidroxila correspondente ao C6); ii) as ligações glicosídicas no xilano são $\beta(1 \rightarrow 4)$, assim como na celulose. Logo, as posições cristalográficas que seriam ocupadas por cadeias de celulose, se o cristalito fosse maior, são sítios de adsorção perfeitos para moléculas de xilano, pois estes dois polímeros são extremamente similares estruturalmente. Esta hipótese é consistente com o padrão par de substituição encontrado experimentalmente por Dupree e colaboradores, mas não com o padrão ímpar: como ilustrado na Figura 1.20, o padrão par é consistente com a adsorção nos sítios cristalográficos, mas o padrão ímpar geraria repulsões estéricas que impediriam a formação de um complexo de adsorção com essa geometria, em algumas das faces.

Utilizando simulações por DM, nesta terceira parte do trabalho de tese investigamos, portanto, se substituintes presentes em padrão par nas cadeias de xilano adsorvidas à celulose permitem que o xilano substituído estabeleça um complexo de adsorção estável com a celulose cristalina $I\beta$ em meio aquoso. Os substituintes

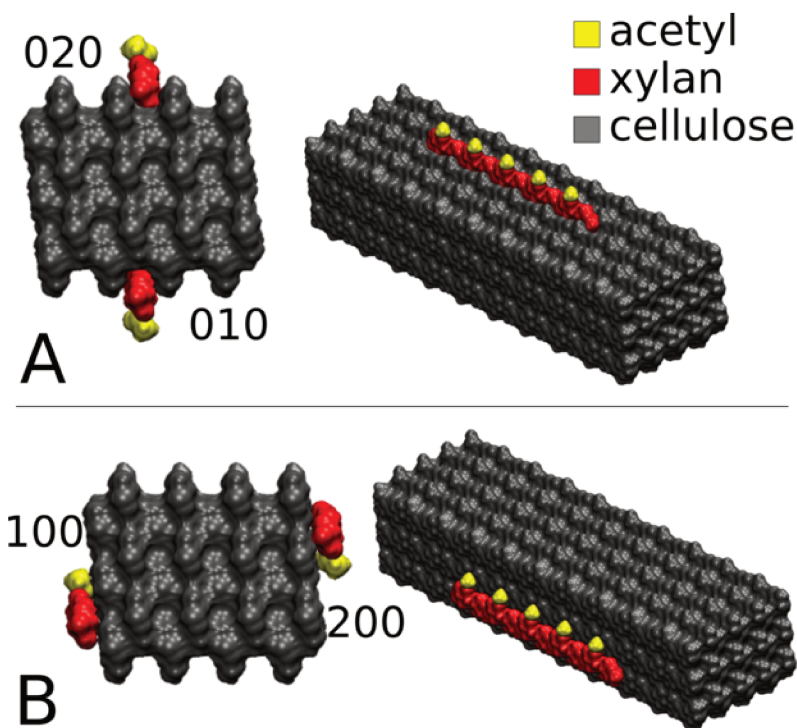


Figura 1.20: Acetilxilano com padrão de substituição par, adsorvido nas faces (A) hidrofílicas e (B) hidrofóbicas da celulose cristalina $I\beta$. As faces cristalográficas são indicadas pelos seus respectivos índices de Miller. Com os substituintes em padrão par (nas posições i , $i + 2$, $i + 4$, *etc.*), os xilanos substituídos podem ocupar os sítios cristalográficos que outra molécula de celulose ocuparia se o cristalito de celulose fosse mais extenso. Se o padrão de substituição fosse ímpar (com substituintes nas posições i , $i + 1$, $i + 3$, *etc.*), os xilanos substituídos desta maneira não poderiam ocupar os sítios cristalográficos nas faces (010) ou (020).

investigados neste trabalho foram o acetil e o ácido glucurônico (GlcA). Com os resultados desta terceira parte do trabalho, que serão expostos no Capítulo 5, estamos redigindo um artigo em colaboração com Dupree e seu grupo de pesquisas, que será submetido brevemente a publicação. Em particular, escolhemos estudar, mediante simulações por DM, grupos substituintes acetil e GlcA uma vez que Dupree já dispõe de dados experimentais para acetilxilanos e glucuronoxilanos. Esta seção encerra o capítulo introdutório desta tese. A seguir é registrada a metodologia utilizada em cada uma das três partes deste trabalho de doutorado.

Capítulo 2

Metodologia

Neste capítulo será apresentada a metodologia utilizada em cada uma das três partes deste trabalho de tese. A primeira e a terceira parte deste trabalho, que tratam respectivamente de termoestabilidade em laminarinases e de interações entre celulose e xilanos, envolveram a realização e análise de simulações por dinâmica molecular, e a metodologia empregada nestes estudos é descrita na seção a seguir. A segunda parte deste trabalho de tese, que consistiu no desenvolvimento do programa *cellulose-builder*, envolveu a utilização de alguns conceitos em teoria de grupos espaciais e a aplicação de algumas técnicas de programação, que serão apresentadas na Seção 2.2.

2.1 Simulações por dinâmica molecular

Simulações por dinâmica molecular consistem na integração numérica das equações de movimento da mecânica clássica, nas quais a energia potencial do sistema é descrita pela forma e pelos parâmetros do campo de força utilizado. Excelentes textos introdutórios sobre dinâmica molecular estão disponíveis em livros⁶¹⁻⁶³, dissertações e teses de outros alunos do grupo de pesquisa. Esta seção limita-se ao registro do protocolo de simulação utilizado para cada sistema estudado.

Para as simulações por DM de laminarinases, as caixas de simulação foram construídas utilizando suas estruturas cristalográficas. A estrutura cristalográfica da RmLam foi resolvida experimentalmente por colaboradores⁴⁴, e as estruturas das enzimas aNLam⁶⁴ e PcLam⁶⁵ já estavam depositadas no *Protein Data Bank*. O programa Packmol⁶⁶ foi utilizado para construir caixas de simulação cúbicas, com aproximadamente 80Å de lado, contendo 15000 moléculas de água, de forma

que a camada de solvente ao redor das enzimas tenha espessura de, no mínimo, 12Å. O estado de protonação dos resíduos ionizáveis (K, R, H, D, e E), em pH=7, foi determinado calculando o pK_a destes resíduos, utilizando o servidor H++⁶⁷. Para garantir a eletroneutralidade dentro da caixa de simulação, foram adicionados, respectivamente, 1, 3 e 13 íons sódio aos sistemas contendo RmLam, aNLam e PcLam. Além destes íons, foram adicionados outros íons sódio e cloreto, em igual quantidade, de maneira a obter uma concentração próxima a 0.16 mol·L⁻¹. Nas três proteínas também há um íon cálcio, característico das GH da família 16, coordenado à estrutura de cada enzima. Para cada sistema simulado, foram realizadas minimizações de energia mediante 500 iterações do método do gradiente conjugado⁶⁸, conforme implementado no programa NAMD⁶⁹. As minimizações de energia foram seguidas de etapas de equilíbrio, a 298 K, e com tempo total de 1 ns, feitas inicialmente fixando o *backbone* das proteínas mas permitindo o movimento das cadeias laterais, e em seguida deixando o sistema inteiro livre para mover-se. Para as simulações em temperatura mais alta (363 K), foram realizadas também mais duas etapas de equilíbrio, uma durante 500 ps em 330.5 K, e outra por 1 ns em 363 K. A partir dos sistemas equilibrados, foram realizadas: 10 simulações independentes, de 12 ns cada, para a RmLam; 8 simulações, de 12 ns cada, para a aNLam; 8 simulações, de 9 ns cada, para a PcLam; todas a 25°C. Entretanto, a 90°C foram realizadas cinco simulações independentes de 9 ns para cada enzima. As simulações foram produzidas utilizando o programa NAMD⁶⁹, aplicando condições periódicas de contorno. Para tratar interações eletrostáticas de longo alcance, foi utilizado o método Particle Mesh Ewald (PME).⁷⁰ O campo de força CHARMM⁷¹⁻⁷³ foi utilizado em todas as simulações (versão *all27* para proteínas e *all36* para carboidratos), e as moléculas de água foram representadas pelo modelo explícito TIP3P.⁷⁴ Os parâmetros utilizados para os íons são aqueles desenvolvidos por Beglov e Roux.⁷⁵ Os sistemas foram simulados em *ensembles* NPT (número de partículas, pressão e temperatura constantes), utilizando o termostato de Langevin e o barostato de Nosé-Hoover^{76;77}, com pressão alvo em 1 bar. Para a integração numérica das equações de movimento foi utilizado o algoritmo *velocity-Verlet*, com passo de integração numérica de 2 fs. Todas as ligações covalentes entre átomos de hidrogênio e átomos mais pesados foram mantidas rígidas, utilizando o algoritmo SHAKE.⁷⁸ Para computar as interações de Van der Waals, descritas pelo modelo de Lennard-Jones, foi aplicado um *cutoff* em 12Å, mediante a utilização de *switching functions* a partir de 10Å. O critério utilizado para contabilizar pontes salinas foi o seguinte: se dois resíduos de carga oposta

têm átomos de sua cadeia lateral a menos de 4 Å de distância, a ponte salina é contabilizada naquele *frame* da simulação.

As simulações por DM de xilanos adsorvidos à celulose foram realizadas utilizando um protocolo bastante semelhante ao utilizado para laminarinases, com algumas diferenças, elencadas a seguir. As simulações envolvendo xilanos foram realizadas apenas a 25°C, mas para cada sistema foram produzidas duas simulações independentes de 50 ns cada. Para poupar esforços computacionais, cada sistema contém duas cadeias de xilano, adsorvidos em faces opostas da celulose I β . Há cerca de 13000 moléculas de água em cada sistema. Os sistemas exigiram até 4000 iterações do método de minimização de energia por gradiente conjugado. Para cada sistema simulado, uma distribuição de velocidades consistente com uma temperatura de 30 K foi atribuída. Em seguida, os sistemas foram aquecidos até 298 K a uma taxa de 10 K a cada 4 ps. A seguir, foi realizada a equilibrção em *ensemble* NPT, com os átomos das cadeias de xilano inicialmente fixos (exceto substituintes, onde houver) durante 1 ns, seguido de 1 ns com todos os átomos do sistema livres para se mover. Em seguida, foi calculado o volume médio do sistema durante o último nanosegundo da equilibrção e a etapa de produção foi então realizada em *ensemble* NVT.

Para caracterizar as interações entre celulose e xilano, foi realizada uma contagem detalhada das ligações de hidrogênio estabelecidas entre estes dois componentes. Os critérios geométricos utilizados para definir ligações de hidrogênio nestes sistemas são apresentados na Figura 2.1. Este critério é um pouco mais

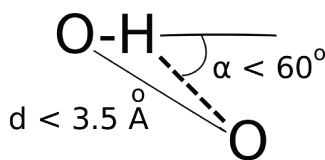


Figura 2.1: Critérios geométricos utilizados para contabilizar ligações de hidrogênio nas simulações por DM de xilanos adsorvidos à celulose. O ângulo α deve ser igual ou menor que 60° e a distância d deve ser menor ou igual a 3.5\AA .

flexível em relação ao que é geralmente utilizado, mas já foi aplicado a estudos, incluindo simulações por DM, de ligações de hidrogênio em celulose.⁷⁹

As análises das simulações produzidas foram realizadas utilizando programas publicamente disponíveis, como o VMD⁸⁰, e também programas e scripts *ad-hoc*, internos ao nosso laboratório, produzidos pelo autor desta tese ou por seus colegas. Todas as análises de trajetória foram realizadas utilizando todas as simulações

independentes produzidas para cada sistema.

2.1.1 Campos de força para simulações de glicanos

Ao realizar extensivas simulações de cristalitos de celulose I β , com até 800 ns de duração, comparando três diferentes campos de força para carboidratos (CHARMM 35, Glycam 06 e Gromos 45a4) Matthews *et al.*⁵¹ observaram que:

- Cada um dos campos de força testados diverge da estrutura cristalográfica experimental em diferentes graus, sob as condições testadas;
- Após longo tempo de simulação, os campos CHARMM 35 e Glycam 06 resultam em estruturas similares àquelas obtidas a 500 K com os mesmos campos de força, e que são consistentes com o comportamento experimental da celulose I a altas temperaturas;
- O campo de força Gromos 45a4 produz um comportamento significativamente diferente em relação ao experimental, aos outros dois campos de força e a outras simulações mais curtas, reportadas anteriormente, com o mesmo campo de força e utilizando restrições das cadeias de celulose mediante condições periódicas de contorno;
- Para o campo de força Glycam 06, a conformação inicial escolhida para a rede de ligações de hidrogênio, e a escolha dos *scaling factors* para o cálculo de interações eletrostáticas, afetam significativamente o taxa de divergência estrutural.

Portanto, diante dos resultados apresentados por Matthews *et al.*⁵¹, consideramos que atualmente o campo de força CHARMM seja o mais adequado para simulações por DM dos sistemas investigados nesta tese.

2.2 Cellulose-builder

Como foi visto na Seção 1.4, na segunda parte deste trabalho de tese foi desenvolvido um programa, denominado *cellulose-builder*, para automatizar a construção de arquivos em formato PDB contendo coordenadas Cartesianas para os átomos de cristalitos de celulose. Nesta seção será apresentada uma breve descrição do procedimento adotado pelo programa *cellulose-builder* para produzir tais

arquivos em formato PDB, que são indispensáveis para a realização de simulações por DM.

Para assegurar que o programa construa arquivos PDB representando cristallitos de celulose de acordo com as determinações experimentais mais recentes e acuradas da estrutura cristalográfica da celulose cristalina, foram utilizados os dados publicados por Nishiyama *et al.*⁴⁶⁻⁴⁹ para as unidades assimétricas dos alomorfos I α , I β , II e III $_I$ da celulose cristalina.

A unidade assimétrica de um cristal é a menor unidade na qual um sólido cristalino pode ser dividido utilizando as operações de simetria do grupo espacial ao qual pertence a estrutura do cristal estudado.⁸¹⁻⁸⁴ Em geral é a unidade assimétrica que é resolvida por cristalógrafos ao analisar a estrutura de substâncias cristalinas mediante experimentos de difração. Nishiyama *et al.* determinaram experimentalmente a unidade assimétrica dos alomorfos I α , I β , II e III $_I$ da celulose cristalina mediante experimentos de difração de raios-X de fonte síncrotron, e também mediante experimentos de difração de neutrons em cristais de celulose deuterada.⁴⁶⁻⁴⁹ Os experimentos realizados por estes pesquisadores permitiram a localização das posições atômicas dentro dos cristais com alta resolução, permitindo inclusive a determinação das posições dos átomos de hidrogênio, exceto para a celulose II (para a qual estes autores realizaram apenas experimentos de difração de raios-X).⁴⁸ A determinação das posições atômicas de átomos leves, como o hidrogênio, é bastante difícil de obter mediante experimentos de difração de raios-X, e para a maioria das estruturas cristalográficas de proteínas determinadas desta forma e depositadas no *Protein Data Bank* (PDB) não há dados sobre as posições dos átomos de hidrogênio. A falta destes dados no caso de proteínas não é grave pois, antes de realizar simulações por DM destas proteínas, pode-se estimar as posições dos hidrogênios faltantes utilizando as outras coordenadas internas disponíveis. No caso da celulose cristalina, entretanto, apenas estimar as posições dos átomos de hidrogênio no cristal não é suficiente. A determinação experimental das posições atômicas dos hidrogênios é muito importante, pois a correta disposição da rede de ligações de hidrogênio na celulose cristalina é determinante para a sua estrutura e propriedades. A determinação experimental das posições atômicas dos átomos de hidrogênio em cristais de celulose, em alta resolução, foi possível a partir dos experimentos de difração de neutrons por celulose deuterada realizados por estes autores. Com isso, o programa *cellulose-builder* não precisa estimar posições para os hidrogênios, e as calcula diretamente a partir das coordenadas fracionárias relatadas por Nishiyama *et al.*, de forma idêntica ao que

é feito por este programa para os átomos mais pesados. Os átomos de hidrogênio para os quais estes autores relataram mais de uma posição possível no cristal são colocados pelo programa *cellulose-builder* em sua posição de maior ocupação.

As posições dos átomos que constituem as unidades assimétricas de cristais resolvidas por cristalógrafos são geralmente reportadas em coordenadas fracionárias. Para o caso de cristais, este sistema de coordenadas apresenta diversas vantagens em relação ao sistema de coordenadas Cartesianas, e uma dessas vantagens é a maior facilidade para aplicar matematicamente operações de simetria de grupo espacial em sistemas descritos no sistema de coordenadas fracionárias.⁸¹⁻⁸⁴ Neste sistema de coordenadas, os vetores de base para o espaço tridimensional \mathbb{R}^3 são as próprias arestas da cela unitária do cristal. Este conjunto de base, portanto, não é ortonormal para a maioria dos retículos de Bravais. A posição de um átomo dentro da cela unitária é descrita por três coordenadas (x, y, z) , denominadas fracionárias porque cada coordenada corresponde a uma fração do comprimento da cela unitária em uma das três direções: um átomo cujas coordenadas fracionárias sejam $(0.5, 0.5, 0.5)$, por exemplo, está localizado exatamente no centro da cela unitária, pois estas coordenadas especificam que a posição do átomo é atingida percorrendo, a partir da origem do sistema de coordenadas, metade do comprimento de cada aresta ao longo de suas respectivas direções. Por outro lado, um átomo cujas coordenadas fracionárias sejam $(0.5, 0, 0)$ está localizado sobre uma aresta da cela unitária, exatamente na metade desta aresta; um átomo com coordenadas fracionárias $(0.4, 0.4, 0)$ está localizado sobre uma das faces da cela unitária, e um átomo cujas coordenadas fracionárias sejam $(1, 1, 1)$ está localizado no vértice da cela unitária mais distante da origem do sistema de coordenadas.

Dado que em trabalhos de cristalografia é comum reportar apenas a unidade assimétrica do cristal cuja estrutura foi determinada, o programa *cellulose-builder* precisa calcular as posições simetricamente equivalentes dos outros átomos da cela unitária: a unidade assimétrica em geral abrange apenas alguns átomos da cela unitária do cristal, cujas posições são relacionadas aos outros átomos dentro da mesma cela unitária por operações de simetria. Portanto, o programa precisa aplicar operações de simetria às coordenadas fracionárias dos átomos na unidade assimétrica para calcular as coordenadas fracionárias dos demais átomos que compõem a cela unitária. Somente após conhecer as coordenadas fracionárias de todos os átomos na cela unitária é possível replicá-la para construir um cristal de tamanho arbitrário. Portanto, a primeira etapa executada pelo *cellulose-builder* é a aplicação de operações de simetria às coordenadas fracionárias determinadas

experimentalmente. Os alomorfos da celulose I β , II e III_I possuem retículos monoclinicos e suas estruturas cristalográficas pertencem ao grupo espacial P2₁, enquanto o alomorfo I α possui retículo triclinico e sua estrutura cristalográfica pertence ao grupo espacial P1. No alomorfo I α , a unidade assimétrica coincide com a cela unitária e não é necessária a aplicação de operações de simetria. Para os alomorfos cuja estrutura pertence ao grupo espacial P2₁, entretanto, a aplicação de um eixo helicoidal 2₁ às coordenadas fracionárias dos átomos da unidade assimétrica origina as coordenadas fracionárias dos demais átomos na cela unitária. Esta etapa é ilustrada na Figura 2.2. Trabalhando em coordenadas fracionárias, a aplicação do eixo helicoidal 2₁ simplifica-se às equações 2.1:

$$\begin{aligned}x' &= -x \\y' &= -y \\z' &= z + 1/2\end{aligned}\tag{2.1}$$

nas quais (x, y, z) são as coordenadas fracionárias de um dado átomo presente na unidade assimétrica, e (x', y', z') são as coordenadas fracionárias simetricamente equivalentes de um outro átomo (do mesmo elemento químico) na cela unitária de um cristal cuja estrutura pertence ao grupo espacial P2₁. O eixo helicoidal 2₁ corresponde a uma rotação de 180° em torno do eixo cristalográfico z seguida de uma translação, ao longo do mesmo eixo, cuja magnitude é metade do comprimento da cela unitária na direção z . * O eixo cristalográfico z é paralelo às cadeias de celulose no cristal, e é denominado *eixo principal*.

De posse das coordenadas fracionárias de todos os átomos que compõem a cela unitária, o programa pode partir para a replicação da cela unitária, como ilustrado na Figura 2.3, aproveitando a facilidade de realizar esta replicação no sistema de coordenadas fracionárias: para replicar a cela unitária uma vez na direção x , basta adicionar 1 à coordenada fracionária x de todos os átomos na cela unitária, de maneira que se na cela unitária original as coordenadas fracionárias são da forma (x, y, z) , na réplica as coordenadas fracionárias terão a forma $(x + 1, y, z)$. Analogamente, adicionando 2 à coordenada fracionária x de todos os átomos na cela unitária original, obtém-se uma segunda réplica na direção x na qual as coor-

*As celas unitárias em cristais são descritas pelos seus *parâmetros de cela*, que especificam o comprimento de suas arestas (a, b, c) e os ângulos entre elas (α, β, γ) . Metade do comprimento da cela unitária na direção z corresponde a $c/2$, mas no sistema de coordenadas fracionárias o comprimento c é o comprimento unitário na direção z , pela definição dos vetores de base, e portanto, nesse sistema de coordenadas, na direção z vale a igualdade $c = 1$.

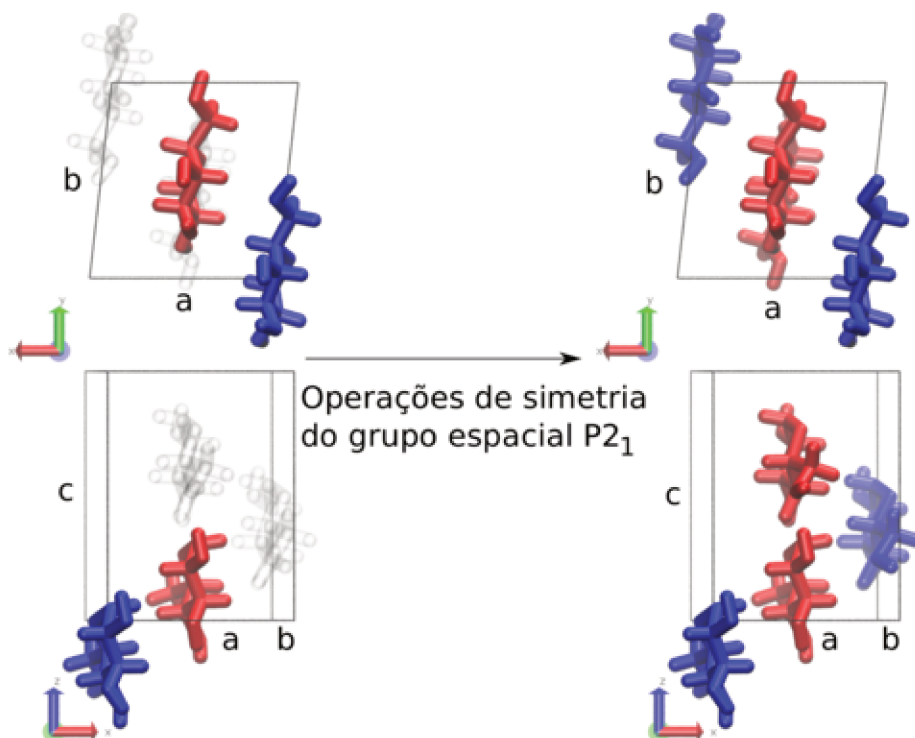


Figura 2.2: Representação da primeira etapa executada pelo programa *cellulose-builder*, que consiste na aplicação de uma operação de simetria às coordenadas fracionárias dos átomos da unidade assimétrica para gerar as coordenadas fracionárias dos átomos simetricamente equivalentes na cela unitária. O alomorfo $I\beta$ da celulose foi empregado nesta figura para exemplificar a aplicação de um eixo helicoidal 2_1 às coordenadas fracionárias dos átomos que compõem a unidade assimétrica da celulose $I\beta$. À esquerda são representados os anéis de anidroglicopiranosose que compõem a unidade assimétrica, denominados *origin* (em azul) e *center* (em vermelho). O anel *origin* está localizado próximo à origem da cela unitária, enquanto o anel *center* está no centro da cela unitária, daí suas denominações. As arestas da cela unitária são representadas como linhas pretas e são denominadas *a*, *b* e *c*. As arestas *a*, *b* e *c* são os vetores de base do sistema de coordenadas fracionárias, e definem as direções cristalográficas *x*, *y* e *z*. As sombras que aparecem à esquerda representam as posições simetricamente equivalentes na cela unitária: a aplicação do eixo helicoidal 2_1 às coordenadas fracionárias dos átomos da unidade assimétrica origina um novo conjunto de coordenadas (à direita), ditas *simetricamente equivalentes*, que correspondem às coordenadas dos outros átomos que compõem a cela unitária. As imagens mostradas acima e abaixo são equivalentes, apenas representadas a partir de diferentes pontos de vista.

denadas fracionárias terão a forma $(x + 2, y, z)$, e assim por diante. Obviamente, o mesmo vale para as coordenadas *y* e *z*. De maneira geral, portanto, se ξ for uma das três coordenadas (x, y, z) basta adicionar um número inteiro *n* à coordenada fracionária ξ de todos os átomos na cela unitária original para gerar uma réplica da cela unitária deslocada *n* vezes na direção ξ . Logo, para uma ré-

plica qualquer da cela unitária original as coordenadas fracionárias terão a forma $(x + l, y + m, z + n)$, onde l, m, n são números inteiros. Portanto, a replicação da cela unitária original para gerar um cristal de tamanho arbitrário é facilmente executada pelo programa no sistema de coordenadas fracionárias, utilizando três laços de repetição aninhados.

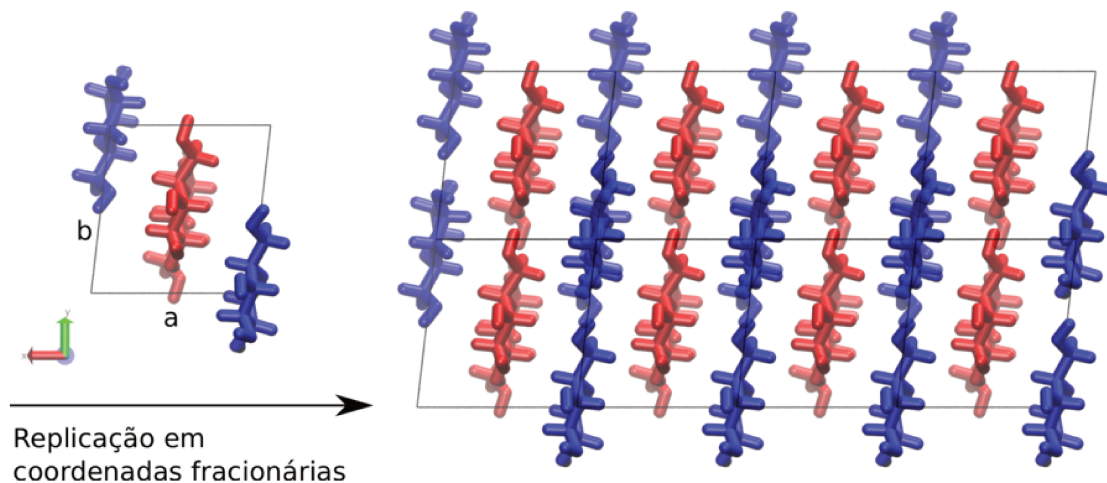


Figura 2.3: Representação da etapa de replicação executada pelo programa *cellulose-builder* após ter construído a cela unitária a partir de sua unidade assimétrica, como foi exemplificado na Figura 2.2. A utilização de coordenadas fracionárias facilita enormemente a implementação computacional de ambas as etapas. Após a replicação, os anéis de anidroglicopiranosose *origin* da cela unitária, em azul, compõem cadeias de celulose *origin* no cristal de celulose I β , e os anéis *center*, em vermelho, formam cadeias de celulose *center* neste alomorfo. Os sentido de crescimento das cadeias de celulose é perpendicular ao plano da página, paralelo à direção cristalográfica z .

Após ter replicado a cela unitária original um número arbitrário de vezes, o *cellulose-builder* terá produzido coordenadas fracionárias para todos os átomos de um cristalito de celulose. Resta agora transformar as coordenadas fracionárias de cada átomo deste cristalito em suas correspondentes coordenadas Cartesianas, mediante a aplicação do operador Jacobiano apropriado para esta transformação de coordenadas. Sejam x, y e z as coordenadas fracionárias de um átomo qualquer, e X, Y e Z as suas coordenadas Cartesianas. A relação entre estes dois conjuntos

de coordenadas é dada, de maneira completamente geral, por:

$$\begin{aligned}
 X &= ax + by \cos \gamma + cz \cos \beta \\
 Y &= by \sin \gamma + \frac{cz(\cos \alpha - \cos \beta \cdot \cos \gamma)}{\sin \gamma} \\
 Z &= \frac{zV}{ab \sin \gamma} \\
 V &= (1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2 \cos \alpha \cdot \cos \beta \cdot \cos \gamma)^{1/2} abc \quad (2.2)
 \end{aligned}$$

onde os parâmetros de cela a , b e c denotam o comprimento das arestas da cela unitária e α , β e γ denotam os ângulos entre as arestas da cela unitária. As equações 2.2 foram implementadas para o alomorfo I α , cuja estrutura cristalográfica pertence ao grupo espacial P1. Para os outros três alomorfos, cuja estrutura cristalográfica pertence ao grupo espacial P2₁, os ângulos α e β são ângulos retos, e as equações 2.2 podem ser simplificadas para:

$$\begin{aligned}
 X &= ax + by \cos \gamma \\
 Y &= by \sin \gamma \\
 Z &= cz \\
 V &= abc \sin \gamma \quad (2.3)
 \end{aligned}$$

portanto as equações 2.3 foram implementadas para os alomorfos I β , II e III_I.

Após ter aplicado as transformações no sistema de coordenadas e ter calculado as coordenadas Cartesianas para todos os átomos do cristalito, o programa faz várias outras operações necessárias para colocar os átomos em uma ordem bastante específica, que será discutida no Capítulo 4, e escrever todo o resultado em formato PDB. Além das coordenadas Cartesianas dos átomos do cristalito em formato PDB, um arquivo de conectividade no formato PSF também é produzido, no qual constam as listas de ligações químicas, ângulos e diedros presentes no sistema. A Figura 2.4 apresenta um resumo esquemático da linha de execução do programa *cellulose-builder*. Para realizar todas as operações necessárias para produzir estes dois arquivos (PDB e PSF), o *cellulose-builder* utiliza uma série de ferramentas, entre elas *Octave*, *sed*, *grep*, *Perl*, *VMD*⁸⁰ e *psfgen*⁶⁹. Antes de partir para a sua linha de execução principal, o *cellulose-builder* testa pela presença destas ferramentas no sistema operacional e avisa ao usuário se alguma delas precisa ser instalada. A maioria destas ferramentas aceita arquivos contendo um conjunto de instruções, que podem ser passados como argumento na linha de comando

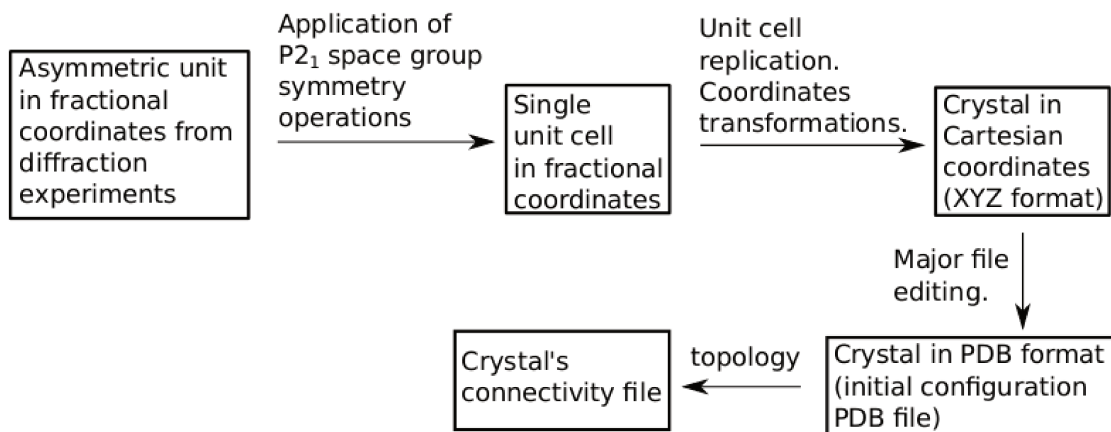


Figura 2.4: Fluxograma que representa simplificadaamente o fluxo de trabalho do programa *cellulose-builder*. Dadas as coordenadas fracionárias e grupos espaciais determinados experimentalmente⁴⁶⁻⁴⁹ ($P2_1$ para celulose I β , II e III $_T$), o programa determina as posições simetricamente equivalentes para todos os átomos na cela unitária, e então replica a cela unitária de acordo com os requisitos fornecidos pelo usuário, explorando a conveniência de trabalhar em coordenadas fracionárias para realizar a replicação. Após a replicação, as coordenadas fracionárias são transformadas em coordenadas Cartesianas usando os parâmetros de cela experimentais, o que resulta em um arquivo em formato XYZ. Várias etapas de edição de arquivos são então realizadas, até chegar a um arquivo PDB adequado para programas de simulação por DM. O programa também produz um script com instruções para a ferramenta *psfgen* e o executa, produzindo um arquivo em formato PSF que contém informações sobre a conectividade no cristalito de celulose, representado pelo arquivo PDB. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, Cellulose-Builder: A toolkit for building crystalline structures of cellulose. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

ou redirecionados para a entrada padrão (*stdin*)⁸⁵⁻⁸⁷, portanto o *cellulose-builder* interage com estas ferramentas produzindo *scripts* secundários *ad-hoc*, contendo instruções escritas na sintaxe apropriada para cada ferramenta. Logo, enquanto cada ferramenta realiza parte do trabalho total, o *cellulose-builder* encomenda que estas tarefas sejam feitas de uma maneira específica, coleta os resultados, manipula arquivos e variáveis, e então encomenda as próximas tarefas. Cada uma destas ferramentas foi escolhida pela sua conveniência e facilidade de codificação do algoritmo encomendado: o *Octave* tem uma sintaxe muito apropriada para a manipulação de dados em vetores e matrizes, o *Perl* permite facilmente a manipulação de expressões regulares^{88;89} complicadas e torna fácil a manipulação dos arquivos em formato PDB, o *VMD* convenientemente reconhece cadeias individuais de celulose, e o *psfgen*, se for alimentado com um arquivo PDB bem organizado, permite a geração automática do arquivo de conectividade em formato PSF. To-

das as ferramentas nas quais o *cellulose-builder* se baseia são já bem estabelecidas e gratuitamente disponíveis.

O *cellulose-builder* foi escrito em linguagem de programação Bash, interpretada nativamente por computadores com sistema operacional GNU/Linux (sistema operacional de computadores amplamente utilizado nos meios científicos) e portanto o programa não precisa ser compilado. O código do programa é repleto de testes e dispositivos lógicos projetados para que o programa siga estritamente uma linha de execução válida ou aborte sua execução informando ao usuário os erros encontrados em cada caso específico. Os usuários são orientados a reportar os eventuais erros ao autor para permitir o refinamento do código. Antes da publicação do código, diversos testes foram realizados para garantir que toda a lógica interna do programa esteja correta e que eventuais erros (do usuário ou de seu sistema, ocasionais ou intencionais) não levem o programa a entregar resultados enganosos ou que possam incorrer em erros nas etapas de trabalho posteriores que utilizem informações geradas a partir do programa. O programa *cellulose-builder* está aberta e gratuitamente disponível sob a licença GNU *General Public Licence, version 3*. As funcionalidades do *cellulose-builder* serão discutidas no Capítulo 4. Nos três capítulos a seguir serão apresentados e discutidos os resultados das três partes que compõem este trabalho de tese.

Capítulo 3

Termoestabilidade em laminarinases

Neste capítulo serão apresentados e discutidos os resultados referentes à primeira parte deste trabalho de tese, que consiste no estudo, mediante simulações por DM, de três enzimas da família 16 das glicosídeo hidrolases, genericamente denominadas *laminarinases*. Os resultados desta parte do trabalho foram publicados² em 2011, em um artigo em co-autoria com outros alunos do nosso grupo de pesquisas e pesquisadores externos à Unicamp. As três enzimas estudadas neste trabalho são:

- uma laminarinase de *Rhodothermus marinus* (GH da família 16, EC 3.2.1.6, código PDB 3ILN), hipertermofílica, com atividade ótima a 88 °C, cuja sigla adotada neste texto é RmLam,
- uma laminarinase de *Nocardiopsis sp.* cepa F96 (GH da família 16, EC 3.2.1.39, código PDB 2HYK), termofílica, com atividade ótima a 77 °C, cuja sigla é aNLam,
- uma laminarinase de *Phanerochaete chrysosporium* (GH da família 16, EC 3.2.1.6, código PDB 2CL2), mesofílica, com atividade ótima abaixo de 50 °C), sigla PcLam.

As três enzimas estudadas têm atividade catalítica máxima em temperaturas bastante diferentes entre elas, e acima da temperatura de desnaturação de muitas proteínas. Apesar das seqüências primárias destas três enzimas não serem muito similares, como ilustra a Figura 3.1, a estrutura terciária nestas três enzimas é

RmLamR	7	--RLPHWELVWSDEFD--YNGLPDPAKWDYDVGGHGWGNQELQYYTRARI
aNLam	1	TESDMRATLVWSDEFDGPAGSAPDPANWNHETGDHGWGNNELQNYTDSRA
PcLam	1	-----ATYHLEDNW---VGS AFLSTFTHEA IADPTHGRVNVVDQATALA
RmLamR	53	ENARVGGGVLIIEARRESYEGREYTSARLVTRGKASWTYGRFEIRARL--
aNLam	51	NSALDGNGNLVITARQEADGG--YTSARLTTQNKVQPQYGRVEASIQI--
PcLam	42	KNLTYASGDTLILRADHTTTLSPSGPGRNSVRIRSIKTYTTHVAVFDVRH
RmLamR	101	-PSGRGTWPAIWMLPDRQTYGSAYWPDNGEIDIMEHVGFNPDVVHGTVHT
aNLam	97	-PRQGQIWPAFWML--GADFPNTPWPDSGEIDIMENIGREPHLVHGSLSHG
PcLam	92	MPQGCCTWPAAW-----ETDEGDWPNNGEVDIIEGVNDQSPNAMTLHTG
RmLamR	150	KAYNHLLGTQRGG-----SIRVPTA---RTDFHVYAIE
aNLam	144	PGY--FGGEPLTG-----SYMHPQGWSFADTFHTFAVD
PcLam	136	ANCAMPASRTMTGHATNNNCDVNTDGNTGCGVQAPTANSYGPSFNANGGG
RmLamR	180	WTPEEIR-----WFVDDSLYYRFPNERLTNPEADWRHWPFDQPFH----
aNLam	175	WRPGSIT-----WSVDGVAYQTY-----TSADTRGNPWVFDQPFH----
PcLam	186	WYAMERTNSFIKWFFPRNAGNVPNDIASGPATINTDNWGTPTAFFPNTN
RmLamR	220	-----LIMNIAVGGTGGQQGVDPFAFPAQLVVDYVVRVYRWVE-
aNLam	210	-----MILNVAVGGDWPYDPGSTQ-FPQEMRVDYVRYELG--
PcLam	236	CDIGSHFDANNIIINLTFCGDWAGQASIFNGAGCPGSCVDYVNNNPSAFA
RmLamR		-----
aNLam		-----
PcLam	286	NAYWDIASVRVYQ

Figura 3.1: Alinhamento, baseado em estrutura terciária, das seqüências primárias das enzimas RmLam, aNLam e PcLam. Em negrito: a tríade catalítica (Glu, Asp, Glu) é conservada. A identidade entre as seqüências primárias de RmLam e aNLam, e entre RmLam e PcLam é de 44% e de 13%, respectivamente. Figura extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

bastante semelhante, e os resíduos catalíticos, situados na fita- β número 8 (denominada S8, do inglês *strand*, para fita), são preservados nas três enzimas. A estrutura terciária das enzimas GH da família 16 é denominada β -*jelly roll*, ou rocambole- β , na qual as fitas- β presentes são dispostas de maneira a formar duas folhas- β que são quase paralelas mas que, ao invés de estarem contidas em planos paralelos, formam superfícies curvas com seção transversal semelhante a uma letra C. A Figura 3.2 ilustra que, nas três enzimas estudadas, o padrão de enovelamento formado pelas fitas- β é essencialmente idêntico. As alças, que são segmentos sem estrutura secundária, são bastante semelhantes entre RmLam e aNLam, mas substancialmente diferentes na PcLam, que contém várias α -hélices em regiões periféricas. A PcLam é a enzima mesofílica da família 16 cuja estrutura terciária é mais próxima da RmLam, embora seja necessário cautela ao comparar o com-

portamento destas duas proteínas, devido à baixa identidade entre as seqüências primárias. Nestas três enzimas, o padrão de enovelamento formado pelas fitas- β é altamente conservado, não por acaso, uma vez que as diversas fitas- β for-

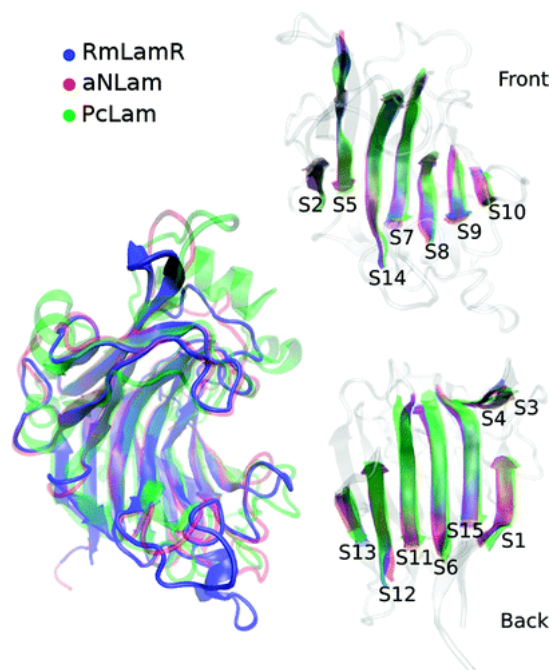


Figura 3.2: Alinhamento das estruturas terciárias da RmLam (azul), aNLam (vermelho) e PcLam (verde). As fitas- β , numeradas de S1 a S15, são as porções de maior sobreposição nestas três estruturas. Figura extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

am duas folhas- β que delimitam as duas regiões mais importantes de cada uma destas enzimas: a fenda catalítica e o núcleo hidrofóbico. Estas duas regiões são evidenciadas na Figura 3.3. Nas três enzimas, os resíduos catalíticos situam-se na fita- β S8, e são voltados para a fenda catalítica, e as três exibem um núcleo hidrofóbico, situado na região entre as duas folhas- β , no qual apenas cadeias laterais de resíduos hidrofóbicos estão presentes.

Ao comparar o número de pontes salinas presentes na estrutura de diversas enzimas GH da família 16, depositadas no *Protein Data Bank*, a RmLam destaca-se pelo alto número destas pontes salinas presentes em sua estrutura se comparada a outras enzimas da mesma família, como mostra a Tabela 3.1. Os dados nesta tabela foram obtidos através das estruturas estáticas, sem a realização de simula-

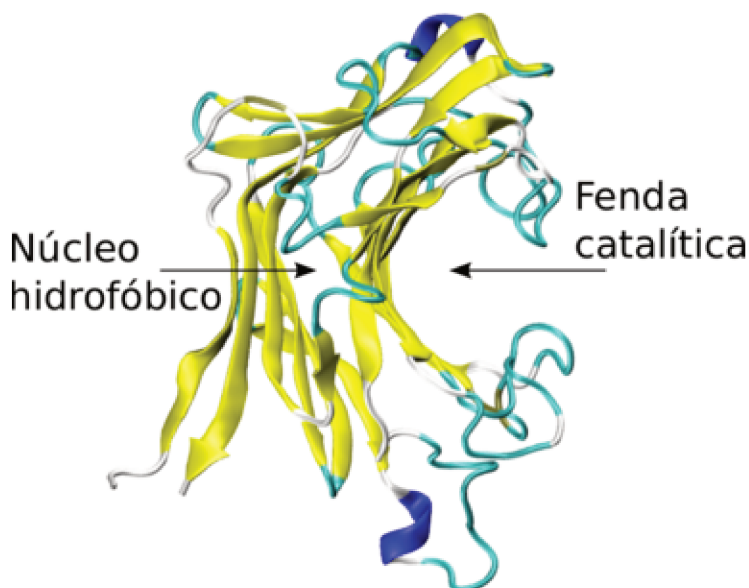


Figura 3.3: As duas regiões mais importantes na estrutura das laminarinases são a fenda catalítica e o núcleo hidrofóbico. O núcleo hidrofóbico compreende uma extensa região entre as duas folhas- β , na qual estão presentes apenas cadeias laterais de resíduos hidrofóbicos.

ções por DM. Na Tabela 3.2, por outro lado, são reportados dados calculados a partir de simulações por DM das enzimas RmLam, aNLam e PcLam.

Foram realizadas simulações a 25°C e a 90°C, para investigar a estabilidade e a mobilidade destas estruturas. Desvios em relação às correspondentes estruturas cristalográficas, mensurados pelos desvios médios quadráticos (RMSD) dos átomos da cadeia principal (*backbone*) da proteína, são próximos a 1Å para todas as enzimas em todas as temperaturas. Para as enzimas RmLam e aNLam, o aumento do RMSD a 90°C é relativamente menor do que o aumento observado no RMSD da PcLam, a 90°C. Experimentalmente, as enzimas RmLam e aNLam preservam (ou até aumentam) a sua atividade enzimática a temperaturas elevadas e, portanto, a sua estrutura também deve ser preservadas a estas temperaturas.

Analisando o comportamento das cadeias laterais das três enzimas durante as simulações por DM, as diferenças em estabilidade entre estas três enzimas tornam-se ainda mais notáveis: o número médio de ligações de hidrogênio entre cadeias laterais de resíduos decresce de 77 a 25°C para 65 a 90°C na RmLam, de 59 para 53 na aNLam, e de 77 para 62 na PcLam (uma grande proporção das ligações de hidrogênio na PcLam encontra-se em regiões periféricas). Ao mesmo tempo, as pontes salinas em todas as estruturas permanecem praticamente intactas quando a temperatura se eleva de 25°C para 90°C. Além disso, o número médio de con-

Tabela 3.1: Número de pontes salinas na RmLam e em outras estruturas de enzimas GH da família 16 depositadas no *Protein Data Bank*. Tabela extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

Estrutura (organismo)	número de pontes salinas
3ILN (<i>R. marinus</i>)	18
2HYK (<i>Nocardiopsis</i> sp.)	7
1DYP (<i>P. carrageenovora</i>)	11
1GBG (<i>B. licheniformis</i>)	5
1MVE (<i>F. succinogenes</i>)	8
1U0A (<i>B. macerans</i>)	5
1UN1 (<i>P. tremula</i>)	11
1O4Y (<i>Z. galactanivorans</i>)	11
2CL2 (<i>P. chrysosporium</i>)	11

tatos hidrofóbicos é aproximadamente 10% maior na RmLam do que nas outras duas enzimas. Em todas as três enzimas, o núcleo hidrofóbico é formado por contatos entre resíduos localizados nas duas folhas- β opostas, entretanto, na RmLam o núcleo hidrofóbico é melhor empacotado e possui mais contatos hidrofóbicos. Em suma, estes resultados revelam que, em comparação com seus homólogos estruturais, a enzima hipertermofílica RmLam exibe um número substancialmente maior de pontes salinas, e um maior número de ligações de hidrogênio intraprotéicas e de contatos hidrofóbicos, fatores que possivelmente contribuem para a sua termoestabilidade. Em termos das propriedades analisadas na Tabela 3.2, as enzimas aNLam e PcLam parecem bastante similares, entretanto, considerando que a PcLam é significativamente maior do que suas homólogas (RmLamR, aNLam, e PcLam têm 251, 245, e 298 resíduos, respectivamente), os conteúdos de ligações de hidrogênio, pontes salinas e contatos hidrofóbicos *por resíduo* é menor na PcLam.

Como o padrão de enovelamento formado pelas fitas- β nas três enzimas, que delimita as regiões do núcleo hidrofóbico e da fenda catalítica, é altamente conservado, é razoável supor que a estabilidade deste padrão de enovelamento seja determinante para a atividade catalítica de uma laminarinase a altas temperaturas. É muito interessante notar que as fitas- β nestas enzimas contêm a maioria dos resíduos envolvidos em pontes (ou clusters) salinos, e que a distribuição das pontes salinas nestas enzimas forma “cintas” ou “vigas” de amarração que conectam fitas- β adjacentes dentro de uma mesma folha- β . Este comportamento é

Tabela 3.2: Dados computados a partir das simulações por DM das três laminarinases homólogas, RmLam, aNLam, and PcLam a 25°C e 90°C. Tabela extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

	RmLam		aNLam		PcLam	
	25 °C	90 °C	25 °C	90 °C	25 °C	90 °C
RMSD (Å·ps ⁻¹)	0.82	0.94	0.92	1.04	0.72	0.94
Estrutura secundária	49.0%	48.8%	45.6%	44.2%	50.5%	49.8%
Ligações de Hidrogênio	77 ± 7	65 ± 7	59 ± 6	53 ± 6	77 ± 8	62 ± 6
Ligações de H por resíduo	0.31	0.26	0.24	0.22	0.26	0.21
Pontes salinas	24 ± 3	24 ± 2	9 ± 1	10 ± 2	11 ± 1	11 ± 1
Contatos hidrofóbicos	166 ± 2	165 ± 3	150 ± 2	151 ± 3	153 ± 3	152 ± 4

particularmente acentuado no RmLam: os mapas de conectividade das pontes salinas, considerando apenas aquelas que persistem pelo menos 50% do tempo de simulação, mostram que nas três enzimas há várias pontes salinas “simples” (ou seja, clusters salinos com tamanho N=1), bem como clusters salinos envolvendo mais de dois resíduos adjacentes eletricamente carregados. Entretanto, apenas a RmLam apresenta grandes clusters salinos formados por cinco ou mais resíduos. Estes clusters salinos interconectam diversas fitas- β e conferem uma estabilização ulterior à estrutura da RmLam.

Uma inspeção mais detalhada da topologia das pontes salinas revela que estas laminarinases diferem em outro aspecto importante: as pontes salinas nas enzimas termofílicas (RmLam e aNLam) conectam fitas- β apenas dentro de uma mesma folha- β , ao contrário da PcLam que também apresenta pontes salinas entre a folha- β interna e a folha- β externa, mediadas por dois clusters salinos (marcados em vermelho na Figura 3.4): o resíduo Glu104 na fita- β S7 estabelece pontes salinas com a His83 na fita S6 e com a Arg191 na fita S11, e o resíduo Lys78 na fita- β S8 forma pontes salinas com o Asp6 e o Glu7 na fita S1. Esta característica revela que na PcLam há interações eletrostáticas entre as folha- β interna e a folha- β externa, enquanto nas enzimas mais termoestáveis RmLam e aNLam o núcleo hidrofóbico é mantido apenas por contatos hidrofóbicos. A presença de pontes salinas “contaminando” o núcleo hidrofóbico da PcLam tem efeito direto em sua termoestabilidade: as Figuras 3.6 e 3.7 mostram que a RmLam e a aNLam

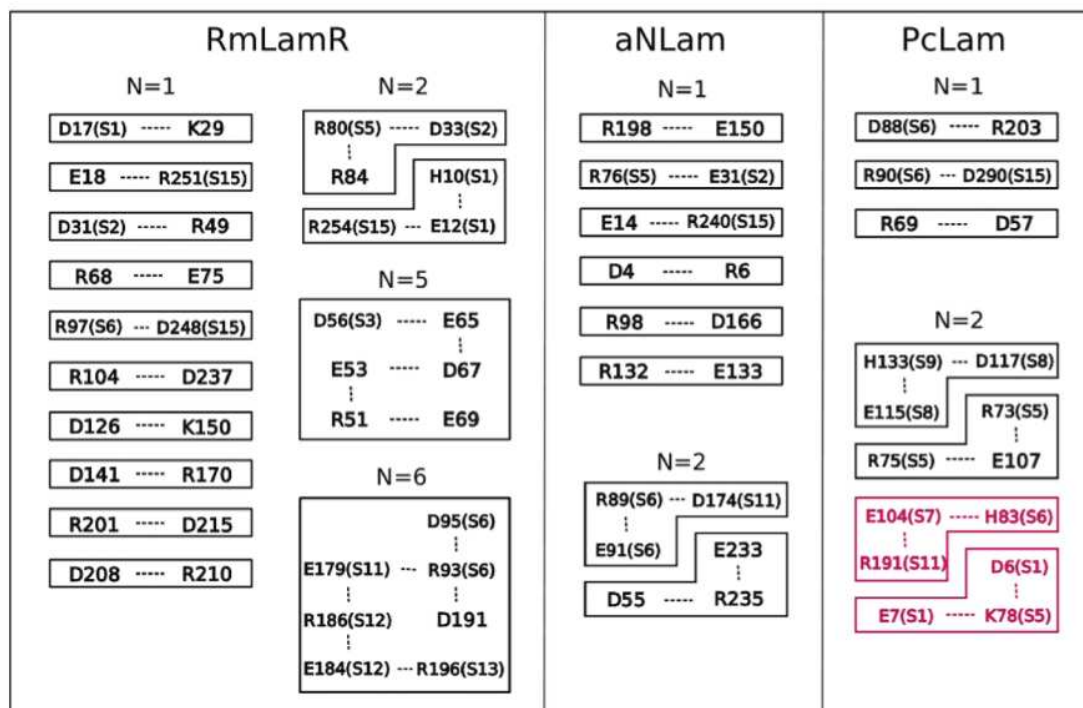


Figura 3.4: Características topológicas das pontes salinas na laminarinases estudadas. Cada linha pontilhada representa uma ponte salina que permanece por pelo menos 50% do tempo de simulação. Em parêntesis estão rótulos para as folhas- β (S de sheet, em inglês). Os resíduos de cada proteína envolvidos em pontes salinas são representados por uma letra maiúscula seguido de seu número na seqüência primária da proteína, segundo as abreviações padrão para aminoácidos. Na enzima RmLam, há clusters salinos contendo até $N = 6$ pontes salinas interconectadas. Estas pontes salinas estabilizam fortemente a estrutura da RmLam. Na enzima PcLam, entretanto, há pontes salinas “contra-producentes” (marcadas em vermelho) entre as duas folhas- β , que facilitam a entrada de moléculas de água no núcleo hidrofóbico da proteína a altas temperaturas (vide Figura 3.6). Figura extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

coordenam, respectivamente, uma e duas moléculas de água * em ambas as temperaturas simuladas, enquanto na PcLam há um aumento significativo no número de águas capazes de penetrar o núcleo hidrofóbico. Este número cresce de uma média de três para até seis moléculas de água quando a enzima PcLam é simulada a 90°C. Esta observação sugere que o começo de um processo de desnaturação pode estar ocorrendo na PcLam, mas a escala de tempo das simulações não é suficiente para capturar todo o desenrolar deste processo. As simulações revelam que as pontes salinas entre as duas folhas- β na PcLam servem como uma porta

*Denominadas águas cristalográficas por estarem presentes nas estruturas cristalográficas depositadas no PDB

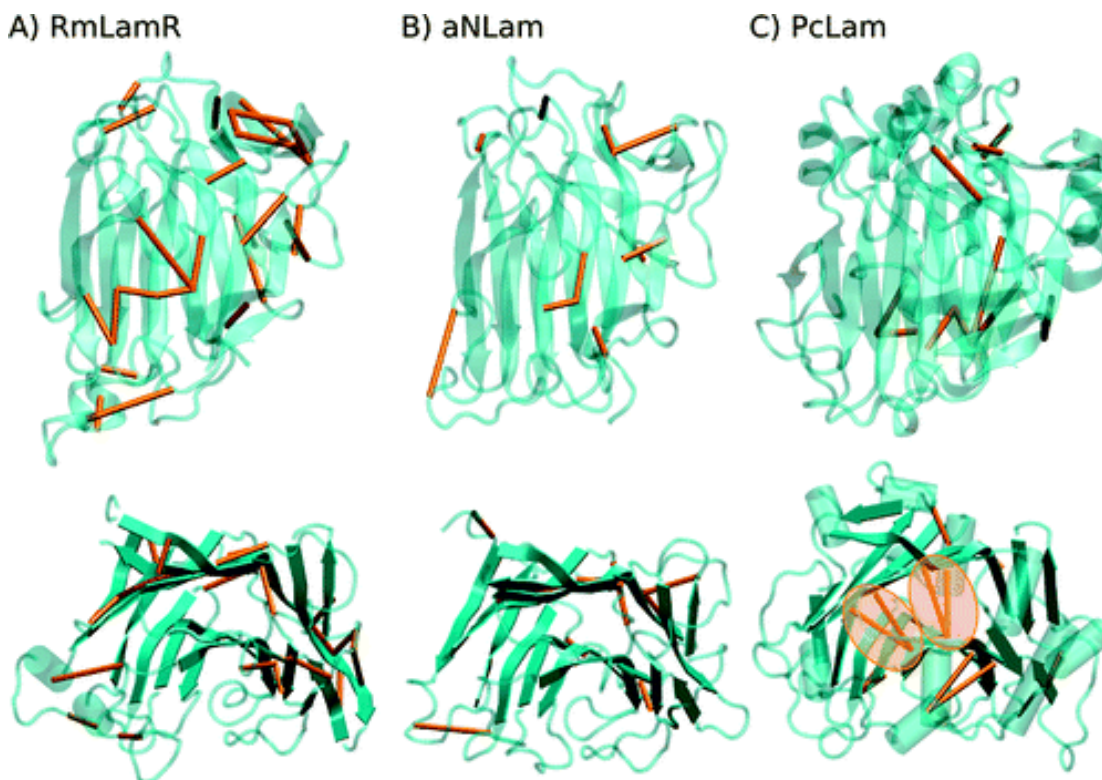


Figura 3.5: Topologia das pontes salinas em laminarinases. (a) RmLam, hipertermofílica. (b) aNLam, termofílica. (c) PcLam, mesofílica. A conectividade das pontes salinas, descrita na Figura 3.4, é representada nas imagens de cima mediante linhas que conectam os carbonos α dos resíduos correspondentes. As imagens de baixo mostram que nas enzimas RmLam e aNLam as pontes salinas conectam apenas fitas- β dentro de uma mesma folha- β . Apenas a enzima mesofílica PcLam apresenta pontes salinas formadas entre as duas folhas- β . Estas pontes salinas desestabilizam o núcleo hidrofóbico da PcLam, facilitando a penetração de moléculas de água a altas temperaturas. Figura extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

de entrada para moléculas de água em seu núcleo hidrofóbico (vide Figura 3.6 D), contribuindo para a desestabilização da PcLam frente ao aumento de temperatura quando comparada às outras duas enzimas mais termoestáveis. Portanto, a disposição das pontes salinas é um fator crucial para a estabilidade térmica nestas laminarinases: pontes salinas inadequadamente localizadas podem ter um papel desestabilizante.

A investigação destas laminarinases, mediante simulações por DM, revelou padrões estruturais e dinâmicos que sugerem explicações em nível microscópico para a hipertermoestabilidade da enzima RmLam. Dois dos principais fatores responsáveis por estabilizar a estrutura da RmLam são: 1) A existência de um

extenso núcleo de resíduos hidrofóbicos na parte interior da proteína, pouco aces-

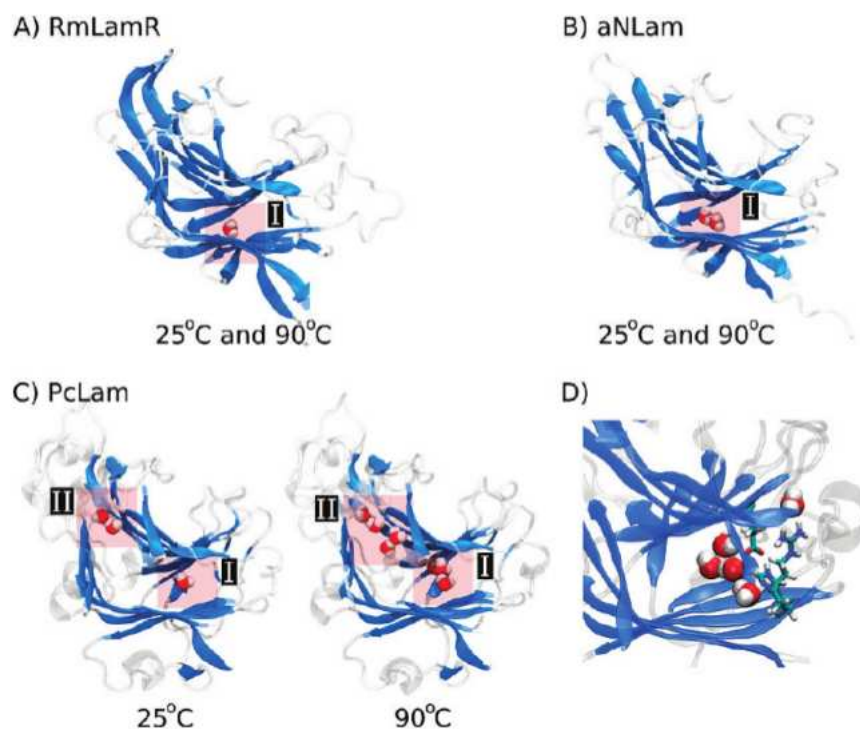


Figura 3.6: Detalhes da penetração de água no núcleo de resíduos hidrofóbicos encerrado entre as folhas β interna e externa das laminarinases. A) Apenas uma moléculas de água está presente no núcleo hidrofóbico da RmLam seja a 25 °C como a 90 °C (região I). A imagem é representativa para ambos os conjuntos de simulações. B) Duas moléculas de água penetram no núcleo hidrofóbico da aNLam a 25 °C e a 90 °C na região I. C) O núcleo hidrofóbico da PcLam é mais acessível à água, uma vez que três moléculas de água penetram no núcleo a 25 °C e até seis moléculas podem ser encontradas nas regiões I e II do núcleo hidrofóbico desta proteína a 90 °C . D) A penetração de água dentro do núcleo hidrofóbico da PcLam é facilitada por pontes salinas na entrada desta região e que estabelecem interações entre as duas folhas- β que delimitam o núcleo hidrofóbico. Figura extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

sível ao solvente (água), mesmo a altas temperaturas. † 2) A presença de uma extensa rede de pontes salinas na parte exterior das folhas- β da enzima, que se mantêm mesmo a altas temperaturas. A localização destas pontes salinas na estrutura da RmLam é outro fator determinante para a estabilização desta estrutura: na enzima mesofílica homóloga PcLam, as pontes salinas que interferem com o

[†]A existência de um núcleo hidrofóbico é um dos fatores que promove o enovelamento e a estabilização da estrutura terciária em diversas proteínas, sendo as globulinas o exemplo mais conhecido.

núcleo hidrofóbico desta enzima facilitam a entrada de moléculas de água em seu núcleo hidrofóbico em temperaturas elevadas, o que indica alterações na estrutura terciária que possivelmente levariam à desnaturação.

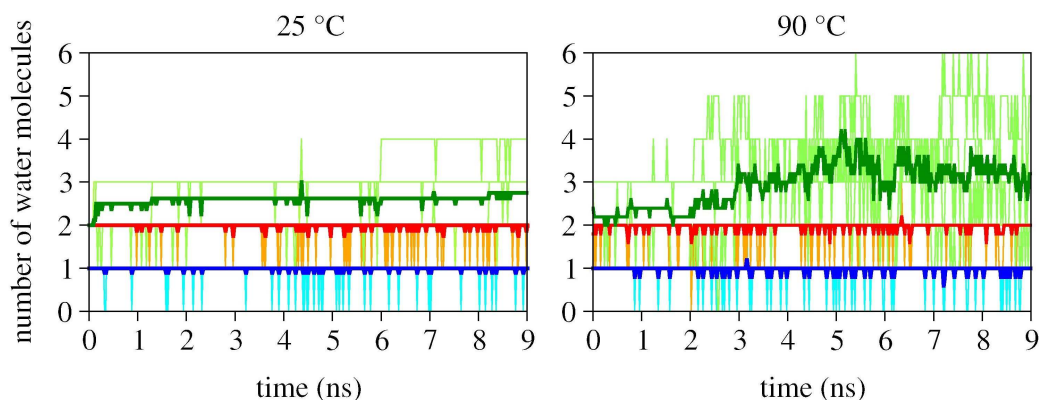


Figura 3.7: Número de moléculas de água detectadas dentro dos núcleos hidrofóbicos da RmLam (azul/ciano), aNLam (vermelho, laranja), e PcLam (verde claro e verde escuro) em função do tempo. Cores claras correspondem a resultados de simulações individuais enquanto cores escuras correspondem à média. Figura extraída de Bleicher *et al.*², Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011, reproduzida com permissão da American Chemical Society (copyright 2011).

Capítulo 4

Cellulose-Builder

Neste capítulo serão apresentados e discutidos os resultados referentes à segunda parte deste trabalho de tese, na qual foi desenvolvido o programa *cellulose-builder*. Este programa foi publicado em um artigo no *Journal of Computational Chemistry*³ em 2012, e suas funcionalidades serão discutidas a seguir. Este artigo contém a maior parte da documentação do *cellulose-builder*, e um guia rápido à sua utilização foi disponibilizado em <http://code.google.com/p/cellulose-builder/>, endereço no qual o programa encontra-se publica e gratuitamente disponível.

Como foi explicado na Seção 2.2, o programa *cellulose-builder* produz arquivos em formato PDB que representam cristaltos de alguns alomorfos da celulose cristalina. Arquivos PDB são arquivos texto, nos quais cada linha corresponde a um átomo presente no sistema representado pelo arquivo PDB: em cada linha constam informações referentes àquele átomo, como sua posição em coordenadas Cartesianas, o tipo de elemento químico, entre outras. Para que os arquivos PDB produzidos pelo *cellulose-builder* sejam úteis para utilização em simulações por DM não basta calcular as coordenadas Cartesianas de cada átomo no cristalito (procedimento que foi descrito na Seção 2.2): é necessário também ordenar a lista de átomos e suas coordenadas de maneira bastante específica, na qual primeiro são listados todos os átomos do primeiro resíduo da cadeia de celulose M0, a seguir são listados todos os átomos do segundo resíduo da mesma cadeia de celulose, e assim por diante para todos os resíduos da primeira cadeia de celulose, M0, e para todas as cadeias de celulose subsequêntes. A Figura 4.1 ilustra o formato necessário para que o arquivo PDB seja utilizável. Além da necessidade de dispor as linhas que representam cada átomo nesta ordem específica, em cada linha deve constar a que tipo de resíduo o átomo pertence (neste caso β -D-glucopiranosose, abreviado como BGLC na Figura 4.1), qual o número deste resíduo em sua cadeia de celulose

ATOM	1	C1	BGLCX	1	0.148	-0.342	0.449	1.00	0.00	M0	C
ATOM	2	H1	BGLCX	1	1.093	-0.153	0.621	1.00	0.00	M0	H
ATOM	3	C5	BGLCX	1	0.055	1.296	-1.279	1.00	0.00	M0	C
ATOM	4	H5	BGLCX	1	1.004	1.484	-1.131	1.00	0.00	M0	H
ATOM	5	O5	BGLCX	1	-0.528	0.807	-0.055	1.00	0.00	M0	O
ATOM	6	C2	BGLCX	1	-0.032	-1.499	-0.536	1.00	0.00	M0	C
ATOM	7	H2	BGLCX	1	-0.972	-1.774	-0.567	1.00	0.00	M0	H
ATOM	8	O2	BGLCX	1	0.766	-2.559	-0.003	1.00	0.00	M0	O
ATOM	9	H02	BGLCX	1	0.581	-2.603	0.955	1.00	0.00	M0	H
ATOM	10	C3	BGLCX	1	0.439	-1.116	-1.918	1.00	0.00	M0	C
ATOM	11	H3	BGLCX	1	1.420	-1.078	-1.910	1.00	0.00	M0	H

Figura 4.1: Formato necessário para que um arquivo PDB seja utilizável por programas de simulação por DM. Cada linha no arquivo corresponde a um átomo do sistema representado. Além de estarem na ordem correta, cada linha deve conter as informações necessárias aos programas de DM sobre o átomo ao qual a linha se refere.

e a qual cadeia de celulose o resíduo pertence. Além disso cada átomo deve ser rotulado com um “tipo atômico”, que é uma informação topológica utilizada para determinar a conectividade entre os átomos (ou seja, qual átomo está ligado a qual) e para carregar os parâmetros do campo de força que servem para descrever o movimento daquele átomo durante as simulações por DM. Portanto, o cálculo das coordenadas Cartesianas para os átomos de um cristalito de celulose, com tamanho especificado pelo usuário, é apenas uma parte do trabalho executado pelo programa *cellulose-builder*, que precisa escrever as coordenadas calculadas em um arquivo PDB com formato adequado. Os arquivos PDB produzidos pelo *cellulose-builder* podem representar cristalitos dos alomorfos de celulose $I\alpha$, $I\beta$, II e III_I , como aqueles visualizados nas Figuras 1.16 e 1.17.

O programa *cellulose-builder* permite a construção de cristalitos de tamanho arbitrário, o qual é especificado pelo usuário diretamente na linha de comando. Ao executar o *cellulose-builder* no prompt do Bash, basta passar como argumento três números inteiros positivos, que representam quantas vezes a cela unitária será replicada em cada uma das três direções cristalográficas (x, y, z) . Para construir um cristalito cujo tamanho é equivalente a dez celas unitárias na direção x , duas celas unitárias na direção y e cinco celas unitárias na direção z , basta digitar o comando:

```
$ ./cellulose-builder 10 2 5
```

e um arquivo PDB (junto com seu respectivo arquivo PSF) será produzido, representando um cristalito com estas dimensões. Um cristalito de celulose $I\beta$ com tais dimensões é representado à esquerda na Figura 4.2. A escolha de qual alomorfo da celulose será representado é especificada editando o arquivo *input.inp*. O terceiro número inteiro fornecido como argumento ao *cellulose-builder* também representa

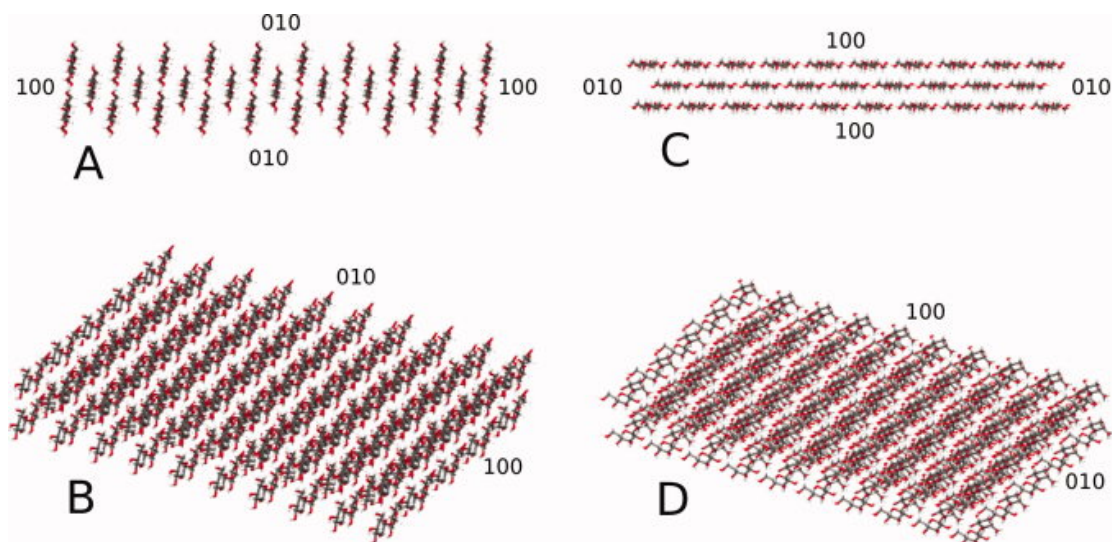


Figura 4.2: Dois cristallitos expondo diferentes superfícies da celulose cristalina $I\beta$. À esquerda: o cristallito expõe predominantemente a face cristalográfica (010), hidrofílica. À direita: o cristallito expõe predominantemente a face cristalográfica (100), hidrofóbica. As imagens A e B representam o mesmo cristallito observado de pontos de vista diferentes. Idem para C e D. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, Cellulose-Builder: A toolkit for building crystalline structures of cellulose. Journal of Computational Chemistry, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

o grau de polimerização das cadeias de celulose no cristallito construído: neste exemplo, cada cadeia de celulose possui cinco unidades de celobiose (a celobiose, a rigor, é o monômero da celulose, conforme foi discutido na Seção 1.2.2), portanto dez unidades de glicose. A Figura 4.2 também ilustra, à direita, um cristallito de celulose $I\beta$ cujo arquivo PDB foi gerado pelo *cellulose-builder* utilizando o comando:

```
$ ./cellulose-builder 2 10 5
```

As Figuras 4.2 e 4.3 são úteis para ilustrar que diferentes faces cristalográficas do alomorfo $I\beta$ da celulose possuem diferentes graus de afinidade pela água: as faces (010) e (020) são denominadas *hidrofílicas* e as faces (100) e (200) são denominadas *hidrofóbicas*. Isto ocorre porque a D-glicose, em sua conformação mais estável, a cadeira 4C_1 , tem todas as suas hidroxilas na posição equatorial, e na celulose as unidades de D-glicose estão unidas por ligações glicosídicas do tipo $\beta(1 \rightarrow 4)$, o que permite às cadeias de celulose empacotarem-se no cristal adotando conformações em fitas ou hélices quase planares, denominadas hélices 2_1 . Portanto, nas faces (010) e (020) as hidroxilas estão expostas, apontando para o exterior do cristal, logo estas faces são mais hidrofílicas. As cadeias de celulose no alomorfo

$I\beta$ formam camadas, compostas exclusivamente por cadeias *origin* ou por cadeias *center*. Cadeias de celulose vizinhas em uma dada camada estabelecem ligações de hidrogênio entre elas, mas não há ligações de hidrogênio entre camadas vizinhas. Nas faces delimitadas por estas camadas, as faces (100) e (200), não há hidroxilas expostas ao meio externo como há nas outras faces: apenas hidrogênios ligados a carbonos, na posição axial nos anéis de glicose, apontam para fora do cristal nas faces (100) e (200) e, portanto, estas faces são denominadas *hidrofóbicas*. É claro que estas superfícies não são tão hidrofóbicas como seria de se esperar para uma superfície de polietileno cristalino (um polímero totalmente apolar), por exemplo, mas é razoável supor que a afinidade pela água das faces (100) e (200) da celulose $I\beta$ seja significativamente menor do que a de suas faces (010) e (020).

Freqüentemente, em simulações por DM, utilizam-se condições periódicas de contorno nas quais a caixa de simulação possui réplicas além de suas extremidades. Para aplicar condições periódicas de contorno em simulações de cristais, eles devem ter a simetria translacional adequada em todas as direções em que o cristal será replicado. O programa *cellulose-builder* permite construir cristalitos com simetria translacional adequada para a aplicação de condições periódicas de contorno em qualquer direção cristalográfica, como é ilustrado na Figura 4.3. As opções a respeito da simetria translacional do cristalito são especificadas editando o arquivo *input.inp*.

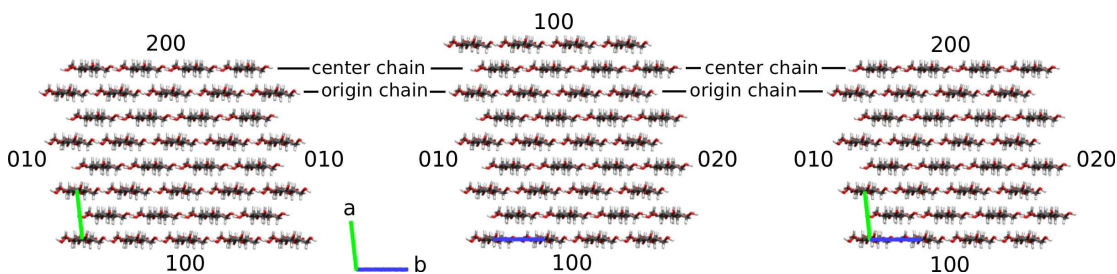


Figura 4.3: Cristalitos de celulose $I\beta$ aptos para a aplicação de condições periódicas de contorno ao longo das direções cristalográficas a (esquerda), b (meio) e ambas a e b (direita). Camadas de cadeias *origin* e *center*⁴⁷ estão indicadas, assim como direções cristalográficas e índices de Miller. A notação adotada para os índices de Miller é a mesma adotada por Himmel *et al.*⁵⁰, portanto as faces nas quais as cadeias *center* estão na superfície são indicadas por (200) ou (020), explicitando que as mesmas estão posicionadas no meio da cela unitária. As linhas coloridas sobre os cristalitos representam as direções ao longo das quais o cristalito possui simetria translacional. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, *Cellulose-Builder: A toolkit for building crystalline structures of cellulose*. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

Portanto, dependendo das opções de *input* fornecidas pelo usuário, podem

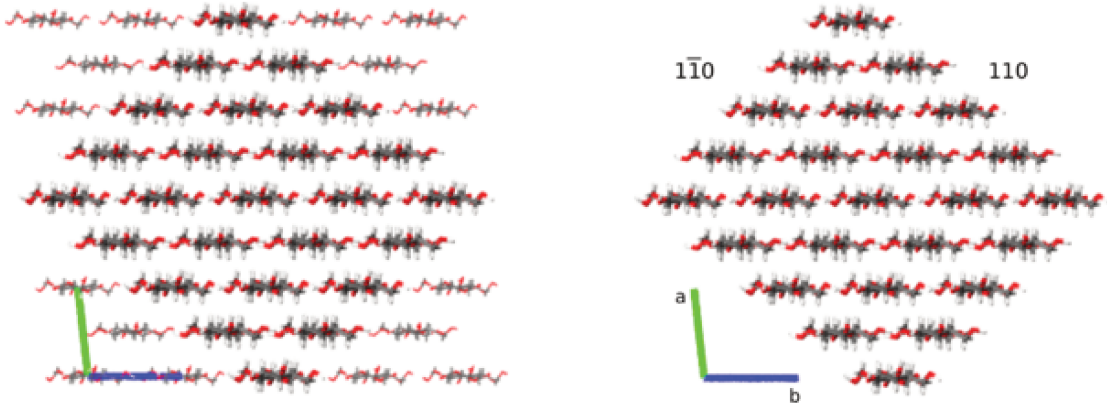


Figura 4.4: O método de esculpir um cristalito a partir de outro cristalito maior é utilizado pelo *cellulose-builder* para construir automaticamente fibrilas com a seção transversal representada na Figura 4.5. Este método também pode ser utilizado para produzir manualmente cristalitos com formato totalmente geral, como por exemplo os cristalitos nesta figura, que expõem as faces cristalográficas (110) e (1-10) da celulose I β . À direita, as faces opostas entre si têm os mesmos índices de Miller. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, Cellulose-Builder: A toolkit for building crystalline structures of cellulose. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

ser expostas as diferentes faces cristalográficas dos diferentes alomorfos suportados pelo *cellulose-builder*. A Figura 4.4 demonstra que, de maneira ainda mais geral, cristalitos de forma e tamanho arbitrários podem ser obtidos mediante a simples eliminação de cadeias de celulose a partir de um cristalito maior. Em

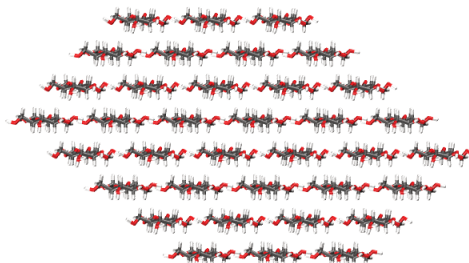


Figura 4.5: Modelo proposto por Ding e Himmel⁹⁰ para fibrila elementar de celulose em parede celular primária de milho, vista a partir de suas extremidades não redutoras. A fibrila elementar é composta de 36 cadeias de celulose. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, Cellulose-Builder: A toolkit for building crystalline structures of cellulose. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

outras palavras, a partir de um arquivo PDB produzido pelo *cellulose-builder* o usuário pode “esculpir” um cristalito de qualquer formato, simplesmente eliminando as cadeias de celulose indesejadas no arquivo PDB, o que é bastante fácil

de se fazer dado que as linhas nos arquivos PDB produzidos pelo *cellulose-builder* são ordenadas de maneira precisamente conhecida, e que cada átomo possui um rótulo que identifica a cadeia de celulose à qual ele pertence. Em particular, um destes formatos especiais é suportado automaticamente pelo cellulose builder, e é mostrado na Figura 4.5. Esta figura ilustra um modelo, proposto por Ding e Himmel⁹⁰, para a seção transversal de fibrilas elementares de celulose em parede celular de milho (*Zea mays*), e cristalitos com esta seção transversal são gerados pelo *cellulose-builder* mediante um comando do tipo:

```
$ ./cellulose-builder fibril k
```

onde k é um número inteiro positivo. Este tipo de comando permite construir estas fibrilas elementares com qualquer grau de polimerização, ou seja, com qualquer número de unidades repetitivas de celobiose, como mostrado na Figura 4.6, que ilustra fibrilas de celulose construídas a partir deste comando com k igual a 5, 25, 50, 100 e 500.

Ainda que, no caso do *cellulose-builder*, a performance do código não seja um fator crítico para a utilidade deste tipo de programa *, na Tabela 4.1 são apresentados os resultados de aferição do tempo e recursos computacionais utilizados pelo *cellulose-builder* para construir os cristalitos representados na Figura 4.6. Os dados nesta tabela demonstram que, mesmo com recursos de hardware bastante modestos, as coordenadas Cartesianas para fibrilas elementares de celulose como aquelas mostradas na Figura 4.6 podem ser obtidas em um intervalo de tempo que varia de poucos segundos a cerca de três minutos, dependendo do comprimento da fibrila, a uma taxa de 2.86 unidades de celobiose por segundo. O *wall clock time* é linearmente correlacionado com o número de unidades de celobiose na fibrila, com coeficiente de correlação $R^2 = 0.99992$.

O programa *cellulose-builder* mostrou-se estável, de fácil utilização e portabilidade para outros sistemas operacionais que se atêm ao padrão POSIX. Os testes realizados demonstram que é possível obter os cristalitos descritos na Figura 4.6 em questão de segundos ou minutos, dependendo do tamanho do sistema. A maior fibrila mostrada na Figura 4.6, por exemplo, é obtida em cerca de três minutos em um simples PC (rodando sob sistema operacional GNU/Linux Ubuntu 8.10, kernel 2.6.27-7-generic, i686 com 2GB de memória RAM e processador AMD Athlon(tm)

*Caso a performance fosse determinante para a utilidade do programa, a linguagem de programação escolhida teria sido C ou Fortran, mas como esse não é o caso, o uso da linguagem Bash é muito mais conveniente pois é possível atingir o mesmo resultado com muito menos linhas de código.

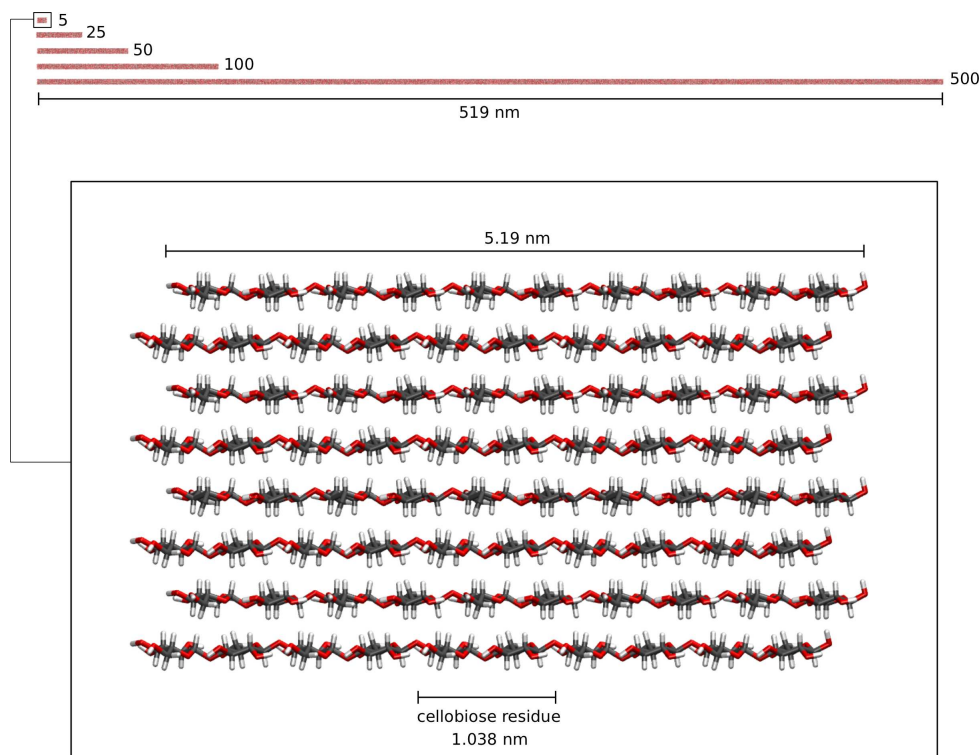


Figura 4.6: Fibrilas elementares de celulose I β com diferentes graus de polimerização (número de unidades de celobiose, k , igual a 5, 25, 50, 100, 500), geradas com o programa *cellulose-builder*. A fibrila com $k = 5$ foi ampliada. Todas as fibrilas tem a seção transversal mostrada na Figura 4.5. Figura extraída de Thiago C. F. Gomes and Munir S. Skaf³, *Cellulose-BUILDER: A toolkit for building crystalline structures of cellulose*. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

64 X2 Dual Core Processor 5200+, frequência CPU: 2613.345 MHz; tamanho do cache: 1024 KB). Portanto, o programa permite poupar um tempo enorme na construção de cristalitos de celulose em relação à execução da mesma tarefa manualmente, e deve possibilitar maior celeridade na investigação da celulose mediante simulações por DM. O programa já tem sido utilizado por pesquisadores ao redor do mundo, como mostra a Figura 4.7, alguns dos quais inclusive colaboraram reportando pequenos erros que ocorriam ao tentar executar o *cellulose-builder* em seus sistemas operacionais. Após cada uma destas notificações, pequenas adaptações foram feitas no código do programa para permitir que ele seja executado sem problemas na maior variedade possível de sistemas operacionais. Até o presente momento, os usuários que reportaram problemas gentilmente testaram posteriormente as versões atualizadas após as suas notificações, que funcionaram corretamente em seus computadores e, portanto, foram disponibilizadas ao pú-

Tabela 4.1: Tempos de execução do *cellulose-builder* para produzir arquivos PDB e PSF para as fibrilas elementares de celulose representadas na Figura 4.6, em função do número de unidades monoméricas de celobiose em cada fibrila. Tempo real decorrido (*wall clock time*), tempo de CPU que o processo utiliza diretamente (em *user mode*), tempo de utilização da CPU pelo sistema operacional por causa do processo (em *kernel mode*), e porcentagem de utilização da CPU pelo *cellulose-builder*, conforme fornecidos pelo */usr/bin/time*. Todos os tempos estão em segundos. A porcentagem de utilização de CPU equivale aos tempos (*user + system*) divididos pelo tempo real decorrido. A quantidade de memória RAM utilizada para construir o maior sistema foi apenas 360MB. Tabela extraída de Thiago C. F. Gomes and Munir S. Skaf³, Cellulose-Builder: A toolkit for building crystalline structures of cellulose. Journal of Computational Chemistry, 33(14):1338–1346, 2012, reproduzida com permissão de John Wiley & Sons (copyright 2012).

Celobioses	<i>Wall clock time</i>	<i>User CPU-time</i>	<i>System CPU-time</i>	%CPU
5	5.75	2.93	0.81	65
25	11.28	9.08	1.56	94
50	20.06	16.80	2.72	97
100	37.51	32.21	4.86	98
500	178.06	154.60	23.11	99

blico em geral, de forma que o código fonte do *cellulose-builder* tem sido mantido satisfatoriamente até o momento. Contudo, futuramente o programa será disponibilizado para execução remota em um servidor na rede mundial de computadores, no qual o usuário poderá executar o *cellulose-builder* e realizar o *download* dos arquivos então produzidos. Esta medida evitará eventuais problemas de portabilidade entre sistemas operacionais, permitirá a obtenção de maiores informações a respeito da utilização do código e dará ao usuário a opção de utilizar o programa em qualquer computador com acesso à *internet*, de maneira imediata, sem precisar instalar as dependências do *cellulose-builder*, como o Octave, VMD e psfgen.

Dado o crescente interesse de pesquisa sobre sistemas lignocelulósicos, estudos computacionais inclusive, o programa por nós apresentado consiste em uma importante ferramenta para cientistas interessados em realizar simulações por DM em sistemas contendo domínios cristalinos de celulose. Além disso, ainda que a finalidade deste programa seja a de fornecer arquivos de entrada (*inputs*) para programas de DM, os arquivos gerados pelo *cellulose-builder* também podem ser úteis para estudos em cristalografia ou mesmo para visualização de domínios cristalinos dos diferentes alomorfos de celulose suportados pelo programa.

O *cellulose-builder* foi utilizado para construir os arquivos PDB necessários

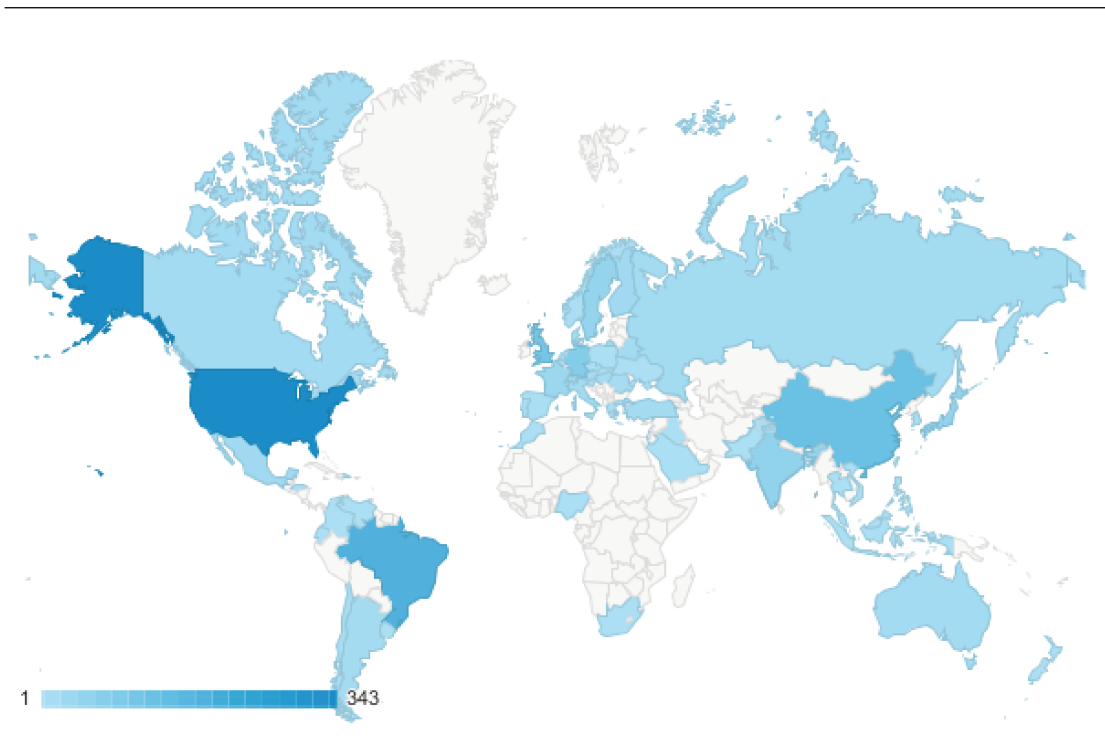


Figura 4.7: Representação do número de acessos ao sítio <http://code.google.com/p/cellulose-builder/> em escala de cores, por país. Entre Fevereiro de 2012 e Novembro de 2013 o sítio foi acessado 1412 vezes e foram realizados 493 *downloads* do *cellulose-builder*. Estes dados não representam o número total de pessoas que acessaram o sítio ou baixaram o programa: uma mesma pessoa pode acessar o sítio várias vezes, em diferentes momentos, e baixar o programa mais de uma vez, em computadores diferentes por exemplo. Fonte: Google Analytics.

para as simulações por DM discutidas no próximo capítulo. Este programa também já foi utilizado e citado em outros trabalhos, um deles feito por outro aluno em nosso grupo de pesquisas⁹¹ e o outro realizado por pesquisadores⁹² no exterior, que não fazem parte da nossa rede de colaborações.

Capítulo 5

Interações entre celulose e xilanos

Neste capítulo, os resultados referentes à terceira parte do trabalho de tese serão apresentados e discutidos. Nesta terceira parte, foram estudados, mediante simulações por DM, complexos de adsorção entre celulose e xilano, simples e substituídos. Alguns dos sistemas estudados nesta parte do trabalho foram representados nas Figuras 1.18 e 1.20. Para facilitar a discussão sobre estes sis-

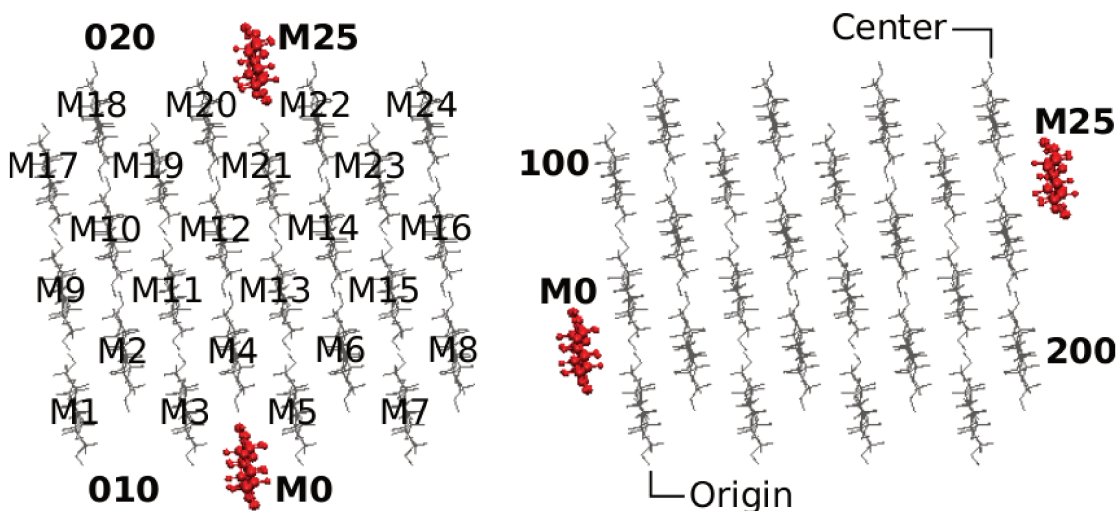


Figura 5.1: Conjunto de rótulos para os sistemas estudados, no qual cada cadeia polimérica recebe um rótulo. As cadeias M0 e M25 são moléculas de xilano. Por clareza, os rótulos das cadeias de celulose foram omitidos no sistema à direita, entretanto o formato do cristal de celulose e os rótulos são idênticos nos dois sistemas. Cadeias de celulose do tipo *origin* têm rótulos ímpares (M1, M3, etc.), cadeias de celulose do tipo *center* têm rótulos pares (M2, M4, etc.). As faces hidrofílicas (010) e (020) são opostas entre si, assim como as faces hidrofóbicas (100) e (200). As faces (010) e (100) são sempre delimitadas por cadeias de celulose do tipo *origin* e, nos sistemas estudados, xilanos com rótulo M0 estão adsorvidos nestas faces. As faces (020) e (200) são sempre delimitadas por cadeias de celulose do tipo *center*, e xilanos com rótulo M25 estão adsorvidos nestas outras faces.

temas, será introduzido um conjunto de rótulos, apresentado na Figura 5.1, em que cada molécula de glicano presente no sistema recebe um nome. Em cada sistema, as moléculas de xilano são as que foram estudadas em maior detalhe e são denominadas M0 e M25. As simulações compreendem xilanos adsorvidos às faces cristalográficas (010), (020), (100) e (200) do alomorfo $I\beta$ da celulose cristalina. Os cristais de celulose sobre os quais as moléculas de xilano estão adsorvidas são compostos por 24 moléculas de celulose, e têm seção transversal denominada “quadrada”⁹³. Evidências experimentais para a existência de cristais de celulose com esta morfologia em plantas foi obtida por Jarvis *et al.*⁹³ Como visto no Capítulo 4, as faces (010) e (020) da celulose $I\beta$ são chamadas *hidrofílicas*, enquanto as faces (100) e (200) são denominadas *hidrofóbicas*. Nas seções a seguir serão apresentados os resultados de análises das simulações por DM de xilanos adsorvidos sobre faces hidrofílicas e hidrofóbicas de cristais de celulose $I\beta$.

5.1 Xilanos adsorvidos sobre faces hidrofílicas

Uma representação da dinâmica das moléculas de xilano adsorvidas sobre as faces hidrofílicas da celulose $I\beta$ é apresentada na Figura 5.2. Nesta figura, observa-se que o complexo de adsorção formado na face (010) mostrou-se consideravelmente mais estável, mantendo-se próximo à sua configuração inicial durante as simulações, enquanto a molécula de xilano adsorvida à face (020) apresentou grande mobilidade nos resíduos mais próximos à sua extremidade não redutora.

Experimentalmente⁴⁷, na celulose $I\beta$ as cadeias *origin* e *center* estão dispostas em camadas, conforme mostrado nas Figuras 1.16 e 4.3. Só há ligações de hidrogênio dentro de cada camada, mas não entre camadas, portanto não há ligações de hidrogênio entre cadeias de celulose *origin* e *center*. Isto ocorre porque a D-glicose, em sua conformação mais estável, a cadeira 4C_1 , tem todas as suas hidroxilas na posição equatorial, permitindo que ligações de hidrogênio sejam estabelecidas com cadeias de celulose vizinhas ao longo da direção cristalográfica *b*, mas não com cadeias vizinhas ao longo da direção *a*. Como a D-xilose equivale à D-glicose sem o carbono C6, as cavidades que existem nas faces cristalográficas (010) e (020), que seriam ocupadas por outras moléculas de celulose se o cristal fosse mais extenso, são sítios de adsorção perfeitos também para moléculas de xilano. No xilano, as hidroxilas também estão na posição equatorial, portanto as interações entre xilano e as faces hidrofílicas da celulose $I\beta$ devem ocorrer mediante ligações de hidrogênio entre o xilano e a cadeia de celulose adjacente na

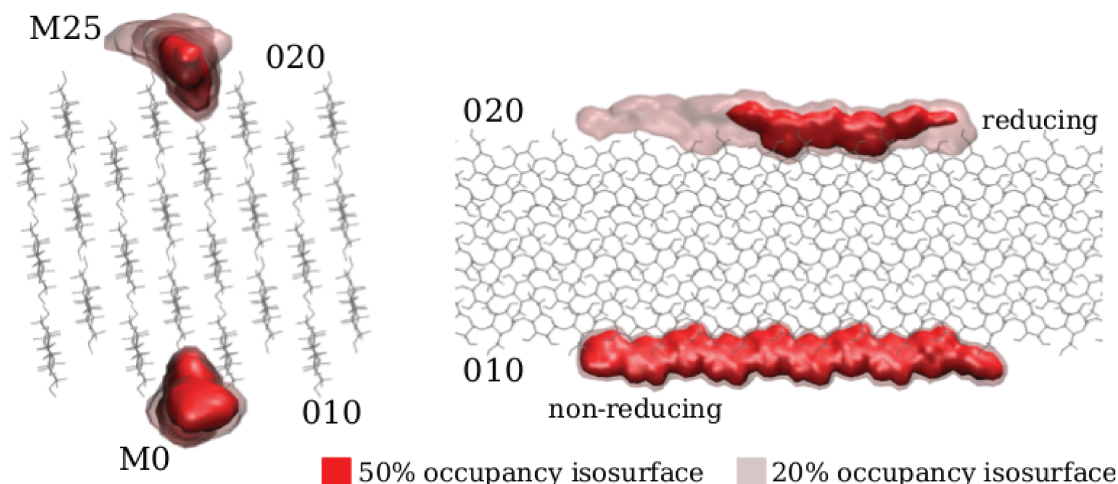


Figura 5.2: Representação estatística da dinâmica de moléculas de xilano adsorvidas sobre as faces hidrofílicas da celulose $I\beta$, utilizando superfícies de nível, ou *isosuperfícies*, de ocupação. O cristal de celulose está representado em cinza, na sua configuração inicial. As isosuperfícies em vermelho representam os pontos nos quais algum átomo da molécula de xilano está presente em 50% do tempo de simulação, enquanto as isosuperfícies transparentes em rosa são compostas pelos pontos no espaço em que algum átomo da molécula de xilano está presente em 20% do tempo de simulação. A molécula de xilano na face (010) não deixa seu sítio de adsorção, permanecendo todo o tempo de simulação oscilando próximo a ele. Já a molécula de xilano na face (020) apresentou grande mobilidade, especialmente em resíduos próximos à sua extremidade não redutora.

mesma camada (*origin* ou *center*), de forma análoga ao que ocorre com cadeias de celulose vizinhas em uma mesma camada.

Para caracterizar as interações entre celulose e xilano nestas duas faces hidrofílicas, foi realizada uma contagem detalhada das ligações de hidrogênio estabelecidas entre estes dois componentes nas simulações. Os critérios geométricos utilizados para definir ligações de hidrogênio foram apresentados na Figura 2.1. Analisando as contagens de ligações de hidrogênio apresentadas nas Tabelas 5.1 e 5.2, pode-se afirmar que uma descrição estatisticamente representativa das interações estabelecidas entre celulose e xilano neste estudo é aquela apresentada na Figura 5.3: nas nossas simulações, celulose e xilano interagem majoritariamente através de ligações de hidrogênio entre os oxigênios exocíclicos O6 dos resíduos de glicose e os oxigênios O2 dos resíduos de xilose. Analisando em detalhe a Tabela 5.1, observamos que a tendência geral para a molécula de xilano M0 é que apenas seus resíduos pares (resíduos 0*, 2, 4, 6, 8) estabelecem ligações de hidrogênio com a molécula de celulose M4, enquanto resíduos ímpares (resíduos 1, 3, 5, 7, 9) não o fazem. Esta diferença ocorre porque nos resíduos de xilose os melhores candidatos

*O resíduo 0 é a extremidade não-redutora

para participação em ligações de hidrogênio são os oxigênios O2 e O3, que nos resíduos ímpares do xilano M0 apontam para longe do cristal de celulose, o que pode ser visto claramente na Figura 5.3. Na Tabela 5.1 observa-se que na face (010) o modo mais freqüente de ligação de hidrogênio entre M0 e M4 é aquele que envolve os oxigênios O2 das xiloses como doadores de próton e os oxigênios O6 das glicoses como aceptores de prótons. Em princípio, esta relação também poderia ser invertida, com os O2 das xiloses atuando como aceptores e os O6 das glicoses como doadores de prótons, mas na Tabela 5.1 nota-se que esta inversão raramente ocorre. Há também um modo secundário, em que são os oxigênios O3 dos resíduos de xilose que estabelecem ligações de hidrogênio com os oxigênios O6 dos resíduos de glicose, mas este modo não é predominante. Estas observações

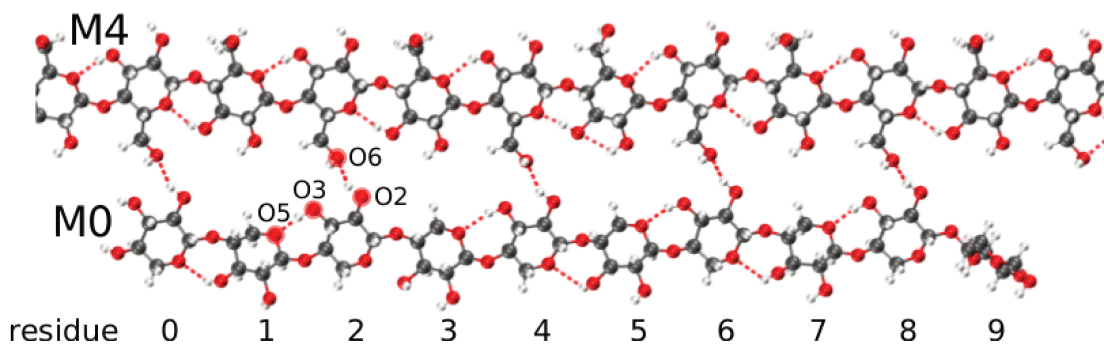


Figura 5.3: *Frame* estatisticamente representativo das ligações de hidrogênio entre a cadeia de xilano M0 e a cadeia de celulose M4, posicionada imediatamente acima (vide Figura 5.1), na face (010). As ligações de hidrogênio entre M0 e M4 ocorrem predominantemente com o oxigênios O2 dos resíduos de xilose como doadores de próton e os oxigênios O6 dos resíduos de glicose com aceptores de próton. Em M0, apenas os resíduos de xilose pares (0, 2, 4, 6, 8) podem estabelecer ligações de hidrogênio com M4. Na maior parte do tempo de simulação, o oxigênios O3 no xilano estabelecem ligações de hidrogênio intramoleculares com um oxigênio O5, endocíclico, no resíduo de xilose precedente.

sugerem que os oxigênios exocíclicos O6 da celulose sejam imprescindíveis para as interações estabelecidas entre celulose I β e xilano nas faces hidrofílicas, uma vez que os outros oxigênios na molécula M4 não estabelecem ligações de hidrogênio com M0 por tempo significativo. Nos curtos intervalos de tempo em que um determinado oxigênio O6 da celulose não está envolvido em ligação de hidrogênio com o correspondente O2 do xilano, ele estabelece ligação de hidrogênio com o O3 no mesmo anel de xilose. Portanto, na maior parte do tempo de simulação há cinco ou mais ligações de hidrogênio entre M0 e M4, como pode-se observar na Figura 5.4.

Entretanto, a mesma figura também evidencia que a cadeia de xilano M25,

Tabela 5.1: Porcentagem do tempo de simulação em que ligações de hidrogênio ocorrem entre a molécula de xilano M0 e a molécula de celulose M4, na face (010) (vide Figuras 5.1 e 5.3). As ligações de hidrogênio estão divididas por resíduo de xilose. Apenas os resíduos pares em M0 participam de ligações de hidrogênio com M4. Em negrito: para a maioria dos resíduos de xilose, o tipo de ligação de hidrogênio predominante é aquela em que os oxigênios O2 das xiloses atuam como doadores de prótons para os oxigênios O6 das glicoses, que atuam como aceptores de próton. Para o resíduo 0, que é a extremidade não redutora do xilano M0, o padrão é diferente devido à maior mobilidade do resíduo. Os oxigênios O3 das xiloses estabelecem ligações de hidrogênio menos freqüentemente com M4, atuando principalmente como aceptores de próton para os oxigênios O6 das glicoses, que atuam como doadores de próton neste caso.

Resíduo 0			Resíduo 2		
Doador	Aceptor	Ocupância (%)	Doador	Aceptor	Ocupância (%)
Xyl0 O2	Glc5 O6	30.86	Xyl2 O2	Glc7 O6	91.78
Glc5 O6	Xyl0 O2	49.89	Glc7 O6	Xyl2 O2	0.28
Xyl0 O3	Glc5 O6	16.79	Xyl2 O3	Glc7 O6	6.76
Glc5 O6	Xyl0 O3	2.58	Glc7 O6	Xyl2 O3	9.79
Resíduo 4			Resíduo 6		
Doador	Aceptor	Ocupância (%)	Doador	Aceptor	Ocupância (%)
Xyl4 O2	Glc9 O6	91.57	Xyl6 O2	Glc11 O6	87.61
Glc9 O6	Xyl4 O2	3.22	Glc11 O6	Xyl6 O2	0.74
Xyl4 O3	Glc9 O6	2.74	Xyl6 O3	Glc11 O6	4.38
Glc9 O6	Xyl4 O3	11.53	Glc11 O6	Xyl6 O3	13.45
Resíduo 8					
Doador	Aceptor	Ocupância (%)			
Xyl8 O2	Glc13 O6	87.87			
Glc13 O6	Xyl8 O2	2.30			
Xyl8 O3	Glc13 O6	1.90			
Glc13 O6	Xyl8 O3	13.31			

sobre a face hidrofílica (020), estabelece ligações de hidrogênio menos freqüentemente com o cristal de celulose, se comparada à cadeia de xilano M0 sobre a face hidrofílica (010). Esta observação concorda com a fato de M25 ser mais móvel que M0 durante as simulações, como foi mostrado na Figura 5.2. A Tabela 5.2 mostra que, embora M25 estabeleça ligações de hidrogênio menos freqüentemente do que M0 com o cristal de celulose, quando as interações ocorrem os tipos de ligações de hidrogênio formadas são idênticos: i) predominantemente oxigênios O2 das xiloses atuando como doadores de próton para os oxigênios O6 das glicoses, que agem com aceptores de prótons; ii) um modo secundário no qual são os oxigênios O3 dos resíduos de xilose que estabelecem ligações de hidrogênio com os oxigênios O6

dos resíduos de glicose. Entretanto, para a cadeia de xilano M25 são os resíduos de xilose ímpares (resíduo 1, 3, 5, 7, 9) aqueles capazes de estabelecer ligações de hidrogênio com a celulose, uma vez que nos resíduos pares os oxigênios O2 e O3 apontam para longe do cristal. Portanto, ao caracterizar-se quantitativamente as

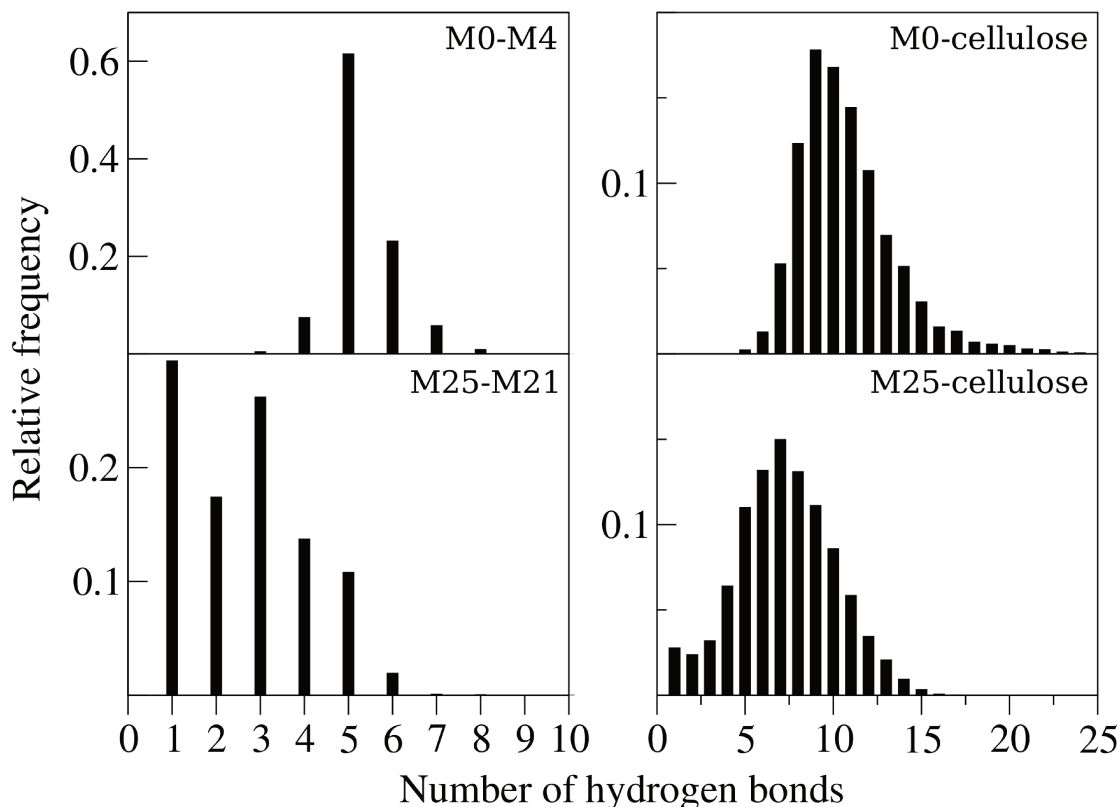


Figura 5.4: Distribuição do número de ligações de hidrogênio estabelecidas entre cadeias de xilano e as faces hidrofílicas da celulose. À esquerda: distribuição do número de ligações de hidrogênio entre cadeias de xilano e cadeias de celulose posicionadas imediatamente acima ou abaixo, vide Figura 5.1. Na maior parte do tempo de simulação há cinco ou mais ligações de hidrogênio entre M0 e M4, mas há cinco ou menos ligações de hidrogênio entre M25 e M21. À direita: distribuição do número de ligações de hidrogênio estabelecidas entre cadeias de xilano e o cristal de celulose por inteiro. A molécula de xilano M0 estabelece ligações de hidrogênio com o cristal de celulose mais freqüentemente do que a molécula de xilano M25. Em todos os histogramas, os valores na ordenada foram normalizados para que indiquem a porcentagem do tempo de simulação em que ocorre o seu valor correspondente na abscissa.

ligações de hidrogênio entre xilano e as faces hidrofílicas da celulose nestas simulações, percebe-se que as estas interações são qualitativamente idênticas em ambas as faces (010) e (020), embora haja diferenças quantitativas quanto à freqüência com que as ligações de hidrogênio ocorrem nas duas faces. Estas diferenças estão relacionadas à mobilidade das cadeias de xilano sobre cada face: a cadeia

M25 é mais móvel que M0, e durante as simulações estabelece menos ligações de hidrogênio com o cristal de celulose.

Tabela 5.2: Porcentagem do tempo de simulação em que ligações de hidrogênio ocorrem entre a molécula de xilano M25 e a molécula de celulose M21, na face (020) (vide Figura 5.1). As ligações de hidrogênio estão divididas por resíduo de xilose. Apenas os resíduos ímpares em M25 participam de ligações de hidrogênio com M21. Em negrito: o tipo de ligação de hidrogênio predominante é aquela em que os oxigênios O2 das xiloses atuam como doadores de prótons para os oxigênios O6 das glicoses, que atuam como aceptores de próton. Ligações de hidrogênio envolvendo o oxigênio O3 das xiloses ocorrem menos freqüentemente que aquelas envolvendo o O2 no mesmo resíduo de xilose.

Resíduo 1			Resíduo 3		
Doador	Aceptor	Ocupância (%)	Doador	Aceptor	Ocupância (%)
Xyl1 O2	Glc6 O6	7.96	Xyl3 O2	Glc8 O6	19.71
Glc6 O6	Xyl1 O2	0.16	Glc8 O6	Xyl3 O2	2.02
Xyl1 O3	Glc6 O6	2.32	Xyl3 O3	Glc8 O6	6.84
Glc6 O6	Xyl1 O3	0.74	Glc8 O6	Xyl3 O3	1.50
Resíduo 5			Resíduo 7		
Doador	Aceptor	Ocupância (%)	Doador	Aceptor	Ocupância (%)
Xyl5 O2	Glc10 O6	64.14	Xyl7 O2	Glc12 O6	56.77
Glc10 O6	Xyl5 O2	0.66	Glc12 O6	Xyl7 O2	0.56
Xyl5 O3	Glc10 O6	6.12	Xyl7 O3	Glc12 O6	12.37
Glc10 O6	Xyl5 O3	4.46	Glc12 O6	Xyl7 O3	24.57
Resíduo 9					
Doador	Aceptor	Ocupância (%)			
Xyl9 O2	Glc14 O6	19.51			
Glc14 O6	Xyl9 O2	1.70			
Xyl9 O3	Glc14 O6	1.98			
Glc14 O6	Xyl9 O3	3.04			

Uma maneira eficiente de caracterizar os movimentos realizados por cada cadeia de xilano durante as simulações é realizar a análise conformacional destes glicanos, de acordo com o exposto Seção 1.2.2, acompanhando a trajetória de seus diedros glicosídicos ϕ e ψ , que foram apresentados na Figura 1.10. Os diedros glicosídicos são importantes pois representam torções em torno das ligações glicosídicas, que são os elos que unem os monossacarídeos em glicanos. A maior parte da mobilidade conformacional em glicanos compostos por piranoses, como é o caso também do xilano, advém de torções em torno dos diedros glicosídicos. O conjunto de pares ordenados $(\phi_1, \psi_1; \phi_2, \psi_2; \dots \phi_i, \psi_i)$ nos permite fazer afirmações acerca da estrutura secundária adotada por um glicano: por exemplo, se em um glicano os diedros ϕ forem iguais em cada uma das ligações glicosídicas, e o mesmo

ocorrer para os ψ , o glicano assume uma forma helicoidal. Em particular, se a soma ($\phi + \psi$) for igual a 120° o glicano assume uma conformação em hélice denominada 2_1 ^{57;94}, como aquela mostrada na Figura 5.5. Se a soma ($\phi + \psi$) for igual a 190° o glicano assume uma conformação em hélice 3_1 levógira, enquanto que se a soma ($\phi + \psi$) for igual a 50° o glicano assume uma conformação em hélice 3_1 dextrógira.^{57;95}

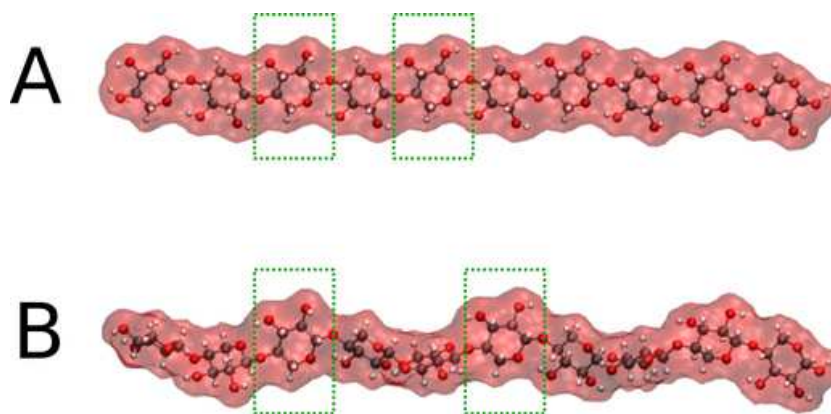


Figura 5.5: Cadeias de xilano assumindo conformações em hélices. A) Hélice 2_1 . B) Hélice 3_1 . A partir de um resíduo de xilose escolhido arbitrariamente, na conformação em hélice 2_1 é necessário andar dois resíduos para a frente para encontrar um resíduo no qual as posições atômicas sejam análogas ao do resíduo originalmente escolhido, diferindo apenas por uma translação ao longo do eixo paralelo ao sentido de crescimento da cadeia polimérica. Na conformação em hélice 3_1 é necessário andar três resíduos para a frente na cadeia de xilano para atingir a mesma condição.

A molécula de xilano M0, sobre a face hidrofílica (010), que praticamente não se afasta do seu sítio de adsorção durante as simulações, apresenta apenas conformações próximas a hélices 2_1 , como mostrado na Figura 5.6. Nessa figura, a tendência geral é que os máximos dos histogramas se alternam entre estarem localizados em valores de ($\phi + \psi$) ligeiramente maiores que 120° (entre 120° e 150°) e ligeiramente menores (entre 90° e 120°), o que é consistente com uma conformação próxima a uma hélice 2_1 , com exceção dos diedros glicosídicos nas extremidades, que apresentam tendências diferentes devido à maior mobilidade dos resíduos de xilose terminais.

Já a molécula de xilano M25, sobre a face hidrofílica (020), que é bastante móvel e se afasta de seu sítio de adsorção, indo em direção à fase aquosa em alguns momentos da simulação, apresenta uma coexistência de conformações similares às hélices 2_1 e também 3_1 , como pode-se observar na Figura 5.7. Nos primeiros quatro histogramas desta figura, que correspondem aos pares de diedros glicosídicos mais

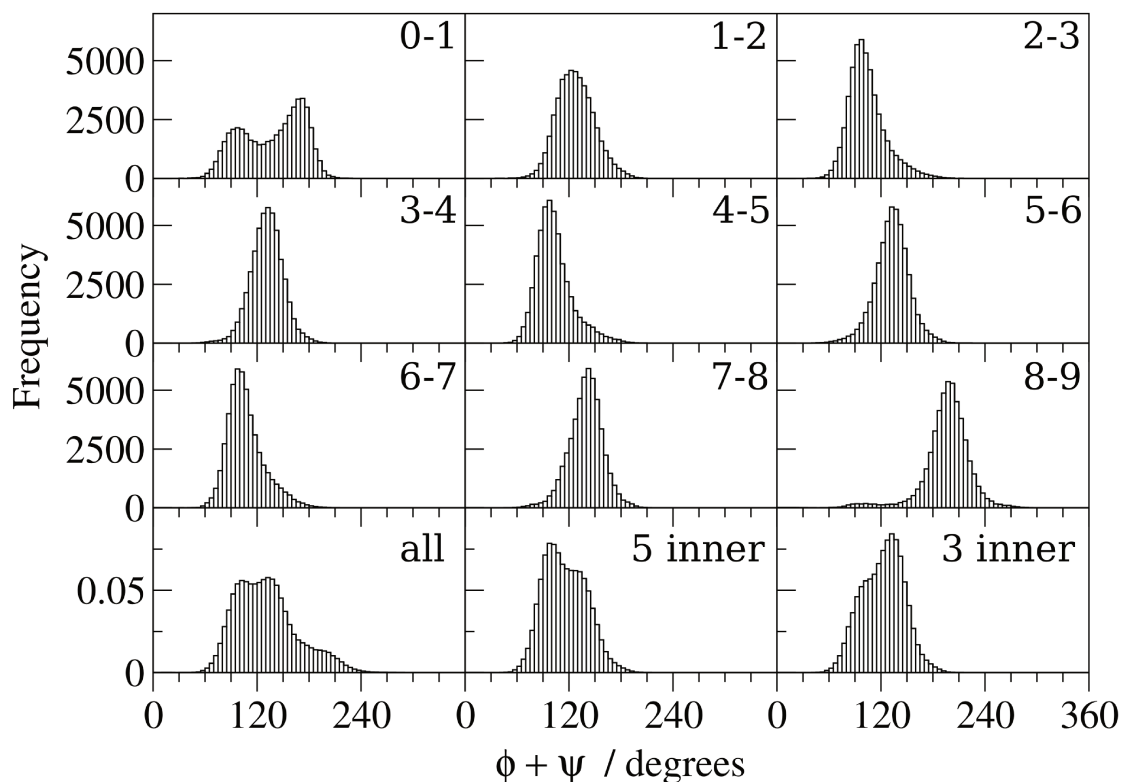


Figura 5.6: Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M0, sobre a face (010). Se a soma é igual a 120° então o par correspondente de resíduos de xilose está em uma conformação em hélice 2_1 , enquanto que se a soma é igual a 50° ou a 190° então o par correspondente de resíduos de xilose está em uma conformação em hélice 3_1 , dextrógira e levógira respectivamente. Os primeiros nove histogramas correspondem aos pares de diedros glicosídicos entre os resíduos 0 e 1, 1 e 2, e assim por diante, e cada histograma tem uma população total de 50 mil. Os últimos três histogramas abrangem mais de um par de diedros glicosídicos: i) todos ; ii) os cinco mais internos ; iii) os três mais internos. Para estes três histogramas, os valores na ordenada foram normalizados de maneira a indicar a frequência relativa com que ocorre o seu valor correspondente na abscissa.

próximos à extremidade não redutora e mais móvel do xilano M25 (vide Figura 5.2), os máximos ocorrem em, ou próximo a, 190° , o que é consistente com um trecho de xilano assumindo conformações em hélice 3_1 , que é justamente a conformação assumida por uma cadeia de xilano livre em água pura, como se vê na Figura 5.9. Além das populações próximas a 190° , os próximos cinco histogramas da Figura 5.7 também possuem máximos próximos a 120° , o que corresponde a conformações em hélice 2_1 , que é a conformação apresentada quando o xilano está adsorvido em seu sítio cristalográfico.

Na Figura 5.8 pode-se confirmar que, de fato, as conformações em hélice 2_1 apresentadas pelo xilano M25 sobre a face hidrofílica (020) correspondem a ins-

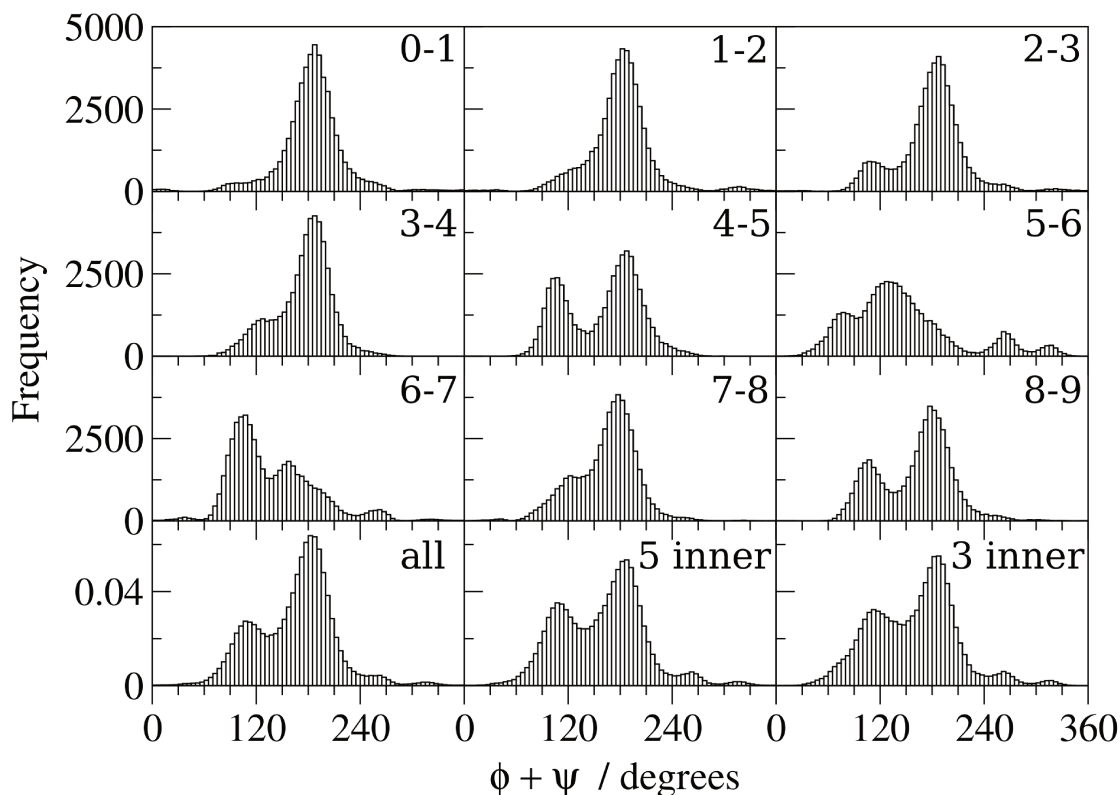


Figura 5.7: Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M25, sobre a face (020). Se a soma é igual a 120° então o par correspondente de resíduos de xilose está em uma conformação em hélice 2_1 , enquanto que se a soma é igual a 50° ou a 190° então o par correspondente de resíduos de xilose está em uma conformação em hélice 3_1 , dextrógira e levógira respectivamente. Os primeiros nove histogramas correspondem aos pares de diedros glicosídicos entre os resíduos 0 e 1, 1 e 2, e assim por diante, e cada histograma tem uma população total de 50 mil. Os últimos três histogramas abrangem mais de um par de diedros glicosídicos: i) todos ; ii) os cinco mais internos ; iii) os três mais internos. Para estes três histogramas, os valores na ordenada foram normalizados para que indiquem a freqüência relativa com que ocorre o seu valor correspondente na abscissa.

tantes nas simulações em que a cadeia de xilano está adsorvida em seu sítio cristalográfico, e que as conformações em hélice 3_1 correspondem a instantes nos quais a cadeia de xilano está mais distante da face (020). Nesta figura, a distância entre o oxigênio glicosídico que une os resíduos 3 e 4 do xilano M25 e o oxigênio glicosídico logo abaixo dele na cadeia de celulose M21 foi traçada em função da soma ($\phi + \psi$) para os diedros glicosídicos entre os mesmos resíduos 3 e 4 do xilano M25. Na figura, percebe-se que na grande maioria das vezes em que estes resíduos assumem conformações próximas a hélices 2_1 , os resíduos estão próximos ao seu sítio de adsorção na face (020). Quando este par de resíduos assume conformações próximas a hélices 3_1 , por outro lado, eles estão mais distantes do cristal de celulose.

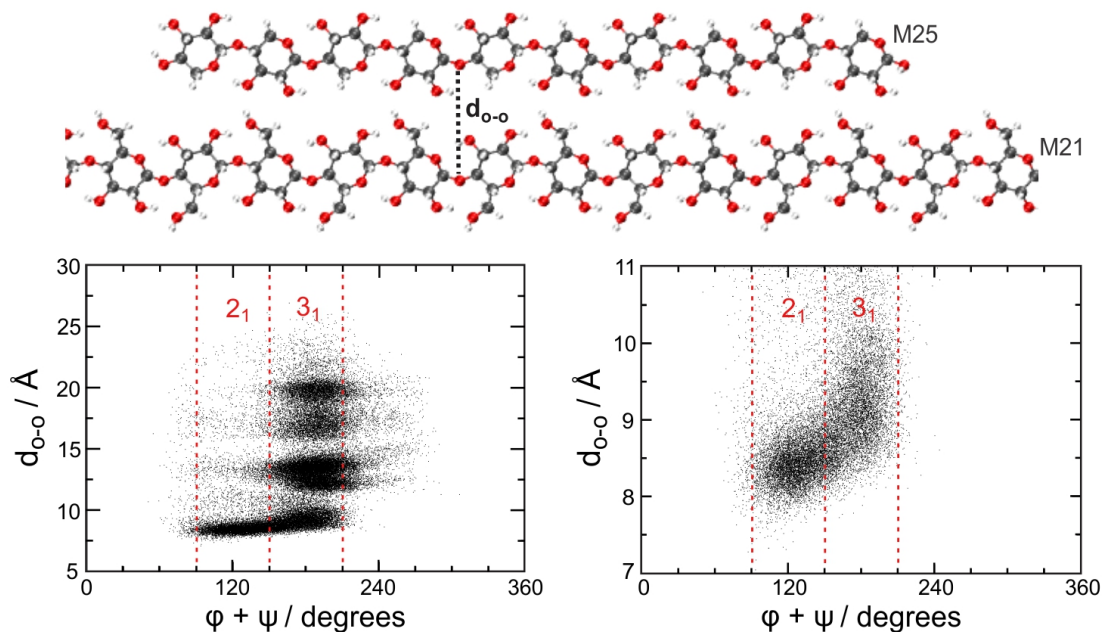


Figura 5.8: Distância entre o oxigênio glicosídico que une os resíduos 3 e 4 do xilano M25 e o oxigênio glicosídico logo abaixo dele na cadeia de celulose M21, em função da soma ($\phi + \psi$) para o par de resíduos 3 e 4. O vetor que une estes dois oxigênios corresponde à aresta b da cela unitária do cristal de celulose, e no cristal o seu módulo é 8.20 \AA .⁴⁷ Quando o par de resíduos 3 e 4 assume conformações similares a hélices 2_1 , estes resíduos estão próximos a seu sítio cristalográfico de adsorção, com $d_{O-O} \sim 8 \text{ \AA}$, mas quando assumem conformações similares a hélices 3_1 eles podem estar muito mais distantes. Não há um número significativo de instantes em que os resíduos estejam a distâncias grandes da face (020) ao mesmo tempo que assumem conformações em hélices 2_1 , e os poucos pontos esparsos em que isso ocorre correspondem a instantes em que o par de resíduos está transitando de ou para a conformação mais estável em água, que é a hélice 3_1 . Para os estados em que os resíduos apresentam conformação em hélice 3_1 e estão próximos à face (020), o detalhe à direita mostra que estes estados são mais distantes da face (020) do que os estados em que a conformação em hélice 2_1 é adotada.

Os resultados deste estudo mediante simulações de DM dos complexos de adsorção de xilano nas faces hidrofílicas (010) e (020) da celulose $I\beta$ demonstram, portanto, que está correta a hipótese levantada na Seção 1.5, de que as cadeias de xilano poderiam adsorver a superfícies da celulose exatamente nas posições cristalográficas em que outra cadeia de celulose ocuparia se o cristalito fosse mais extenso. Os resultados demonstram, entretanto, que há diferenças entre as dinâmicas do xilano adsorvido na face (010) e aquela do xilano adsorvido na face (020), já que na face (010) o xilano permanece estável em seu sítio de adsorção e assume exclusivamente conformações próximas a hélices 2_1 , enquanto o xilano na face (020) tem maior mobilidade e assume conformações similares a hélices 3_1 quando está

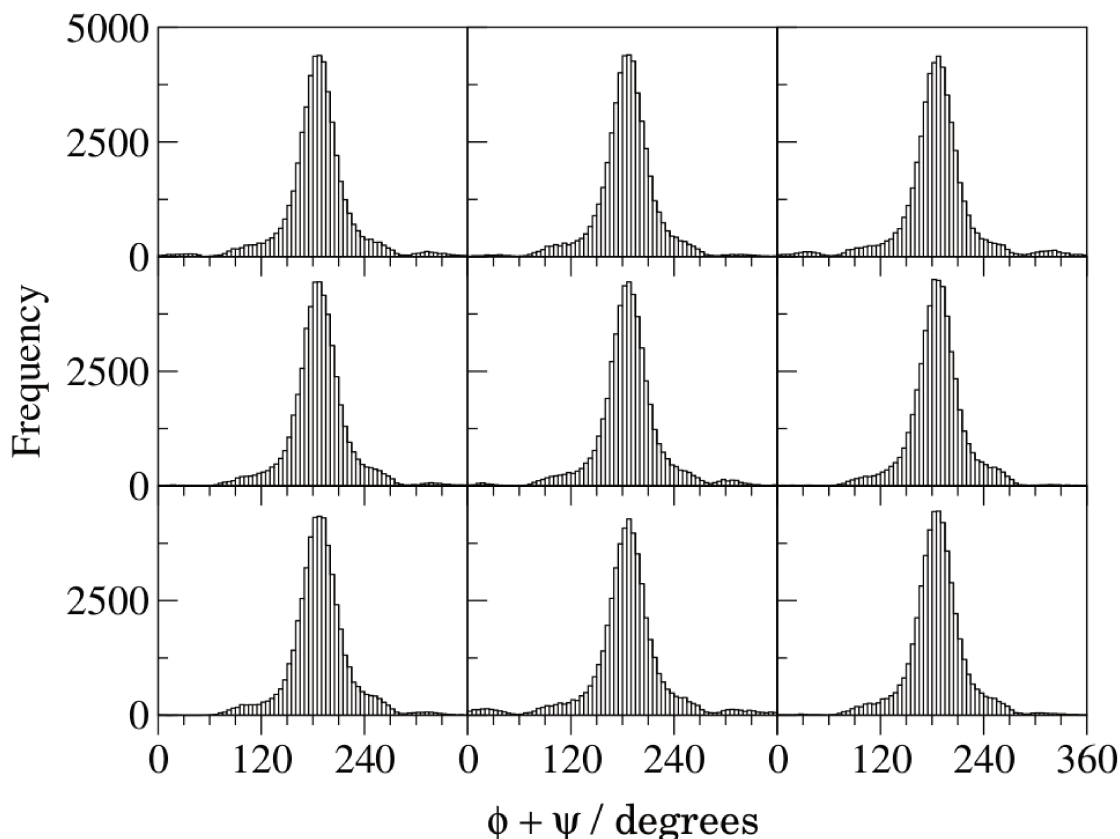


Figura 5.9: Distribuição da soma dos diedros glicosídicos ϕ e ψ dos nove pares de diedros glicosídicos em uma molécula de $\beta(1 \rightarrow 4)$ -xilano composta por dez unidades de D-xilose, em água pura. Os máximos de todos os histogramas ocorrem em $(\phi + \psi) = 190^\circ$, o que mostra que o xilano em água adota preferencialmente conformações em hélice 3_1 . Cada histograma tem uma população total de 50 mil.

longe desta face. Na seção a seguir esta mesma hipótese será aplicada a xilanos adsorvidos nas faces hidrofóbicas da celulose, e nas próximas seções estes estudos serão estendidos para alguns xilanos substituídos.

5.2 Xilanos adsorvidos sobre faces hidrofóbicas

Nesta seção serão apresentados os resultados das simulações por DM de xilanos adsorvidos às faces hidrofóbicas (100) e (200) do cristal de celulose cristalina $I\beta$ que foi representado à direita na Figura 5.1. A dinâmica das moléculas de xilano sobre estas faces pode ser visualizada na Figura 5.10. Nesta figura observa-se que os complexos formados nas duas faces são estáveis, mas o xilano M25 sobre a face (200) tem mobilidade maior. Assim como observado entre as duas faces hidrofílicas na seção anterior, notam-se diferenças entre as duas faces hidrofóbicas

da celulose no que se refere à dinâmica de moléculas de xilano adsorvidas sobre elas. Estas diferenças, inicialmente não esperadas para faces com propriedades tão semelhantes, seja para o par de faces hidrofílicas, seja para o par de faces hidrofóbicas, têm origem em fatores que serão explicados na Seção 5.3. Por ora, serão discutidas as diferenças encontradas entre as faces hidrofóbicas.

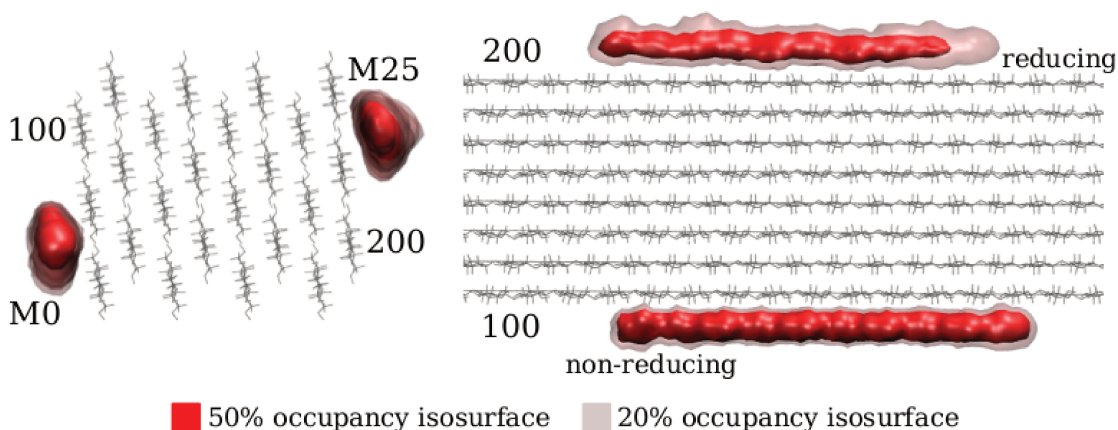


Figura 5.10: Representação estatística da dinâmica de moléculas de xilano adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$, utilizando superfícies de nível, ou *isosuperfícies*, de ocupação. O cristal de celulose está representado em cinza, na sua configuração inicial. As isosuperfícies em vermelho representam os pontos nos quais algum átomo da molécula de xilano está presente em 50% do tempo de simulação, enquanto as isosuperfícies transparentes em rosa são compostas pelos pontos no espaço em que algum átomo da molécula de xilano está presente em 20% do tempo de simulação. A molécula de xilano na face (100) se move menos em torno de seu sítio de adsorção, já a molécula de xilano na face (200) apresentou maior mobilidade ao longo de toda a cadeia, principalmente em resíduos próximos à sua extremidade redutora.

As diferenças na mobilidade das moléculas de xilano M0 e M25, adsorvidas sobre as faces (100) e (200), determinam diferenças no número de ligações de hidrogênio estabelecidas entre estas cadeias de xilano e o cristal de celulose durante as simulações. A Figura 5.11 ilustra a distribuição do número de ligações de hidrogênio estabelecidas nas faces (100) e (200). A rigor, se as moléculas de xilano mantivessem posições bastante próximas da sua configuração inicial, no sítio cristalográfico de adsorção, nas faces hidrofóbicas não deveria haver ligações de hidrogênio entre xilano e celulose, em decorrência da estrutura em camadas da celulose $I\beta$ (vide Figuras 1.16 e 4.3), na qual só há ligações de hidrogênio entre cadeias poliméricas em uma mesma camada, mas não entre camadas de cadeias *origin* e *center*. Entretanto, as moléculas de xilano nestas simulações são relativamente pequenas, têm dez xiloses, e percorrem pontos em seu espaço de configurações nos quais estabelecem ligações de hidrogênio com a celulose. As

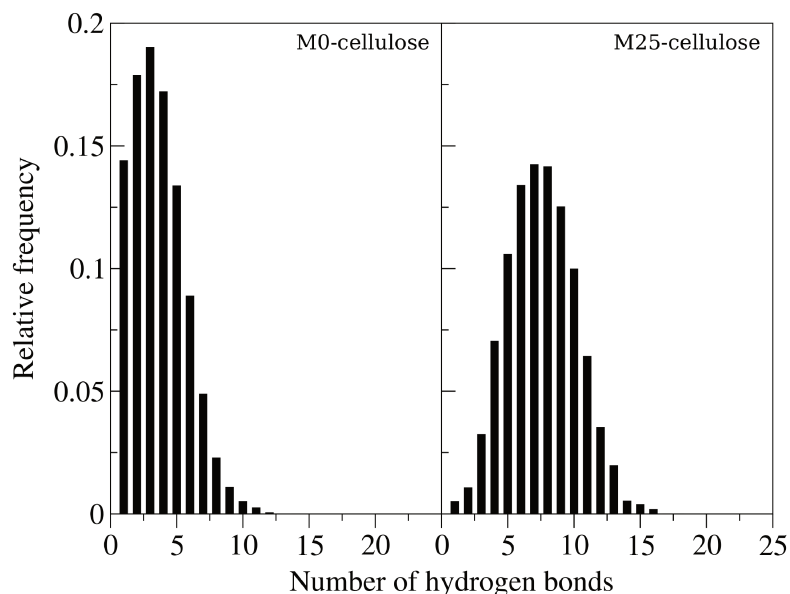


Figura 5.11: À esquerda: distribuição do número de ligações de hidrogênio entre a cadeia de xilano M0 e o cristal de celulose $I\beta$. À direita: distribuição do número de ligações de hidrogênio estabelecidas entre a cadeia de xilano M25 e a celulose. A molécula de xilano M0 estabelece, em média, menos ligações de hidrogênio com o cristal de celulose do que a molécula de xilano M25. Em todos os histogramas, os valores na ordenada foram normalizados para que indiquem a porcentagem do tempo de simulação em que ocorre o seu valor correspondente na abscissa.

diferenças de mobilidade entre as cadeias de xilano M0 e M25 também determinam diferenças nas conformações assumidas por estas cadeias de xilano. Na Figura 5.12 nota-se que o xilano M0, que se apresenta menor mobilidade, adota conformações em hélice 2_1 , o que é consistente com uma cadeia de xilano que se mantém próxima ao seu sítio de adsorção na face (010). Na figura observa-se que, exceto para o par de resíduos 8 e 9 na extremidade redutora, que são mais sujeitos a efeitos de borda, a soma dos diedros glicosídicos ($\phi + \psi$) assume valores próximos a 120° , o que corresponde a conformações em hélice 2_1 . A cadeia de xilano M25, que é mais móvel durante as simulações, e se afasta mais frequentemente de seu sítio na face (020), pode apresentar conformações em hélice 2_1 e em hélice 3_1 , como nota-se na Figura 5.13. Nesta figura observa-se que para todos os pares de resíduos da cadeia de xilano M25 os histogramas correspondentes apresentam máximos próximos a 190° para o valor da soma ($\phi + \psi$), o que corresponde a conformações em hélice 3_1 . Na maioria dos histogramas, os máximos próximos a 190° são inclusive mais altos do que máximos apresentados próximos a 120° , que por sua vez correspondem a conformações em hélice 2_1 .

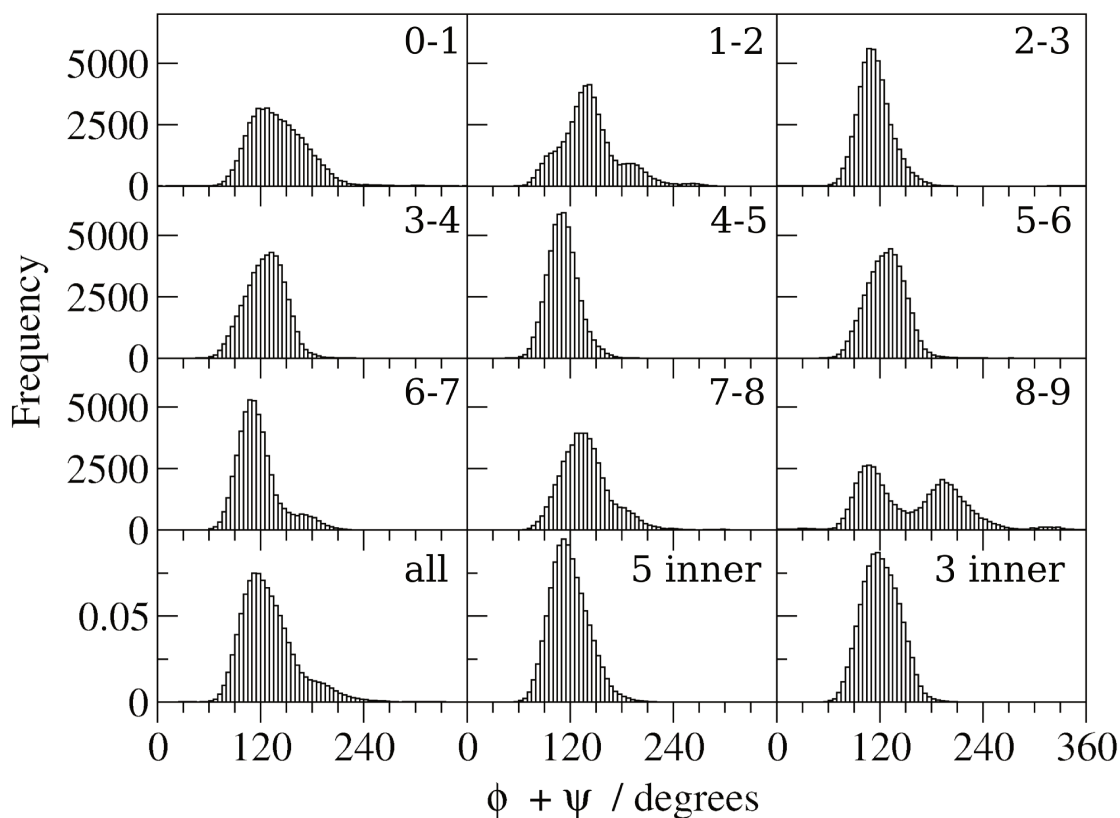


Figura 5.12: Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M0, sobre a face (100). Os primeiros nove histogramas correspondem aos pares de diedros glicosídicos entre os resíduos 0 e 1, 1 e 2, e assim por diante, e cada histograma tem uma população total de 50 mil. Os últimos três histogramas abrangem mais de um par de diedros glicosídicos: i) todos ; ii) os cinco mais internos ; iii) os três mais internos. Para estes três histogramas, os valores na ordenada foram normalizados de maneira a indicar a frequência relativa com que ocorre o seu valor correspondente na abscissa.

Em suma, os complexos de adsorção formados por xilanos sobre as quatro faces cristalográficas da celulose $I\beta$ estudadas neste trabalho são estáveis durante as simulações por DM que foram realizadas, indicando que é plausível supor que na parede celular vegetal os xilanos interajam com a celulose cristalina mediante processos de adsorção física que ocorrem nos próprios sítios cristalográficos que seriam ocupados por uma próxima cadeia de celulose em um cristal de celulose mais extenso. As simulações mostram, entretanto, que ocorrem diferenças não antecipadas entre as duas faces hidrofílicas e também entre as duas faces hidrofóbicas estudadas. As causas destas diferenças foram investigadas e serão discutidas na seção a seguir. Contudo, os resultados nos motivaram a estender a investigação desta hipótese de adsorção em sítios cristalográficos para determinar se xilanos substituídos, os mais comuns em vegetais superiores, também poderiam sofrer ad-

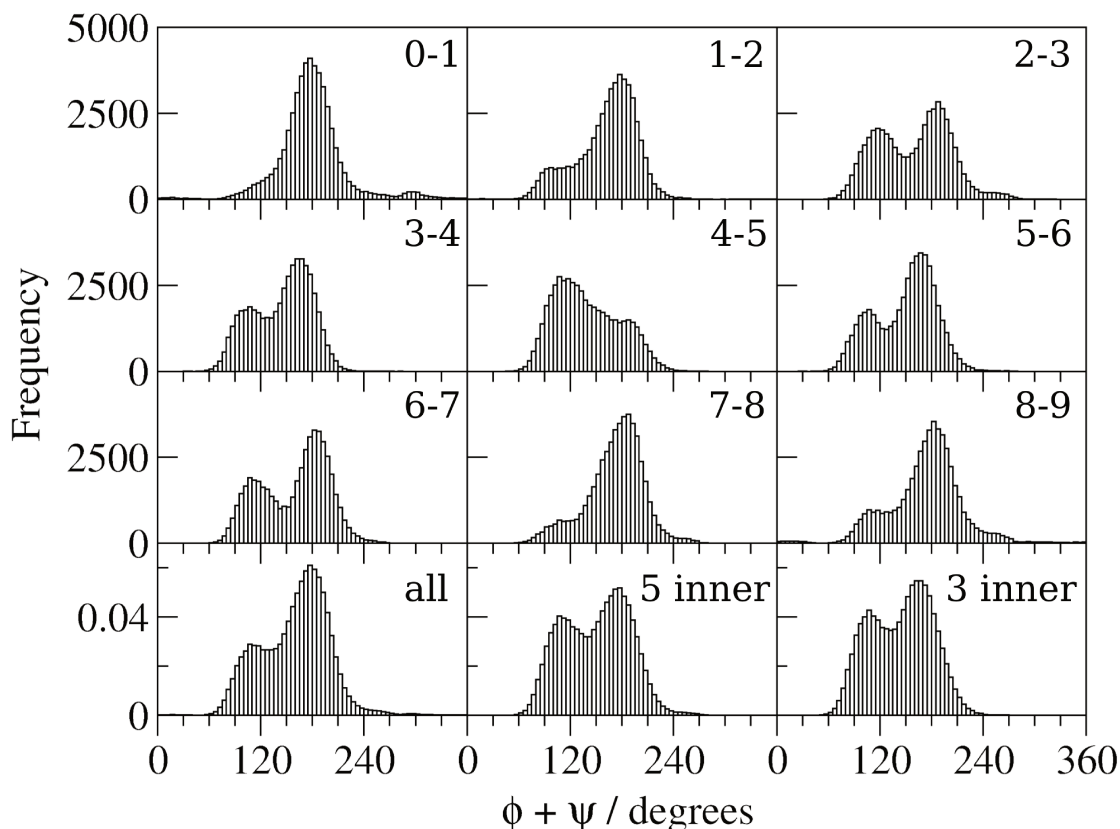


Figura 5.13: Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas na molécula de xilano M25, sobre a face (200). Os primeiros nove histogramas correspondem aos pares de diedros glicosídicos entre os resíduos 0 e 1, 1 e 2, e assim por diante, e cada histograma tem uma população total de 50 mil. Os últimos três histogramas abrangem mais de um par de diedros glicosídicos: i) todos ; ii) os cinco mais internos ; iii) os três mais internos. Para estes três histogramas, os valores na ordenada foram normalizados de maneira a indicar a freqüência relativa com que ocorre o seu valor correspondente na abscissa.

sorção às faces da celulose $I\beta$ desta mesma maneira. Os resultados das simulações com xilanos substituídos são apresentadas na Seção 5.4.

5.3 O cristal de celulose

Embora o foco desta terceira parte do trabalho de tese seja a investigação de xilano adsorvido a diferentes faces cristalográficas da celulose $I\beta$, a investigação da dinâmica dos cristais de celulose em si também mostrou-se bastante instrutiva e serviu para elucidar as diferenças não antecipadas, mencionadas nas duas seções anteriores, observadas entre faces cristalográficas semelhantes dos cristais de celulose estudados. Estas diferenças acontecem devido à uma distorção na estrutura

dos cristais, que ocorre durante as simulações. A Figura 5.14 mostra instantes aleatórios das simulações realizadas, e eles representam a distorção que ocorre na estrutura cristalina da celulose I β . Já na etapa de equilíbrio, e durante todas as simulações, os cristais de celulose sofrem uma distorção em sua geometria, na qual as cadeias *origin* desviam consideravelmente de suas posições cristalográficas. As cadeias *center*, entretanto, mantêm geometrias próximas às reportadas experimentalmente⁴⁷ para o alomorfo I β da celulose cristalina. Esta diferença fica mais clara ao comparar a Figura 5.14, que corresponde às simulações por DM, com as Figuras 1.16 ou 4.3, que correspondem às estruturas determinadas experimentalmente.

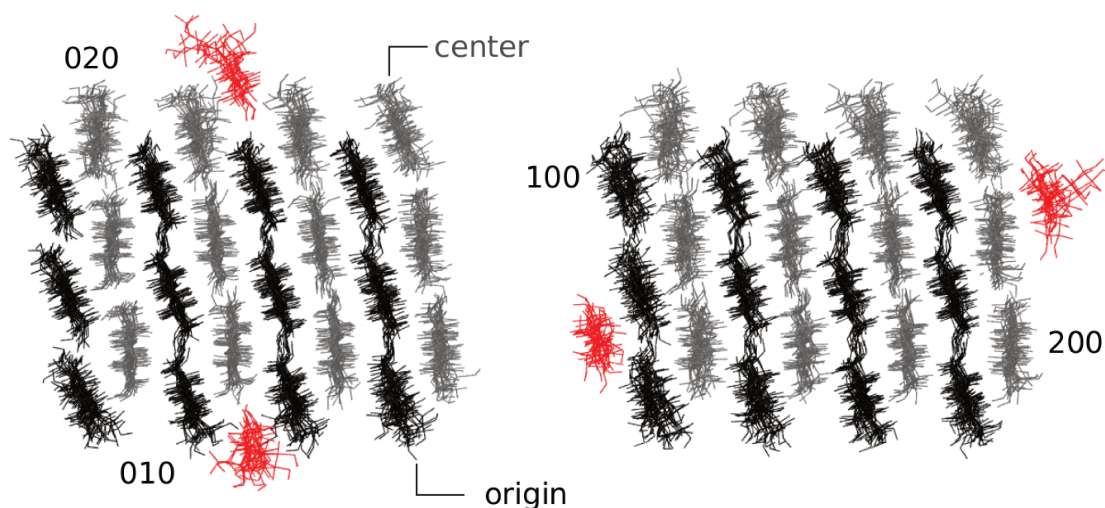


Figura 5.14: Representação da distorção na geometria dos cristais de celulose I β , que ocorre durante as simulações por DM. Cada imagem corresponde a um único *frame* da simulação, tomado aleatoriamente. As cadeias do tipo *origin*, em preto, tornam-se sistematicamente mais inclinadas à esquerda do que de fato estariam em um cristal real. As cadeias do tipo *center*, em cinza, mantêm-se próximas às suas posições cristalográficas, exceto as da face (020), que são mais móveis durante as simulações, em ambos os sistemas representados, devido a efeito de borda. Em vermelho, as cadeias de xilano.

Esta distorção provoca alterações significativas na rede de ligações de hidrogênio estabelecidas dentro e na superfície do cristal durante as simulações, uma vez que a distorção resulta na formação de ligações de hidrogênio entre camadas de cadeias *origin* e *center*, e experimentalmente as ligações de hidrogênio só ocorrem dentro de cada camada. A distorção ocasiona alterações não só dentro do próprio cristal, como também nas interações com as cadeias de xilano adsorvidas sobre as diversas faces. Uma forma de descrever as alterações dentro do cristal é monitorar as conformações dos grupos exocíclicos hidroximetil, correspondentes ao carbono C6 em cada anel de glicose das cadeias de celulose. As conformações do grupo

hidroximetil em piranoses foram apresentadas na Figura 1.8. A Figura 5.15 é um painel que mostra a distribuição destas conformações em cada cadeia de celulose durante as simulações, e os resultados reportados nessa figura de fato mostram uma diferença entre cadeias *origin* e *center*.

A diferença observada entre as cadeias *origin* e *center* se deve, em parte, às condições periódicas de contorno impostas aos sistemas estudados: para minimizar efeitos de borda, os cristais de celulose simulados são infinitos na direção de crescimento das cadeias de celulose, com cada cadeia de celulose covalentemente ligada às suas imagens periódicas além das extremidades da caixa de simulação. O tamanho da caixa de simulação na direção das cadeias (a direção cristalográfica c) foi mantido fixo, com valor igual a um múltiplo inteiro do tamanho, determinado experimentalmente⁴⁷, da cela unitária da celulose $I\beta$ nesta mesma direção.⁴⁷ A configuração inicial dos cristais de celulose nas simulações corresponde, portanto, à estrutura cristalográfica determinada experimentalmente. Entretanto, como nenhum campo de força é capaz de descrever perfeitamente todas as propriedades de um sistema qualquer, ocorre que, devido às condições periódicas de contorno impostas, o sistema fica sob tensão. Esta tensão ocorre porque, com o campo de força utilizado nas simulações, as cadeias de celulose seriam mais estáveis se a distância entre seus monômeros fosse ligeiramente diferente daquela experimental e, portanto, se o comprimento da cadeia inteira fosse um pouco diferente. Desvios nos parâmetros de cela são comuns em simulações por DM de cristais e, em simulações feitas para a celulose $I\beta$, Himmel *et al.* obtiveram o valor de 10.512Å para o comprimento da cela unitária na direção cristalográfica c , contra 10.380Å determinados experimentalmente⁴⁷, utilizando o campo de força para carboidratos desenvolvido por Kuttel, Brady e Naidoo⁹⁶. Por menor que seja a diferença entre a distância experimental entre resíduos adjacentes e a distância mais estável descrita pelo campo de força, este efeito se amplificará a cada monômero presente na cadeia e quanto maior for o polímero, maior será a tensão a que a cadeia polimérica estará submetida em simulações por DM se estiver confinada ao comprimento experimental, diferente daquele descrito como mais estável, ou relaxado, pelo campo de força utilizado. Como as cadeias de celulose não podem ajustar seu comprimento por causa das condições periódicas de contorno impostas, o sistema relaxa da maneira que pode ao distorcer a geometria das cadeias *origin*. Não está claro se esta diferença entre cadeias *origin* e *center* se deve apenas aos efeitos descritos acima, ou se é uma diferença a ser esperada também experimentalmente para um cristalito real.

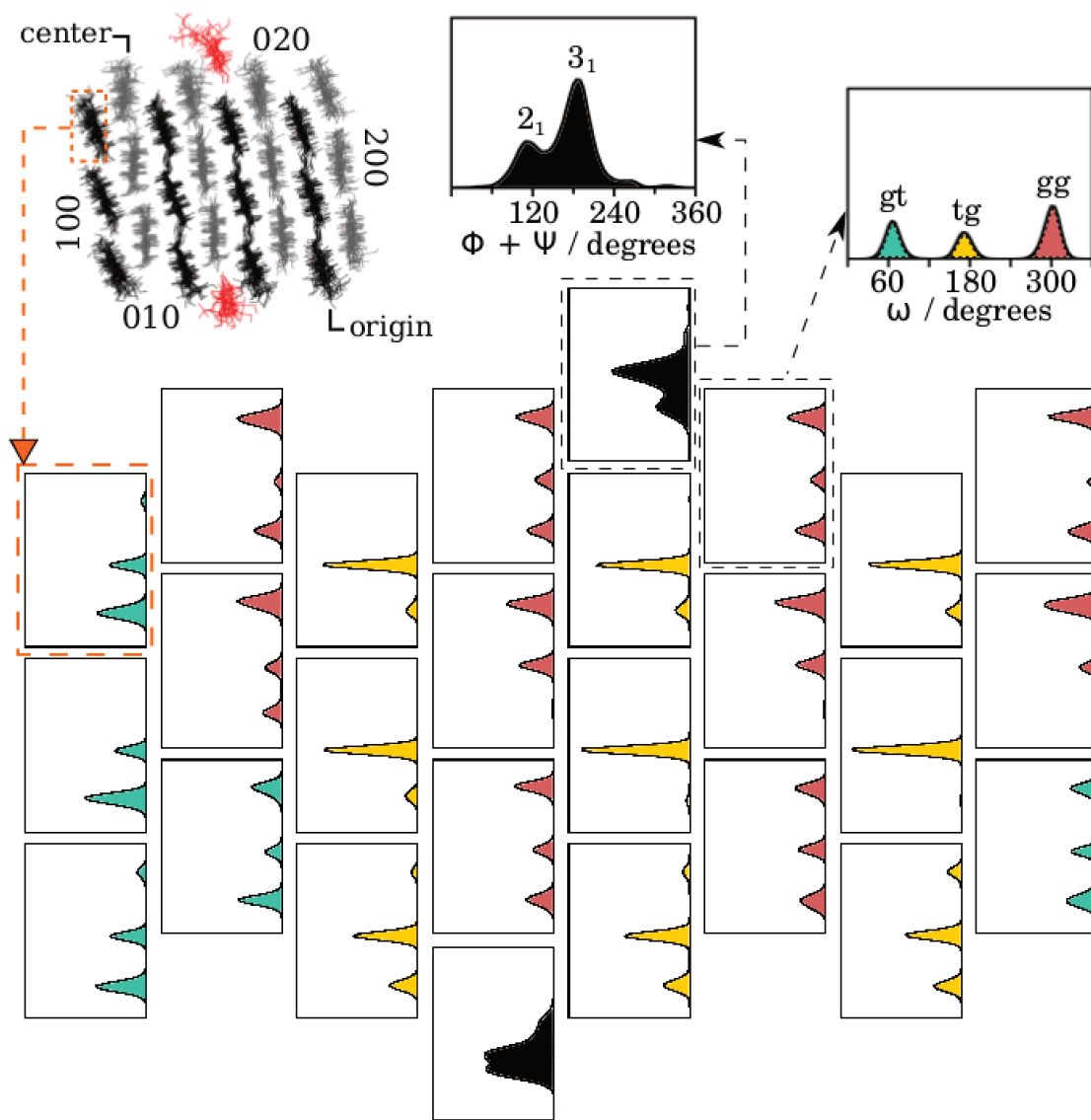


Figura 5.15: Pannel que ilustra histogramas que representam a distribuição das conformações do grupo exocíclico hidroximetil, correspondente ao carbono C6 nos anéis de glicose (vide Figura 1.8), em cada uma das 24 cadeias de celulose presentes no cristal. Cada histograma se refere a uma cadeia de celulose, e representa a distribuição obtida para a união dos conjuntos de dados de todos os vinte anéis de glicose presentes em uma dada cadeia. A área abaixo da curva em cada histograma foi colorida de acordo com a conformação cuja população foi maior naquela cadeia, durante as simulações: histogramas em ciano correspondem a cadeias de celulose em que a conformação predominante é *gt*, em amarelo *tg* e em vermelho *gg*. Apesar da distorção que sofrem em sua geometria, as cadeias *origin* mantêm a conformação *tg* observada experimentalmente⁴⁷, exceto cadeias *origin* na face (100), em que predominam conformações *gt*. Apesar de apresentarem predominantemente a conformação *tg* observada experimentalmente, as cadeias *origin* na face (010) adotam outras conformações mais freqüentemente, devido a efeitos de borda, se comparadas a cadeias *origin* no interior do cristal. Nas cadeias *center* predominam as conformações *gg*, diferentemente das *tg* encontradas experimentalmente, embora em duas das cadeias *center* as conformações *gt* sejam predominantes. Também para as cadeias *center* pode-se notar certos efeitos de borda: nos histogramas das cadeias mais internas a população de conformações *gg* é maior e quase não há conformações *gt*. A xilose não possui carbono C6, e os histogramas correspondentes às cadeias de xilano, em preto, representam a distribuição da soma dos diedros ($\phi + \psi$).

Esta distorção é a principal origem das diferenças observadas entre cadeias *origin* e *center* em nossas simulações e, por extensão, entre as faces cristalográficas constituídas por estes dois tipos de cadeia: as faces (010) e (020) são ambas hidrofílicas, mas a primeira é constituída por cadeias *origin* e a segunda por cadeias *center*. Considerando a estrutura cristalográfica experimental, estas duas faces deveriam ter propriedades muito semelhantes, e as diferenças observadas nas simulações devem-se à descrição inexata da geometria das cadeias *origin* fornecida pelo campo de força utilizado. O mesmo se aplica ao par de faces hidrofóbicas (100) e (200).

Apesar das distorções que ocorrem na geometria das cadeias *origin*, o modelo computacional ainda assim prevê dois tipos de cadeias diferentes, em acordo com o que é observado experimentalmente, embora as diferenças sejam exacerbadas pelo modelo computacional. Estas diferenças não invalidam o fato que as simulações mostram que a adsorção de cadeias de xilano com conformação em hélice 2_1 a seus sítios cristalográficos nas diversas faces da celulose é uma hipótese plausível. As simulações claramente indicam que esse tipo de complexo de adsorção é viável e estável, e o mecanismo de adsorção às faces hidrofílicas, por meio de ligações de hidrogênio (Tabelas 5.1 e 5.2), foi caracterizado, qualitativamente ao menos.

As condições periódicas de contorno que causam esta distorção poderiam ser removidas, mas ao simular um cristalito de celulose finito a estrutura do cristal sofre distorções ainda mais graves, apresentando uma torção que abrange toda a fibrila de celulose, inclusive para cristalitos com quarenta unidades de glicose em cada cadeia de celulose.^{50;51} Contudo, o campo de força utilizado nas simulações descritas nesta tese parece ser o mais adequado entre os campos de força mais modernos desenvolvidos para glicanos^{72;73;97;98}, considerando a comparação entre três campos de força realizada por Himmel *et al.* em simulações de cristais celulose $I\beta$.⁵¹ Outra alternativa para contornar este problema é discutida na Seção 6.1.

Embora o modelo computacional apresentado tenha suas limitações, ele ainda permite testar em xilanos substituídos a mesma hipótese verificada nas seções anteriores para xilanos simples. Na seção a seguir são apresentados resultados de simulações por DM de alguns xilanos substituídos adsorvidos sobre celulose $I\beta$.

5.4 Xilanos substituídos

Nesta última seção, serão apresentados alguns resultados de simulações por DM de xilanos substituídos adsorvidos sobre as faces cristalográficas da celulose

$I\beta$. A derivatização química de glicanos mediante adição de grupos químicos substituintes às suas hidroxilas é bastante usual, e muitos compostos de uso corriqueiro contêm este tipo de glicano derivatizado. Os exemplos mais comuns são o acetato de celulose, os xantanos (ou viscose) utilizados por exemplo na produção industrial de rayon e celofane, a nitrocelulose, entre outros. As substituições em glicanos em geral ocorrem em suas hidroxilas e no caso de xilanos as hidroxilas disponíveis para substituição são aquelas correspondentes aos oxigênios O2 e O3 nos anéis de xilose. Estas são as mesmas hidroxilas que participam de ligações de hidrogênio entre xilano e celulose, como foi mostrado na Figura 5.3 e nas Tabelas 5.1 e 5.2. Portanto, pode-se esperar que a presença de substituintes interfira na forma como o xilano interage com a celulose, dependendo do padrão com que os substituintes estão dispostos ao longo da cadeia de xilano.

Dois grupos substituintes muito comuns em xilanos presentes em parede celular vegetal são os grupos acetil e glucuronoil. A substituição por acetil envolve a formação de uma ligação éster entre a carbonila do grupo acetil e o oxigênio O2, ou O3, de um anel de xilose. Já o grupo glucuronoil é um resíduo de ácido glucurônico (GlcA) [†] ligado ao oxigênio O2 ou ao O3 de um anel de xilose, mediante ligações glicosídicas $\beta(1 \rightarrow 2)$ ou $\beta(1 \rightarrow 3)$. Investigando a composição de xilanos em parede celular de plantas, Dupree *et al.* descobriram que as substituições glucuronoil em glucuronoxilanos de parede celular exibem um padrão em que as substituições ocorrem majoritariamente com espaçamento par entre elas.⁵⁸⁻⁶⁰ Dados ainda não publicados dos mesmos pesquisadores indicam que em acetilxilanos as acetilações ocorrem em aproximadamente 50% dos resíduos de xilose, e que também obedecem o padrão de espaçamento par e, portanto, estão alternadamente presentes ou ausentes em resíduos de xilose adjacentes. Ou seja, se um dado resíduo de xilose de uma cadeia de acetilxilano é substituído, o próximo resíduo não será substituído, o seguinte será substituído e assim por diante. Xilanos com este padrão de substituição foram mostrados na Figura 1.20 adsorvidos a faces cristalográficas da celulose $I\beta$, e os sistemas representados naquela figura foram investigados mediante simulações por DM nesta última parte do trabalho de tese.

Embora a função desempenhada pelos substituintes em xilanos ainda não seja bem compreendida, a existência de um padrão par no espaçamento entre os substituintes sugere que haja uma razão para esta disposição tão específica dos substituintes ao longo da cadeia de xilano. A Figura 5.16 ilustra representações da

[†]O ácido glucurônico é um monossacarídeo derivado da glicose, com o carbono exocíclico C6 oxidado a ácido carboxílico.

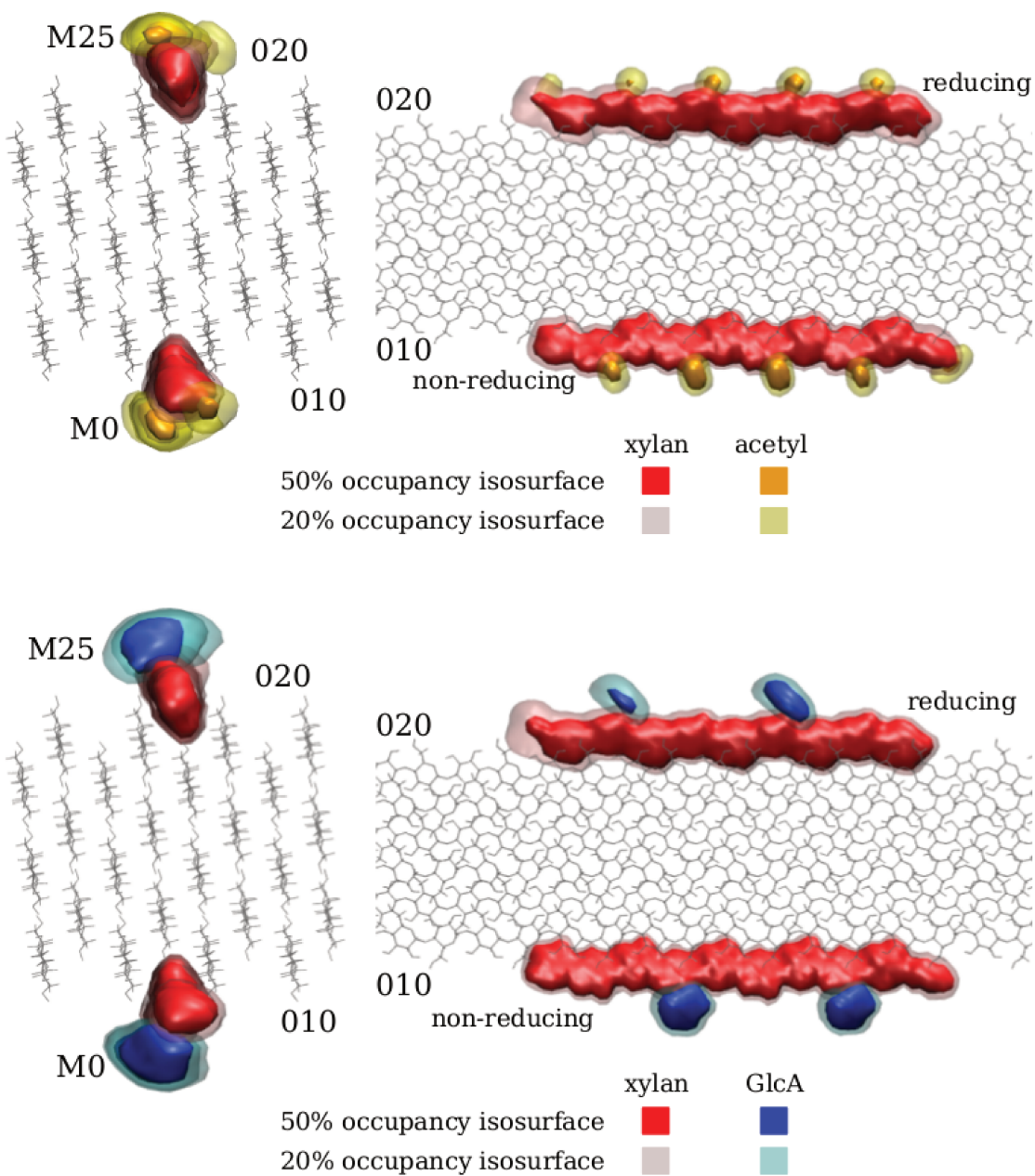


Figura 5.16: Acima: representação da dinâmica de moléculas de acetilxilano adsorvidas sobre as faces hidrofílicas da celulose $I\beta$, utilizando isosuperfícies de ocupação. O cristal de celulose está representado em cinza, na sua configuração inicial. As isosuperfícies em vermelho e em laranja representam os pontos nos quais algum átomo da molécula está presente em 50% do tempo de simulação, enquanto as isosuperfícies transparentes em rosa e em amarelo são compostas pelos pontos no espaço em que algum átomo da molécula está presente em 20% do tempo de simulação. As cores vermelho e rosa referem-se à cadeia principal, laranja e amarelo aos substituintes acetil. Abaixo: representação da dinâmica de moléculas de glucuronoxilano adsorvidas sobre as faces hidrofílicas da celulose $I\beta$, utilizando isosuperfícies de ocupação. O cristal de celulose está representado em cinza, na sua configuração inicial. As isosuperfícies em vermelho e em azul representam os pontos nos quais algum átomo da molécula está presente em 50% do tempo de simulação, enquanto as isosuperfícies transparentes em rosa e em ciano são compostas pelos pontos no espaço em que algum átomo da molécula está presente em 20% do tempo de simulação. As cores vermelho e rosa referem-se à cadeia principal, azul e ciano aos substituintes glucuronóil.

dinâmica das moléculas de xilano com padrão de substituição par adsorvidas sobre as faces hidrofílicas da celulose I β . Como também já foi observado para o caso de xilanos simples, nesta figura nota-se que os complexos de adsorção formados na face (010) são menos móveis, mantendo-se mais próximos à sua configuração inicial durante as simulações, enquanto os xilanos substituídos adsorvidos à face (020) apresentaram maior mobilidade, em especial nos resíduos mais próximos de sua extremidade não redutora. Como discutido na Seção 5.3, as diferenças observadas entre faces com mesma hidrofílicidade podem ser atribuídas às distorções na geometria do cristal de celulose observada durante as simulações.

O padrão de ligações de hidrogênio estabelecidas entre xilanos substituídos e as faces hidrofílicas da celulose é muito similar ao padrão que foi encontrado nas Tabelas 5.1 e 5.2 e na Figura 5.3. De fato, as substituições não afetam as hidroxilas envolvidas em ligações de hidrogênio com a celulose, pois os substituintes encontram-se apenas em resíduos de xilose cujos oxigênios O2 e O3 apontam para longe do cristal de celulose. Isto decorre do padrão par de substituição nos xilanos, pois se os substituintes estivessem aleatoriamente presentes a formação deste tipo de complexo de adsorção com as faces hidrofílicas da celulose não seria possível. Estas observações sugerem que o padrão par de substituição, encontrado experimentalmente em xilanos presentes em plantas, seja uma maneira de permitir a funcionalização das cadeias de xilano sem comprometer a sua capacidade de adsorção às faces hidrofílicas da celulose.

Para caracterizar a dinâmica destes xilanos substituídos sobre as faces da celulose em termos de suas conformações, foram construídos histogramas para a soma ($\phi + \psi$), análogos aos mostrados em figuras anteriores. Assim como ocorre para xilanos simples, quando as cadeias de xilano substituído se afastam de seu sítio de adsorção em direção à água eles adotam conformações em hélice 3_1 , mas quando estão no sítio de adsorção sua conformação é próxima a hélices 2_1 . Nota-se na Figura 5.17 que conformações consistentes com ambos os tipos de hélice ocorrem durante as simulações. Para o acetilxilano M0 sobre a face hidrofílica (010), os picos correspondentes a conformações similares a hélices 2_1 ocorrem à esquerda e à direita da marca dos 120° , mais afastados deste valor se comparados aos dois picos análogos em xilanos simples. O mesmo se aplica à cadeia M25 sobre a outra face hidrofílica, a (020). Para ambas as cadeias M0 e M25 há também populações próximas a 190° , correspondentes a conformações em hélice 3_1 .

Complexos de adsorção de xilanos substituídos sobre as faces hidrofóbicas da celulose I β também foram examinados. A Figura 5.18 ilustra representações da

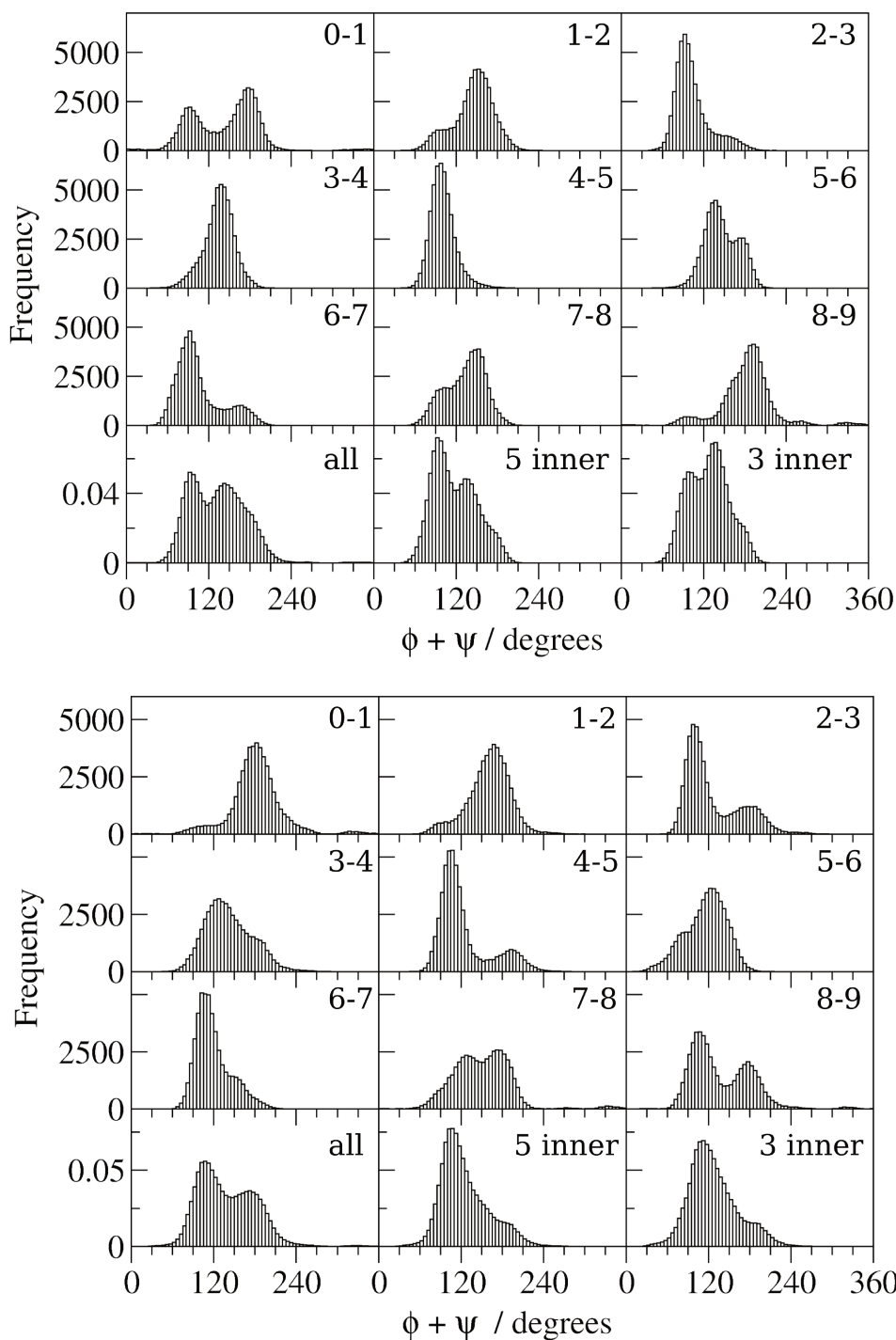


Figura 5.17: Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas nas moléculas de acetilxilano adsorvidas sobre as faces hidrofílicas da celulose I β . Acima: acetilxilano M0 sobre a face (010). Abaixo: acetilxilano M25 sobre a face (020). Para cada cadeia de acetilxilano os primeiros nove histogramas correspondem aos pares de diedros glicosídicos entre os resíduos de xilose 0 e 1, 1 e 2, e assim por diante, e cada histograma tem uma população total de 50 mil. Os últimos três histogramas abrangem mais de um par de diedros glicosídicos: i) todos; ii) os cinco mais internos; iii) os três mais internos. Para estes três histogramas, os valores na ordenada foram normalizados de maneira a indicar a frequência relativa com que ocorre o seu valor correspondente na abscissa.

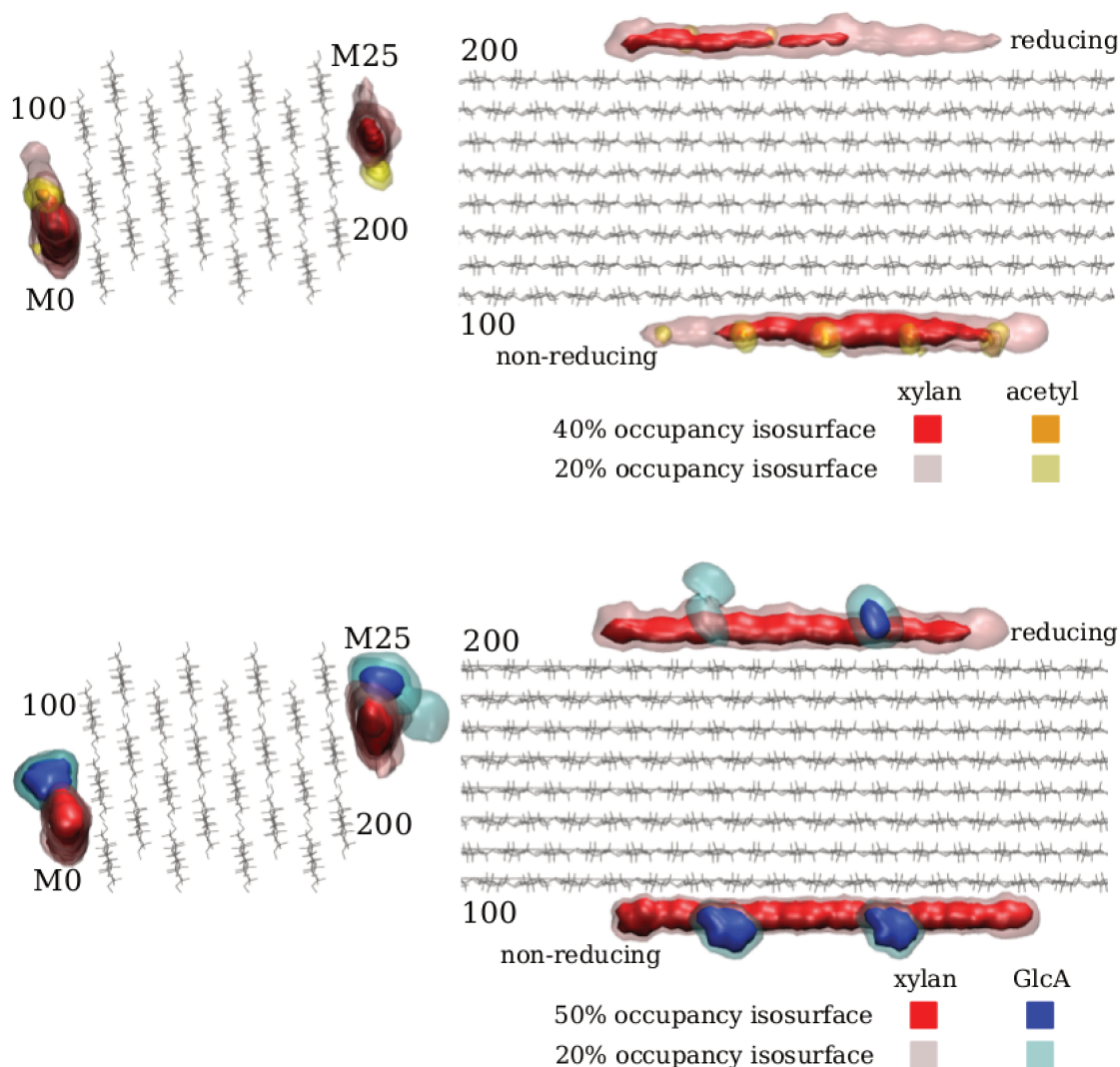


Figura 5.18: Acima: representação da dinâmica de moléculas de acetilxilano adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$, utilizando isosuperfícies de ocupação. O cristal de celulose está representado em cinza, na sua configuração inicial. As isosuperfícies em vermelho e em laranja representam os pontos nos quais algum átomo da molécula está presente em 40% do tempo de simulação, enquanto as isosuperfícies transparentes em rosa e em amarelo são compostas pelos pontos no espaço em que algum átomo da molécula está presente em 20% do tempo de simulação. As cores vermelho e rosa referem-se à cadeia principal, laranja e amarelo aos substituintes acetil. Abaixo: representação da dinâmica de moléculas de glucuronoxilano adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$, utilizando isosuperfícies de ocupação. O cristal de celulose está representado em cinza, na sua configuração inicial. As isosuperfícies em vermelho e em azul representam os pontos nos quais algum átomo da molécula está presente em 50% do tempo de simulação, enquanto as isosuperfícies transparentes em rosa e em ciano são compostas pelos pontos no espaço em que algum átomo da molécula está presente em 20% do tempo de simulação. As cores vermelho e rosa referem-se à cadeia principal, azul e ciano aos substituintes glucuronil.

dinâmica das moléculas de xilano com padrão de substituição par adsorvidas sobre as faces hidrofóbicas da celulose $I\beta$. Nesta figura observa-se que, durante as simulações, as cadeias de xilano substituído sobre a face hidrofóbica (100) são menos móveis do que aquelas sobre a face hidrofóbica (200), tanto para acetilxilanos como para glucuronoxilanos. Ambas as cadeias M25 apresentaram maior mobilidade em resíduos próximos à sua extremidade redutora. Esta é a mesma tendência que foi observada para xilanos simples. No caso de acetilxilanos sobre faces hidrofóbicas, ambas as cadeias M0 e M25 são significativamente mais móveis do que seus análogos não substituídos. Entretanto, os resultados demonstram que a formação de complexos de adsorção de xilanos substituídos sobre as faces hidrofóbicas da celulose $I\beta$ é factível se o padrão de substituição for par. Contudo, no caso de faces hidrofóbicas estes resultados não permitem descartar o padrão ímpar, que em princípio também seria viável nestas faces. Na Figura 5.19 observa-se que ambas as moléculas de acetilxilano M0 e M25, que são bastante móveis, adotam conformações similares tanto a hélices 2_1 como a hélices 3_1 , e que as hélices 3_1 predominam. Comparando com os resultados para os análogos não substituídos nas Figuras 5.12 e 5.13, nota-se que esta tendência é igual à observada para o xilano simples M25 na face hidrofóbica (200), mas não para o xilano simples M0 sobre a face hidrofóbica (100), o qual teve menor mobilidade e apresentou conformações em hélice 3_1 apenas nas extremidades da cadeia.

É possível afirmar, portanto, que a presença de substituintes glucuronil dispostos em padrão par provoca poucas alterações na dinâmica dos glucuronoxilanos adsorvidos sobre as faces cristalográficas da celulose $I\beta$, quando comparados aos xilanos não substituídos. Esta conclusão também é verdadeira para acetilxilanos adsorvidos à faces hidrofílicas, mas acetilxilanos em faces hidrofóbicas apresentaram mobilidade sensivelmente superior àquela de seus análogos não substituídos.

Contudo, estas simulações por DM mostram que xilanos substituídos com o padrão de espaçamento par entre os substituintes também formam complexos de adsorção estáveis com as faces cristalográficas da celulose $I\beta$, assim como foi observado nas simulações com xilanos não substituídos. O padrão par de substituição permite à molécula de xilano estabelecer o mesmo tipo de interação com a celulose cristalina, com o mesmo padrão de ligações de hidrogênio nas faces hidrofílicas, que é observado para xilano não substituído durante as simulações. Mais do que isso, este padrão de substituição permite a funcionalização química das cadeias de xilano, ao mesmo tempo em que mantém as características estruturais que permitem

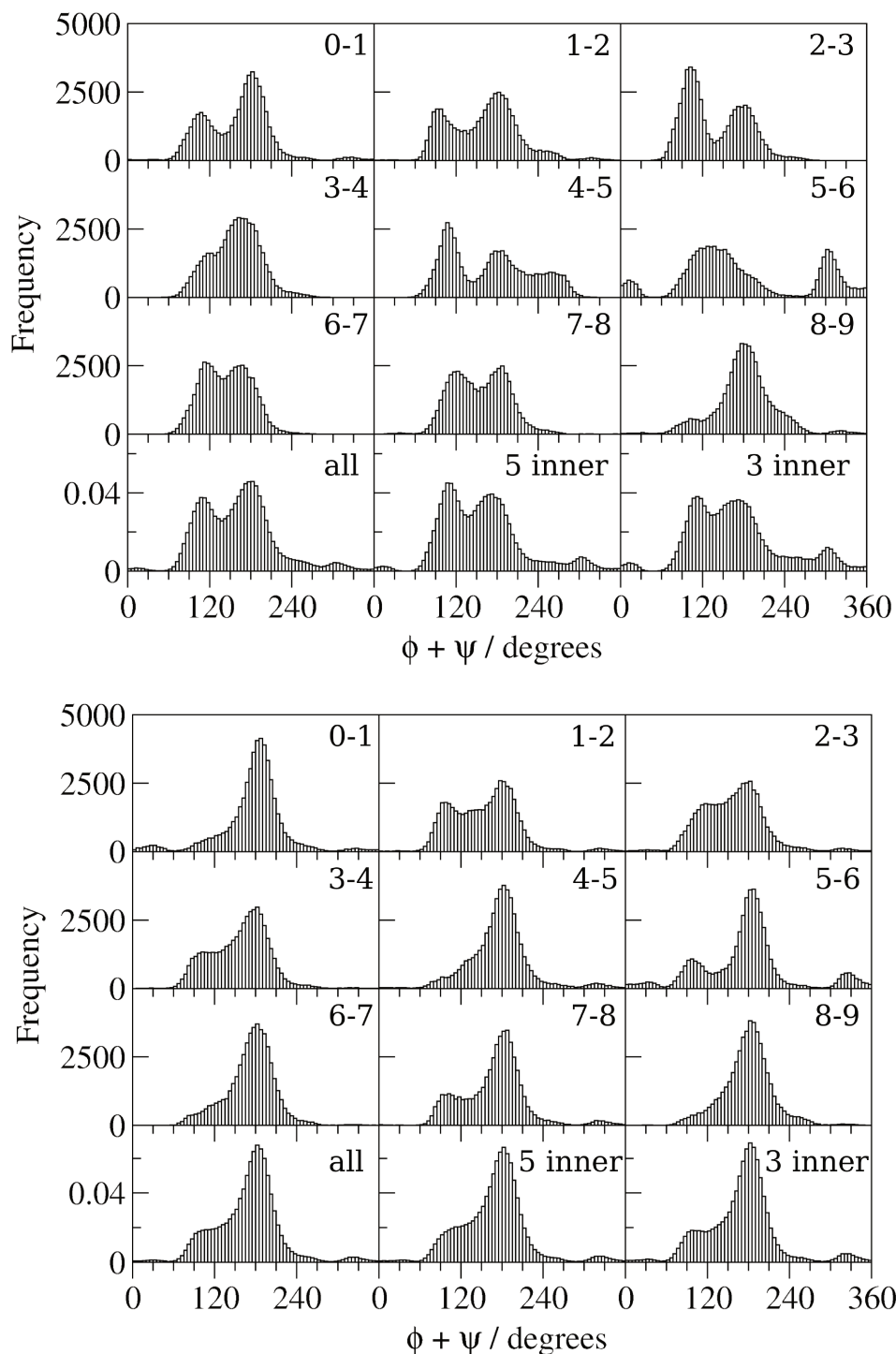


Figura 5.19: Distribuição da soma dos diedros glicosídicos ϕ e ψ durante as simulações, de todas as ligações glicosídicas nas moléculas de acetilxilano adsorvidas sobre as faces hidrofóbicas da celulose I β . Acima: acetilxilano M0 sobre a face (100). Abaixo: acetilxilano M25 sobre a face (200). Para cada cadeia de acetilxilano os primeiros nove histogramas correspondem aos pares de diedros glicosídicos entre os resíduos de xilose 0 e 1, 1 e 2, e assim por diante, e cada histograma tem uma população total de 50 mil. Os últimos três histogramas abrangem mais de um par de diedros glicosídicos: i) todos ; ii) os cinco mais internos ; iii) os três mais internos. Para estes três histogramas, os valores na ordenada foram normalizados de maneira a indicar a frequência relativa com que ocorre o seu valor correspondente na abscissa.

ao xilano interagir bem com a celulose cristalina. A funcionalização química de cadeias de xilano que ocorre em plantas pode ter diversas funções, ainda não totalmente esclarecidas: especula-se que substituintes glucuronóil poderiam ser pontos de reticulação entre a matriz de lignina e o glucuronoxilano, mediante ligações éster envolvendo o grupo carboxílico do GlcA. A substituição por acetil, por sua vez, diminui o número de hidroxilas - não envolvidas em ligações de hidrogênio com a celulose se o padrão de substituição for par - o que obviamente torna o acetilxilano menos polar, ou hidrofílico, mas sem comprometer a sua capacidade de interação com a celulose, se o padrão de substituição for par. O padrão par de acetilação em xilanos encontrado experimentalmente por Dupree *et al.* poderia servir para aumentar a afinidade da lignina, apolar, pelo xilano e pelas superfícies hidrofílicas da celulose recobertas por ele: em plantas, os xilanos podem ter várias centenas de unidades de xilose, e longos trechos de xilano poderiam estar adsorvidos à celulose cristalina enquanto outros trechos da mesma cadeia de xilano poderiam estar imersos na matriz de lignina. Compostos químicos que atuam de forma a aumentar a afinidade entre polímeros hidrofílicos e hidrofóbicos, em misturas como blendas e compósitos, ou em misturas com certos copolímeros, por exemplo, são denominados *compatibilizantes*.

Dentro deste contexto, propomos nesta tese que o padrão par de acetilação faz do acetilxilano um compatibilizante muito apropriado para aumentar a afinidade entre a celulose e a matriz de lignina, diminuindo a tensão interfacial entre estes dois polímeros que compõem a parede celular vegetal. O conceito e a importância de compatibilizantes são bem conhecidos na área de química de polímeros, com diversas aplicações em blendas e compósitos de uso corriqueiro, e uma recente revisão a respeito de compatibilizantes foi produzida por Utracki.⁹⁹ A presença de substituintes que aumentem a hidrofobicidade do xilano, como o grupo acetil, é uma estratégia de compatibilização diferente daquela em que os substituintes no xilano ligam-se covalentemente à lignina, como postula-se que ocorra com o grupo GlcA. A rigor, esta última seria classificada como reticulação, não compatibilização. Em ambos os casos, entretanto, os substituintes nas cadeias de xilano serviriam para reforçar as interações entre os componentes da matriz lignocelulósica de forma a aumentar a estabilidade e a coesão da parede celular vegetal.

Os resultados obtidos nesta terceira parte do trabalho de doutorado, na qual foram estudados complexos de adsorção de xilanos simples e substituídos adsorvidos à celulose I β , serão publicados em conjunto com dados experimentais produzidos por Dupree *et al.*, em um trabalho de colaboração que está em fase de

redação. Esta seção encerra a apresentação e discussão dos resultados obtidos durante este trabalho de tese. Na próxima seção serão expostas as considerações finais e perspectivas para pesquisas futuras.

Capítulo 6

Conclusões

No presente trabalho de tese, foram estudados alguns sistemas relacionados à hidrólise de biomassa lignocelulósica. Na primeira parte do trabalho foram estudadas, mediante simulações por DM, três enzimas da família 16 das glicosídeo hidrolases, e foram elucidados alguns fatores estruturais e dinâmicos que são responsáveis pela hipertermoestabilidade na enzima RmLam. Foi mostrado que a termoestabilidade na RmLam é devida à presença de um núcleo hidrofóbico bastante coeso e de uma extensa rede de conexões salinas adequadamente arranjadas, que reforçam e estabilizam a estrutura da RmLam sem interferir com seu núcleo hidrofóbico. Os resultados das simulações por DM destas enzimas fornecem *insights* úteis para pesquisadores interessados em projetar enzimas GH com propriedades adequadas à sua aplicação industrial.

Na segunda parte do trabalho de tese, foi desenvolvido o programa *cellulose-builder*. Este programa será uma ferramenta útil para cientistas interessados em realizar simulações por DM em sistemas contendo domínios cristalinos de celulose. O programa fornece uma maneira fácil e automática de gerar coordenadas Cartesianas em formato PDB para os alomorfos $I\alpha$, $I\beta$, II e III_I da celulose cristalina, e permite o total controle das dimensões dos cristalitos, e do comprimento das fibras produzidas, diretamente a partir do prompt de comando Bash. Os cristalitos podem ser construídos possuindo simetria translacional, o que permite a aplicação de condições periódicas de contorno ao longo de qualquer direção cristalográfica. As estruturas cristalinas geradas pelo *cellulose-builder* podem ser combinadas com outras moléculas, como por exemplo solventes ou proteínas, utilizando outros programas disponíveis⁶⁶ para gerar configurações iniciais para simulações por DM.

Na terceira parte do trabalho, foram investigadas as interações que moléculas de xilano, simples e substituídas, estabelecem com diferentes faces cristalográficas

da celulose I β . As simulações demonstram que as moléculas de xilano são capazes de estabelecer complexos de adsorção com a celulose cristalina, utilizando como sítios de adsorção as posições cristalográficas que eventualmente seriam ocupadas por cadeias de celulose em um cristal maior. As simulações também sugerem que o padrão par de substituição, observado experimentalmente em xilanos de plantas, não interfere com a capacidade do xilano em estabelecer complexos de adsorção estáveis com a celulose cristalina, ao mesmo tempo que permite a funcionalização química da cadeia de xilano.

6.1 Perspectivas

O estudo das interações entre celulose e xilano será futuramente estendido para abranger outros grupos substituintes na cadeia de xilano. Também serão realizadas tentativas de contornar o problema da distorção do cristal de celulose, apresentado na Seção 5.3, introduzindo constantes de força harmônicas que limitem o movimento dos átomos do cristal de celulose a regiões próximas a suas posições cristalográficas determinadas experimentalmente. Além disso, os resultados obtidos até agora encorajam a realização de estudos para estimar a energia livre de adsorção dos diferentes xilanos a faces da celulose cristalina.

Bibliografia

- [1] Francis A. Carey. *Organic Chemistry*. McGraw-Hill Co., New York, 4th international edition, 2000.
- [2] Lucas Bleicher, Erica T. Prates, Thiago C. F. Gomes, Rodrigo L. Silveira, Alessandro S. Nascimento, Adriana L. Rojas, Alexander Golubev, Leandro Martinez, Munir S. Skaf, and Igor Polikarpov. Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *Journal of Physical Chemistry B*, 115(24):7940–7949, 2011. URL <http://dx.doi.org/10.1021/jp200330z>.
- [3] Thiago C. F. Gomes and Munir S. Skaf. Cellulose-Builder: A toolkit for building crystalline structures of cellulose. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012. URL <http://dx.doi.org/10.1002/jcc.22959>.
- [4] Erica T. Prates, Ivana Stankovic, Rodrigo L. Silveira, Marcelo V. Liberato, Flavio Henrique-Silva, Nei Pereira, Jr., Igor Polikarpov, and Munir 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(3), 2013. URL <http://dx.doi.org/10.1371/journal.pone.0059069>.
- [5] Larissa C. Textor, Francieli Colussi, Rodrigo L. Silveira, Viviane Serpa, Bruno L. de Mello, Joao Renato C. Muniz, Fabio M. Squina, Nei Pereira, Jr., Munir S. Skaf, and Igor Polikarpov. Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from *Trichoderma harzianum*: deciphering the structural features of cellobiohydrolase catalytic activity. *Febs Journal*, 280(1):56–69, 2013. URL <http://dx.doi.org/10.1111/febs.12049>.
- [6] Stephen K. Ritter. Lignocellulose: a complex biomaterial. *Chemical & Engineering News*, 86(49):15, 2008. URL <http://cen.acs.org/articles/86/i49/Lignocellulose-Complex-Biomaterial.html>.
- [7] Tuula T. Teeri. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology*, 15(5):160 – 167, 1997. URL [http://dx.doi.org/10.1016/S0167-7799\(97\)01032-9](http://dx.doi.org/10.1016/S0167-7799(97)01032-9).
- [8] Anke Schmitz, Heike Schaetzel, and Helmut Schmitz. Distribution and functional morphology of photomechanic infrared sensilla in flat bugs of the genus *Aradus* (Heteroptera, Aradidae). *Arthropod Structure & Development*, 39(1):17–25, 2010. URL <http://dx.doi.org/10.1016/j.asd.2009.10.007>.

- [9] Mikyung Seo, Norberto Castillo, Robert Ganzynkowicz, Charlisa R. Daniels, Robert J. Woods, Todd L. Lowary, and Pierre-Nicholas Roy. Approach for the simulation and modeling of flexible rings: Application to the alpha-D-arabinofuranoside ring, a key constituent of polysaccharides from *Mycobacterium tuberculosis*. *Journal of Chemical Theory and Computation*, 4(1):184–191, 2008. URL [dx.doi.org/10.1021/ct700284r](https://doi.org/10.1021/ct700284r).
- [10] Inmaculada Sanchez-Medina, Martin Frank, Claus-Wilhelm von der Lieth, and Johannes P. Kamerling. Conformational analysis of the neutral exopolysaccharide produced by *Lactobacillus delbrueckii* ssp *bulgaricus* LBB.B26. *Organic & Biomolecular Chemistry*, 7(2):280–287, 2009. URL <http://dx.doi.org/10.1039/b810468a>.
- [11] David B. Kony, Wolfgang Damm, Serge Stoll, Wilfred F. van Gunsteren, and Philippe H. Huenenberger. Explicit-solvent molecular dynamics simulations of the polysaccharide schizophyllan in water. *Biophysical Journal*, 93(2):442–455, 2007. URL [dx.doi.org/10.1529/biophysj.106.086116](https://doi.org/10.1529/biophysj.106.086116).
- [12] B. T. Costa-Carvalho, R. M. A. Martinez, A. T. N. Dias, C. A. Kubo, P. Barros-Nunes, L. Leiva, D. Sole, M. M. S. Carneiro-Sampaio, C. K. Naspitz, and R. U. Sorensen. Antibody response to pneumococcal capsular polysaccharide vaccine in Down syndrome patients. *Brazilian Journal of Medical and Biological Research*, 39(12):1587–1592, 2006.
- [13] Leen Moens, Axel Jeurissen, Stefan Nierkens, Louis Boon, Luc Van Kaer, Ahmad Kasran, Greet Wuyts, Jan L. Ceuppens, and Xavier Bossuyt. Generation of antibody responses to pneumococcal capsular polysaccharides is independent of CD1 expression in mice. *Infection and Immunity*, 77(5):1976–1980, 2009. URL [dx.doi.org/10.1128/IAI.01091-08](https://doi.org/10.1128/IAI.01091-08).
- [14] George W. Huber, Sara Iborra, and Avelino Corma. Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. *Chemical Reviews*, 106(9):4044–4098, 2006. URL <http://dx.doi.org/10.1021/cr068360d>.
- [15] Edward L. Kunkes, Dante A. Simonetti, Ryan M. West, Juan Carlos Serrano-Ruiz, Christian A. Gartner, and James A. Dumesic. Catalytic conversion of biomass to monofunctional hydrocarbons and targeted liquid-fuel classes. *Science*, 322(5900):417–421, 2008. URL <http://dx.doi.org/10.1126/science.1159210>.
- [16] Donald Voet and Judith G. Voet. *Biochemistry*. John Wiley & Sons, 2006.
- [17] Yuriy Roman-Leshkov, Juben N. Chheda, and James A. Dumesic. Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science*, 312(5782):1933–1937, 2006. URL <http://dx.doi.org/10.1126/science.1126337>.
- [18] Haibo Zhao, Johnathan E. Holladay, Heather Brown, and Z. Conrad Zhang. Metal chlorides in ionic liquid solvents convert sugars to 5-hydroxymethylfurfural. *Science*, 316(5831):1597–1600, 2007. URL <http://dx.doi.org/10.1126/science.1141199>.

- [19] V. S. Raghavendra Rao, Pradman K. Qasba, Petety V. Balaji, and Resgaswami Chandrasekaran. *Conformations of carbohydrates*. Harwood Academic Publishers, Amsterdam, 1st edition, 1998.
- [20] Carolyn R. Bertozzi and David Rabuka. *Essentials of Glycobiology*, chapter Structural Basis of Glycan Diversity. Cold Spring Harbor Laboratory Press, 2nd edition, 1999. URL <http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=glyco>.
- [21] A. R. Ionescu, A. Berces, M. Z. Zgierski, D. M. Whitfield, and T. Nukada. Conformational pathways of saturated six-membered rings. A static and dynamical density functional study. *Journal of Physical Chemistry A*, 109(36):8096–8105, 2005. URL <http://dx.doi.org/10.1021/jp052197t>.
- [22] A. Berces, D. M. Whitfield, and T. Nukada. Quantitative description of six-membered ring conformations following the IUPAC conformational nomenclature. *Tetrahedron*, 57(3):477–491, 2001.
- [23] James B. Hendrickson. Molecular geometry. vii. modes of interconversion in the medium rings. *Journal of the American Chemical Society*, 89(26):7047–7061, 1967. URL <http://dx.doi.org/10.1021/ja01002a038>.
- [24] D. Cremer and J. A. Pople. General definition of ring puckering coordinates. *Journal of the American Chemical Society*, 97(6):1354–1358, 1975.
- [25] P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole-moment derivatives. *Journal of the American Chemical Society*, 101(10):2550–2560, 1979.
- [26] G. Fogarasi, X. F. Zhou, P. W. Taylor, and P. Pulay. The calculation of ab initio molecular geometries - Efficient optimization by natural internal coordinates and empirical correction by offset forces. *Journal of the American Chemical Society*, 114(21):8191–8201, 1992.
- [27] Karl N. Kirschner and Robert J. Woods. Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences*, 98(19):10541–10545, 2001. URL <http://dx.doi.org/10.1073/pnas.191362798>.
- [28] C. Altona and M. Sundaralingam. Conformational analysis of sigar ring in nucleosides and nucleotides - New description using concept of pseudorotation. *Journal of the American Chemical Society*, 94(23):8205, 1972.
- [29] Hashem A. Taha, Norberto Castillo, Devin N. Sears, Roderick E. Wasylshen, Todd L. Lowary, and Pierre-Nicholas Roy. Conformational analysis of arabinofuranosides: Prediction of (3)J(H,H) using MD Simulations with DFT-derived spin-spin coupling profiles. *Journal of Chemical Theory and Computation*, 6(1):212–222, 2010. URL <http://dx.doi.org/10.1021/ct900477x>.
- [30] Tongye Shen, Paul Langan, Alfred D. French, Glenn P. Johnson, and S. Gnanakaran. Conformational flexibility of soluble cellulose oligomers: chain length and temperature dependence. *Journal of the American Chemical Society*, 131(41):14786–14794, 2009. URL <http://dx.doi.org/10.1021/ja9034158>.

- [31] Sara Guglieri, Milos Hricovini, Rahul Raman, Laura Polito, Giangiacomo Torri, Benito Casu, Ram Sasisekharan, and Marco Guerrini. Minimum FGF2 binding structural requirements of heparin and heparan sulfate oligosaccharides as determined by NMR spectroscopy. *Biochemistry*, 47(52):13862–13869, 2008. URL <http://dx.doi.org/10.1021/bi801007p>.
- [32] Marit Sletmoen and Bjorn T. Stokke. Review : Higher order structure of (1,3)-beta-D-glucans and its influence on their biological activities and complexation abilities. *Biopolymers*, 89(4):310–321, 2008. URL <http://dx.doi.org/10.1002/bip.20920>.
- [33] B. H. Falch and B. T. Stokke. Structural stability of (1-3)-beta-d-glucan macrocycles. *Carbohydrate Polymers*, 44(2):113–121, 2001. URL [http://dx.doi.org/10.1016/S0144-8617\(00\)00214-9](http://dx.doi.org/10.1016/S0144-8617(00)00214-9).
- [34] G. Lee, W. Nowak, J. Jaroniec, Q. M. Zhang, and P. E. Marszalek. Molecular dynamics simulations of forced conformational transitions in 1,6-linked polysaccharides. *Biophysical Journal*, 87(3):1456–1465, 2004. URL <http://dx.doi.org/10.1529/biophysj.104.042879>.
- [35] Zhenyu Lu, Hao Hu, Weitao Yang, and Piotr E. Marszaleky. Simulating force-induced conformational transitions in polysaccharides with the SMD replica exchange method. *Biophysical Journal*, 91(6):L57–L59, 2006. URL <http://dx.doi.org/10.1529/biophysj.106.090324>.
- [36] P. E. Marszalek, A. F. Oberhauser, Y. P. Pang, and J. M. Fernandez. Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring. *Nature*, 396(6712):661–664, 1998.
- [37] Edwin C. Webb. *Enzyme Nomenclature*. International Union of Biochemistry and Molecular Biology. Nomenclature Committee. Academic Press, San Diego, CA, 1992. URL <http://www.chem.qmul.ac.uk/iubmb/enzyme/>.
- [38] Brandi L. Cantarel, Pedro M. Coutinho, Corinne Rancurel, Thomas Bernard, Vincent Lombard, and Bernard Henrissat. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, 37:D233–D238, 2009. URL <http://dx.doi.org/10.1093/nar/gkn663>.
- [39] B. Henrissat. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. *Biochemical Journal*, 280(Part 2):309–316, 1991.
- [40] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon, and G. Davies. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences*, 92(15):7090–7094, 1995. URL <http://www.pnas.org/content/92/15/7090.abstract>.
- [41] Claire Dumon, Alexander Varvak, Mark A. Wall, James E. Flint, Richard J. Lewis, Jeremy H. Lakey, Carl Morland, Peter Luginbühl, Shaun Healey, Thomas Todaro, Grace DeSantis, May Sun, Lilian Parra-Gessert, Xuqiu Tan, David P. Weiner, and Harry J. Gilbert. Engineering hyperthermostability into a gh11 xylanase is

- mediated by subtle changes to protein structure. *Journal of Biological Chemistry*, 283(33):22557–22564, 2008. URL <http://dx.doi.org/10.1074/jbc.M800936200>.
- [42] Khanok Ratanakhanokchai, Rattiya Waeonukul, Patthra Pason, Chakrit Tachapaikoon, Khin Lay Kyu, Kazuo Sakka, Akihiko Kosugi, and Yutaka Mori. *Biomass Now - Cultivation and Utilization*, chapter Paenibacillus curdolanolyticus strain B-6 multienzyme complex: a novel system for biomass utilization. InTech, 2013. URL <http://dx.doi.org/10.5772/51820>.
- [43] Nina Aro, Tiina Pakula, and Merja Penttilä. Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiology Reviews*, 29(4):719–739, 2005. URL <http://dx.doi.org/10.1016/j.femsre.2004.11.006>.
- [44] Alexander M. Golubev, Adriana L. Rojas, Alessandro S. Nascimento, Lucas Bleicher, Anna A. Kulminskaya, Elena V. Eneyskaya, and Igor Polikarpov. Crystallization and preliminary crystallographic analysis of laminarinase from rhodothermus marinus: a case of pseudomerohedral twinning. *Protein and Peptide Letters*, 15(10):1142–1144, 2008. URL <http://dx.doi.org/doi:10.2174/092986608786071139>.
- [45] Martin Krah, Rolf Misselwitz, Oliver Politz, Karl Kristian Thomsen, Heinz Welfle, and Rainer Borriss. The laminarinase from thermophilic eubacterium rhodothermus marinus. *European Journal of Biochemistry*, 257(1):101–111, 1998. URL <http://dx.doi.org/10.1046/j.1432-1327.1998.2570101.x>.
- [46] Yoshiharu Nishiyama, Junji Sugiyama, Henri Chanzy, and Paul Langan. Crystal structure and hydrogen bonding system in cellulose I α from synchrotron x-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 125(47):14300–14306, 2003. URL <http://dx.doi.org/10.1021/ja037055w>.
- [47] Y. Nishiyama, P. Langan, and H. Chanzy. Crystal structure and hydrogen-bonding system in cellulose I β from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 124(31):9074–9082, 2002. URL <http://dx.doi.org/10.1021/ja0257319>.
- [48] Paul Langan, Yoshiharu Nishiyama, and Henri Chanzy. X-ray structure of mercerized cellulose II at 1 Å resolution. *Biomacromolecules*, 2(2):410–416, 2001. URL <http://dx.doi.org/10.1021/bm005612q>.
- [49] Masahisa Wada, Henri Chanzy, Yoshiharu Nishiyama, and Paul Langan. Cellulose III_I crystal structure and hydrogen bonding by synchrotron x-ray and neutron fiber diffraction. *Macromolecules*, 37(23):8548–8555, 2004. URL <http://dx.doi.org/10.1021/ma0485585>.
- [50] J. F. Matthews, C. E. Skopec, P. E. Mason, P. Zuccato, R. W. Torget, J. Sugiyama, M. E. Himmel, and J. W. Brady. Computer simulation studies of microcrystalline cellulose I β . *Carbohydrate Research*, 341(1):138–152, 2006. URL <http://dx.doi.org/10.1016/j.carres.2005.09.028>.
- [51] James F. Matthews, Gregg T. Beckham, Malin Bergenstrahle-Wohlert, John W. Brady, Michael E. Himmel, and Michael F. Crowley. Comparison of cellulose I β

- simulations with three carbohydrate force fields. *Journal of Chemical Theory and Computation*, 8(2):735–748, 2012. URL <http://dx.doi.org/10.1021/ct2007692>.
- [52] Yoshiharu Nishiyama, Glenn P. Johnson, Alfred D. French, V. Trevor Forsyth, and Paul Langan. Neutron crystallography, molecular dynamics, and quantum mechanics studies of the nature of hydrogen bonding in cellulose I β . *Biomacromolecules*, 9(11):3133–3140, 2008. URL <http://dx.doi.org/10.1021/bm800726v>.
- [53] Christina M. Payne, Michael G. Resch, Liqun Chen, Michael F. Crowley, Michael E. Himmel, Larry E. Taylor II, Mats Sandgren, Jerry Stahlberg, Ingeborg Stals, Zhongping Tan, and Gregg T. Beckham. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proceedings of the National Academy of Sciences*, 110(36):14646–14651, 2013. URL <http://dx.doi.org/10.1073/pnas.1309106110>.
- [54] Mark R. Nimlos, Gregg T. Beckham, James F. Matthews, Lintao Bu, Michael E. Himmel, and Michael F. Crowley. Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *Journal of Biological Chemistry*, 287(24):20603–20612, 2012. URL <http://dx.doi.org/10.1074/jbc.M112.358184>.
- [55] Landry Charlier and Karim Mazeau. Molecular modeling of the structural and dynamical properties of secondary plant cell walls: influence of lignin chemistry. *The Journal of Physical Chemistry B*, 116(14):4163–4174, 2012. URL <http://dx.doi.org/10.1021/jp300395k>.
- [56] Qiong Zhang, Harry Brumer, Hans Agren, and Yaoquan Tu. The adsorption of xyloglucan on cellulose: effects of explicit water and side chain variation. *Carbohydrate Research*, 346(16):2595–2602, 2011. URL <http://dx.doi.org/10.1016/j.carres.2011.09.007>.
- [57] Karim Mazeau and Landry Charlier. The molecular basis of the adsorption of xylans on cellulose surface. *Cellulose*, 19(2):337–349, 2012. URL <http://dx.doi.org/10.1007/s10570-011-9643-7>.
- [58] David M. Brown, Florence Goubet, Vicky W. Wong, Royston Goodacre, Elaine Stephens, Paul Dupree, and Simon R. Turner. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *The Plant Journal*, 52(6):1154–1168, 2007. URL <http://dx.doi.org/10.1111/j.1365-313X.2007.03307.x>.
- [59] Jennifer R. Bromley, Marta Busse-Wicher, Theodora Tryfona, Jennifer C. Mortimer, Zhinong Zhang, David M. Brown, and Paul Dupree. GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. *Plant Journal*, 74(3):423–434, 2013. URL <http://dx.doi.org/10.1111/tpj.12135>.
- [60] Jennifer C. Mortimer, Godfrey P. Miles, David M. Brown, Zhinong Zhang, Marcelo P. Segura, Thilo Weimar, Xiaolan Yu, Keith A. Seffen, Elaine Stephens, Simon R. Turner, and Paul Dupree. Absence of branches from xylan in *Arabidopsis*

- gux mutants reveals potential for simplification of lignocellulosic biomass. *Proceedings of the National Academy of Sciences*, 107(40):17409–17414, 2010. URL <http://dx.doi.org/10.1073/pnas.1005456107>.
- [61] A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education, Harlow, Great Britain, 2001.
- [62] M. P. Allen and D. J. Tildesley. *Computer simulations of liquids*. Oxford University Press, New York, 1987.
- [63] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press, San Diego, CA, 1996.
- [64] Guntur Fibriansah, Sumiko Masuda, Naoya Koizumi, Satoshi Nakamura, and Takashi Kumasaka. The 1.3 Å crystal structure of a novel endo- β -1,3-glucanase of glycoside hydrolase family 16 from alkaliphilic nocardiosis sp. strain f96. *Proteins: Structure, Function, and Bioinformatics*, 69(3):683–690, 2007. URL <http://dx.doi.org/10.1002/prot.21589>.
- [65] Jonas Vasur, Rie Kawai, Anna M. Larsson, Kiyohiko Igarashi, Mats Sandgren, Masahiro Samejima, and Jerry Ståhlberg. X-ray crystallographic native sulfur SAD structure determination of laminarinase Lam16A from *Phanerochaete chrysosporium*. *Acta Crystallographica Section D*, 62(11):1422–1429, 2006. URL <http://dx.doi.org/10.1107/S0907444906036407>.
- [66] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, 2009. URL <http://dx.doi.org/10.1002/jcc.21224>.
- [67] Ramu Anandkrishnan, Boris Aguilar, and Alexey V. Onufriev. H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, 40(W1):W537–W541, 2012. URL <http://dx.doi.org/10.1093/nar/gks375>.
- [68] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154, 1964. URL <http://dx.doi.org/10.1093/comjnl/7.2.149>.
- [69] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005. URL <http://dx.doi.org/10.1002/jcc.20289>.
- [70] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993. URL <http://dx.doi.org/10.1063/1.464397>.

- [71] 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. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. URL <http://dx.doi.org/10.1021/jp973084f>.
- [72] Olgun Guvench, Shannon N. Greene, Ganesh Kamath, John W. Brady, Richard M. Venable, Richard W. Pastor, and Alexander D. Mackerell. Additive empirical force field for hexopyranose monosaccharides. *Journal of Computational Chemistry*, 29(15):2543–2564, 2008. URL <http://dx.doi.org/10.1002/jcc.21004>.
- [73] Olgun Guvench, Elizabeth Hatcher, Richard M. Venable, Richard W. Pastor, and Alexander D. MacKerell, Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation*, 5(9):2353–2370, 2009. URL <http://dx.doi.org/10.1021/ct900242e>.
- [74] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983. URL <http://dx.doi.org/10.1063/1.445869>.
- [75] Dmitrii Beglov and Benoît Roux. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *The Journal of Chemical Physics*, 100(12):9050–9063, 1994. doi: <http://dx.doi.org/10.1063/1.466711>. URL <http://scitation.aip.org/content/aip/journal/jcp/100/12/10.1063/1.466711>.
- [76] T. Schneider and E. Stoll. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B*, 17:1302–1322, 1978. URL <http://dx.doi.org/10.1103/PhysRevB.17.1302>.
- [77] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994. URL <http://dx.doi.org/10.1063/1.467468>.
- [78] Jean-Paul Ryckaert, Giovanni Ciccoliti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327 – 341, 1977. URL [http://dx.doi.org/10.1016/0021-9991\(77\)90098-5](http://dx.doi.org/10.1016/0021-9991(77)90098-5).
- [79] Mark R. Nimlos and Michael F. Crowley. *Computer modeling in lignocellulosic biofuel production*. American Chemical Society, Washington, DC, 2010.
- [80] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [81] J. P. Glusker, M. Lewis, and M. Rossi. *Crystal Structure Analysis for Chemists and Biologists*. Wiley-VCH, New York, 1994.

-
- [82] M. F. C. Ladd and R. A. Palmer. *Structure Determination by X-ray Crystallography*. Plenum Press, New York, 1994.
- [83] Michael Glazer and Gerald Burns. *Space Groups for Solid State Scientists*. Academic Press, Oxford, 2013.
- [84] T. Hahn. *International Tables for Crystallography*, volume A. Springer, Dordrecht, 2005.
- [85] C. Newham and B. Rosenblatt. *Learning the Bash Shell*. O'Reilly & Associates, Cambridge, 1998.
- [86] C. Albing, J. P. Vossen, and C. Newham. *Bash Cookbook*. O'Reilly Media, Sebastopol, CA, 2007.
- [87] A. Robbins and N. H. F. Beebe. *Classic Shell Scripting*. O'Reilly Media, Beijing, 2005.
- [88] L. Wall, T. Christiansen, and J. Orwant. *Programming Perl*. O'Reilly Media, Sebastopol, CA, 2000.
- [89] J. E. F. Friedl. *Mastering Regular Expressions*. O'Reilly Media, Sebastopol, CA, 2006.
- [90] S. Y. Ding and M. E. Himmel. The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural and Food Chemistry*, 54(3):597–606, 2006. URL <http://dx.doi.org/10.1021/jf051851z>.
- [91] Rodrigo L. Silveira, Stanislav R. Stoyanov, Sergey Gusarov, Munir S. Skaf, and Andriy Kovalenko. Plant biomass recalcitrance: effect of hemicellulose composition on nanoscale forces that control cell wall strength. *Journal of the American Chemical Society*, 135(51):19048–19051, 2013. URL <http://dx.doi.org/10.1021/ja405634k>.
- [92] David C. Adler and Markus J. Buehler. Mesoscale mechanics of wood cell walls under axial strain. *Soft Matter*, 9:7138–7144, 2013. URL <http://dx.doi.org/10.1039/C3SM50183C>.
- [93] Anwasha N. Fernandes, Lynne H. Thomas, Clemens M. Altaner, Philip Callow, V. Trevor Forsyth, David C. Apperley, Craig J. Kennedy, and Michael C. Jarvis. Nanostructure of cellulose microfibrils in spruce wood. *Proceedings of the National Academy of Sciences*, 108(47):E1195–E1203, 2011. URL <http://dx.doi.org/10.1073/pnas.1108942108>.
- [94] Alfred D. French and Glenn P. Johnson. Cellulose and the twofold screw axis: modeling and experimental arguments. *Cellulose*, 16(6):959–973, 2009. URL <http://dx.doi.org/10.1007/s10570-009-9347-4>.
- [95] Karim Mazeau, Charlotte Moine, Pierre Krausz, and Vincent Gloaguen. Conformational analysis of xylan chains. *Carbohydrate Research*, 340(18):2752 – 2760, 2005. URL <http://dx.doi.org/10.1016/j.carres.2005.09.023>.

- [96] Michelle Kuttel, J. W. Brady, and Kevin J. Naidoo. Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *Journal of Computational Chemistry*, 23(13):1236–1243, 2002. URL <http://dx.doi.org/10.1002/jcc.10119>.
- [97] Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño, Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. Glycam06: A generalizable biomolecular force field. carbohydrates. *Journal of Computational Chemistry*, 29(4):622–655, 2008. ISSN 1096-987X. doi: 10.1002/jcc.20820. URL <http://dx.doi.org/10.1002/jcc.20820>.
- [98] Roberto D. Lins and Philippe H. Hünenberger. A new GROMOS force field for hexopyranose-based carbohydrates. *Journal of Computational Chemistry*, 26(13):1400–1412, 2005. URL <http://dx.doi.org/10.1002/jcc.20275>.
- [99] Leszek A. Utracki. Compatibilization of polymer blends. *The Canadian Journal of Chemical Engineering*, 80(6):1008–1016, 2002. URL <http://dx.doi.org/10.1002/cjce.5450800601>.