



ÉRICA TEIXEIRA PRATES

**DINÂMICA MOLECULAR DE HIDROLASES PARA
SACARIFICAÇÃO DE CELULOSE E PROTEÍNAS CORRELATAS**

**CAMPINAS
2013**



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**

ÉRICA TEIXEIRA PRATES

**DINÂMICA MOLECULAR DE HIDROLASES PARA SACARIFICAÇÃO DE
CELULOSE E PROTEÍNAS CORRELATAS**

ORIENTADOR: PROF. DR. MUNIR SALOMÃO SKAF

**TESE DE DOUTORADO APRESENTADA AO
INSTITUTO DE QUÍMICA DA UNICAMP PARA
OBTENÇÃO DO TÍTULO DE DOUTORA EM CIÊNCIAS.**

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA
POR ÉRICA TEIXEIRA PRATES, E ORIENTADA PELO PROF.DR. MUNIR
SALOMÃO SKAF.**

Assinatura do Orientador

**CAMPINAS
2013**

FICHA CATALOGRÁFICA ELABORADA POR SIMONE LUCAS - CRB8/8144 -
BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

P887d Prates, Érica Teixeira (1985-).
Dinâmica molecular de hidrolases para sacarificação
de celulose e proteínas correlatas / Érica Teixeira Prates.
– Campinas, SP: [s.n.], 2013.

Orientador: Munir Salomão Skaf.

Tese (doutorado) - Universidade Estadual de
Campinas, Instituto de Química.

1. Dinâmica molecular. 2. Celulases. 3. Celulose.
4. Endoglucanases. 5. Enzimas termofílicas. I. Skaf,
Munir Salomão. II. Universidade Estadual de Campinas.
Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em inglês: Molecular dynamics study of hydrolases for saccharification of cellulose and related proteins

Palavras-chave em inglês:

Molecular dynamics
Cellulases
Cellulose
Endoglucanases
Thermophilic enzymes

Área de concentração: Físico-Química

Titulação: Doutora em Ciências

Banca examinadora:

Munir Salomão Skaf [Orientador]
Alessandro Silva Nascimento
Paulo Augusto Netz
René Alfonso Nome Silva
Ricardo Aparício

Data de defesa: 24/05/2013

Programa de pós-graduação: Química

*Se você pensa que pequenas coisas não fazem diferença,
tente passar a noite em um quarto com um mosquito!*

Dalai Lama

Agradecimentos

Gosto muito desta parte da tese. Tenho a oportunidade de dizer coisas importantíssimas. Nesta tese, vão ingredientes muito variados. Sonhos, quedas e esforço são alguns deles. Em alguns momentos errei na mão e precisei tentar compensar com um ou outro tempero. Mas não tem graça cozinhar sozinho e somente para si mesmo. O resultado pode ser desastroso. Minha gratidão e meu coração às pessoas que tornaram meus dias neste doutorado significativos. Meu coração a quem os deixou mais saborosos!

Antes de qualquer outra pessoa, mãe, é a você que eu quero agradecer! Eu te agradeço por todos os dias em que você cuidou de mim. Mesmo de longe, você sempre buscou ficar perto e se assegurar que estava tudo bem comigo.

“And love dares you to care for the people on the edge of the night / And loves dares you to change our way of caring about ourselves” (Queen)

Agradeço ao Babico, meu querido pai, por cuidar de mim, por todo apoio e pelas nossas conversas inspiradoras. Você desperta o que eu tenho de melhor. Agradeço ao meu irmãozinho Daniel. Obrigada pela amizade, pelo apoio, pelos exemplos de generosidade e integridade. Te admiro muito e estaremos juntos sempre.

Um agradecimento muito especial ao meu orientador. Munir, obrigada por participar da minha vida nestes anos todos, sendo um misto de chefe-amigo-psicólogo (para cada necessidade, um papel). Obrigada por ter gasto tanta

energia comigo, conhecendo como eu “funciono” e buscando a melhor maneira de me ajudar a melhorar. Obrigada por se fazer tão presente/imprescindível neste trabalho.

“O bom professor explica, mas o grande professor, como você, inspira.”

Agradeço aos meus amigos deste grupo de pesquisa. Eu tenho muita sorte de trabalhar neste grupo, que só tem gente legal. Obrigada às minhas companheiras de “lab noturno”, Tati e Susan! Vocês certamente fizeram minhas madrugadas de trabalho mais alegres, leves e inesquecíveis. Obrigada ao pessoal do “lab de baixo” (Ivana, Paulo, Denise, Clarisse, Melina, Thiago e Rodrigo) pelo coleguismo e cooperatividade. Agradeço também aos que já não estão aqui, Leandro, Júlio e Anders, que, mesmo não fazendo parte do grupo como alunos ou pesquisadores, continuam presentes e sempre se mostram dispostos a ajudar.

Agradeço à Ivana, pelas discussões sobre a teoria da técnica de Dinâmica Molecular e de Mecânica Estatística, bem como pela parceria no estudo de endoglucanases 3 de *Trichoderma harzianum*. Ao Rodrigo e ao Thiago, pela parceria no estudo de laminarinase de *Rhodothermus marinus*. Ao Paulo, pelo apoio e amizade e por me ajudar a iniciar os trabalhos de dinâmica molecular de proteínas.

Aos principais programadores do grupo (Paulo, Rodrigo e Leandro) pela disponibilização de grande parte dos programas de análise de simulações usados neste trabalho.

Agradeço ao pessoal da CPG, por serem muito atenciosos, simpáticos e

competentes.

À FAPESP, pelo suporte financeiro.

Ao Instituto de Química e à UNICAMP, pelo acolhimento e excelência cursos no cursos que oferece. À UNICAMP, pelo bandeirão.

Agradeço aos professores titulares da banca Alessandro Nascimento, Ricardo Aparício, René Nome e Paulo Netz, bem como os professores suplentes Fábio Gozzo, Marcelo Ganzarolli e Carlos Alberto Labate por aceitarem prontamente compor a minha banca e pelo empenho na leitura desta tese. Agradeço também pela importante contribuição dos professores Ricardo, René e Fábio no exame de qualificação de área.

Agradeço muito ao meu pai e às minhas irmãs, Verônica e Bruna, pela amizade e amor, por me motivarem, se preocuparem e pelo esforço em transpor os obstáculos que nos separam. Ao meu pai por todo apoio.

Aos amigos tão especiais Mônica e Maurício, por serem tão presentes, pela preocupação, pelo carinho, pelas conversas e pela amizade que dura para a vida toda.

Ao meu avô Carlos Alberto, pelas palavras de incentivo e por buscar estar sempre presente nos momentos mais marcantes da minha vida. Sua presença é muito importante para mim!

Aos meus amigos. Agradeço pelo conforto de contar com vocês, pelas

conversas instrutivas no bandeirão, pela companhia e muitas risadas. À Isa, pelas “brocolagens” tão carinhosas. À minha nova, queridíssima amiga, Cintia, por todo apoio, paciência e receptividade no nosso apezinho.

Um agradecimento muito especial à minha terapeuta, Ana Luiza, por todo carinho, preocupação e competência nos anos que estivemos juntas. Ana, sua ajuda foi crucial neste trabalho.

Aos amigos dos treinos de Dharma e Kung Fu, por me ajudarem a melhorar todos os dias, a encontrar mais paz e força.

Ao meu amigo Márcio, que me faz tanta falta. Obrigada, Marcinho, pela companhia de trabalho notívago, pela amizade linda que nasceu em tão pouco tempo, por ter sido parte da minha vida e ter me ensinado tanto. Inspiro-me na sua vida, tão breve, e no pensamento de Trotsky:

“A vida é bela. Que as gerações futuras a limpem de todo o mal, de toda opressão, de toda violência e possam gozá-la plenamente.”

Agradeço a todos os meus parentes. Família a gente não escolhe, mas, se eu pudesse, escolheria a cada um de vocês, Teixeiras e Prates.

Por fim, agradeço especialmente ao Pedro pelo apoio crucial em momentos muito difíceis. Agradeço pelo amor, pelo cuidado e pelas palavras doces. Obrigada por me ajudar neste trabalho, com sua companhia que me revigora no computador ao lado, ou mesmo pelo chat.

Curriculum Vitae

Nome Érica Teixeira Prates
Nascimento 22/01/1985 – Resende/RJ - Brasil

Formação acadêmica/titulação

- 2009-2013 Doutorado em Ciências – Instituto de Química, Universidade Estadual de Campinas (IQ-UNICAMP), Campinas – SP.
Título: Dinâmica Molecular de Hidrolases para Sacarificação de Celulose e Proteínas Correlatas.
Orientador: Prof. Dr. Munir S. Skaf (IQ-UNICAMP).
Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Projeto parte dos programas nacionais Bioen e INCT do bioetanol.
- 2007 - 2009 Mestrado em Química (IQ – UNICAMP).
Título: Interações de Anestésicos Locais com Bicamadas Fosfolipídicas.
Orientador: Prof. Dr. Munir Salomão Skaf (IQ - UNICAMP) / Co-orientadora: Prof^a. Dr^a. Mónica Pickholz (Facultad de Farmacia y Bioquímica – UBA, Buenos Aires).
Bolsista da FAPESP.
- 2003 - 2007 Graduação em Bacharelado em Química (IQ – UNICAMP).
Iniciação Científica (2005 – 2007).
Título: Dinâmica Molecular de Aminoácidos e Peptídeos em Solução Aquosa.
Orientador: Munir Salomão Skaf (IQ – UNICAMP).

Formação complementar

- 2012 Estágio docente, PED C (IQ - UNICAMP).
Carga horária: 128h.
- 2011 Curso “Coarse Grain Methods for Biomolecular Simulations”.
Carga horária: 80h.
Institut Pasteur de Montevideo-Uruguai.
- 2011 Avaliadora no Congresso Interno de Iniciação Científica da

- Unicamp.
- 2010 Curso “Introdução à Linguagem C”.
Carga horária: 15h.
Centro Nacional de Processamento de Alto Desempenho.
- 2010 Curso “Advanced School on Biochemistry of Biofuels”.
Carga horária: 80h.
Sociedade Brasileira de Bioquímica e Biologia Molecular, São Paulo –SP. Bolsa integral concedida pela SBBQ.
- 2009 I Workshop de Docking Molecular.
Carga horária: 8h.
Instituto de Química – Universidade Federal de São Carlos - SP.
- 2007 Curso “Aspectos físico-químicos da estrutura de membranas”.
Carga horária: 8h.
Instituto de Física - USP, São Paulo- SP.
- 2007 Curso “Spring College on Water in Physics, Chemistry and Biology”.
Carga horária: 80h.
International Center for Theoretical Physics (ICTP), Trieste, Itália.
Bolsa integral concedida pelo ICTP e IQ – UNICAMP.
- 2006 Curso “Métodos de caracterização estrutural de proteínas”
Carga horária: 16h.
Laboratório Nacional de Luz Síncrotron, LNLS, Campinas - SP.
- 2005 Curso “Introdução ao FORTRAN 90”
Carga horária: 20h.
Centro Nacional de Processamento de Alto Desempenho, Campinas - SP.

Histórico Escolar

Encerra o curso de doutorado com coeficiente de rendimento (CR) máximo, 4.0, assim como no curso de mestrado. Durante a graduação, permaneceu entre os dez alunos com maior CR da turma e sem reprovações.

Distinções

- 2009 Terceira colocada na prova de ingresso no curso de doutorado do IQ – UNICAMP.
- 2007 Foi a única aluna de graduação selecionada para participar,

com bolsa integral, do curso “Spring college on water in Physics, Chemistry and Biology”, realizado no International Center of Theoretical Physics, em Trieste (Itália).

2006

Menção Honrosa pelo painel "Dinâmica Molecular de Biomoléculas em Solução Aquosa”, PIBIC/CNPq (XIV Congresso Interno de Iniciação Científica – UNICAMP).

Artigos completos publicados em periódicos

1. **Érica T. Prates**, Ivana Stanković, Rodrigo L. Silveira, Marcelo V. Liberato, Flávio Henrique Silva, Nei Pereira, Igor Polikarpov, Munir S. Skaf (2013).
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: p.e59069.
2. Lucas Bleicher, **Érica T. Prates**, Igor Poilikarpov, Thiago C. F. Gomes, Rodrigo L. Silveira, Leandro Martínez, Alessandro S. Nascimento, Adriana L. Rojas, Alexander Golubev, Munir S. Skaf (2011).
Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. The Journal of Physical Chemistry B 115: 7940 - 7949.
3. **Érica T. Prates**, Paulo C. T. Souza, Mónica Pickholz, Munir S. Skaf (2011).
CHARMM-based parameterization of neutral articaine—A widely used local anesthetic. International Journal of Quantum Chemistry 111: 1339 - 1345.

Artigos submetidos

1. Marisa A. Lima, Mario de O. Neto, Flávio R. Rosseto, **Érica T. Prates**, Munir S. Skaf, Igor Polikarpov (2013).
***Aspergillus niger* beta-glucosidase has a cellulase-like tadpole molecular shape.** The Journal of Biological Chemistry.

Artigos em preparação

1. **Érica T. Prates**, Munir S. Skaf.
The denaturation of hyperthermophilic laminarinases studied via high temperature molecular dynamics simulations.
2. Fernanda do Carmo Egídio, Denise C. Melo, **Érica T. Prates**, Munir S. Skaf

Edvaldo Sabadini.

Dynamical and structural properties of polyethylene oxide of different lengths in aqueous solution.

3. **Érica T. Prates**, Mónica Pickholz, Munir S. Skaf.

Distribution of neutral articaine in a phospholipid bilayer: insights from molecular dynamics simulations.

Principais apresentações de trabalhos em eventos

1. **Érica T. Prates**, Ivana Stanković, Rodrigo L. Silveira, Marcelo V. Liberato, Igor Polikarpov, Munir S. Skaf.

Molecular Dynamics Simulations of Oligosaccharide Interactions with Endoglucanases that Lack the Cellulose Binding Domain, 2012. (Comunicação, Apresentação oral de trabalho); *Local: Hotel Maresias Beach; Cidade: São Sebastião - SP; Evento: IV Workshop CeProBio / SUNLIBB 2012.*

2. **Érica T. Prates**, Munir S. Skaf.

Hyperthermophilic of Laminarinases Studied via High Temperature Molecular Dynamics Simulations, 2011. (Simpósio)

Local: Centro de Convenções Parque Metalúrgico da UFOP; Cidade: Ouro Preto - MG; Evento: XVI Simpósio Brasileiro de Química Teórica - SBQT 2011.

3. **Érica T. Prates**, Thiago C. F. Gomes, Jorge J. C. Haddad., Rodrigo L. Silveira, Munir S. Skaf.

Molecular Dynamics Study of the Hyperthermophilic Laminarinase from *Rodothermus marinus*, 2010. (Congresso, apresentação oral de trabalho)

Local: Hotel Fazenda Fonte Colina Verde; Cidade: São Pedro - SP; Evento: 2nd Pan American Congress on Plants and Bioenergy.

4. **Érica T. Prates**, Mónica Pickholz, Munir S. Skaf.

Dinâmica molecular de articaína em membranas POPC, 2008. (Conferência, apresentação oral de trabalho)

Local: Instituto de Biologia, UNICAMP; Cidade: Campinas; Evento: I Workshop Temático - Novas Formulações Farmacêuticas de Anestésicos Locais de Ação Prolongada.

5. **Érica T. Prates**, Munir S. Skaf.

Dinámica Molecular de Biomoléculas en Soluciones Acuosas, 2007. (Congresso, apresentação oral de trabalho)

Local: Universidad Nacional de Asunción; Cidade: Assunção; Evento: XV Jornadas de Jóvenes Investigadores.

Resumo

Dinâmica molecular de hidrolases para sacarificação da celulose e proteínas correlatas

A biomassa lignocelulósica proveniente do bagaço da cana-de-açúcar e de outras matérias-primas é um material altamente promissor para a geração de biocombustíveis renováveis e ambientalmente positivos. A melhor opção para a conversão dessa biomassa em açúcares solúveis fermentáveis a etanol, em termos de rendimento e de vantagens ambientais, é a catálise enzimática. Mas esta é também a etapa mais cara do processo de obtenção de etanol de segunda geração devido à baixa eficiência e alto custo dos coquetéis enzimáticos atualmente disponíveis para este fim. Para tornar estes processos mais eficientes e economicamente viáveis, é preciso aprofundar nossa compreensão dos mecanismos de hidrólise celulolítica. Grande investimento em pesquisa tem sido empregado com esta finalidade e, como parte disto, este trabalho consiste em um conjunto de pesquisas desenvolvidas na área de simulação computacional via dinâmica molecular de três enzimas celulolíticas: 1) Laminarinase de *Rodhothermus marinus*; 2) Endoglucanase 3 de *Trichoderma harzianum* e 3) β -glicosidase de *Aspergillus niger*. De modo geral, estes estudos visaram investigar a relação entre o arranjo estrutural e propriedades mensuráveis em laboratório interessantes na avaliação da performance destes biocatalisadores, como afinidade pelo substrato e estabilidade térmica. Como parte do Projeto Temático BioEn, financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), estes estudos computacionais foram realizados em estreita colaboração com grupos de biofísicos estruturais e biólogos moleculares, sendo a escolha dos três temas embasada em resultados experimentais.

Abstract

Molecular dynamics study of hydrolases for saccharification of cellulose and related proteins

The lignocellulosic biomass from sugar cane bagasse and from other raw materials is a highly promising material for the generation of renewable and environmentally positive fuels. In terms of performance and environmental advantages, the best option for converting this biomass into soluble sugars to produce ethanol is the enzymatic catalysis. However, this is also the most expensive step of the second-generation ethanol production due to the low efficiency and high cost of the currently available enzyme cocktails. In order to make the process more efficient and economically viable, it is necessary to deepen the understanding of the cellulolytic hydrolysis mechanisms. Great investment in research has been employed for this purpose, and as part of these efforts, this work consists on a set of molecular dynamics studies of three cellulolytic enzymes, namely: 1) laminarinase from *Rodhothermus marinus*; 2) Endoglucanase 3 from *Trichoderma harzianum* and 3) β -glucosidase from *Aspergillus niger*. In general, these studies aimed to investigate the relationship between the structural arrangement and experimental data that are interesting for the biocatalyst performance evaluation, such as affinity to the substrate and thermal stability. As part of the BioEn Thematic Project, funded by FAPESP (Research Foundation of the State of São Paulo), these computational studies were carried out in close collaboration with structural biophysicists and molecular biologists. The choice of the three proteins considered here was based on these experimental studies.

Sumário

Lista de Abreviaturas	xxv
Lista de Tabelas	xxvii
Lista de Figuras	xxix
1 Introdução	1
1.1 Matérias-primas lignocelulósicas.....	4
1.2 Hidrólise enzimática da biomassa.....	8
1.2.1 Celulases.....	9
1.2.2 Mecanismos de reação de glicosídeo hidrolases.....	15
1.2.3 Classificação de glicosídeo hidrolases.....	16
1.3 Temas estudados.....	18
Laminarinase de <i>Rhodothermus marinus</i>	19
Endoglucanase 3 de <i>Trichoderma harzianum</i>	19
β -glicosidases de <i>Aspergillus niger</i>	20
2 Metodologia	21
2.1 Uma visão geral dos fundamentos de dinâmica molecular.....	24
2.1.1 Do mundo microscópico ao macroscópico.....	24
2.1.2 Campos de força.....	27

2.1.3	Resolvendo as equações de movimento	31
2.2	Simulações de proteínas	33
2.3	Rodando as simulações	36
2.4	Análises das simulações	37
3	Laminarinase de <i>Rhodothermus marinus</i>	43
3.1	Introdução	43
3.2	Descrição e análise estrutural	46
3.3	Metodologia	49
3.3.1	Construção dos sistemas e simulações	49
3.4	Simulações comparativas a 298 e 363 K	54
3.4.1	Estrutura e termoestabilidade	54
3.4.2	Dinâmica na fenda catalítica	60
3.5	Simulações de desnaturação de laminarinases	63
3.6	Conclusões	71
4	Endoglucanase 3 de <i>Trichoderma harzianum</i>	73
4.1	Introdução	73
4.2	Descrição estrutural	74
4.3	Metodologia	76
4.3.1	Construção dos complexos ThEG3-substrato	76
4.3.2	Construção dos sistemas e simulações	78
4.4	Resultados	83
4.4.1	ThEG3 vs TrEG3: estrutura e dinâmica	83
4.4.2	Ligação a oligossacarídeos	87
	Experimentos de <i>docking</i> combinados a	

	simulações de MD	98
	4.4.3 Motivo de CBM no centro catalítico	103
4.5	Conclusões	105
5	β-glicosidase de <i>Aspergillus niger</i>	107
5.1	Introdução	107
5.2	Construção dos sistemas e simulações	110
	5.2.1 Construção dos sistemas	111
	5.2.2 <i>Docking</i>	112
	5.2.3 Simulações de MD	114
5.3	Estrutura do domínio fibronectina III de AnBgl1	115
5.4	Resultados das simulações	118
5.5	O papel do domínio fibronectina III de AnBgl1	122
5.6	Conclusões	124
6	Conclusões gerais e perspectivas	125
7	Referências bibliográficas	129
8	Anexo	147

Lista de Abreviaturas

AnBg11	β -glicosidase de <i>Aspergillus niger</i>
aNLam	Laminarinase de <i>Nocardiosis sp.</i>
CBM	Módulo de ligação à celulose (<i>cellulose-binding module</i>)
CCD	Domínio catalítico (<i>catalytic core domain</i>)
CH	Contatos hidrofóbicos
FnIII	Domínio fibronectina do tipo III
GH	Glicosídeo hidrolase
GHF	Família de glicosídeo hidrolases
LH	Ligação de hidrogênio
MD	Dinâmica molecular
PcLam	Laminarinase de <i>Phanerochaete chrysosporium</i>
PDB	<i>Protein data bank</i>
PS	Ponte salina
R_g	Raio de giração
RmLamR	Laminarinase de <i>Rodothermus marinus</i>
RMSD	Desvio quadrático médio
ThEG3	Endoglucanase 3 de <i>Trichoderma harzianum</i>
TrEG3	Endoglucanase 3 de <i>Trichoderma reesei</i>

Lista de Tabelas

1 Introdução

1.1	Famílias de glicosídeo hidrolases	17
-----	-----------------------------------------	----

2 Laminarinase de *Rhodothermus marinus*

3.1	Composição de aminoácidos relacionados à termoestabilidade de RmLamR e outras GH-16	48
3.2	Resumo dos sistemas simulados	52
3.3	Dados computados a partir das simulações das laminarinases homólogas	56

4 Endoglucanase 3 de *Trichoderma harzianum*

4.1	Modelos de ThEG3 complexada à celopentaose obtidos por <i>docking</i>	77
4.2	Resumo dos sistemas simulados	80
4.3	Frequências de contatos hidrofóbicos e ligações de hidrogênio nas simulações de ThEG3-pt* e CfCBM-pt	104

5 β -glicosidase de *Aspergillus niger*

- 5.1 Descrição dos sistemas simulados..... 112
- 5.2 Energias de afinidade dos modos de ligação gerados por *docking* e os respectivos sítios ocupados..... 113

Lista de Figuras

1 Introdução

1.1	Modelo genérico de estrutura lignocelulósica nas paredes celulares de plantas	6
1.2	Monômeros principais que formam lignina.....	7
1.3	Esquema de mecanismo de hidrólise de celulose de celulases não-complexadas.....	10
1.4	Tipos de fenda catalítica encontradas em glicosídeo hidrolases.....	12
1.5	Esquema ilustrativo do mecanismo processivo de hidrólise de celulose por uma exoglucanase (CBHI de <i>T. reesei</i>)	14
1.6	Mecanismos de hidrólise com retenção (A) e com inversão (B) de configuração anomérica.....	15

2 Metodologia.

2.1	Representação esquemática dos componentes principais de um campo de força.....	29
2.2	(A) Níveis de energia referentes a eventos dinâmicos de proteínas enoveladas. (B) Escala de tempo associadas aos processos dinâmicos em proteínas. (C) Curva tridimensional de estados conformacionais acessíveis a uma proteína.....	34

3 Laminarinase de *Rhodothermus marinus*

3.1	Estrutura global e topologia da laminarinase de <i>R. marinus</i>	46
3.2	Alinhamento com base estrutural das sequências primárias de RmLamR, aNLam e PcLam.....	50
3.3	Alinhamento estrutural de laminarinases de <i>R. marinus</i> , <i>Nocardiosis sp.</i> e <i>P. chrysosporium</i>	51
3.4	A) Organização topológica de pontes salinas de laminarinases. B) Conectividade das pontes salinas.....	57
3.5	Entrada de água entre as camadas de fitas- β	59
3.6	Mobilidades das cadeias laterais dos resíduos computadas por RMSF.....	61
3.7	Análises das modificações estruturais globais ocorridas nas simulações de PcLam, aNLam e RmLamR a 298 K e a 500 K.....	66
3.8	Estruturas das laminarinases em diferentes momentos das simulações de desnaturação.....	67
3.9	Flutuações dos resíduos em RmLamR em relação à estrutura média.....	68
3.10	Flutuações dos resíduos em aNLam em relação à estrutura média.....	69
3.11	Flutuações dos resíduos em PcLam em relação à estrutura média.....	70

4 Endoglucanase 3 de *Trichoderma harzianum*

- 4.1 A estrutura de β -jelly roll da endoglucanase 3 de *Trichoderma harzianum*..... 75
- 4.2 Evolução no tempo de simulação dos desvios quadráticos médios dos átomos da cadeia principal dos modelos ThEG3-substrato, além dos complexos cristalográficos CfCBM-pt e TmEG3-tt..... 82
- 4.3 Alinhamento das sequências primárias de ThEG3 e TrEG3..... 83
- 4.4 Fenda de ligação ao substrato da enzima ThEG3..... 84
- 4.5 Perfil de mobilidade ao longo da sequência primária de ThEG3 e TrEG3..... 86
- 4.6 (A) Posições iniciais dos substratos em ThEG3 e (B) as interações dos resíduos catalíticos com a celotetraose..... 88
- 4.7 (A) Detalhes das fendas de ligação ao substrato de ThEG3 e TmEG3 e suas interações com o substrato. (B) Distribuição de resíduos de diferentes naturezas na fenda catalítica de ambas as enzimas..... 89
- 4.8 Evolução temporal das distâncias entre os resíduos catalíticos e átomos do substrato que são alvos na reação de hidrólise catalítica..... 91
- 4.9 Estruturas dos domínios de ligação à celulose (CBMs) de diferentes famílias..... 93
- 4.10 Sobreposição de *snapshots* das simulações que mostram como variam as posições dos resíduos aromáticos Tyr7, Trp23 e Tyr112, e dos oligossacarídeos celotetraose, celopentaose e celopentaose*..... 94
- 4.11 Evolução no tempo das energias de interação entre os resíduos

	aromáticos Tyr7, Trp23 e Tyr112 e os substratos.....	96
4.12	<i>Snapshots</i> de diferentes estágios ao longo das simulações de ThEG3-tt (A) e ThEG3-pt* (B).....	97
4.13	Modelos Autodock-1 (A), Autodock-2 (B) e Autodock-3 (C).....	99
4.14	(A) Desvio quadrático médio ao longo das simulações dos átomos da cadeia principal de ThEG3 nos modelos de <i>docking</i> , tomando a estrutura cristalográfica como referência. (B) Evolução no tempo das distâncias entre os resíduos catalíticos e os prováveis alvos de reação de hidrólise no substrato. C) Evolução temporal nas energias de interação entre os resíduos aromáticos Tyr7, Trp23 e Tyr112 e os substratos obtidos das simulações dos modelos de <i>docking</i> de ThEG3-oligossacarídeo.....	101
4.15	Evento de deslocamento lateral observado na simulação do modelo Autodock-3.....	102
4.16	(A) Os resíduos de ThEG3 e CfCBM em contato hidrofóbico com a celopentaose e (B) resíduos que interagem via ligações de hidrogênio.....	103

5 β -glicosidase de *Aspergillus niger*

5.1	Representação de lignina de choupo e dímeros de guaiacil e siringil modelados para as simulações de MD.....	111
5.2	Modos de ligação de LGG sugeridos por <i>docking</i>	113
5.3	A) Estrutura de "sanduíche" de folhas- β do domínio fibronectina do tipo III. B) Superfície de van der Waals mostrando a exposição de resíduos de diferentes naturezas.....	116

5.4	Os cinco sítios de ligação apontados pelos cálculos com o programa Autodock Vina.....	117
5.5	Energia de interação entre os dímeros LGG e LSS com os resíduos de FnIII.....	118
5.6	Interações entre fragmentos de lignina e o domínio FnIII verificadas nas simulações. (A) Perfis de energia de interação entre os dímeros LGG e LSS e resíduos do domínio FnIII. Os painéis (B), (C) e (D) são <i>snapshots</i> das simulações representativos dos modos de ligação lignina – FnIII.....	120
5.7	Gráficos que mostram simultaneamente a evolução temporal da energia de interação arginina-LGG e orientação relativa do cátion guanidíneo de R e anel aromático de LGG.....	121

Capítulo 1

Introdução

A forte preocupação em reduzir os impactos ambientais causados pelo consumo global de combustíveis fósseis e a necessidade de evitar a dependência do mercado altamente flutuante do petróleo tem levado a um interesse crescente na utilização do álcool etílico como uma fonte alternativa e renovável de energia. Em países como o Brasil e os Estados Unidos, este biocombustível tem sido produzido em grandes quantidades por fermentação alcoólica dos extratos de matérias-primas como cana-de-açúcar e milho, respectivamente.

O Brasil é o maior produtor de etanol da cana-de-açúcar do mundo e ocupa uma posição de liderança na tecnologia envolvida na produção, calcada em projetos genuinamente nacionais desenvolvidos nas décadas de 70 e 80 com o Pró-álcool e nos massivos investimentos atuais em pesquisa na área (Bon & Ferrara, 2007). Entretanto, apenas um terço desta biomassa é aproveitado na produção de biocombustível. Cada tonelada de cana-de-açúcar processada durante a produção de açúcar ou etanol gera, aproximadamente, 300 kg de bagaço. Um aumento significativo e sustentável na produção de bioetanol no Brasil poderia ser alcançado com o aproveitamento maximizado deste material.

O interesse na diversificação de matérias-primas para fins energéticos e o melhor aproveitamento da biomassa que já é explorada, está impulsionando um novo salto tecnológico: a produção de etanol a partir de material lignocelulósico. No caso da exploração da cana-de-açúcar, por exemplo, estima-se que o aproveitamento de subprodutos como o bagaço e a palha poderia elevar em 30 a 40% a produção de álcool

para uma mesma área plantada (Pacheco, 2011).

O domínio de técnicas da desconstrução da biomassa lignocelulósica para obtenção de açúcares fermentáveis a biocombustíveis ou outros biomateriais é muito vantajosa, pois possibilitaria a exploração de uma enorme diversidade de matérias-primas que muitas vezes são acumuladas, como: palha e sabugo de milho, palha de arroz, palha de coco, resíduos de madeira (lixos de papel, por exemplo), capim e resíduos da produção de sucos (Bon *et al.*, 2008).

A conversão do material lignocelulósico introduz novos desafios que alcançam muitas escalas. Em um extremo, há que se consolidar e regulamentar a exploração da biomassa de forma efetivamente sustentável e economicamente vantajosa. Uma vez consolidada, no outro extremo, a conversão da biomassa consiste também em um desafio considerável, dada a complexidade e resistência do material, associada à própria função da parede celular de proteger as plantas de deterioração.

O uso da biomassa como substrato para a fermentação alcoólica inclui a hidrólise dos polissacarídeos presentes nestes materiais para liberação dos açúcares fermentáveis. A hidrólise ácida leva à formação de produtos tóxicos, principalmente furfurais e substâncias aromáticas derivadas da lignina, que limitam o metabolismo da célula viva. Por esta razão, a hidrólise enzimática tem sido considerada o procedimento mais vantajoso (Bon *et al.*, 2008). Um coquetel enzimático permite a degradação com maior especificidade para cada componente do material, e em condições mais brandas. Entretanto, em escala industrial, a catálise enzimática é uma das etapas mais caras da bioconversão biomassa-etanol celulósico, devido à baixa eficiência e alto custo dos coquetéis enzimáticos disponíveis atualmente. Visando desenvolver estratégias para reduzir os custos deste processo através do aprimoramento destas enzimas para uso em

biorrefinarias¹, grandes esforços têm sido investidos no estudo de micro-organismos celulolíticos e na investigação dos mecanismos de hidrólise enzimática, assim como das propriedades estruturais e dinâmicas gerais das glicosídeo hidrolases (GHs)².

Tanto empresas multinacionais, como a Novozymes³, quanto grandes empresas nacionais, como Petrobrás⁴ e GranBio⁵, já apresentam projetos muito próximos de se tornarem realidade para a exploração massiva da cana-de-açúcar com a finalidade de produção de etanol de segunda geração⁶. Dentre os maiores projetos nacionais de pesquisa na área, encontram-se o Programa BIOEN da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) e os projetos do Instituto de Ciência e Tecnologia do bioetanol (INCT do bioetanol). Nestes programas, estão envolvidos diversos grupos de pesquisa cobrindo estudos que vão desde a transformação gênica de materiais lignocelulósicos, prospecção de microorganismos (fungos e bactérias) capazes de expressar enzimas celulolíticas e proteínas correlatas, a ensaios de catálise enzimática e estudos de tais processos em nível molecular ou atômico por meio de espectroscopia por ressonância magnética nuclear, espectrometria de massas e difração de raios-X. Neste contexto, o Grupo de Simulações por Dinâmica Molecular do IQ/ UNICAMP vem atuando de forma sinérgica junto a pesquisadores das áreas de biofísica e biologia estrutural molecular através do uso de técnicas avançadas de simulação computacional por dinâmica molecular, auxiliando na compreensão das bases moleculares associadas ao comportamento funcional dessas enzimas, conforme veremos adiante.

1 O termo biorrefinaria tem sido utilizado para designar o uso maximizado da biomassa através de processos químicos ou bioquímicos para a produção de energia e/ou substâncias variadas com geração mínima de resíduos e emissões.

2 São encontradas algumas variações na literatura em português para designar a família de enzimas GH (do inglês, glycoside hydrolases; Cantarel et al., 2009). São usados, principalmente, glicosil hidrolases, hidrolases de glicosídeo, e glicosídeo hidrolases. Aqui, julgaram-se mais apropriadas a segunda e terceira variantes para referenciar a função enzimática de clivagem de ligações glicosídicas.

3 http://novozymes.com.br/produtos/producoes_produtos_etanol.htm, acesso em junho de 2013.

4 <http://www.petrobras.com.br/pt/energia-e-tecnologia/fontes-de-energia/biocombustiveis/>, acesso em junho de 2013.

5 <http://www.granbio.com.br/produtos/biocombustiveis/>, acesso em junho de 2013.

6 Combustíveis de segunda geração são os combustíveis gerados a partir de biomassa.

Neste trabalho, foram utilizadas técnicas de simulação computacional por dinâmica molecular em três estudos distintos de enzimas de grande interesse envolvidas na sacarificação de biomassa celulósica:

- Laminarinase de *Rodthotermus marinus*
- Endoglucanase de *Trichoderma harzianum*
- β -glicosidase de *Aspergillus niger*

Cada um dos três estudos, abordados detalhadamente em diferentes capítulos da tese, será comentado brevemente no final deste capítulo, mas, antes, é feita uma pequena introdução dos aspectos gerais relevantes ao uso de enzimas como catalisadores para sacarificação do material lignocelulósico.

1.1 Matérias-primas lignocelulósicas

Os materiais lignocelulósicos são os compostos orgânicos mais abundantes da biosfera, representando 50% da biomassa terrestre (Sarko, 1997). Estes materiais nada mais são do que as paredes celulares das plantas, que dão rigidez às células e previnem o ataque de patógenos.

Nos modelos recentes, as paredes celulares são descritas como microfibrilas de celulose (20-50% em massa em uma base seca) cercadas por hemicelulose (15-35%) e embebidas em uma matriz de lignina (10-30 %), enquanto que proteínas, lipídios, açúcares solúveis e minerais são componentes minoritários (Bon *et al.*, 2008; Nookaraju *et al.*, 2013; D'Almeida, 1988; Wayman & Parekh, 1990). Há ainda as pectinas, também polissacarídeos, que controlam a porosidade da parede celular. As composições químicas diferem entre as espécies de plantas, o que influencia na susceptibilidade à desconstrução da parede celular – também conhecida como recalcitrância da biomassa. Muitos estudos têm sido direcionados à manipulação genética da composição da parede

celular de cana-de-açúcar para facilitar a sacarificação (Nookaraju *et al.*, 2013). Um modelo genérico de uma parede celular de planta, bem como os elementos que a compõem são apresentados na Figura 1.1.

Por ser o mais abundante e homogêneo, a celulose é o componente do material lignocelulósico mais estudado para a bioconversão. A molécula de celulose é um homopolissacarídeo linear que contém até 15.000 unidades de *D*-glicose unidas por ligações glicosídicas β -1,4, sendo sua unidade monomérica a união de duas glicoses, a celobiose. As cadeias individuais de celulose interagem umas com as outras através de ligações de hidrogênio e por forças dispersivas, formando as microfibrilas de celulose (compostas por 30 a 36 cadeias do glicano). O arranjo de cadeias celulósicas na microfibrila permite um empacotamento paralelo bem justo, formando regiões de elevado grau de cristalinidade, alternadas por regiões de menor grau de ordenação, chamadas regiões amorfas (Chundawat *et al.*, 2011).

A morfologia cristalina da celulose é responsável pela baixa taxa de sacarificação das cadeias. Este arranjo difere do encontrado em outra fonte muito importante de glicose, o amido. Neste caso, a ligação α -1,4 entre unidades de glicose faz com que estes polímeros assumam estruturas helicoidais (Nelson & Cox, 2005). Devido a isso, a hidrólise enzimática do amido, com as hidrolases amilases, é um processo facilitado, relacionado ao seu papel de reserva energética em plantas.

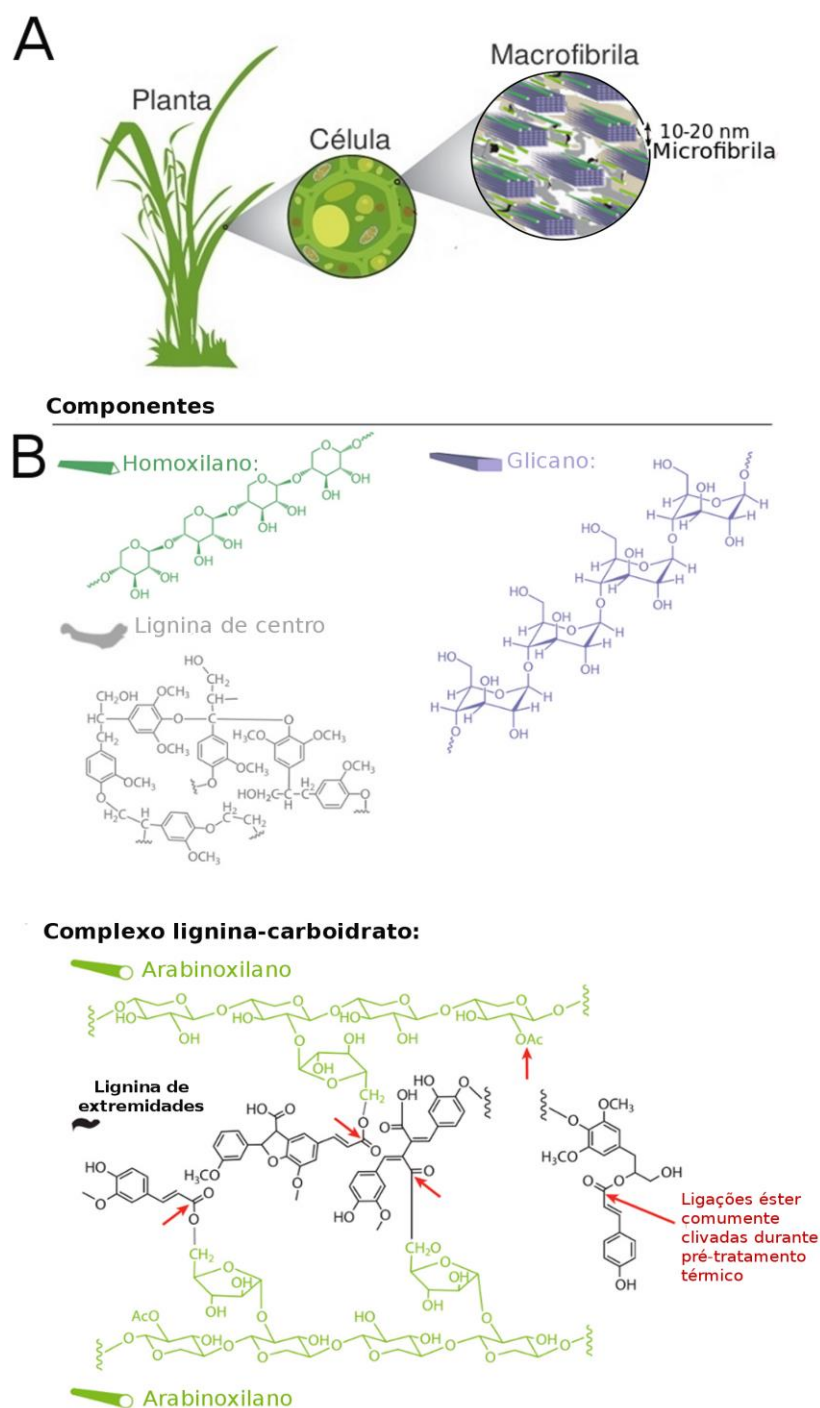


Figura 1.1. Modelo genérico de estrutura lignocelulósica nas paredes celulares de plantas. Figura adaptada das referências (Chundawat *et al.*, 2011; Rubin, 2008). A) Material lignocelulósico de plantas. A ilustração demarca diferentes escalas espaciais envolvidas na rota de degradação da biomassa. B) Componentes da parede celular. No painel inferior, ligações covalentes formadas entre extremidades de polímeros de lignina e cadeias de hemicelulose, as quais formam uma complexa rede de ligações cruzadas na parede celular.

Cercando as fibrilas de celulose, encontram-se as hemiceluloses. As cadeias não ramificadas de hemicelulose (xiloglucanos, homoxilanos e mananos) formam ligações de hidrogênio com as fibrilas de celulose, ao passo que cadeias ramificadas (com ácido urônico e arabinose, por exemplo), acredita-se, podem estar ligadas covalentemente a outras cadeias de hemicelulose ou à lignina (Chundawat *et al.*, 2011). As hemiceluloses são polímeros bem menores do que as celulosas, apresentado entre 100 e 200 unidades glicosídicas. Mesmo assim, a complexidade deste heterossacarídeo torna necessária a participação de diversas enzimas para uma hidrólise enzimática. O pré-tratamento em ácido diluído da biomassa celulósica, entretanto, possibilita a remoção da hemicelulose logo nesta etapa. Assim, apesar do interesse em elaborar um processo completamente enzimático na degradação da biomassa, os estudos de hemicelulases não estão tão avançados quanto os de celulasas.

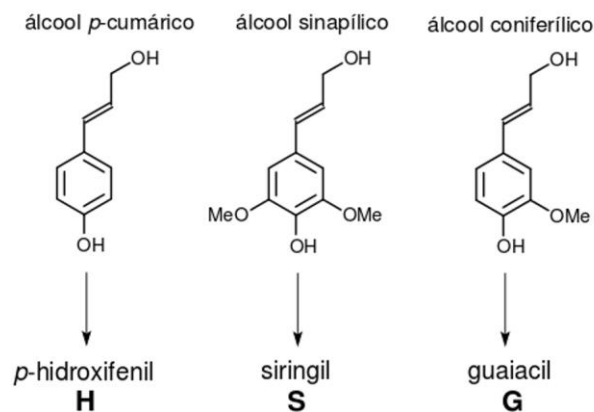


Figura 1.2. Monômeros principais que formam lignina.

Por fim, toda a matriz hemicelulósica encontra-se embebida em uma complexa rede polimérica de natureza aromática, a lignina. Este polímero interage não só via ligações de hidrogênio, interações de van der Waals, mas também forma ligações covalentes com hemicelulose. A lignina é uma macromolécula opticamente inativa,

composta principalmente de três diferentes unidades monoméricas: *p*-hidroxifenil (H), guaiacil (G), e siringil (S), os quais têm em comum um centro de álcool cumárico, mas variam no grau de metoxilação (Figura 1.2). Tais resíduos são unidos entre si por ligações do tipo éter e estabelecem ligações cruzadas entre si. Embora o foco atual de processos de pré-tratamento da biomassa lignocelulósica seja aumentar a acessibilidade da celulose, há interesse na preservação da lignina para evitar a liberação de resíduos tóxicos e para que também ocorra sua biodigestão por complexos enzimáticos.

Os polímeros da parede celular arranjam-se, então, como um compósito nano-estruturado. Para “desmontar” completamente esta estrutura, as principais enzimas utilizadas são as glicosídeo hidrolases, que clivam as ligações glicosídicas dos polissacarídeos deste material para obter açúcares solúveis.

1.2 Hidrólise enzimática da biomassa

Os coquetéis enzimáticos que degradam a biomassa são compostos de enzimas que trabalham sinergicamente, como as endoglucanases, exoglucanases e β -glicosidades que degradam a celulose e múltiplas hemicelulases. São encontradas também nesta mistura, proteínas auxiliares não-hidrolíticas, chamadas de suoleninas (do inglês, *swollenins*) e expansinas, que atuam diminuindo o empacotamento e, portanto, o grau de cristalinidade da matriz celulósica (Lynd *et al.*, 2002).

A maioria dos fungos e bactérias capazes de digerir material lignocelulósico secretam enzimas não complexadas. Entretanto, algumas bactérias apresentam grandes complexos enzimáticos chamados celulosomos (Bayer *et al.*, 2004). As enzimas de fungos são as mais estudadas devido a maior disponibilidade comercial, uma vez que secretam quantidades mais elevadas de proteínas. Atualmente, a maioria das enzimas disponíveis comercialmente são produzidas por cepas geneticamente modificadas dos

fungos *Trichoderma reesei* (*Hypocrea jeecorina*) e *Aspergillus niger* (Kumar *et al.*, 2008).

Como o foco deste trabalho incide nas celulases, na Seção seguinte, descreve-se como essas enzimas atuam na degradação da celulose.

1.2.1 Celulases

Atualmente, as celulases são o terceiro tipo de enzima mais utilizada no mundo (Wilson, 2009). Tem sido utilizada no processamento de algodão, em reciclagem de papel, em produção de detergente, extração de néctares, e em aditivos de alimentos de animais (Bhat, 2000). Em médio prazo, é provável que sejam amplamente empregadas na produção de etanol de segunda geração.

A degradação da celulose envolve a participação de uma mistura de enzimas que clivam ligações β -1,4 e se complementam em suas funções. Elas são geralmente divididas em enzimas endoglucanases (EC 3.2.1.4), exoglucanases ou celobiohidrolases (EC 3.2.1.91) e β -glicosidases (EC 3.2.1.91) (Zhang *et al.*, 2006; Henrissat, 1994; Lynd *et al.*, 2002; Teeri, 1997). As endoglucanases clivam ligações glicosídicas aleatoriamente nas cadeias celulósicas em regiões amorfas, gerando novas extremidades redutoras e não-redutoras, e liberando celo-oligossacarídeos solúveis. As exoglucanases atuam nestas extremidades, processivamente, liberando moléculas de celobiose na solução. As exoglucanases também podem ser específicas para o tipo de terminal em que atuam: as enzimas do tipo I atuam em terminais redutores (CBHI), enquanto que as do tipo II (CBHII) hidrolisam terminais não-redutores. Finalmente, as β -glicosidases (β -Gase) completam o processo com a hidrólise das moléculas de celobiose, liberando glicose (Figura 1.3).

Uma única enzima destas não é capaz de hidrolisar a celulose com eficiência, sendo necessária à ação sinérgica e complementar destes três tipos de enzima. Quando

atuam em conjunto, o rendimento é muito melhor do que a soma dos rendimentos individuais (Lynd *et al.*, 2002). O sinergismo entre endoglucanases e exoglucanases é o tipo mais estudado e tem importância destacada na hidrólise da celulose, mas outros tipos de sinergismo, como entre endoglucanases ou entre exoglucanases, também ocorrem (Tuka, Zverlov & Velikodvorskaya 1992; Walker, Wilson & Irwin 1992; Din, Damude, *et al.* 1994; Zhang & Lynd, 2004). O consumo de celobiose pelas β -glicosidases, por exemplo, é um fator crítico, não apenas para finalizar a degradação da celulose em glicose, como para diminuir o efeito inibidor da celobiose sobre as endo e exoglucanases, sendo este mais um exemplo de sinergia (Bon *et al.*, 2008).

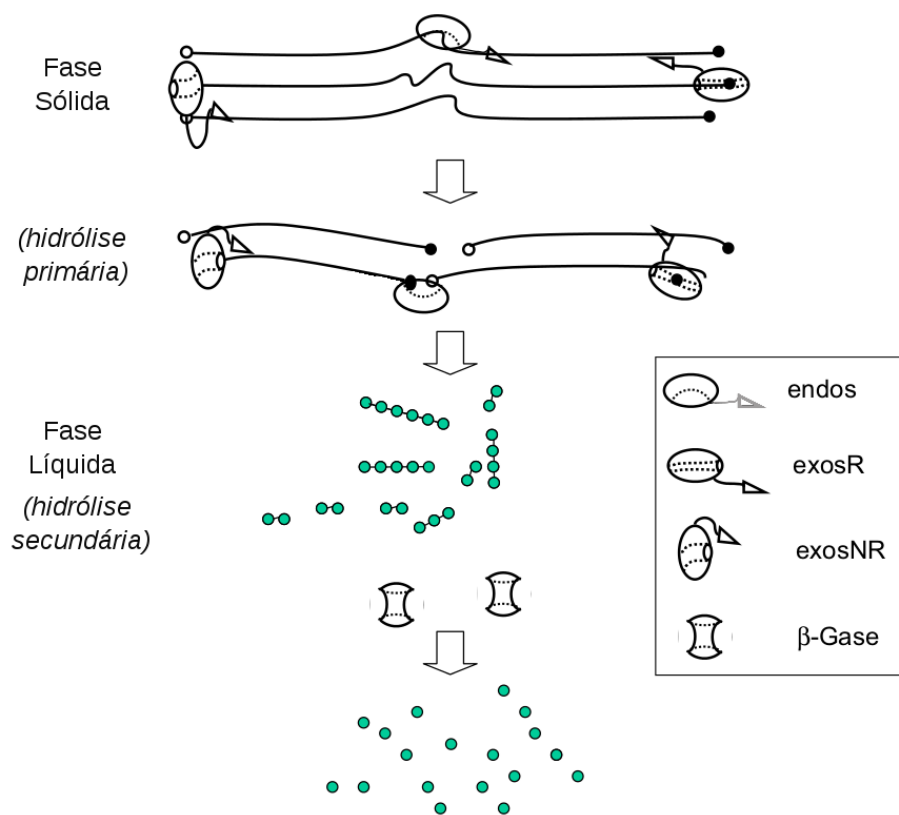


Figura 1.3. Esquema de mecanismo de hidrólise de celulose por enzimas não-complexadas (Figura adaptada da referência (Zhang *et al.*, 2006)). O substrato celulósico é representado com linhas paralelas nas regiões cristalinas, e com linhas irregulares nas regiões amorfas representada com curvas. As unidades de glicose são representadas com pequenos círculos em verde.

Substratos diferentes utilizados na caracterização da atividade enzimática são ferramentas para identificar o modo de atuação (endo, exo, β -Gase) de uma enzima, mas ressalta-se que a sinergia entre elas dificulta uma quantificação precisa. Os ensaios de atividade de endoglucanases são feitos, geralmente, com o uso de carboximetilcelulose como substrato (CMC) e oligossacarídeos parcialmente solúveis (Zhang *et al.*, 2006; Generoso *et al.*, 2012; Karlsson *et al.*, 2002). A CMC é um polímero cuja estrutura molecular é similar a de cadeias de celulose, mas apresenta hidroxilas substituídas por grupos acetato nas unidades de glicose, em graus variados de substituição, que tornam este polímero solúvel em água. A estrutura livre permite o ataque enzimático de endoglucanases randomicamente, produzindo uma mudança rápida no grau de polimerização. Já a celulose microcristalina avicel é um substrato adequado para verificar a atividade de exoglucanases (Zhang *et al.*, 2006; Wood & Bhat, 1988). A medida de atividade global enzimática utilizando como substrato o papel filtro Whatman N^o 1 (Fpase) é muito comum também (Ghose, 1987).

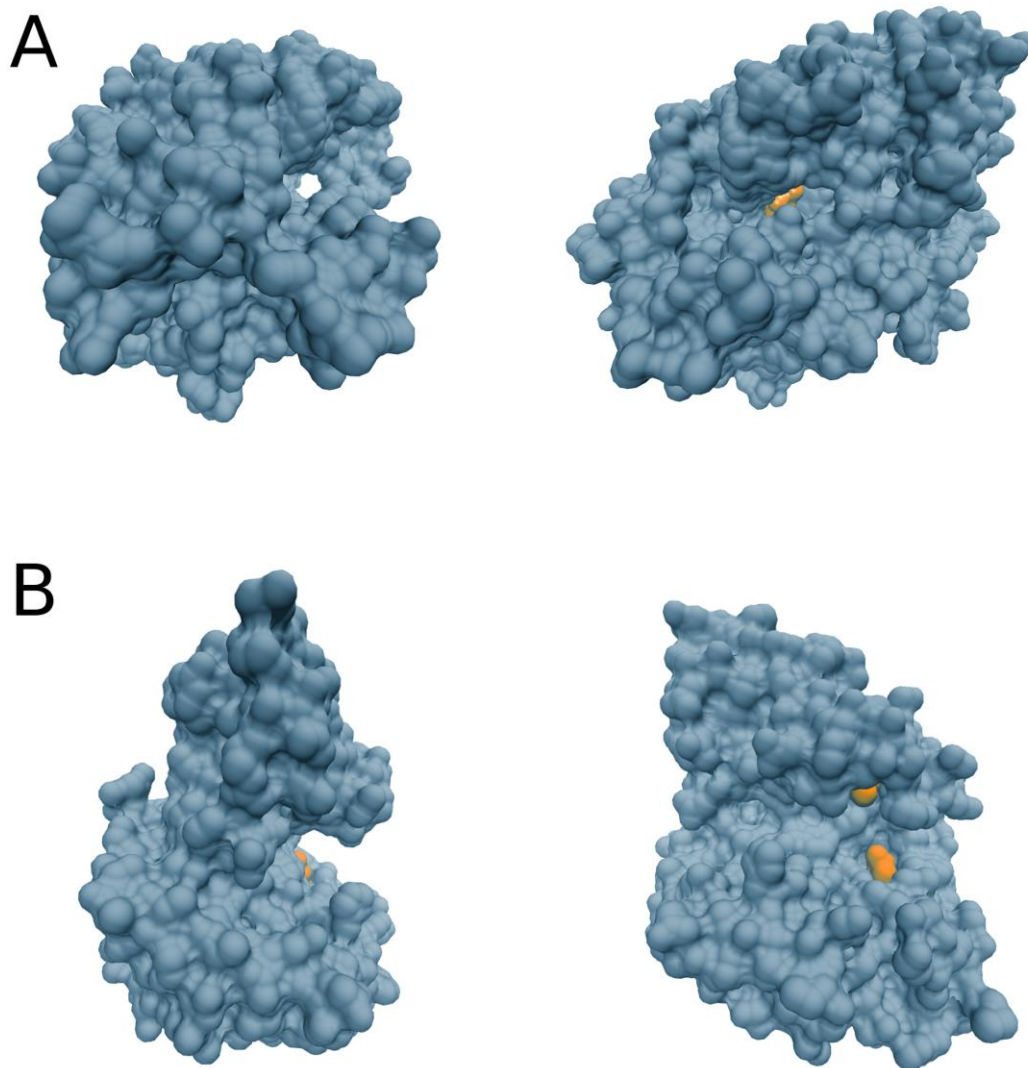


Figura 1.4. Tipos de fenda catalítica encontradas em glicosídeo hidrolases. A) Fenda tubular em celobiohidrolase II de *Trichoderma reesei* (PDB id: 2V31; Fagerstrom *et al.*, 2009); (B) Fenda aberta em endoglucanase de *Thermobifida fusca* (PDB id 2BOE; Larsson *et al.*, 2005). Os resíduos catalíticos são destacados em laranja.

A forma de ataque e o tipo de substrato preferencial destas enzimas são determinados por algumas características estruturais típicas de cada categoria. Assim, enzimas processivas, ou seja, capazes de realizar múltiplos eventos hidrolíticos sem liberar a cadeia celulósica apresentam uma fenda catalítica com um formato tubular. Os subsítios, incluindo os resíduos catalíticos, estão contidos neste tubo, dificultando a

liberação da cadeia polissacarídica após um evento catalítico. Este tipo de fenda é característico de exoglucanases. As endoglucanases, por sua vez, dispõem de um aparato catalítico localizado em uma fenda aberta, que permite ligações randômicas ao longo da cadeia polimérica. Note que esta diferenciação é representada na Figura 1.4. Outra característica estrutural que deve ser mencionada é a de que estas enzimas que atuam através da adsorção à superfície celulósica geralmente apresentam estrutura bimodular, formada por um domínio maior onde ocorre a hidrólise, chamado de domínio catalítico (CCD, do inglês, *catalytic core domain*) e por um domínio menor, o domínio de ligação à celulose (CBM, do inglês, *cellulose-binding domain*). Este domínio é responsável por aumentar a concentração das celulasas na superfície celulósica. O domínio CBM é presente na maioria das endo e exoglucanases de estrutura conhecida. A Figura 1.5 ilustra como ocorre a hidrólise de celulose através da catálise da enzima CBHI de *T. reesei*. Sugere-se que as etapas envolvidas incluam a adsorção ao substrato através dos domínios CBM e CCD em conjunto, reconhecimento da cadeia celulósica livre, encaixe inicial desta cadeia no interior do tubo catalítico, reação de hidrólise, expulsão da celobiose resultante, e reestruturação de um complexo cataliticamente ativo. Tais etapas devem ser comuns na atuação de celulasas processivas. Uma exceção ao arranjo multimodular é tratada no Capítulo 4, que relata o estudo da endoglucanase 3 de *Trichoderma harzianum*, constituída apenas pelo centro catalítico.

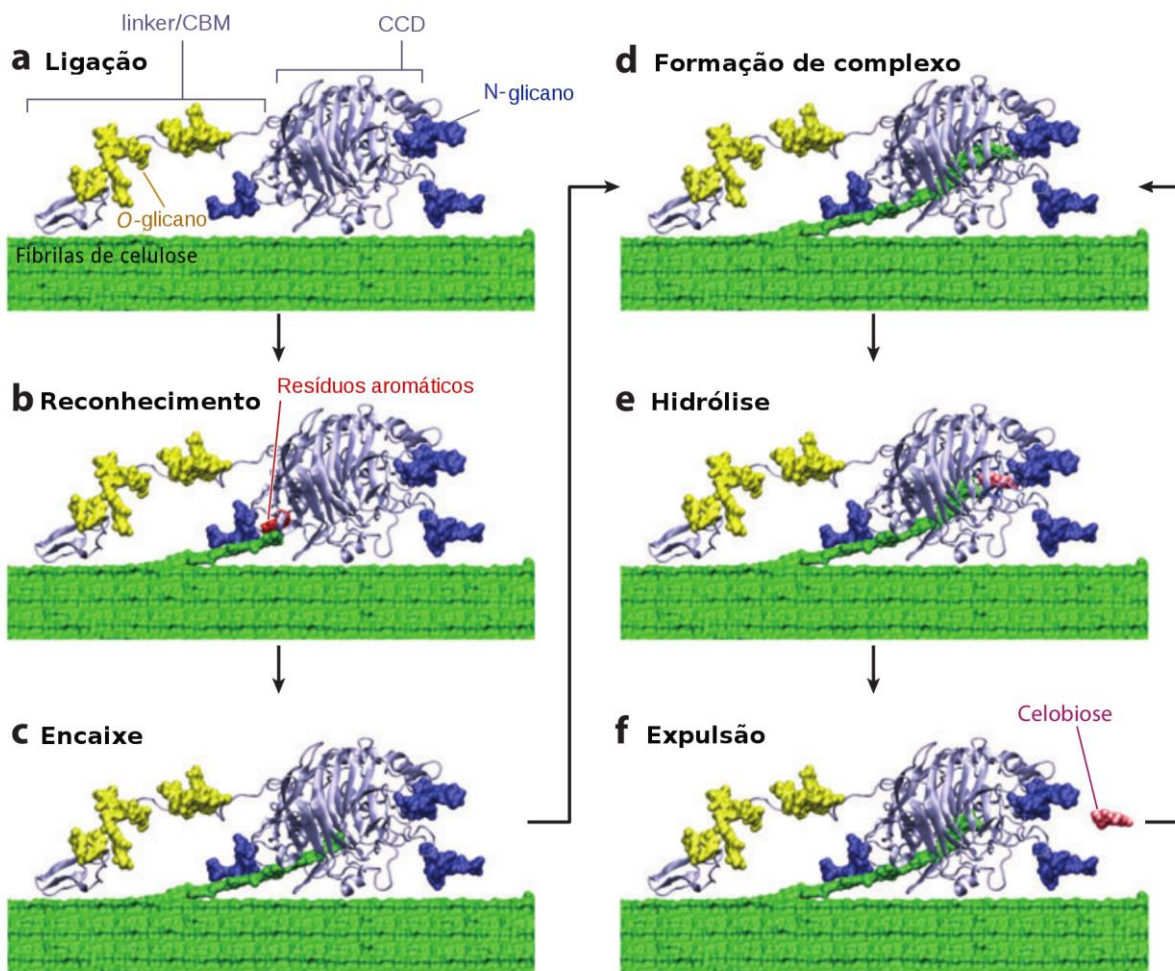


Figura 1.5. Esquema ilustrativo do mecanismo processivo de hidrólise de celulose por uma exoglucanase (CBHI de *T. reesei*). Figura adaptada da referência Chundawat, *et al.*, 2011.

A distinção endo e exo de enzimas, em geral, é muito eficiente. No entanto, algumas vezes ela torna-se de difícil determinação, uma vez que muitas enzimas exibem propriedades intermediárias entre exo e endo (Tomme *et al.*, 1996). Este é o caso, por exemplo, de enzimas que atacam de uma forma típica endo, mas exibem atividade processiva.

1.2.2 Mecanismos de reação de glicosídeo hidrolases

A hidrólise enzimática de uma ligação glicosídica ocorre através de uma catálise ácida que requer dois resíduos essenciais: um doador de próton e um nucleófilo/base. Além disso, há o consumo de uma molécula de água na reação.

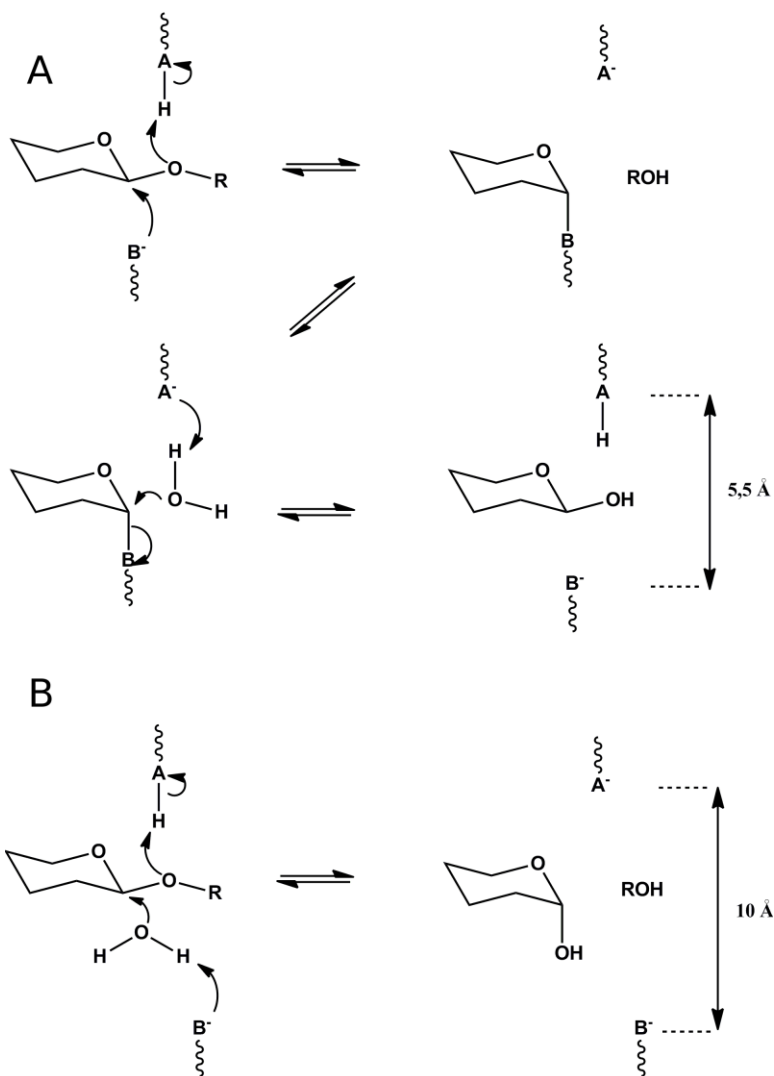


Figura 1.6. Mecanismos de hidrólise com retenção (A) e com inversão (B) de configuração anomérica.

São descritos dois mecanismos gerais de hidrólise, um em que a configuração

anomérica do produto mantém-se após a reação e outro, em que a configuração termina invertida (Figura 1.6). Considerando que estas reações ocorrem de forma concertada, a distância entre os resíduos catalíticos é o aspecto chave que define que tipo de mecanismo deve ocorrer. Assim, o mecanismo em que há inversão ocorre somente quando há espaço suficiente para uma molécula de água penetrar a região entre o resíduo básico e o sítio de ataque nucleofílico no substrato. As distâncias médias entre os dois resíduos são de $\sim 5,5 \text{ \AA}$ em enzimas de retenção e 10 \AA , em enzimas de inversão (Davies & Henrissat, 1995).

1.2.3 Classificação das glicosídeo hidrolases

Existem diversos sistemas utilizados para classificar as glicosídeo hidrolases, incluindo o que especifica o modo de ataque enzimático (enzimas endo *versus* exo). A categorização mais simples é baseada na especificidade do substrato considerando as recomendações da *International Union of Biochemistry and Molecular Biology* (IUBMB) (<http://www.chem.qmul.ac.uk/iubmb/>, último acesso em abril de 2013) e é expressa pelo número EC (*Enzyme Commission*) de uma dada enzima. Às *O*-glicosídeo hidrolases atribui-se o código EC 3.2.1.x, em que x representa a especificidade do substrato (em alguns casos, representa também o mecanismo molecular ou o tipo de ligação a ser clivada, isto é, β -1,4, β -1,3, etc.). As numerações anteriores (3, 2, 1), gradualmente especificam o fato de serem hidrolases, glicosilases e *O*-glicosilases, respectivamente. Dado sua simplicidade, este tipo de classificação é amplamente utilizado para descrever glicosídeo hidrolases (como já foi feito ao longo deste texto mesmo). A desvantagem deste sistema é que ele não acomoda apropriadamente enzimas que atuam em substratos variados. Endoglucanases, por exemplo, que são consideradas tipicamente celulases, são ativas também em vários níveis, em diferentes substratos, como xilano, xiloglicano, β -glicano e outros substratos artificiais.

Há outro sistema que classifica as glicosídeo hidrolases com base na similaridade na sequência primária (Henrissat, 1991, Henrissat & Davies, 1997). O princípio aqui é o de que, uma vez que sequência e estrutura são relacionadas, informações úteis de estrutura e mecanismo podem ser inferidas apenas através do conhecimento da sequência primária. A Tabela 1.1 exemplifica algumas classes deste sistema. No banco de dados CAZy (*Carbohydrate-Active enZymes database*, webpage: www.cazy.org, Cantarel *et al.*, 2009) constam, atualmente, 132 famílias de GHs definidas. Como estruturas tridimensionais podem ser conservadas apesar de uma baixa identidade, algumas destas famílias cujas estruturas estão relacionadas por homologia são agrupadas em clãs.

Tabela 1.1. Exemplos de famílias de glicosídeo hidrolases baseadas na semelhança de sequências primárias (Henrissat & Davies, 1997; Cantarel *et al.*, 2009).

Família	Mecanismo*	Clã	Enovelamento
1	RET	GH-A	(β/α) ₈
2	RET	GH-A	(β/α) ₈
3	RET	-	
4	RET	-	
5	RET	GH-A	(β/α) ₈
6	INV	-	
7	RET	GH-B	β -jelly roll
8	INV	GH-M	(α/α) ₆
9	INV	-	(α/α) ₆
10	RET	GH-A	(β/α) ₈
11	RET	GH-C	β -jelly roll
12	RET	GH-C	β -jelly roll
13	RET	GH-H	(β/α) ₈
14	INV	-	(β/α) ₈
15	INV	GH-L	(α/α) ₆
16	RET	GH-B	β -jelly roll

* INV, mecanismo de inversão; RET, mecanismo de retenção.

Como exemplo, vejamos a classificação nos sistemas apresentados de uma das enzimas estudadas neste trabalho, a laminarinase de *Rhodothermus marinus* (Capítulo 3). Esta enzima é classificada como uma endoglucanase, mas atua especialmente sobre a laminarina (ou laminarano), um polissacarídeo encontrado em algas marrons formado de glicoses unidas por ligações β -1,3 e β -1,6, que é substrato de enzimas do grupo EC 3.2.1.6. De acordo com sua sequência primária, esta enzima pertence à família GH-16, apresentando um enovelamento do tipo β -jelly roll.

Há, ainda, uma classificação mais atual de celulases que tem sido muito empregada (Henrissat, 1998), em que os grupos são denominados CelNx, em que N é o número da família segundo (Henrissat, 1991) e x refere-se à ordem de descoberta da enzima. Por exemplo, Cel7A e Cel7B de *T. reesei* são as antigas CBHI e EGI, ambas pertencentes à família GH-7.

1.3. Temas estudados

Durante os últimos oito anos, os grupos de pesquisa supervisionados pelo Prof. Munir S. Skaf e pelo Prof. Igor Polikarpov têm sido bem sucedidos na combinação de pesquisas experimentais e de simulações computacionais no estudo de várias proteínas (Textor *et al.*, 2013; Prates *et al.*, 2013; Puhl *et al.*, 2012; Silveira *et al.*, 2012, são exemplos recentes). Foram obtidas informações importantes a nível molecular destes sistemas, as quais não seriam acessíveis pelo método experimental ou teórico isoladamente. Do mesmo modo, os estudos reportados neste texto foram realizados em estreita colaboração entre os dois grupos de pesquisa, especialmente vinculada ao Programa BIOEN da Fapesp. Sendo assim, a escolha das enzimas foi feita em conjunto e as principais estruturas de enzimas envolvidas foram resolvidas e fornecidas pela equipe experimentalista.

Como já mencionado, foram três os principais estudos desenvolvidos nesta Tese, resumidos abaixo:

Laminarinase de *Rhodothermus marinus*: Bases estruturais de sua termoestabilidade e termofilicidade (Capítulo 3)

As glicosídeo hidrolases são de grande utilidade industrial e biotecnológica. Devido a estas aplicações, é frequentemente desejável que as glicosídeo hidrolases apresentem estabilidade e atividade em condições extremas, como altas temperaturas e *p*Hs muito ácidos ou muito alcalinos. Neste trabalho, a estrutura da enzima hipertermofílica laminarinase de *Rhodothermus marinus* (RmLamR), recém-resolvida por cristalografia de raios-X, foi submetida a simulações de dinâmica molecular visando compreender as bases moleculares da estabilidade térmica desta classe de enzimas. Como a maioria das enzimas termoestáveis, a RmLamR contém um número relativamente grande de pontes salinas, as quais não se encontram distribuídas aleatoriamente na estrutura. Os resultados que serão apresentados fornecem novas informações acerca do papel de pontes salinas e da flexibilidade no sítio ativo para a estabilidade térmica da proteína, e devem ser relevantes para outras glicosídeo hidrolases.

Endoglucanase de *Trichoderma harzianum*: Organização estrutural e mecanismos de reconhecimento do substrato por endoglucanases que não possuem o módulo de ligação à celulose (Capítulo 4)

É conhecido o fato de que a maioria das celulasas apresentam pelo menos dois domínios, um domínio catalítico (CCD) e um domínio de ligação à celulose (CBM), sem o qual a atividade enzimática é drasticamente reduzida. No entanto, os membros da família Cel12A (GH-12) não apresentam um domínio CBM e, mesmo assim, são

capazes de hidrolisar celulose amorfa com eficiência. Neste trabalho, utilizam-se simulações de dinâmica molecular para detalhar as bases moleculares envolvidas na capacidade da endoglucanase 3 de *Trichoderma harzianum* (ThEG3), um membro da família GH-12, que não possui CBM. Uma análise comparativa com o CBM de endoglucanase C de *Cellulomonas fimi* (CfCBM) identifica os resíduos importantes na interação de Cel12A com cadeias de celulose que compensam a ausência de CBM. Os resultados apresentados podem ser guias para o desenho racional de enzimas com a finalidade de aprimorar as interações de enzimas Cel12A com substratos celulósicos.

β -glicosidase de *Aspergillus niger*: A função do domínio fibronectina do tipo III (Capítulo 5)

O fungo *Aspergillus niger* é conhecido pelo fato de secretar grandes quantidades de β -glicosidases. Este trabalho parte de um modelo gerado por experimentos de espalhamento a baixos ângulos de raios-X (SAXS), que identificaram que a enzima β -glicosidase de *Aspergillus niger* (AnBgl1) apresenta estrutura bimodular com um longo *linker* estendido, uma novidade entre as β -glicosidases. Outra peculiaridade é a de que um dos domínios não se identifica como um CBM, mas com um domínio fibronectina de tipo III (FnIII). As razões para a AnBgl1, que atua em pequenos substratos solúveis, apresentar este enovelamento ainda não estão claras. No entanto, ensaios de adsorção desta enzima com diferentes substratos poliméricos sugerem que a AnBgl1 não pode ligar-se à celulose, xilano e amido, mas apresenta alta afinidade com a lignina. As simulações de dinâmica molecular elucidaram como são as interações entre lignina e o domínio FnIII. Os resultados indicam que o domínio FnIII pode ser importante na imobilização da enzima na parede celular.

Capítulo 2

Metodologia

Simulações de Dinâmica Molecular

Certamente, nenhum tema ou área está progredindo em tantas frentes neste momento como a biologia, e se tivéssemos que citar a hipótese mais poderosa de todas, a que nos impulsiona a uma tentativa de entender a vida, seria a de que 'todas as coisas são feitas de átomos', e tudo que o que o é vivo faz pode ser compreendido em termos do agito e balanço dos átomos.
(The Feynman Lectures in Physics, 1963)

A compreensão de como funcionam as proteínas ganhou forte impulso com o advento de técnicas que possibilitam a descrição estrutural em alta resolução, como cristalografia de raios-X e ressonância magnética nuclear. As figuras estáticas das proteínas possibilitam mapear contatos de diferentes naturezas, átomo a átomo. Entretanto, para compreender como estas proteínas agem em condições realistas, é preciso expandir a relação estrutura-função em mais uma dimensão, o tempo. Mesmo que se mantenham determinadas condições termodinâmicas, as proteínas não assumem uma única conformação. Diferente disso, assim como qualquer outra molécula, as proteínas exibem movimentos internos, isto é, elas navegam em uma superfície de energia, com “vales e morros” que definem as diferentes conformações e as barreiras energéticas entre elas. Uma cascata de eventos dinâmicos que compreendem desde a movimentação local nas cadeias laterais geradas pelas flutuações de energia, até os movimentos coletivos de grandes domínios, é o que rege as funções biológicas das proteínas. Henzler-Wildman e Kern (Henzler-Wildman & Kern, 2007) usaram o termo “personalidade dinâmica” muito apropriadamente para descrever a paisagem de energia que define a dinâmica de uma proteína, a qual está codificada em seu enovelamento.

Múltiplas técnicas experimentais avançaram muito na inclusão da caracterização dinâmica de proteínas. Algumas técnicas de baixa resolução espacial, espectroscopias de absorção ou emissão, podem fornecer informações estruturais em função do tempo. Um exemplo que se destaca é a fluorescência resolvida no tempo, que é muito sensível a modificações de estrutura secundária e terciária da proteína com resolução de nanossegundos a picossegundos (Jameson *et al.*, 2013; Beechem & Brand, 1985). Outras técnicas, geralmente mais conhecidas para a descrição estrutural de proteínas em alta resolução, como cristalografia de raios-X e ressonância magnética nuclear, também podem ser utilizadas para obter informações dinâmicas, sendo que na primeira as informações sobre os movimentos internos da proteína são, em parte, fornecidas pelos parâmetros de mobilidade dos resíduos que se inferem do fator de Debye-Waller associado às flutuações térmicas dos centros espalhadores dos raios-x, isto é, das posições atômicas, também conhecido como fator B ou fator de temperatura. Outras limitações restringem seu uso, como por exemplo, o dano causado ao cristal devido a incidência da radiação. Ainda, é preciso que o evento que se deseja estudar envolva modificações estruturais de amplitudes pequenas o suficiente para que possam ocorrer no confinamento de um retículo cristalino (Schoote *et al.*, 2004; Wilmouth *et al.*, 2000). Mais recentemente, o desenvolvimento de técnicas de difração baseadas em *lasers* de raios-X provenientes de pulsos intensos de luz síncrotron (*free electron laser x-ray diffraction*) tem permitido obter informações de alta resolução espacial utilizando cristais muito menores e sem os problemas do dano por radiação (Boutet *et al.*, 2012; Chapman *et al.*, 2011), mas até onde se sabe, ainda não foram feitos estudos para obter informações de dinâmica de proteínas através desta técnica. Há outras formas mais indiretas muito interessantes de obter informações da dinâmica de proteínas através de experimentos de cristalografia de raios X. Existem casos em que são resolvidas diversas estruturas para uma mesma proteína, nas quais pequenas diferenças revelam alguns

aspectos de mobilidade. Por exemplo, uma hélice pode aparecer em posições ligeiramente deslocadas quando diferentes estruturas de uma proteína são superpostas, indicando uma variação conformacional que pode ser relevante no funcionamento da proteína. Ainda, posições múltiplas de cadeias laterais de uma mesma estrutura cristalográfica revelam a mobilidade de cadeias laterais.

Por outro lado, técnicas de ressonância magnética nuclear têm sido empregadas para caracterizar modificações conformacionais envolvendo catálise, enovelamento de proteínas e outros processos que ocorrem na escala de mili a microssegundos. Dentre os pontos fortes da técnica, está a possibilidade de fazer medidas de movimentos com resolução quasi-atômica (Palmer, 2001; Zhu *et al.*, 2000; Foster *et al.*, 2007).

A ferramenta computacional de dinâmica molecular (MD) vem preencher a lacuna de um método que consegue descrever a posição de cada átomo de uma molécula de proteína individualmente e precisamente, em intervalos de tempo que chegam à ordem de femtossegundos (teoricamente, qualquer intervalo desejado pode ser usado), dado que pelo menos uma estrutura de alta resolução seja conhecida como ponto de partida. Embora estados conformacionais possam ser detectados via técnicas experimentais, o caminho percorrido até eles (os caminhos de transição) descrito segundo uma resolução atômica está fora do alcance de métodos experimentais devido ao tempo de vida curto dos confôrmeros. Técnicas de MD, quando amparadas por um campo de força que descreve adequadamente as interações proteína-solvente, podem suprir esta limitação. Além disso, a técnica revela não somente como os movimentos de átomos e domínios ocorrem, mas o que causam estes movimentos, uma vez que as forças atuantes e energias de interação estão inclusas nas simulações, como veremos adiante.

Neste capítulo, apresenta-se uma breve descrição dos fundamentos básicos da técnica e dos seus atributos que fazem dela uma ferramenta valiosa no estudo da tríade estrutura-dinâmica-função de proteínas, como os três sistemas abordados neste trabalho.

Finalmente, são descritos os parâmetros gerais das simulações de cada tema. Os aspectos particulares são abordados nos respectivos capítulos.

2.1 Uma visão geral dos fundamentos de dinâmica molecular

2.1.1 Do mundo microscópico ao macroscópico

Considere que se deseje conhecer propriedades como a temperatura e a pressão no seio de uma solução aquosa de açúcar. Se não existissem termômetros ou barômetros, alguém poderia cogitar utilizar uma abordagem teórica que possibilitasse chegar a uma boa estimativa. Esta abordagem teórica existe, embora seja de aplicação direta inviável, conforme será discutido. Tais propriedades podem ser obtidas da evolução temporal nas posições e momentos lineares das partículas que compõem o sistema fechado. A propriedade medida experimentalmente, M , pode ser entendida como uma média temporal, que poderia ser descrita como:

$$M_{\text{média}} = \frac{1}{\tau} \int_{t=0}^{\tau} M(\vec{p}^N(t)\vec{r}^N(t))dt, \quad (2.1)$$

em que $\vec{p}^N(t)$ e $\vec{r}^N(t)$ representam os momentos lineares e as posições de N partículas no instante t . Para aplicar esta fórmula seria preciso, então, conhecer de antemão uma configuração do sistema, isto é, posições, momentos de cada partícula, bem como todas as forças as quais cada uma delas está sujeita. A aplicação da segunda lei de Newton, em tese, possibilitaria conhecer a trajetória destas partículas e, então, o cálculo da propriedade desejada.

Esta ideia geral bastante simples é a base da técnica de dinâmica molecular. Sabe-se, no entanto, que a mera descrição de uma das configurações de um sistema macroscópico fechado, composto por moléculas de água e sacarose, exige que se saiba posições e momentos lineares de N átomos, sendo N da ordem de 10^{23} . Além disso, é preciso conhecer as forças de interação entre todos estes átomos, em cada instante. Apesar de tamanha complexidade, a dinâmica molecular reúne algumas aproximações que tornam possível investigar a nível atômico estes sistemas macroscópicos.

Para começar, a técnica faz uso da Mecânica Estatística de Boltzmann e Gibbs, em que um único sistema macroscópico evoluindo no tempo é substituído por um grande número de réplicas representativas deste sistema e que são consideradas simultaneamente. A média no tempo de um observável é aproximada, então, como uma média de *ensemble*:

$$\langle M \rangle = \sum_{\vec{p}^N, \vec{r}^N} M(\vec{p}^N, \vec{r}^N) P(M(\vec{p}^N, \vec{r}^N)), \quad (2.2)$$

em que $P(\vec{p}^N, \vec{r}^N)$ é a probabilidade do *ensemble*, ou seja, a probabilidade de encontrar uma configuração com momentos \vec{p}^N e posições \vec{r}^N , com soma estendendo-se sobre todos os valores possíveis de momenta e posições.⁷ Sob as condições de número de partículas, volume e temperatura constantes, a probabilidade é a conhecida distribuição de Boltzmann para o ensemble canônico (Leach, 2001; McQuarrie, 2000):

$$P(\vec{p}^N, \vec{r}^N) = \exp(-E(\vec{p}^N, \vec{r}^N) / k_B T) / Q \quad (2.3)$$

⁷A soma na Eq. 2.2 é mais frequentemente representada por uma integral pois \vec{p}^N e \vec{r}^N são variáveis contínuas.

em que $E(\vec{p}^N, \vec{r}^N)$ é a energia da configuração e Q é a função de partição do ensemble. Na prática, as trajetórias de MD, especialmente de sistemas grandes (como os protéicos, de interesse neste trabalho) constituem-se de dezenas ou centenas de nanossegundos discretizados em passos de tempo equivalentes a 1 ou 2 femtossegundos (uma simulação de 20 ns usando passo de tempo 2 fs requer 10.000.000 passos). Como um *experimento* no computador, a média de uma propriedade é calculada como média no tempo da simulação (repare na semelhança com a Equação 2.1):

$$\langle M \rangle = \frac{1}{S} \sum_{i=1}^S M_i(p^N, r^N), \quad (2.4)$$

sendo S o número de passos de tempo. A média no tempo da simulação é um valor estimado, que se torna mais próximo à média no *ensemble* quanto maior for o número de configurações exploradas. Em outras palavras, se todo o espaço de fase pudesse ser percorrido na simulação, a média nessa simulação seria equivalente à média no *ensemble*. Esta é a hipótese ergódica. Neste caso ideal, os resultados das simulações seriam os exatamente os mesmos para qualquer configuração inicial. Percorrer todo o espaço de fase de um sistema como alguns milhares de moléculas de água hidratando dezenas de moléculas de sacarose custaria tempo inestimável. Mesmo assim, simulações de MD rotineiras podem fornecer estimativas muito similares ao valor “verdadeiro” da propriedade que se quer medir por gerar número suficiente de configurações/conformações representativas do espaço de fase.

Uma questão central na dinâmica molecular, entretanto, ainda precisa discutida aqui: o tratamento das forças entre os átomos.

2.1.2 Campos de Força

Em tese, a dinâmica de moléculas poderia ser descrita pela resolução da equação de Schrödinger dependente do tempo. Felizmente, no entanto, as propriedades eletrônicas dos sistemas muitas vezes não precisam ser tratadas explicitamente para estudar inúmeros problemas relacionados a sistemas moleculares. Segundo a conhecida aproximação de Born-Oppenheimer, a dinâmica de elétrons e núcleos pode ser desacoplada, e a dinâmica molecular incorpora este pressuposto. Apenas a trajetória dos núcleos é descrita, e esta é definida por uma superfície de energia potencial gerada pelos elétrons. Na prática, nem mesmo os cálculos de estrutura eletrônica são feitos nos passos de simulação, até porque isto seria extremamente caro computacionalmente, mesmo para sistemas relativamente pequenos, de algumas centenas de átomos. Ao invés disso, funções analíticas podem ser ajustadas a curvas de potencial geradas para pequenos grupos de átomos (às vezes, duas ou três moléculas) e funcionam muito bem.

Ao conjunto de funções escolhido para descrever as interações interatômicas (ou a superfície de energia potencial por onde se movem os núcleos) é dado o nome de campo de força efetivo, que nada mais é do que um modelo de potencial de interação tipicamente construído com base em cálculos *ab initio* de Mecânica Quântica e em dados experimentais. Estes campos de força descrevem as interações entre diferentes átomos (ou grupos específicos como CH₂ e CH₃) pela soma de interações de curto alcance do tipo Lennard-Jones, termos de interação eletrostática coulombianos e potenciais internos de estiramento de ligação, deformação angular e torção de diedros. Ou seja, fragmenta-se a energia de interação em um somatório de diferentes funções, agrupadas em potenciais de pares de átomos não ligados e potenciais de átomos ligados. O potencial toma a forma genérica:

$$V = V^{ligado} + V^{nãoligado} \quad (2.5)$$

$$V^{ligado} = \sum_{\text{ligações}} K_r (r - r_0)^2 + \sum_{\text{ângulos}} K_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{n, \text{diédro}} V_n [1 + \cos(n\phi + \gamma_n)] \quad (2.6)$$

$$V^{nãoligado} = \sum_{i,j} \left\{ 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}, \quad (2.7)$$

isto é, são usados potenciais harmônicos para as ligações e ângulos, e os ângulos diédrais formados pelas ligações entre quatro átomos consecutivos são tratados com potenciais periódicos. A interação entre átomos separados por mais do que três ligações covalentes é descrita através do potencial dispersivo e eletrostático. O campo de força CHARMM (Mackerell *et al.*, 1998), utilizado neste trabalho, apresenta dois termos adicionais para descrever o potencial entre átomos ligados, a saber, o potencial de Urey-Bradley e o potencial de diedros impróprios. O potencial de Urey-Bradley tem a forma

$$V^{UB} = k_u (u - u_0)^2,$$

que descreve a função harmônica da distância entre átomos separados por duas ligações covalentes (átomos 1,3). Como um termo cruzado, este termo tem a finalidade de refletir o acoplamento entre as coordenadas internas (estiramento de ligação e deformação angular) e otimizar o ajuste a espectros vibracionais (Leach, 2001). Os parâmetros de Urey-Bradley não são usados com grande frequência no campo de força, sendo aplicados apenas em casos especiais. Um exemplo de aplicação no campo de força CHARMM é na descrição dos modos vibracionais de anéis aromáticos. Já os diedros impróprios têm a forma

$$V^{imp} = \sum_{\text{impróprio}} k_\omega (\omega - \omega_0)^2,$$

que descreve o ângulo de torção para átomos que não estão ligados na sequência 1-2-3-4. Diferente do componente de Urey-Bradley, diedros impróprios são mais corriqueiramente empregados e também são utilizados para adequação às medidas vibracionais experimentais. Os diedros impróprios descrevem termos de deformação fora do plano de grupo de átomos. A coplanaridade de grupos de átomos que se deve a efeitos de ressonância eletrônica, por exemplo, muitas vezes só pode ser alcançada com a incorporação do termo de diedro impróprio. No campo de força CHARMM de proteínas, por exemplo, diedros impróprios são usados para descrever apropriadamente o grupo imidazol de histidinas e o grupo acetato de C-terminais (Mackerell *et al.*, 1998).

Os principais componentes descritos do campo de força são representados na Figura 2.1 com modelos esféricos, ao lado de curvas que ilustram o formato de cada componente do potencial.

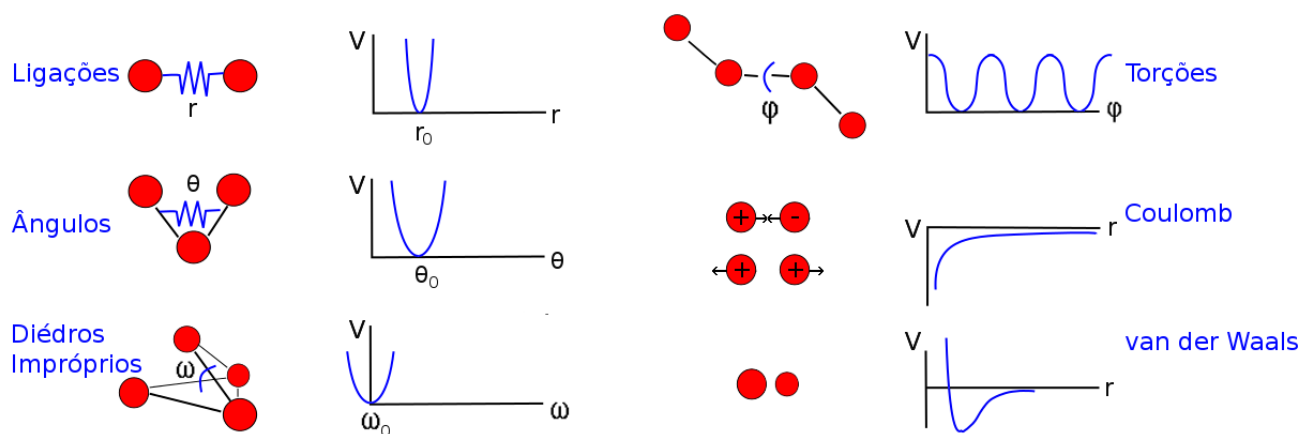


Figura 2.1. Reresentação esquemática dos componentes principais de um campo de força.

Determinar os diversos parâmetros do campo de força (*eg.* Constantes de força, valores de referência de ligação e ângulo, altura de barreiras e ângulos da fase diedrais) é um trabalho desafiador e uma etapa determinante da qualidade das simulações.

Tipicamente, estes parâmetros são definidos de modo a reproduzir propriedades estruturais ou termodinâmicas conhecidas através de experimentos. Alguns dados visados podem ser, por exemplo, a entalpia de vaporização, densidade do líquido e a curva de compressibilidade *versus* pressão de gases. Se as simulações são capazes de reproduzir tais propriedades, também serão capazes de prever outras que também sejam fortemente relacionadas a interações atrativas e repulsivas entre átomos.

Quando não há medidas experimentais que possam ser usadas para o ajuste dos parâmetros, muitas vezes, lança-se mão de cálculos *ab initio*. O campo de força CHARMM de proteínas, por exemplo, foi desenvolvido pela combinação de dados experimentais e cálculos quânticos (Brooks *et al.*, 1983; MacKerell *et al.*, 1998).

A transferabilidade é um atributo importante de campos de força. Isto é, é desejável que os parâmetros de campo de força possam ser utilizados por uma grande variedade de moléculas que apresentem grupos funcionais similares (Leach, 2001). Parâmetros de estiramento geralmente são prontamente transferíveis, assim como parâmetros de Lennard-Jones, os quais geralmente são idênticos para a maioria dos átomos de um mesmo elemento. Outros termos do potencial de interação, os que têm efeitos de maior alcance, algumas vezes precisam ser determinados para uma molécula individualmente, ou para uma parte dela. Este é frequentemente o caso de diedros e de cargas, por exemplo. Diedros são comumente determinados a partir de cálculos de Mecânica Quântica para uma sequência de ângulos de torção, de modo a ajustar o potencial diedral a curva obtida. Já as cargas atômicas são ajustadas pelo potencial eletrostático calculado com uma função de base conveniente (se possível, é melhor que se adote o mesmo protocolo do campo de força escolhido).

Petridis e Smith (Petridis & Smith, 2008), por exemplo, estenderam a parametrização do campo de força CHARMM de polissacarídeos para a inclusão de lignina. Apenas os parâmetros faltantes no campo de força foram otimizados. Por

exemplo, apesar de a lignina apresentar éteres lineares, estes não poderiam ser descritos com os parâmetros de éteres derivados por Vorobyov e colaboradores (Vorobyov *et al.*, 2007), uma vez que estes átomos de oxigênio em lignina encontram-se ligados a grupos fenólicos, que, por ressonância de elétrons π , influencia ligações, ângulos e a distribuição de cargas. As interações ligadas e não ligadas foram obtidas essencialmente a partir de cálculos quânticos para pequenos modelos representativos de lignina. O modelo de potencial foi então validado através da comparação de parâmetros de cela obtidos por simulações de MD e resultados experimentais de cristalografia de raios-X.

2.1.3 Resolvendo as equações de movimento

O campo de força define a superfície de potencial por onde os átomos propagam suas trajetórias. Por serem relativamente pesados, a descrição das trajetórias dos átomos pode ser feita através de mecânica clássica. Os algoritmos utilizados em MD poderiam ser simplificados em uma formulação bem simples, baseada na segunda lei de Newton:

$$\vec{v}\left(t + \frac{1}{2}\Delta t\right) = \vec{v}(t) + \frac{\Delta t}{2} m^{-1} \vec{F}(t) \quad (2.8)$$

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \Delta t \vec{v}\left(t + \frac{\Delta t}{2}\right) \quad (2.9)$$

$$\vec{F}(t + \Delta t) = \vec{F}(\vec{x}(t + \Delta t)) \quad (2.10)$$

$$\vec{v}(t + \Delta t) = \vec{v}\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2} m^{-1} \vec{F}(t + \Delta t) \quad (2.11)$$

em que dispõe-se, para um passo de tempo t , das posições, $\vec{x}(t)$, velocidades, $\vec{v}(t)$, e forças atuantes em cada átomo, $\vec{F}(t)$, derivadas do potencial de interação. A aplicação destas equações de modo iterativo em cada passo permite a obtenção destes dados em tempos sucessivos, fornecendo a trajetória de todos os átomos do sistema.

Existem muitos algoritmos para a resolução das equações de movimento e, para decidir qual é o mais apropriado, devem ser feitas considerações como o esforço computacional requerido e a conservação de energia. Dentre os algoritmos mais utilizados atualmente nas simulações, estão o algoritmo Verlet (Verlet, 1967) e Verlet-velocidade (Feynman *et al.*, 1963; Swope *et al.*, 1982). Este último é o algoritmo implementado no programa de simulação NAMD (Philips *et al.*, 2005) que foi utilizado neste trabalho e é descrito brevemente a seguir.

Todos os algoritmos assumem que as posições e as funções velocidade e aceleração podem ser aproximadas como séries de Taylor. No caso do algoritmo Verlet-velocidade, no cálculo das posições, utiliza-se a expansão de Taylor truncada na segunda ordem.

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \Delta t \vec{v}(t) + \frac{1}{2} \Delta t^2 \vec{a}(t) \quad (2.12)$$

Através das posições no tempo $t + \Delta t$ e do potencial de interação, deriva-se as acelerações de cada átomo neste instante. Já as velocidades são computadas a partir das velocidades no instante anterior t somada à média das acelerações dos instantes t e $t + \Delta t$:

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{a}(t) \Delta t}{2} + \frac{\vec{a}(t + \Delta t) \Delta t}{2} \quad (2.13)$$

A escolha do algoritmo de integração deve considerar, principalmente, se há conservação de energia e momento, reversibilidade no tempo e se passos de tempo grandes podem ser utilizados. Outras considerações a serem feitas são a velocidade de cálculo e memória requerida. No entanto, estas são questões mais secundárias, uma vez

que a parte que demanda muito mais tempo computacional é o cálculo de forças de interação entre os átomos, e não a integração das equações de movimento.

2.2 Simulações de proteínas

O primeiro estudo de simulação de dinâmica molecular de uma macromolécula de interesse biológico foi feita há 36 anos para a pequena proteína BPTI (do inglês, *bovine pancreatic trypsin inhibitor*), consistindo em uma simulação da proteína em vácuo, com a duração de apenas 9,2 ps (McCammon, 1977). Atualmente, a evolução dos recursos computacionais e das técnicas de MD permite que simulações de proteínas muito maiores, de vários milhares de átomos (em geral, da ordem de 10^4) ao invés de algumas centenas, sejam rotineiramente feitas, incluindo solvente explícito, e podendo alcançar a escala de microssegundos de duração. Uma busca rápida pelas palavras 'molecular dynamics' e 'proteins' no ISI Web of Science (<http://apps.webofknowledge.com>) mostra 4387 artigos publicados apenas no ano de 2012, um resultado comparável aos 4494 artigos apontados para as palavras 'crystal structure' e 'proteins', o que revela a importância atual da técnica no estudo de proteínas.

A técnica de MD é apropriada para estudar os sistemas de proteína em níveis variados. A dinâmica molecular convencional geralmente é aplicada para descrever o sistema no equilíbrio e estudar suas propriedades estruturais e dinâmicas, além de obter valores de grandezas termodinâmicas. Também é empregada para observar processos de relaxação, isto é, a evolução no tempo do sistema. Com o estabelecimento de novas aproximações à técnica, processos mais lentos ou propriedades que implicariam em custos computacionais altíssimos, agora podem ser estudados. Por exemplo, técnicas de amostragem para eventos raros (eg. *Replica exchange*, *Locally enhanced sampling*, *Steered molecular dynamics* e várias outras) (Sugita & Okamoto, 1999; Roitberg & Elber, 1991; Israelowitz *et al.*, 2001; Adcock & McCammon, 2006) e técnicas para

computar diferenças de energia livre (Chipot & Pohorille, 2007) têm sido constantemente aprimoradas e empregadas com sucesso. No outro extremo da escala temporal, há os métodos híbridos – como QM/MM, o qual permite sejam estudadas as reações ocorrendo na região catalítica de uma enzima (Field, 2001; Senn & Thiel, 2009).

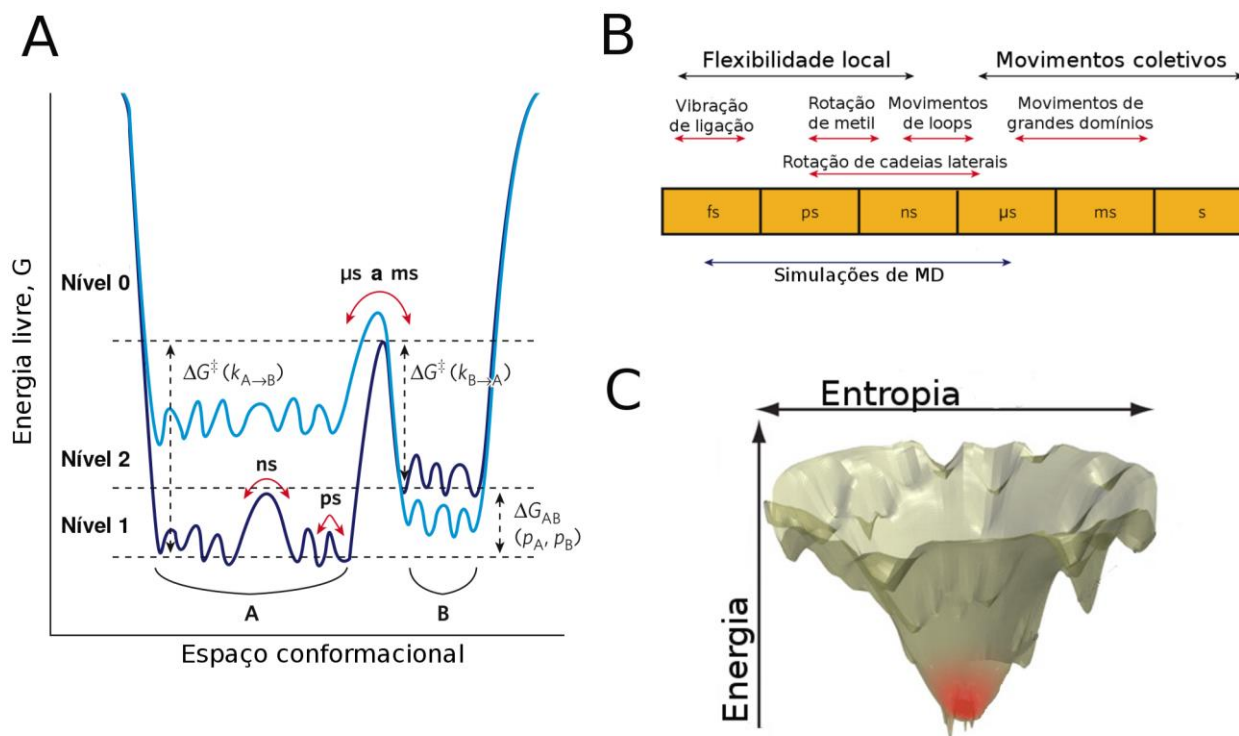


Figura 2.2 (A) Níveis de energia referentes a eventos dinâmicos de proteínas enveladas. (B) Escala de tempo associadas aos processos dinâmicos em proteínas. (C) Curva tridimensional de estados conformacionais acessíveis a uma proteína. Em geral, a curva tem um formato que semelhante a um funil rugoso. O aspecto da curva é modificado por condições como a associação a um ligantes, o tipo de solvente e o pH. Os painéis (A) e (B) são figuras adaptadas da referência Henzler-Wildman e Kern (Henzler-Wildman & Kern, 2007), e o painel (C), da referência (Papoian, 2008).

A Figura 2.2 mostra como podem ser divididas estas abordagens em relação à escala de tempo dos processos. A dinâmica na escala mais lenta (nível 0, Figura 2.2A) relaciona-se a processos que só ocorrem ao ultrapassar barreiras de vários kT . A população dos estados inicial e final (A e B) é definida pela distribuição de Boltzmann

que reflete a estabilidade relativa destes estados. São tipicamente modificações conformacionais na escala de micro a milisegundos, que envolvem a movimentação de grandes domínios, como as que ocorrem no encaixe entre proteínas em um complexo (Figura 2.2B).

As escalas de processos mais rápidos (níveis 1 e 2, Figura 2.2A) englobam flutuações em estados separados por barreiras de energia menores do que kT . São movimentos de cadeia lateral e de *loops*, que ocorrem na escala de pico e nanossegundos (Figura 2.2B). É esta a dinâmica de proteínas abordada neste trabalho. A Figura 2.2A também ilustra a alteração do perfil de energia livre, da curva azul escuro para a azul claro, após transformações como a associação de um ligante à proteína ou a alteração do pH da solução.

Outra forma, mais completa, de pensar na dinâmica de proteínas, é através de um gráfico multidimensional, que lembra um funil rugoso (Fig. 2.2C). De acordo com a Teoria da Paisagem Energética, este é um mapa que relaciona a conformação da cadeia polipeptídica à energia interna. Este mapa define quais configurações são acessíveis a partir de uma dada configuração e como a proteína pode transitar de um estado a outro (Plotkin & Onuchic, 2002). Quanto mais fina a base do gráfico, mais bem definido o estado nativo⁸. As secções horizontais refletem a entropia do sistema. Quanto mais próximo da borda do funil, maior o número de microestados possíveis, isto é, maior o número de conformações de mesma energia. Em processos de desnaturação, tal qual a desnaturação térmica de uma enzima descrito no Capítulo 3, envolve o percurso da base do funil até a borda.

⁸ Em proteínas globulares, o estado nativo (condição em que a proteína apresenta-se funcional) é o mínimo global.

2.3 Rodando as simulações

Embora a composição dos sistemas seja diferente, foi adotado um mesmo protocolo básico para as simulações, descrito a seguir. Alguns detalhes específicos de cada sistema são descritos nos Capítulos correspondentes.

Todos os sistemas passaram por uma etapa de minimização, em 500 passos, usando o método de gradientes conjugados (CG) (Fletcher & Reeves, 1964; Fletcher, 2000), conforme implementado no NAMD (Philips *et al.*, 2005) para eliminar maus contatos. Após a minimização, foram conduzidas rodadas de pré-equilibração (Bleicher & Prates *et al.*, 2011), composta por três fases: Nos primeiros 10 ps, as posições de todos os resíduos, exceto os modelados foram mantidas fixas; nos próximos 100 ps, as cadeias laterais foram deixadas livres para moverem-se; nos próximos 890 ps que completam o intervalo aqui referenciado como fase de pré-equilibração, todos os átomos podiam mover-se.

As simulações foram conduzidas através do programa NAMD (Philips *et al.*, 2005), usando condições periódicas de contorno para evitar efeitos de fronteira. O campo de força CHARMM foi usado para proteínas, oligossacarídeos e lignina (Mackerell, Jr. *et al.*, 1998; Hatcher, *et al.*, 2009; Guvench *et al.*, 2009; Guvench *et al.*, 2011, Petridis & Smith, 2008) e o modelo TIP3P foi usado para moléculas de água (Jorgensen *et al.*, 1983). Temperatura e pressão foram mantidas fixas a 298 K e 1 bar, exceto nas simulações de laminarinases, em que também foram adotadas temperaturas mais altas (363 e 500 K), para o estudo da termoestabilidade e termofilicidade destas enzimas. No controle de temperatura e pressão, foram usadas dinâmica de Langevin (Schneider & Stoll, 1978) e Métodos de Pistão de Nosé-Hoover (Martyna *et al.*, 1994), respectivamente. Todas as ligações envolvendo o átomo de hidrogênio foram mantidas com comprimento fixo usando o algoritmo SHAKE (Ryckaert *et al.*, 1977). Um raio de corte de 12 Å com decaimento suave a partir de 10 Å foi usado para o tratamento das

interações de van der Waals, enquanto que interações eletrostáticas foram tratadas pelo método de somas de Ewald, pelo algoritmo *Particle Mesh Ewald* (Tradução: “Malha de Partículas de Ewald”; Darden *et al.*, 1993).

2.4 Análises das simulações

Algumas análises globais das simulações feitas neste trabalho são descritas aqui⁹. Trata-se do cálculo de médias e flutuações de propriedades estruturais e dinâmicas microscópicas, bem como da evolução temporal destas propriedades em processos de relaxação estrutural. É interessante observar como, na técnica de dinâmica molecular, estas análises podem ser combinadas entre medidas globais, como o raio de giração, e as modificações em interações locais, como a formação/rompimento de ligações de hidrogênio e pontes salinas.

1. Raio de giração: É um parâmetro que traz a informação da dimensão espacial, ou raio médio, da proteína globular, tratada como uma esfera hipotética, sendo definido por:

$$R_g(t) = \sqrt{\frac{\sum_{j=1}^N [r_j(t) - r_{CM}(t)]^2}{N}} \quad (2.17)$$

Esta fórmula poderia ser aplicada com, por exemplo, r_j sendo as posições dos átomos de C_α da cadeia polipeptídica, r_{CM} , a posição do centro de massa definido por estes átomos em um instante t e N , o número total destes átomos. Um gráfico de variação do raio de giração ao longo do tempo mostra se ocorrem processos de compactação ou expansão do

⁹ Outras análises mais específicas deste trabalho, entrada de água no caroço hidrofóbico e empilhamento de estruturas planares, são discutidas nos capítulos em que estas análises aparecem (Cap. 3 e 5, respectivamente). Tais análises foram feitas utilizando os programas SOLVATION-MAP e PACKING.

volume da proteína. Experimentalmente, R_g médio da proteína em solução poderia ser medido através de SAXS (*Small-Angle X-ray Scattering*). Neste trabalho, esta análise foi feita através do programa GYRATION desenvolvido no grupo.

2. Conteúdo de estrutura secundária: Cálculo da porcentagem de resíduos que se mantêm em estrutura secundária regular, como α -hélices, folhas- β e 3-10-hélices. Neste trabalho, foi utilizado o algoritmo STRIDE (Frishman & Argos, 1995), que identifica em um conjunto de coordenadas os elementos de estrutura secundária a partir de critérios geométricos que definem ângulos diedrais além de energia de ligações de hidrogênio. Esta análise poderia ser efetuada a partir de experimentos de dicroísmo circular. O programa utilizado neste caso, foi o VMD (Humphrey *et al.*, 1996) em combinação com o programa GETSECONDARY desenvolvido no grupo.

3. Contatos hidrofóbicos: Cálculo da porcentagem de resíduos hidrofóbicos em contato entre si. Este parâmetro é um indicador da compacidade do centro hidrofóbico da proteína. Usamos como critério para considerar que um par de resíduos apolares (Ala, Val, Leu, Met, Ile, Pro, Trp e Phe) encontra-se em contato, a distância de 5.0 Å entre qualquer átomo (não-hidrogênio) de suas cadeias laterais. Esta análise foi feita através do programa CONTACTS desenvolvido no grupo.

4. Estatística de ligações de hidrogênio: a contagem de ligações de hidrogênio em um passo de simulação é feita segundo critérios geométricos. Neste trabalho, foi usado o *plugin* HBonds do programa VMD (Humphrey *et al.*, 1996). Uma ligação de hidrogênio é contada se a distância entre o átomo doador (D) e o aceptor (A) for menor do que 3.0 Å e o ângulo D-H-A for menor do que 20°.

5. Estatística de pontes salinas: considera-se que há formação de ponte salina quando um par de átomos da cadeia lateral (não-hidrogênio) de um par de resíduos de cargas opostas (Lys, Arg e His – positivos; Asp e Glu – negativos) encontra-se a uma distância menor que 4 Å. Esta análise foi feita através do programa CONTACTS desenvolvido no grupo.

6. Evolução no tempo do desvio quadrático médio (RMSD – *Root mean square deviation*) de posições: mostra o quanto a posição de um átomo ou de um grupo de átomos se modifica em relação à posição correspondente em uma estrutura de referência. A equação abaixo mostra como é computado este valor:

$$\langle RMSD \rangle(t) = \sqrt{\frac{\sum_{j=1}^N [r_j(t) - r_j^0(t)]^2}{N}}, \quad (2.18)$$

sendo N o número de átomos considerados.

Esta análise é comumente utilizada para acompanhar a relaxação estrutural no processo de equilibração da simulação. Poderiam, então, ser armazenadas as posições dos átomos de C_α em intervalos de tempo regulares para compará-las com a estrutura cristalográfica da proteína. Variações de grande amplitude podem significar o movimento de um domínio maior (nível 0, Figura 2.2A), ao passo que pequenas flutuações refletem a energia térmica da proteína (níveis 1 e 2). A unidade aqui é Å por átomo. Neste trabalho, foi utilizado o programa RMSD desenvolvido no grupo.

7. Mobilidade: Trata-se de um modo específico de aplicar o desvio médio quadrático de posições, com a finalidade de analisar a mobilidade de regiões específicas da molécula. Em geral, é chamada de flutuações quadráticas médias (RMSF – *Root mean square fluctuations*). A equação abaixo mostra como é computado este indicador:

$$\langle RMSF_j \rangle = \sqrt{\frac{\sum_{i=1}^M [r_j(t_i) - r_0]^2}{M \times \Delta p}}, \quad (2.19)$$

em que M é o número de configurações consideradas e Δp é o intervalo de passos de tempo em que as configurações são salvas (para proteínas, tipicamente, $\Delta p = 1000$, ou seja, as configurações são salvas a cada 2 ps para passos de tempo de 2 fs). Assim, a multiplicação $M \times \Delta p$ corresponde ao número total de configurações abrangido na análise. Esta análise pode ser feita, por exemplo, considerando o deslocamento dos átomos de C_α da proteína em relação a uma estrutura média da simulação. Eventos de maior amplitude podem ocorrer algumas vezes durante a simulação (por exemplo, o deslocamento de uma hélice) e podem “contaminar” o resultado. Uma estratégia que este grupo de pesquisa tem adotado para evitar este problema, é a de dividir a simulação em pequenos blocos (de centenas de picossegundos). A equação é aplicada para cada bloco, tomando como referência a estrutura média do respectivo intervalo. Para utilizar esta abordagem, o programa MOBILITY, desenvolvido no grupo, foi utilizado.

A mobilidade computada a partir de simulações encontra um correspondente experimental direto em cristalografia, o fator B (também conhecido como fator de temperatura ou fator Debye-Waller). No capítulo 3, é feita uma análise comparativa entre RMSF experimental e fator B. Entretanto, o fator B reflete deslocamentos atômicos originados tanto da desordem estrutural (um ensemble de subestados presentes em solução está confinado no cristal) quanto da desordem dinâmica (flutuações que ocorrem no cristal). Deste modo, há de se ter cautela, uma vez que o fator B não pode ser considerado apenas reflexo de flutuações atômicas. Além disso, contatos cristalinos também afetam o fator B (Henzler-Wildman & Kern, 2007).

8. Energia média de interação: Neste trabalho, a análise foi utilizada para verificar a recorrência de interação de determinados grupos funcionais do substrato a resíduos específicos (Capítulo 5). Estas energias são computadas usando os termos de potencial de interação entre átomos não ligados (Lennard-Jones e interações coulombicas, Equação 2.7), com os parâmetros do campo de força CHARMM. A energia média é calculada com:

$$\langle E \rangle = \frac{1}{N} \sum_i^N E_i, \quad (2.20)$$

em que N é o número de passos de simulação considerados. Nesta análise, foi utilizado o programa ENERGY e o *script* ENERGY-PER-RESIDUE desenvolvidos no grupo.

Capítulo 3

Laminarinase de *Rhodothermus marinus*:

Bases estruturais de sua termoestabilidade e termofilicidade

3.1 Introdução

A bactéria gram-negativa *Rodothermus marinus* pode ser encontrada em *habitats* de grandes flutuações nas condições ambientais. Estas bactérias são apreciadoras do ambiente salino e quente, isto é, são simultaneamente halófilas¹⁰ e termófilas. Assim sendo, este micro-organismo desenvolveu proteínas resistentes ao calor – proteínas termofílicas e hipertermofílicas – que apresentam atividade ótima a altas temperaturas, entre 55 e 100 °C (Bjornsdottir *et al.*, 2006). A elevada estabilidade térmica é de grande interesse na indústria para otimização de processos, e a bioengenharia de enzimas capazes de resistir a condições extremas de temperatura e pH consiste em um grande desafio científico e tecnológico. Por esta razão, as enzimas que catalisam a hidrólise de polissacarídeos em *R. marinus* revelam-se promissoras em aplicações biotecnológicas.

A laminarinase de *R. marinus* (RmLamR) é uma glicosídeo hidrolase da família 16 (GH-16) e, como tal, hidrolisa aleatoriamente β -1,3-glicanos, β -1,3-1,6-glicanos ramificados e β -1,3-1,4-glicanos (por exemplo, lichenina). A enzima exibe atividade ótima a pH 5,5 e 88 °C, e sua temperatura de desnaturação foi estimada em cerca de 97 °C (Xiao Honig, 1999). Por suas temperaturas de funcionamento ótimo e de desnaturação serem superiores a 80 °C, ela é classificada como uma enzima

¹⁰ Halófilos são organismos que apreciam ambientes salinos.

hipertermofílica e hipertermoestável, respectivamente.

A temperatura ótima de atividade de cada enzima resulta de um compromisso complexo das interações entre cargas, resíduos polares ou apolares e interações com o solvente, que são afetados de maneira distinta pela temperatura. Em termos gerais, verifica-se que proteínas termoestáveis apresentam mais interações polares e maiores centros hidrofóbicos (Jaenicke & Bohm, 1998; Goldstein, 2007). Entretanto, o balanço energético é sutil, e algumas interações podem desestabilizar a estrutura enovelada em determinadas faixas de temperatura (Argos *et al.*, 1979). Por exemplo, pontes salinas localizadas na superfície não contribuem para a estabilidade da proteína à temperatura ambiente, mas estudos experimentais e computacionais mostraram que elas são fatores estabilizadores em temperaturas elevadas devido ao decréscimo na penalidade por dessolvatação dos resíduos carregados (Vogt *et al.*, 1997; Priyakumar *et al.*, 2010). Assim, a compreensão das relações entre os diferentes componentes das interações proteína-solvente e resíduo-resíduo é um fator fundamental para desvendar as razões moleculares e estruturais para a termoestabilidade e termofilicidade verificada em uma determinada enzima.

Muitas das proteínas termofílicas de estrutura conhecida apresentam homólogas¹¹ mesofílicas, estáveis apenas em condições brandas, que são muito similares em termos de suas sequências primárias e de suas estruturas tridimensionais. Esta similaridade, apesar dos perfis térmicos completamente distintos, facilita a identificação dos determinantes responsáveis pela hiperestabilidade térmica e é nesta consideração que se baseia este trabalho.

¹¹ O termo é recorrentemente empregado para designar enzimas que apresentam estruturas tridimensionais similares. Alguns autores, no entanto, argumentam que o significado mais preciso é o de enzimas que compartilham a mesma origem evolucionária (Reeck *et al.*, 1987). Nesta tese, o termo foi utilizado de acordo com o significado atribuído usualmente, de semelhança estrutural.

No presente trabalho, estudaram-se as bases moleculares da estabilidade térmica e termofilicidade de laminarinases da família 16 através de simulações comparativas de MD de três homólogas estruturais: a hipertermofílica RmLamR, uma laminarinase termoestável e alcalifílica de *Nocardiopsis sp.*, e uma laminarinase mesofílica de *Phanerochaete chrysosporium*. Através de análises estruturais e dinâmicas, além do estudo de processos de desnaturação, foi possível propor as principais características que conferem estabilidade térmica e termofilicidade à RmLamR, as quais podem ser úteis no desenvolvimento de novas estruturas com aplicabilidade aprimorada.

Primeiramente, será descrita a estrutura da RmLamR cristalizada e resolvida pelo grupo de experimentalistas com o qual colaboramos (Grupo de Cristalografia IFSC/USP, São Carlos). Discutem-se quais aspectos estruturais geralmente relacionados à termoestabilidade podem ser vislumbrados logo na estrutura estática. Em seguida, são discutidas as análises das simulações de MD em duas temperaturas (25 e 90 °C), que fornecem novas informações estruturais relevantes para a termoestabilidade e termofilicidade, a partir de médias temporais e da inspeção visual das simulações. A pesquisa prosseguiu com estudo dos processos de desnaturação das laminarinases através de simulações a temperatura bastante elevada, 227 °C. As simulação a 90 °C forneceram indícios relevantes de quais seriam as primeiras etapas da desnaturação térmica da enzima mesofílica, mas a estratégia de elevar ainda mais a temperatura tornou possível acompanhar o avanço do processo de desnaturação das três enzimas sob um custo computacional viável.

3.2 Descrição e análise estrutural

A laminarinase isolada de *R. marinus* é composta por 250 resíduos e possui massa molecular de 32 kDa. As laminarinases desta família têm, no sítio ativo, a sequência linear Glu(nucleófilo)-H_a-Asp(assistente)-H_a-Met-Glu(catalisador ácido/base), em que

H_a é um aminoácido hidrofóbico. No caso da RmLamR, H_a corresponde à isoleucina. Os grupos catalíticos são o nucleófilo Glu129 e o catalisador ácido/base Glu134, separados por um aspartato que assiste à reação. Este resíduo “assistente” é encontrado em muitas proteínas, e seu papel tem sido atribuído à modulação dos estados de ionização dos grupos nucleofílico e catalisador ácido/base, além da estabilização eletrostática dos estados de transição catiônicos.

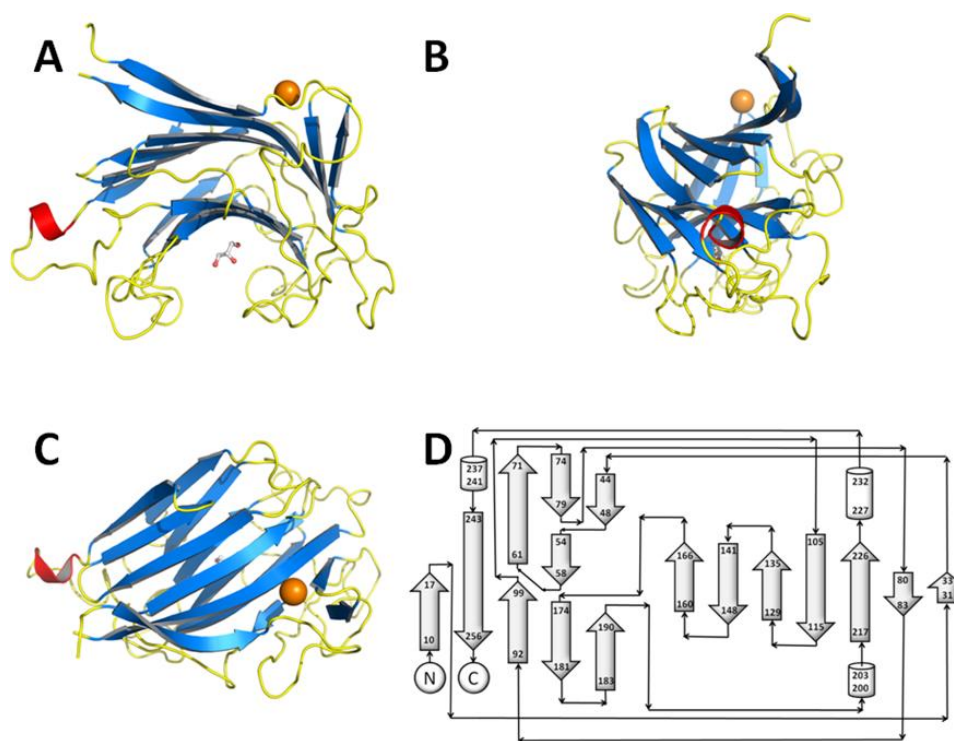


Figura 3.1. (A) Estrutura global da laminarinase de *R. Marinus*, mostrando o íon cálcio coordenado (laranja) e uma molécula de glicerol na região catalítica. (B,C) O mesmo que A, após rotações de 90 °C ao redor dos eixos *y* e *x*. (D) Topologia da enzima RmLamR (Figura copiada dos autores cristalógrafos de Bleicher & Prates *et al.*, 2011).

Na Figura 3.1, apresenta-se a unidade biológica da estrutura cristalográfica da RmLamR, resolvida por Bleicher e colaboradores (Bleicher & Prates *et al.*, 2011; PDB id 3ILN) com resolução 1,95 Å. Como todas as enzimas incluídas na família 16 de GHs,

RmLamR é composta por duas camadas de folhas- β antiparalelas em topologia conhecida como *β -jelly roll* e que inclui longos *loops*, fitas- β segmentadas e três pequenos segmentos de hélice. Na estrutura determinada, encontra-se um íon cálcio interagindo fortemente com átomos na superfície da enzima. Sugere-se que este íon tenha importância na estabilidade das glicosídeo hidrolases da família GH-16 (Tsai *et al.*, 2003; Borris *et al.*, 1989; Keitel *et al.*, 1994; Wefle *et al.*, 1994). Além disso, verifica-se a presença de uma molécula de glicerol ligada fortuitamente ao sítio ativo da estrutura, interagindo por ligações de hidrogênio com os resíduos catalíticos.

A análise da estrutura cristalográfica de RmLamR já sugere alguns fatores que contribuem para sua alta estabilidade térmica. Uma característica comum de proteínas que apresentam elevada estabilidade térmica é a pouca quantidade de *loops* (Thompson & Eisenberg, 1999). A RmLamR, ao contrário, apresenta uma grande quantidade de *loops* e alças, e tais regiões puderam ser modeladas a partir dos mapas de densidade eletrônica. Deste modo, os longos *loops* não representam abundância de regiões desordenadas, uma vez que não adotam múltiplas conformações no cristal.

Bleicher e colaboradores mostram que a sequência da RmLamR contém maior número de resíduos de arginina e glutamato, se comparado a outros membros da GH-16 (Bleicher & Prates *et al.*, 2011). A abundância de resíduos carregados e resíduos aromáticos pode ser correlacionada com a termofilicidade, conforme proposto por Jaenicke e Böhm (Jaenicke & Bohm, 1998).

Tabela 3.1. Número de PS na RmLamR e em outras estruturas GH-16 dispostas no PDB.

Enzima (organismo)	Número de OS
RmLamR (<i>R. marinus</i>)	18
2HYK (<i>Nocardiopsis sp.</i>)	7
1DYP (<i>P. carrageenovora</i>)	11
1GBG (<i>B. licheniformis</i>)	5
1MVE (<i>F. succinogenes</i>)	8
1UPS (<i>C. perfringens</i>)	17(22) ^a
1U0A (<i>B. macerans</i>)	5
1UN1 (<i>P. tremula</i>)	11
1O4Y (<i>Z. galactanivorans</i>)	11
2CL2 (<i>P. chrysosporium</i>)	11

^a A estrutura 1UPS possui um domínio, resíduos 17-286, que correspondem ao domínio catalítico GH-16. Este tem 17 PS, ao passo que a molécula inteira tem um total de 22 PS.

Propõe-se que a relação entre termoestabilidade e a abundância de resíduos carregados ocorra devido à formação de pontes salinas entre estes resíduos, um fator constantemente mencionado como crucial para a estabilidade de proteínas termofílicas (Kumar & Nussinov, 2002; Dominy *et al.*, 2004). A abundância de pontes salinas (PS) foi comparada em um conjunto de estruturas de homólogas da RmLamR, como mostrado na Tabela 3.1. A RmLamR tem o maior número de PS, reforçando, deste modo, o papel delas na estabilização térmica. O único valor comparável de PS é o da endo- β -glucanase de *Clostridium perfringens*. Trata-se de uma bactéria anaeróbica Gram-positiva que pode ser encontrada em abundância em diversos ambientes, como em solos ou no trato intestinal de animais (Petit *et al.*, 1999). Este é um indicativo de que as PS nesta enzima podem ser estabilizadoras desta proteína na ampla variedade de condições fisiológicas em que este organismo vive. Entretanto, a comparação de outras laminarinases, por exemplo, de *Nocardiopsis sp.* e *Phanerochaete chrysosporium* mostra valores muito parecidos de PS, apesar de uma ser termofílica, exibindo atividade ótima a

77 °C (Fibriansah *et al.*, 2007) e outra mesofílica, respectivamente. Esta diferença motivou a comparação do comportamento dinâmico destas enzimas a partir de simulações de MD. Embora alguns aspectos estruturais relacionados à termoestabilidade e termofilicidade possam ser identificados mesmo na estrutura estática, as simulações de MD são muito úteis para avaliar como estes parâmetros variam com diferentes condições de interesse, com o objetivo de compreender melhor a natureza estrutural da estabilidade térmica e termofilicidade da RmLamR.

3.3 Metodologia

3.3.1 Construção dos sistemas e simulações

Três modelos estruturais, as laminarinases de RmLamR, de *Nocardiopsis sp.* cepa F96 (aNLam, PDB id 2HYK) e *Phanerochaete chrysosporium* (PcLam, PDB id 2CL2), foram submetidos a simulações de MD. Este conjunto compõe-se de três estruturas com diferentes características térmicas: RmLamR é hipertermofílica, aNLam é termofílica e PcLam é mesofílica, sendo as temperaturas ótimas de atividade identificadas como 88 °C (pH 5,5) (Krah *et al.*, 1998), 70 °C (pH 6,0) (Masuda *et al.*, 2006) e 30-40 °C (Bumpus & Aust, 1987), respectivamente.

Apesar da pouca identidade de sequência primária (Figura 3.2), as estruturas terciárias destas enzimas são muito similares, e os resíduos catalíticos dispostos na fita- β S8 são altamente preservados. O enovelamento global de camadas de folhas- β é essencialmente idêntico entre as três estruturas (Figura 3.3). Os *loops* são bastante similares entre RmLamR e aNLam, mas diferem substancialmente em PcLam, a qual apresenta numerosas α -hélices em regiões periféricas. Devido à significativa diferença na sequência primária e no tamanho de RmLamR e PcLam, é preciso ter cautela na

comparação do comportamento destas duas proteínas. No entanto, entre os membros mesofílicos desta família de proteínas de estrutura cristalográfica resolvida, PcLam é a mais próxima de RmLamR.

RmLamR	--RLPHWELVWSDEFD--YNGLPDPAKWDYDVGGHGWGNQELQYYTRARIENARVGGGV	56
aNLam	TESDMRATLVWSDEFDGPAGSAPDPANWNHETGDHGWGNELQNYTDSRANSALDGNL	60
PcLam	-----ATYHLEDNWVG-SAFLSTFTHEAIADPTHGRVN-YVDQATALAKNLTYASGDTL	52
	. * : : . : : * * * : : * : : . . . *	
RmLamR	IIEARRE---SYEGREYTSARLVTRGKASWTYGRFEIRARLPSGRGTWPAIWMLPDRQTY	113
aNLam	VITARQE---ADGG--YTSARLTTQNKVQPQYGRVEASIQIPRGGIWPAPFWMLG--ADF	113
PcLam	ILRADHTTTLSPSGPGRNSVRIRSIKTYTTHVAVFDVR-HMPQGCCTWPAAWETD-----	106
	:: * : : * . * . * : : . . : : * * * * * * *	
RmLamR	GSAYWPDNGE IDIME HVGFNP-----DVVHGTVHTKAYNHLLGTQRGG	156
aNLam	PNTPWPDSGE IDIMEN IGREP-----HLVHGSLHGPGY--FGGEPLTG	154
PcLam	-EGDWPNNGE VDIIE GVNDQSPNAMTLHTGANCAMPASRTMTGHATNNNCDVNTDGNTGC	165
	. * : . * : * : * : . . . : *	
RmLamR	SIRVPTAR---TDFHVYAIIEWTPPEIR-----WFVDDSLYYRFPNERLTNPEADWRHWP	207
aNLam	SYMHPQGSFADTFHTFAVDWRPGSIT-----WSVDGVAYQTY-----TSADTRGNPWV	203
PcLam	GVQAPTANSYGPSFNANGGGWYAMERTNSFIKVVFFPRNAGNVPNDIASGPATINTDNWG	225
	. * . * : . * . . * . . * . . *	
RmLamR	FDQPFH-----LIMNIAVGGTGGQGVDPFAFPAQLVVDYVRVYRWVE-	251
aNLam	FDQPFH-----MILNVAVGGDWPQYD--DGSTQFPQEMRVDYVRVYELG--	245
PcLam	TPTAFFPNTNCDIGSHFDANNIIINLTFCDWAGQASIFNGAGCPGSCVDYVNNNPSAFA	285
	. * . : * : * : . * * * * * * *	
RmLamR	-----	
aNLam	-----	
PcLam	NAYWDIASVRVYQ	298

Figura 3.2. Alinhamento com base estrutural das sequências primárias de RmLamR, aNLam e PcLam. A tríade catalítica conservada (Glu, Asp, Glu) é destacada em negrito. A porcentagem de identidade entre RmLamR e aNLam, e entre RmLamR e PcLam são 44% e 13%, respectivamente.

Na construção das caixas de simulação, alguns procedimentos específicos foram tomados. A estrutura tridimensional de aNLam foi completada com resíduos faltantes nas extremidades, de acordo com sua sequência primária (sequência TESDMR na extremidade N-terminal e LG na extremidade C-terminal). As coordenadas destes resíduos foram modeladas a partir do alinhamento de 2HYK à cadeia principal de

RmLamR. As posições dos outros átomos faltantes foram modeladas de acordo com o arquivo de topologia CHARMM27 (MacKerell *et al.*, 1998). As enzimas foram hidratadas por 15.000 moléculas de água em uma caixa cúbica com lados de aproximadamente 80 Å, usando o programa Packmol (Martínez & Martínez, 2003; Martínez *et al.*, 2009), de forma que a camada de solvente ao redor da superfície da proteína tivesse, pelo menos, 12 Å de espessura.

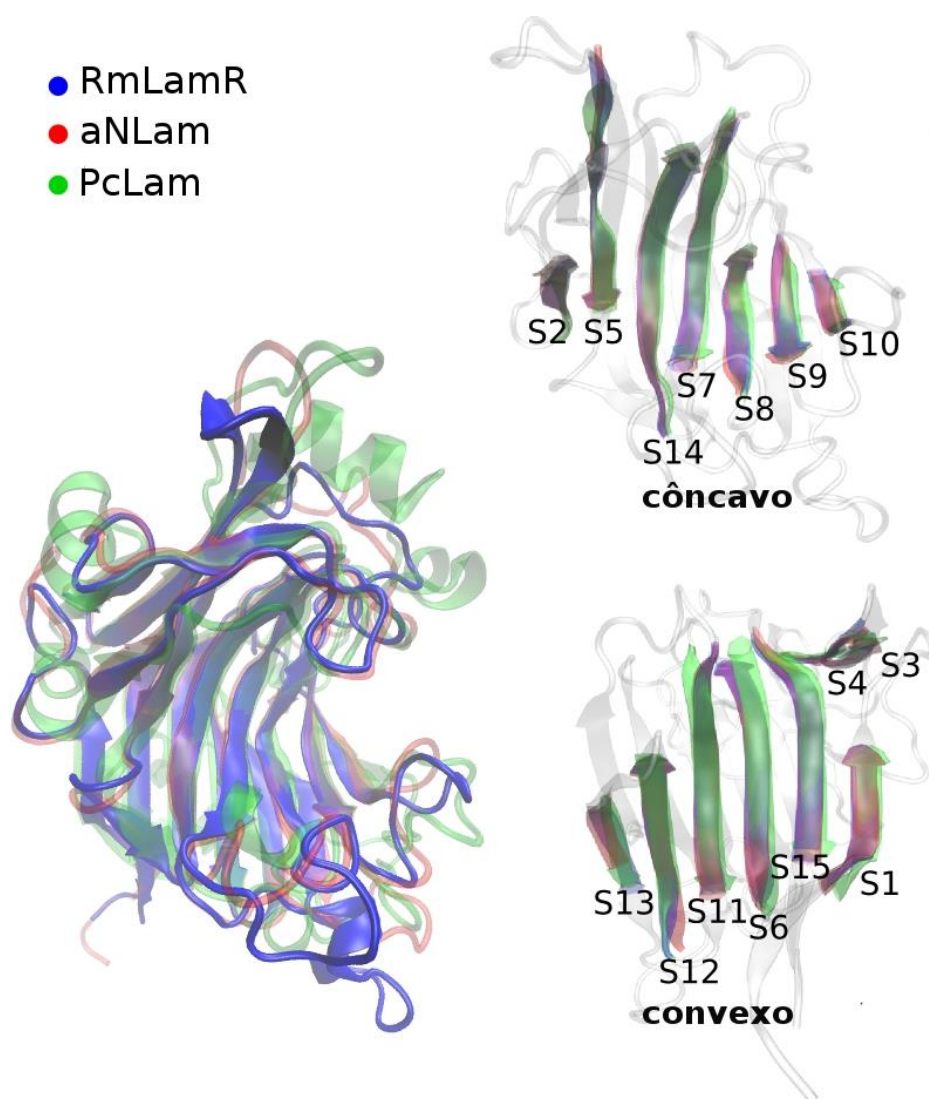


Figure 3.3 Alinhamento estrutural de laminarinases de *R. marinus* (RmLamR - azul), *Nocardiosis sp.* (aNLam - vermelho) e *P. chrysosporium* (PcLam - verde). As fitas- β , numeradas de 1 a 15, são as porções mais sobreponíveis das três enzimas.

Para determinar os estados de ionização dos resíduos ionizáveis (Lys, Arg, His, Asp, Glu), os valores de pK_a correspondentes foram determinados de acordo com o pH neutro e o ambiente molecular (alta constante dielétrica na superfície da proteína e baixa constante dielétrica no seu interior), usando o servidor H++ (<http://biophysics.cs.vt.edu/H++>, último acesso em janeiro de 2013; Anandakrishnan & Onufriev, 2012; Myers *et al.*, 2006; Gordon *et al.*, 2005). Foi dedicada atenção especial na escolha dos estados de ionização dos resíduos no sítio catalítico. Os resíduos Asp131 e Glu134 de RmLamR foram considerados protonados, de forma que eles podem interagir entre eles e com o substrato por meio de ligações de hidrogênio. O resíduo nucleofílico Glu129 foi mantido em sua forma carregada. Em todos os sistemas, foi adicionada a quantidade basal de 50 íons cloreto e 50 íons sódio, além do íon cálcio cristalográfico. A fim de manter a eletroneutralidade dos sistemas, foram adicionados excessos de 1, 13 e 3 íons sódio aos sistemas RmLamR, aNLam e PcLam, respectivamente. A concentração salina resultante é de aproximadamente 0,16 M.

Tabela 3.2. Resumo dos sistemas simulados. Todas as "caixas de simulação" consistem-se da enzima imersa em água e estão eletricamente neutras pela presença de íons dissolvidos. Outros componentes secundários advêm da estrutura cristalina e serão comentados no texto. Todos os sistemas apresentam pH 7,0.

Sistemas	Enzima	Condições	Simulações
RmLamR	Laminarinase de <i>R. Marinus</i>	A) 298 K, 1 bar	A) 10 simulações x 12 ns
		363 K, 1 bar	5 simulações x 9 ns
		B) 500 K	B) 3 simulações x 36 ns
aNLam	Laminarinase de <i>Nocardiosis sp.</i>	A) 298 K, 1 bar	A) 8 simulações x 12 ns
		363 K, 1 bar	5 simulações x 9 ns
		B) 500 K	B) 3 simulações x 36 ns
PcLam	Laminarinase de <i>P. chrysosporium</i>	A) 298 K, 1 bar	A) 8 simulações x 9 ns
		363 K, 1 bar	5 simulações x 9 ns
		B) 500 K	B) 3 simulações x 36 ns

O estudo da termoestabilidade e termofilicidade através de simulações de MD de laminarinases foi conduzido em duas etapas:

a) Simulações comparativas a 298 K (25 °C) e 363 K (90 °C). A temperatura mais alta foi adotada por ser próxima à de atividade máxima da enzima mais estável termicamente.

Para rodar as simulações, o protocolo de minimização de energia e equilibração descrito no Capítulo 2, foi adotado. Para as simulações a maiores temperaturas (363 K), o período de pré-equilibração descrito no Capítulo 2 foi estendido em duas etapas: 500 ps de equilibração em um estado intermediário de temperatura (330,5 K) e 1 ns de equilibração a 363 K. A partir dos sistemas equilibrados, foram conduzidas 10 simulações independentes de 12 ns para RmLamR, 8 simulações de 12 ns para aNLam e 9 corridas de 9 ns para PcLam a 25 °C. A 90 °C, foram conduzidas cinco simulações independentes de 9 ns para cada proteína.

b) Simulações de desnaturação de laminarinases. Uma vez que eventos de desnaturação de proteínas ocorrem majoritariamente na escala de microssegundos e que este estudo trata de enzimas termoestáveis e hipertermoestáveis, as simulações foram conduzidas à temperatura de 500 K (227 °C), de modo a acelerar o processo e torná-lo acessível em termos de custo computacional, seguindo os moldes de vários estudos interessantes de MD de desnaturação de proteínas (Vendruscolo & Paci, 2003; Daggett & Fersht, 2003a; Daggett & Fersht, 2003b; Dinner *et al.*, 2000; Fersht & Daggett, 2002; White *et al.*, 2005; Day & Daggett, 2005; Daggett, 2006; Martínez *et al.*, 2010).

As simulações nesta etapa do trabalho foram conduzidas tendo desligado do controle de pressão, estendendo as simulações no *ensemble* canônico (NVT). Nas simulações a 500 K, portanto, os sistemas estão pressurizados. A pressão média destas simulações é da ordem de 4 kbar. Esta última fase foi reproduzida em triplicata para

cada sistema, tendo duração de 36 ns, cada uma. Além disso, conduziu-se uma simulação de referência, a 298 K, para cada sistema, com a mesma extensão, de 36 ns, com o controle de pressão a 1,01 bar. A Tabela 3.2 apresenta o resumo dos sistemas simulados.

3.4 Simulações comparativas a 298 e 363 K

3.4.1 Estrutura e termoestabilidade

A fim de obter maior compreensão acerca de como fatores estruturais afetam as propriedades enzimáticas sob condições termodinâmicas de interesse, utilizou-se a técnica de simulações de MD como ferramenta. Inicialmente, foram conduzidas simulações a baixa (298 K) e alta (363 K) temperaturas, visando verificar a estabilidade e mobilidade dos três modelos. A partir das simulações, verificou-se que o enovelamento global, expresso por parâmetros como o raio de giração (R_g), é, essencialmente, preservado nas estruturas independentemente da variação na temperatura (Tabela 3.3). O valor R_g de PcLam, a maior das enzimas, é ligeiramente maior, mas também se mostra independente da temperatura. A diferença das estruturas médias das simulações em relação às estruturas cristalográficas correspondentes, medida através de desvio médio quadrático (RMSD) dos átomos da cadeia principal, encontra-se próximo a 1 Å para todas as estruturas e em todas as temperaturas. Esta estabilidade estrutural de RmLamR mostra-se consistente com sua hipertermoestabilidade, uma vez que esta enzima preserva atividade a altas temperaturas e, portanto, sua estrutura também mostra-se preservada. As enzimas aNLam e PcLam são completamente inativas a esta temperatura e, embora não fique evidente através destas médias globais, outras análises que serão apresentadas mais adiante sugerem os primeiros sinais de desnaturação térmica destas

proteínas.

No nível das cadeias laterais, os efeitos do aumento na temperatura tornam-se notáveis. O número médio de ligações de hidrogênio (LH) decresce de 77 a 298 K para 65 a 363 K para RmLamR, de 59 a 53 para aNLam, e de 77 a 62 para PcLam. Ao mesmo tempo, as pontes salinas de todas as estruturas permanecem praticamente intactas quando a temperatura aumenta de 298 K a 363 K. Destaca-se que o número médio de contatos hidrofóbicos (CH), formados entre resíduos localizados nas camadas opostas de folhas- β antiparalelas, é aproximadamente 10% maior em RmLamR do que nas outras duas proteínas, indicando o melhor "empacotamento" hidrofóbico de RmLamR.

Em resumo, estes resultados de MD mostram principalmente que, nas condições termodinâmicas avaliadas e em comparação com suas homólogas estruturais, a enzima termofílica RmLamR exibe um número substancialmente maior de pontes salinas e níveis mais elevados de ligações de hidrogênio e contatos hidrofóbicos intraproteína, os quais são os prováveis fatores contribuintes para a sua estabilidade térmica. Em termos dos parâmetros PS, LH e CH, aNLam e PcLam são muito similares. Entretanto, considerando que PcLam é significativamente maior do que suas homólogas (RmLam, aNLam e PcLam têm 251, 245 e 298 resíduos, respectivamente), o conteúdo de PS, LH e CH por resíduo é menor em PcLam. Os gráficos que geraram os valores médios destas propriedades encontram-se no Anexo I. Então, como estas características realmente influenciam na termoestabilidade de laminarinases?

Tabela 3.3. Dados computados a partir das simulações das laminarinases homólogas (RmLamR, aNLam and PcLam) a 298 K e a 363 K. Os parâmetros estruturais são: raio de giração (R_g), conteúdo de estrutura secundária (Est. Sec.), número de contatos hidrofóbicos (CH), número de ligações de hidrogênio (LH), número médio de pontes salinas (PS), e desvio médio quadrático de átomos na cadeia principal (RMSD).

	RmLamR		aNLam		PcLam	
	298 K	363 K	298 K	363 K	298 K	363 K
R_g (Å)	17.7 ± 0.1	17.7 ± 0.1	17.3 ± 0.1	17.4 ± 0.1	18.3 ± 0.1	18.4 ± 0.1
RMSD (Å.ps ⁻¹)	0.82	0.94	0.92	1.04	0.72	0.94
Est. Sec.	49.0 %	48.8 %	45.6 %	44.2 %	50.5 %	49.8 %
LH	77 ± 7	65 ± 7	59 ± 6	53 ± 6	77 ± 8	62 ± 6
PS	24 ± 3	24 ± 2	9 ± 1	10 ± 2	11 ± 1	11 ± 1
CH	166 ± 2	165 ± 3	150 ± 2	151 ± 3	153 ± 3	152 ± 4

* RmLamR, aNLam, and PcLam têm 251, 245, e 298 resíduos, respectivamente.

Chama a atenção o fato de as pontes salinas não estarem aleatoriamente distribuídas em RmLamR. Os mapas de conectividade de PSs mostram que, considerando apenas aquelas que persistem por mais de 50% do tempo de simulação, nas três estruturas, há várias pontes salinas isoladas (isto é, *clusters* de tamanho N=1), assim como *clusters* contendo N=2 ou mais pontes salinas adjacentes. Entretanto, notavelmente, apenas RmLamR exhibe *clusters* de 5 ou até 6 PS interconectadas, como pode ser visto na Figura 3.4A. Estas formam padrões como vigas conectando fitas- β adjacentes (Figura 3.4B). Sugere-se que estes agrupamentos de pontes salinas em RmLamR lhe confirmam estabilidade adicional à RmLamR.

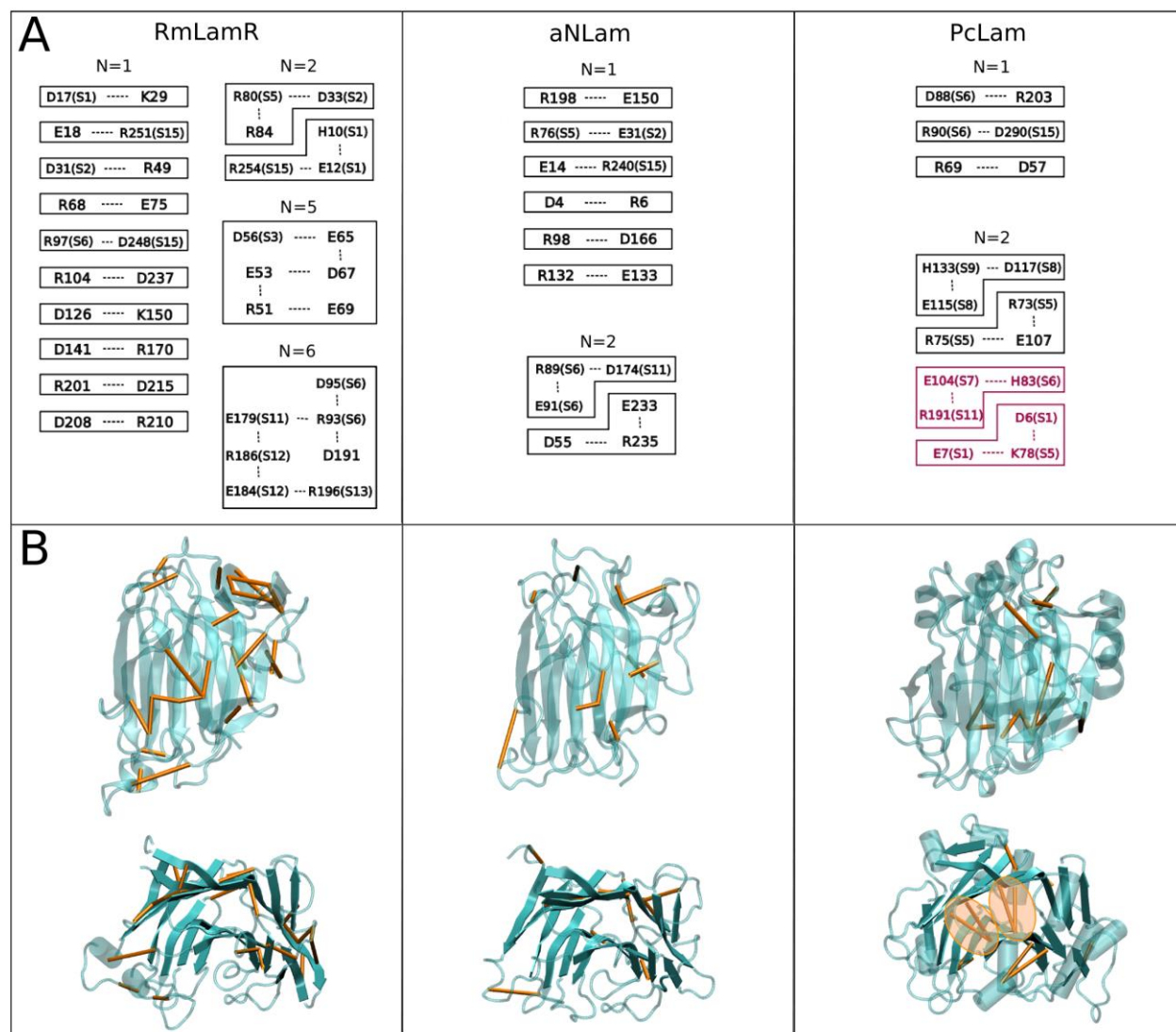


Figura 3.4. A) Organização topológica de pontes salinas de laminarinases. Cada linha tracejada corresponde a uma ponte salina que dura, pelo menos, 50 % do tempo de simulação. Em parênteses, encontram-se os nomes das fitas- β . Em RmLamR, há aglomerados contendo até N=6 PS interconectadas. Em PcLam, há pontes salinas conectando resíduos localizados nas folhas externa e interna (marcadas em rosa). B) A conectividade das pontes salinas, descritas em A é representada nas estruturas superiores por traços que conectam os átomos C_{α} dos resíduos correspondentes. As estruturas mais abaixo mostram que apenas na enzima mesofílica, as pontes salinas são formadas entre as camadas de folhas- β (PS “inter-folhas”).

Uma análise mais minuciosa da topologia de PS revela que as homólogas diferem

em um outro aspecto importante: as pontes salinas nas estruturas termofílicas não conectam resíduos de diferentes camadas de folhas- β ao contrário da mesofílica PcLam, em que quatro PS conectam as duas camadas. Assim, poderia ser dito que, nesta enzima, além de pontes salinas “intra-folhas”, encontram-se pontes salinas “inter-folhas”. Na PcLam (Figura 3.4B, painéis inferiores), Glu104 da fita S7 liga-se simultaneamente à His83 de S6 e à Arg191 de S11; e Lys78 de S8 faz pontes salinas com Asp6 e Glu7 de S1. Esta é uma característica muito reveladora: indica que em PcLam há interações hidrofílicas conectando as camadas de folhas- β interna e externa, enquanto que, nas mais termoestáveis RmLamR e aNLam, o centro hidrofóbico é formado exclusivamente por contatos hidrofóbicos entre resíduos.

A presença de pontes salinas no caroço hidrofóbico de proteínas tem efeitos diretos na resistência térmica destas. As figuras de momentos das simulações das três proteínas a 298 e 373 K (Figura 3.5) mostra que RmLamR e aNLam coordenam fortemente uma e duas moléculas de água cristalográficas, respectivamente, em seu interior hidrofóbico em ambas as temperaturas, enquanto que, na menos termoestável PcLam, há um aumento considerável no número de moléculas de água que penetra o interior hidrofóbico da proteínas.

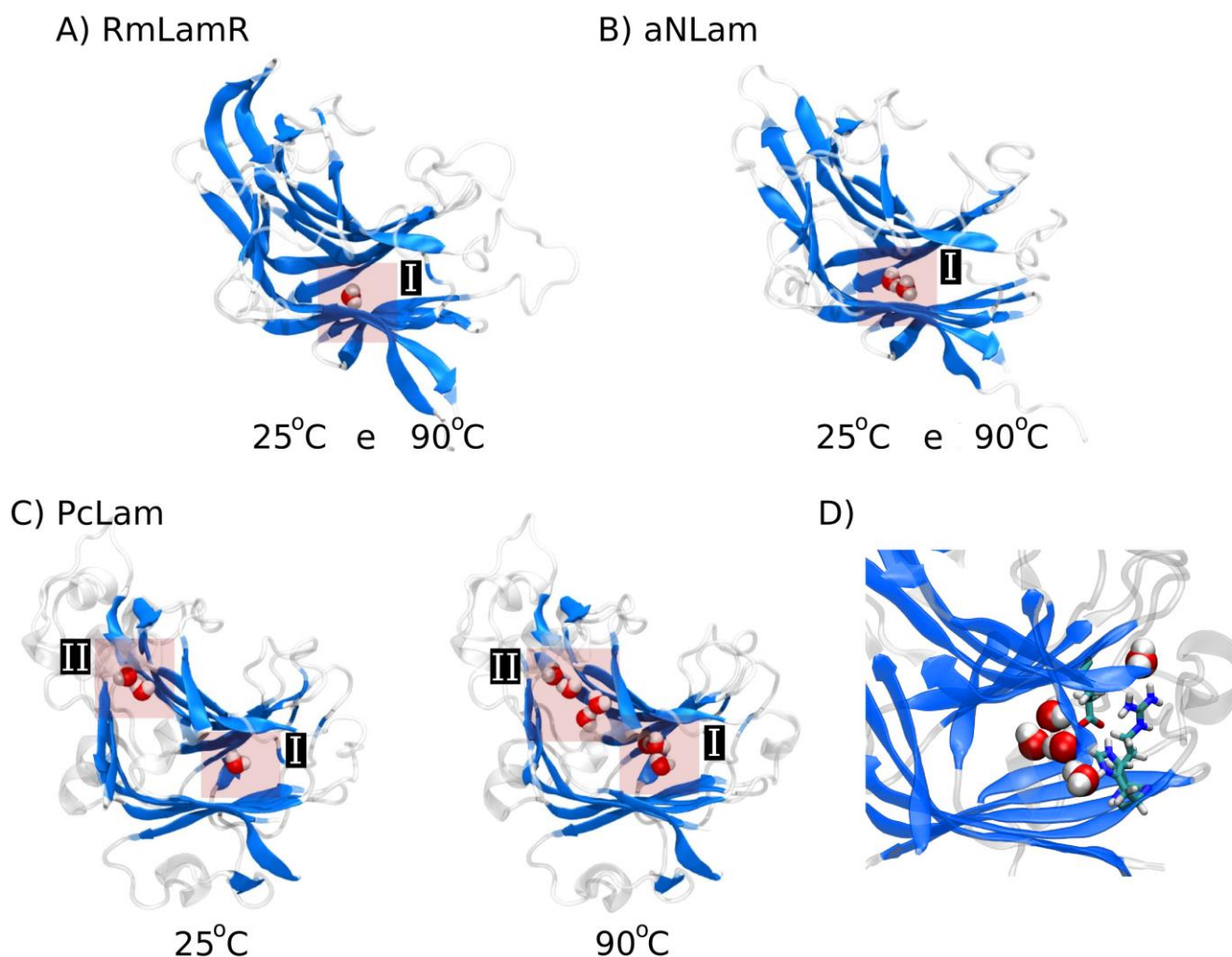


Figura 3.5. Entrada de água entre as camadas de fitas- β . A) Uma única molécula de água encontra-se presente no interior hidrofóbico de RmLamR a 298 (25 °C) e 363 K (90 °C) (região I). B) Duas moléculas de água penetram no carço hidrofóbico de aNLam a 298 e 363 K (região I). C) O carço hidrofóbico de PcLam é mais acessível à água, uma vez que três moléculas de água são encontradas no interior a 298 K e até seis, a 363 K (regiões I e II). D) Entrada de água no interior hidrofóbico de PcLam é facilitada por pontes salinas entre as camadas de folhas- β .

Um dos autores deste trabalho, Thiago C. F. Gomes, fez uma análise quantitativa mais detalhada da entrada de água no centro hidrofóbico. O programa que o Thiago elaborou contava quantas moléculas de água adentrava um volume que ele definiu para cada sistema como sendo o volume do centro hidrofóbico da proteína. Ele mostrou, então, que nas simulações de PcLam a 25 °C há uma média de três moléculas de água

em seu interior hidrofóbico ao longo da simulação, ao passo que, a 90 °C, este número sobe para seis.

As simulações também revelam que as PS entre camadas opostas de folhas- β presentes em PcLam funcionam como uma porta para a entrada de moléculas de água em seu interior hidrofóbico (Figura 3.5D), contribuindo, assim, com a desestabilização térmica de PcLam em comparação com suas homólogas mais estáveis termicamente. Deste modo, a disposição de PS é um fator crucial para a estabilidade térmica: PS orientadas de forma inapropriada na estrutura podem ter um papel desestabilizador.

3.4.2 Dinâmica na fenda catalítica e acessibilidade

Uma questão central na adaptação de biomoléculas a condições extremas é a conservação do seu estado funcional, o que significa o compromisso balanceado entre flexibilidade e estabilidade estrutural (Tsou, 1988; Jaenicke, 1991a). A Figura 3.6 mostra a mobilidade relativa de cadeias laterais ao longo de toda a proteína, com vermelho e azul indicando maior e menor mobilidade, respectivamente, a 363 K em relação às mobilidades a 298 K. Todas as estruturas exibem regiões em que a mobilidade das cadeias laterais aumentou com o aumento da temperatura, particularmente os *loops*, e regiões que permaneciam estáveis a 363 K. Interessantemente, as três cadeias laterais da tríade catalítica (Glu126, Asp128 e Glu131; numeração da RmLamR) estão entre as regiões que foram estabilizadas ou muito pouco perturbadas nas enzimas RmLamR e aNLam. Ao mesmo tempo, um aumento bastante significativo na mobilidade destes resíduos é verificado em PcLam.

É interessante observar que a mobilidade da tríade catalítica é não apenas muito pouco afetada pelo aumento da temperatura em RmLamR e aNLam, mas é também quantitativamente similar à mobilidade dos resíduos correspondentes em PcLam a 298 K, como é mostrado na Figura 3.6B. PcLam é a única que apresenta grandes alterações

nos resíduos catalíticos, Asp117 e Glu120 na temperatura mais elevada.

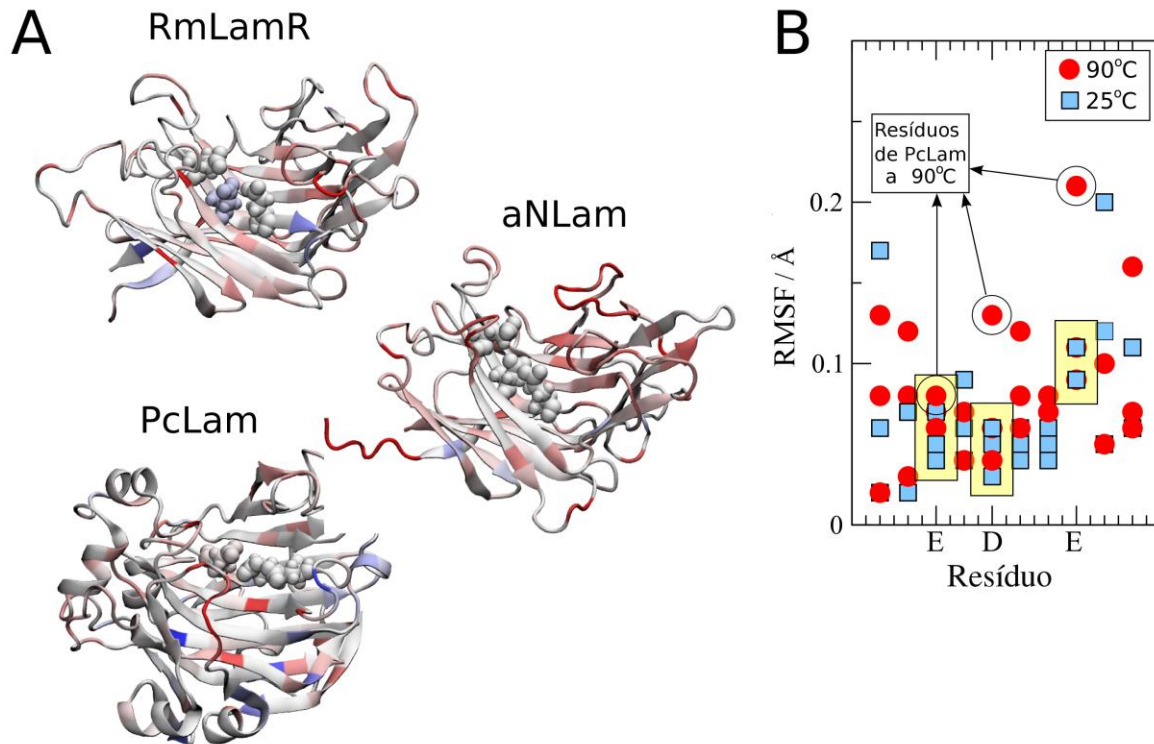


Figura 3.6. Mobilidades das cadeias laterais dos resíduos computadas por RMSF. (A) Diferenças em mobilidades observadas para as estruturas a 298 e 363 K. De azul a vermelho, as cores indicam um aumento na mobilidade a temperatura mais elevada em relação à mais baixa. Branco indica que não houve mudança. (B) Mobilidades das cadeias laterais dos resíduos adjacentes à tríade catalítica (Glu126, Asp128 e Glu131 para RmLamR); cada símbolo representa a mobilidade de um resíduo em uma temperatura. A mobilidade da tríade catalítica aumenta para a mesófila PcLam, mas não é sensível à variação de temperatura nas termófilas aNLam e RmLamR.

Costumava-se associar termofilicidade e termoestabilidade a uma maior rigidez da enzima em relação a uma homóloga mesófila (Vieille & Zeikus, 1996). O que este resultado indica é um pouco diferente disto, isto é, indica que a enzima termófila é capaz de preservar a flexibilidade e conformação do sítio ativo que caracterizam o estado funcional. O compromisso entre estabilidade estrutural e flexibilidade de proteínas em diferentes regiões é um aspecto chave na função biomolecular em

ambientes extremos (Tsou, 1988; Tsou, 1998). As alterações adaptativas essenciais tendem a regular características comuns destas proteínas a favor das respectivas condições ótimas de funcionamento, de forma que estas proteínas mantenham-se nos, assim chamados, “estados correspondentes”, considerando a topologia de enovelamento, a flexibilidade estrutural e a exposição ao solvente (Jaenicke, 1991). Seguindo o princípio dos estados correspondentes, RmLamR, aNLam e PcLam mostram valores de mobilidade no sítio ativo que são similares nas temperaturas correspondentes à atividade ótima.

A mobilidade dos resíduos nas vizinhanças do sítio ativo é outro aspecto chave diferencial entre enzimas mesofílica e termofílicas em altas temperaturas. Rodrigo L. Silveira, um dos autores do trabalho publicado, referência (Bleicher & Prates *et al.*, 2011), mostrou um resultado muito interessante que vale ser comentado. A exposição dos sítios ativos das três enzimas é notavelmente afetada de modos diferentes pelos movimentos das cadeias laterais próximas nos *loops* que cercam a fenda catalítica. Nas três estruturas, a fenda catalítica pode ser categorizada em três conformações, de acordo com a sua acessibilidade a um substrato: estado aberto, fechado/obstruído ou tubular. Verificou-se que a população conformacional de PcLam a 90 °C é dominada por estado obstruídos, enquanto que a conformação aberta prevalece em aNLam, mesmo à alta temperatura. Deste modo, dado que a obstrução do canal de ligação ao substrato gera baixo níveis de atividade enzimática, as simulações fornecem uma razão plausível, a nível molecular, para a maior termofilicidade de aNLam em comparação com PcLam. A distribuição das conformações da fenda catalítica de RmLamR, por outro lado, é notavelmente menos sensível ao aumento de temperatura. A população conformacional é dominada pela formação da estrutura tubular a 298 K, e este estado continua predominante a 363 K. Presume-se que a conformação tubular, a de maior ocorrência em simulações de ambas as temperaturas de RmLamR é, de fato, a mais ativa. Esta

conformação possibilitaria a entrada do substrato polissacarídico no canal do sítio ativo e restringiria sua liberação ao solvente, estabilizando assim um complexo que promove mais interações efetivas enzima--substrato, potencializando a catálise.

3.5 Simulações de desnaturação de laminarinases

Como uma continuidade natural ao estudo da termoestabilidade e termofilicidade de RmLamR, nesta etapa, buscou-se encontrar evidências mais claras de fatores importantes para a diferença na estabilidade térmica de laminarinases através de simulações de desnaturação de laminarinases, conforme descrito na Seção 3.3.1.

As simulações a 500 K foram longas o suficiente para que processo de desenovelamento avançasse bastante nas homólogas de diferentes níveis de termoestabilidade, como pôde ser verificado pela análise visual das simulações e através da evolução temporal de algumas propriedades (Figura 3.7). Por exemplo, verifica-se que os RMSDs dos átomos C_{α} em relação às suas posições cristalográficas aumentam sistematicamente nas simulações a 500 K. No final das simulações, alcança-se valores entre 7 e 16 Å, com PcLam e aNLam exibindo maiores desvios. Observa-se que, nos primeiros nanossegundos, há maior inclinação nos gráficos de RMSD de PcLam e aNLam em relação à homóloga hipertermofílica, o que é condizente com o fato de PcLam, como enzima mesofílica, perder sua atividade com aumentos mais moderados de temperatura. Sugere-se que, a rápida alteração estrutural de aNLam decorra de seu menor tamanho, enquanto que, a menor variação nas posições dos átomos deva estar estritamente ligada às características estruturais que conferem a propriedade de hipertermoestabilidade à RmLamR.

Ao mesmo tempo, o raio de giração (R_g) das enzimas revelou uma expansão lenta e contínua de PcLam e RmLamR. Ao final, as estruturas aumentaram de tamanho em, no

máximo, 8 e 7%, respectivamente. Já aNLam apresentou um aumento muito mais significativo, alcançando o máximo de 16% de aumento.

Nas três estruturas, a evolução na quantidade de interações via ligação de hidrogênio entre átomos das cadeias laterais dos resíduos (Figura 3.7c), ocorreu de forma similar, isto é, houve um rápido decréscimo já nos primeiros picossegundos das simulações. Curiosamente, nos momentos seguintes, as simulações independentes de cada sistema mostram o número de LH oscilando ao redor de valores médios muito próximos. As interações de resíduos carregados, as pontes salinas, são mais fortes, e o valor médio das simulações a 298 e 500 K acaba por ter comportamento similar em todos os sistemas, não havendo decréscimo aparente. Observando em maior detalhe as interações via LH e PS, nota-se que as interações mais fortes perduram até o final da simulação, enquanto que as mais fracas são menos persistentes e podem ser restabelecidas entre resíduos diferentes. Em nosso estudo descrito na Seção anterior, sugeriu-se que a quantidade elevada de interações por pontes salinas e ligações de hidrogênio na RmLamR fosse um fator determinante no caráter hipertermoestável desta enzima. De fato, mesmo nas simulações a 500 K, o número destas interações permanece superior ao das demais.

Por fim, a porcentagem de estrutura secundária nas três estruturas diminuiu significativamente em todos os sistemas, com RmLamR preservando-se um pouco mais estruturada. Os trabalhos de simulação de desnaturação reportados na literatura que tratam de peptídeos ou pequenas proteínas indicam que a desnaturação destas moléculas inicia-se com a expansão de seus centros hidrofóbicos (os trabalhos no tema são numerosos. Boas referências sobre o assunto são os *reviews* (Daggett, 2002) e (Daggett 2006), por exemplo). Entretanto, a sequência de eventos no processo de desenovelamento é dependente da proteína. Um trabalho do nosso grupo de pesquisa, por exemplo, mostrou que este processo ocorre de forma diferente para uma proteína

maior, o domínio de ligação de ligantes (*ligand binding domain* - LBD) de receptores do hormônio tireoideano, composto por cerca de 260 resíduos (Martínez *et al.*, 2010). Isto é, para o LBD, o desenovelamento inicia-se com a perda de contatos nativos e de elementos de estrutura secundária, enquanto a estrutura permanece essencialmente compacta, assemelhando-se a um estado globular fundido. Em conjunto, os resultados para o caso das laminarinases, que também são moléculas relativamente grandes e com um núcleo hidrofóbico muito coeso, mostram que o processo ocorre de modo similar ao LBD, com a perda estrutura secundária ocorrendo antes ou simultaneamente a uma expansão vagarosa do raio de giração.

Na medida em que ocorre perda de estrutura secundária e terciária destas enzimas, as interações entre os resíduos são substituídas por interações com a água. A Figura 3.7f mostra como ocorre o aumento gradual no número de moléculas de água que solvata os resíduos no caroço hidrofóbico das três homólogas. Para fazer a contagem no decorrer das simulações, foram selecionados aproximadamente 40 resíduos do “sanduíche” de folhas- β , que se encontram voltados para o interior hidrofóbico de cada uma das enzimas. As triplicatas para cada sistema apresentam boa reprodutibilidade ao final das simulações em relação a este aspecto. Verifica-se um aumento notável no número de moléculas de água das enzimas PcLam e aNLam, ao passo que, em RmLamR, este acréscimo ocorre lentamente. Isto é condizente com a perda mais lenta na porcentagem de estrutura secundária e aumento também lento no raio de giração desta enzima. Em PcLam, a hidratação ocorre acentuadamente nas vizinhanças das pontes salinas.

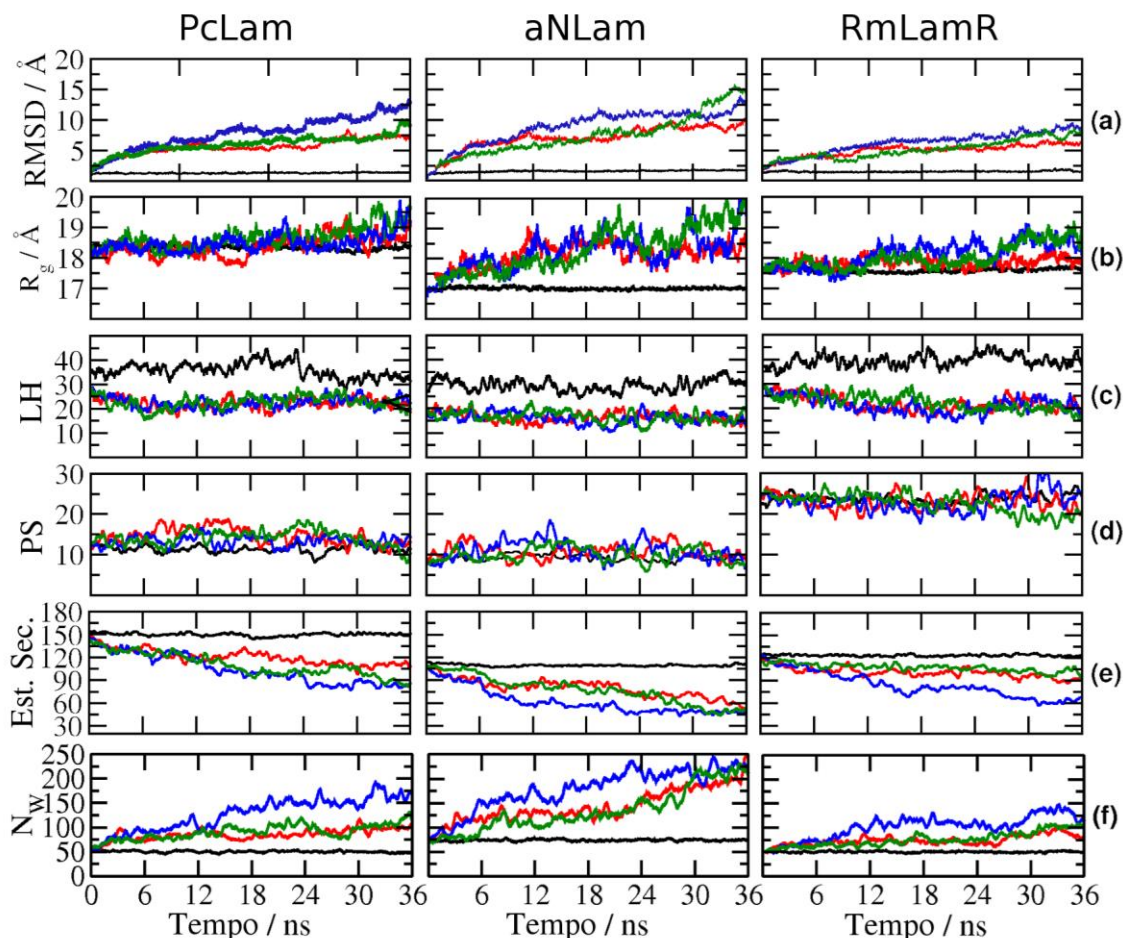


Figura 3.7. Análises das modificações estruturais globais ocorridas nas simulações de PcLam, aNLam e RmLamR a 298 K (linhas pretas) e a 500 K (linhas coloridas). São apresentadas a evolução temporal (a) do desvio médio quadrático (RMSD) dos C_α das enzimas em relação à estrutura cristalográfica, (b) do raio de giração (R_g), (c) do número de ligações de hidrogênio entre átomos das cadeias laterais dos resíduos (LH), (d) do número de pontes salinas (PS), (e) do número de resíduos que formam estrutura secundária (basicamente, hélices ou fitas) e (f) do número de moléculas de água (N_w) na primeira camada de solvatação ao redor das cadeias laterais dos resíduos no carço hidrofóbico das enzimas.

À primeira vista, as triplicatas das simulações não evidenciam etapas bem definidas no mecanismo de desnaturação ou particularidades que pudessem ser atribuídas à meso *versus* termoestabilidade das enzimas. Ocorre que são múltiplos os caminhos de desnaturação. Vale lembrar o perfil de energia afunilado associado a conformações acessíveis de uma proteína, mencionado no Capítulo 2 (Figura 2.2).

Apesar desta multiplicidade de vias de desenovelamento, suas etapas apresentam características que se reproduzem nos diferentes caminhos, pois estes são direcionados por fatores como efeito hidrofóbico, ligações de hidrogênio, pontes salinas, entre outros, e é justamente o papel destes fatores na estabilidade térmica que se pretendia estudar aqui.

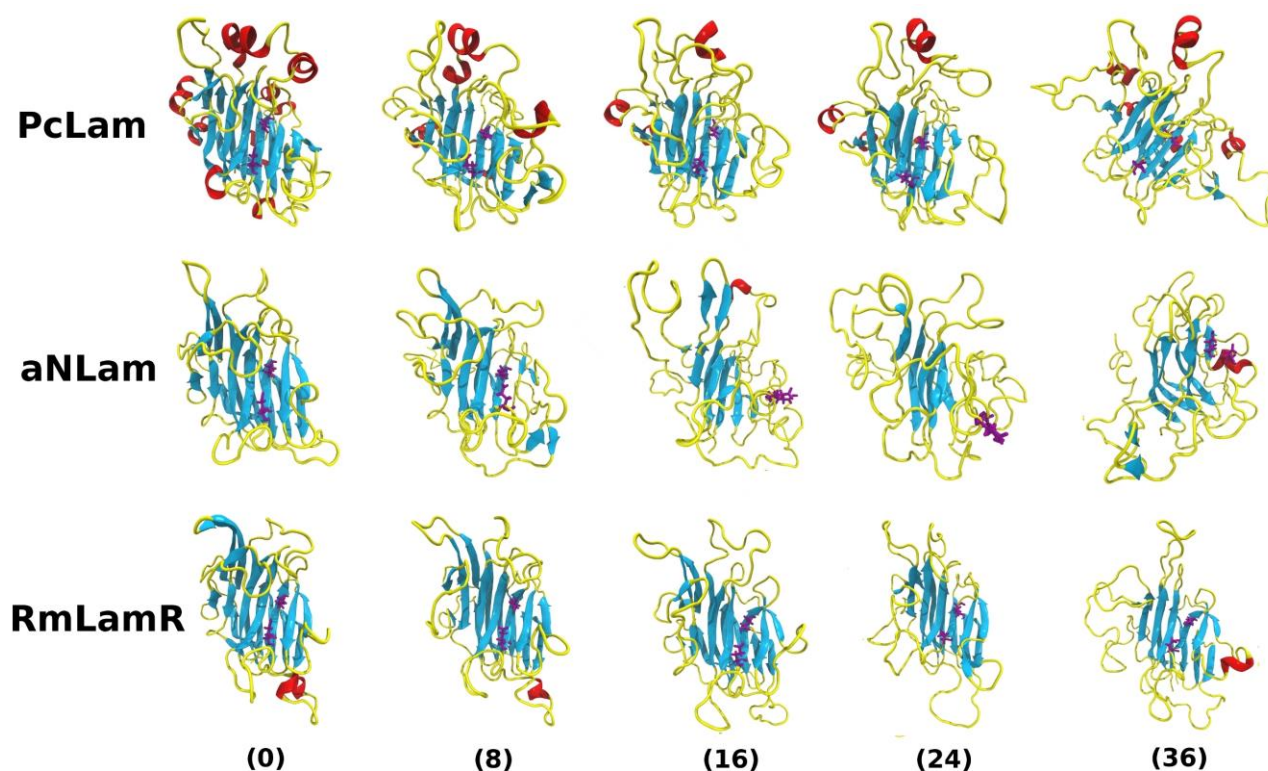


Figura 3.8. Estruturas das laminarinases em diferentes momentos das simulações (especificados entre parênteses, em nanossegundos). Os resíduos catalíticos (nucleófilo e catalisador ácido/base) são representados em violeta, as folhas- β em azul, hélices, em vermelho, e *loops* e porções desestruturadas são representadas em amarelo.

A Figura 3.8 mostra como ocorre a perda gradual de estrutura secundária e terciária das homólogas através de *snapshots* de uma das simulações de cada enzima em determinados momentos. Esta Figura ilustra o que foi explicado nas análises globais: A perda de estrutura secundária e contatos nativos precede a expansão do centro hidrofóbico das enzimas PcLam e RmLamR, ao passo que, em aNLam, tais processos

ocorrem simultaneamente. Em até 8 ns, modificações estruturais ocorreram especialmente nas regiões periféricas das três enzimas. No caso de PcLam, tais modificações já seriam suficientes para a inativação da enzima, pela obstrução da cavidade catalítica pelo *loops* laterais (Bleicher & Prates *et al*, 2011). O colapso do sítio catalítico ocorre rapidamente para aNLam antes dos 16 ns da simulação representada. Ao final das simulações, as estruturas da PcLam e aNLam aparecem acentuadamente distorcidas, ao passo que RmLamR, ainda deixa preservada boa parte de sua estrutura de folhas- β e mantém os resíduos catalíticos em posições parecidas com as iniciais.

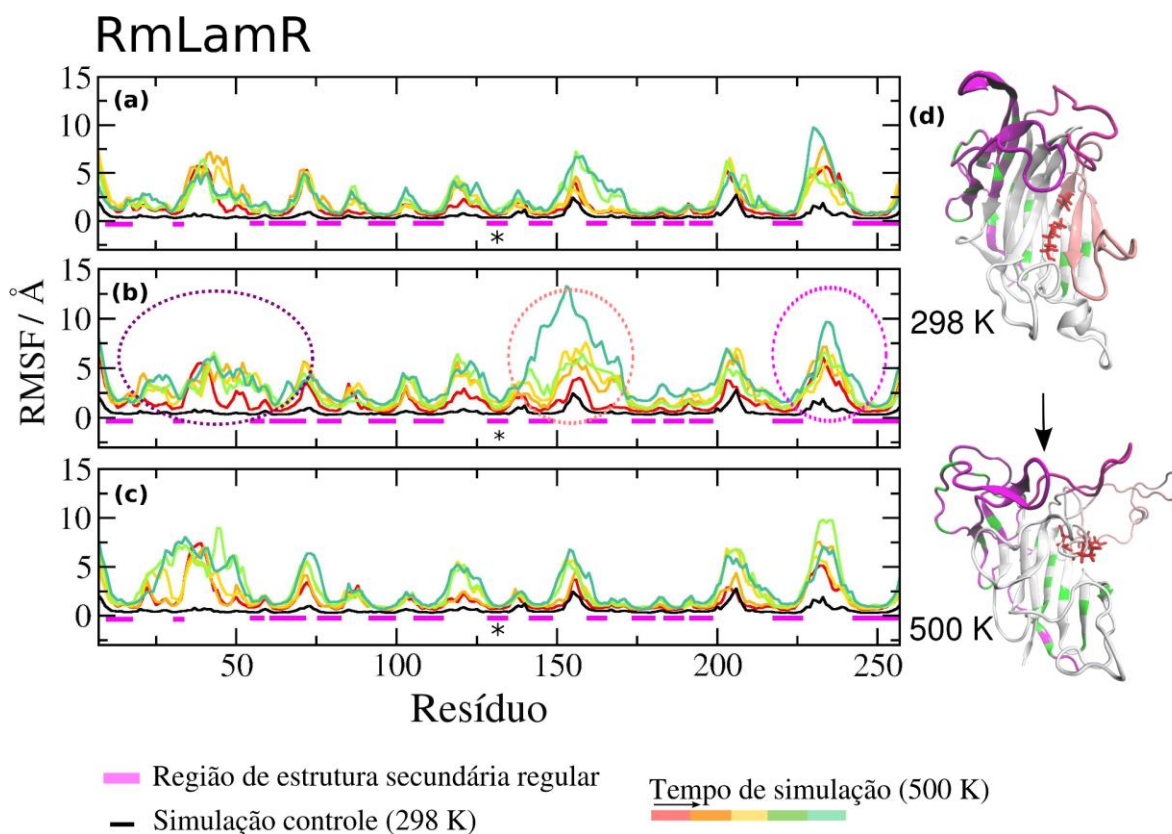


Figura 3.9. Flutuações dos resíduos em RmLamR em relação à estrutura média em cinco trechos de 7 ns de simulação de 500 K. A linha em preto refere-se à simulação a 298 K. (a), (b) e (c) correspondem às três simulações independentes. As regiões circuladas em (b) são as que mais sofreram alterações na mobilidade com o aumento na temperatura. Estas regiões estão destacadas em tons de rosa nas estruturas tridimensionais em (d), onde também se apresenta a tríade catalítica em cor vermelha e os resíduos que interagem via ponte salina nas duas temperaturas, em verde. As estruturas das simulações a 500 K foram extraídas do final das simulações.

Algumas características do processo de desnaturação ficaram mais claras na análise das flutuações das posições dos resíduos em relação à estrutura média em trechos sucessivos das simulações de desnaturação (Figuras 3.9-11). Os perfis de mobilidade mostram que, em todos os sistemas, as principais regiões afetadas estão em porções periféricas das enzimas, constituídas especialmente de *loops*. Estas regiões foram assinaladas nos gráficos e destacadas nas estruturas tridimensionais das enzimas nas simulações a diferentes temperaturas.

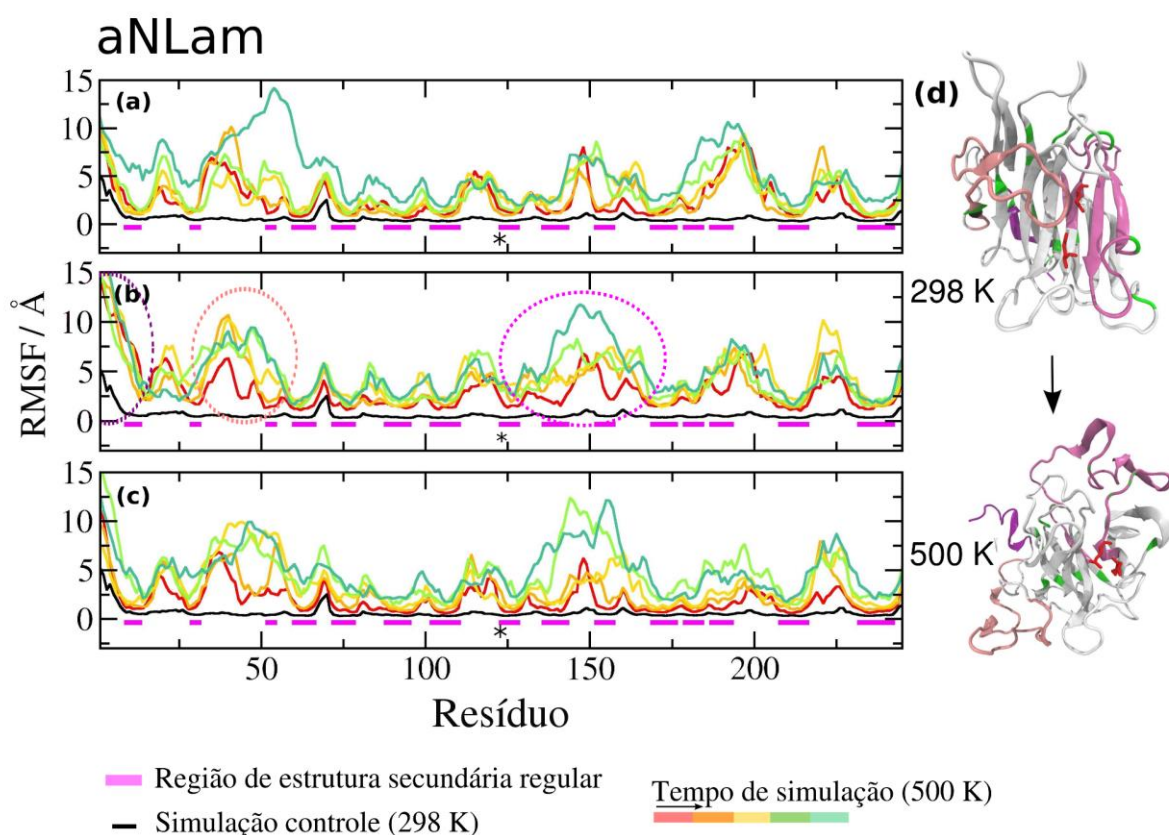


Figura 3.10. Flutuações dos resíduos em aNLam em relação à estrutura média em cinco trechos de 7 ns de simulação de 500 K. A linha em preto refere-se à simulação a 298 K. (a), (b) e (c) correspondem às três simulações independentes. As regiões circuladas em (b) são as que mais sofreram alterações na mobilidade com o aumento na temperatura. Estas regiões estão destacadas em tons de rosa nas estruturas tridimensionais em (d), onde também se apresenta a tríade catalítica em cor vermelha e os resíduos que interagem via ponte salina nas duas temperaturas, em verde. As estruturas das simulações a 500 K foram extraídas do final das simulações.

As curvas de RMSD por resíduo na Figura 3.9 são pouco distantes em trechos

iniciais das simulações de RmLamR, afastando-se gradativamente da curva da simulação controle, ao contrário do que ocorre nas simulações de PcLam e aNLam, em que o ganho na mobilidade ocorre rapidamente (Figuras 3.10-11). Além disso, para a RmLamR, as curvas referentes aos diferentes trechos de simulação mantêm um perfil mais bem definido e parecido com o da simulação controle do que no caso da PcLam, em que os perfis de RMSD logo perdem as características da curva da simulação de referência.

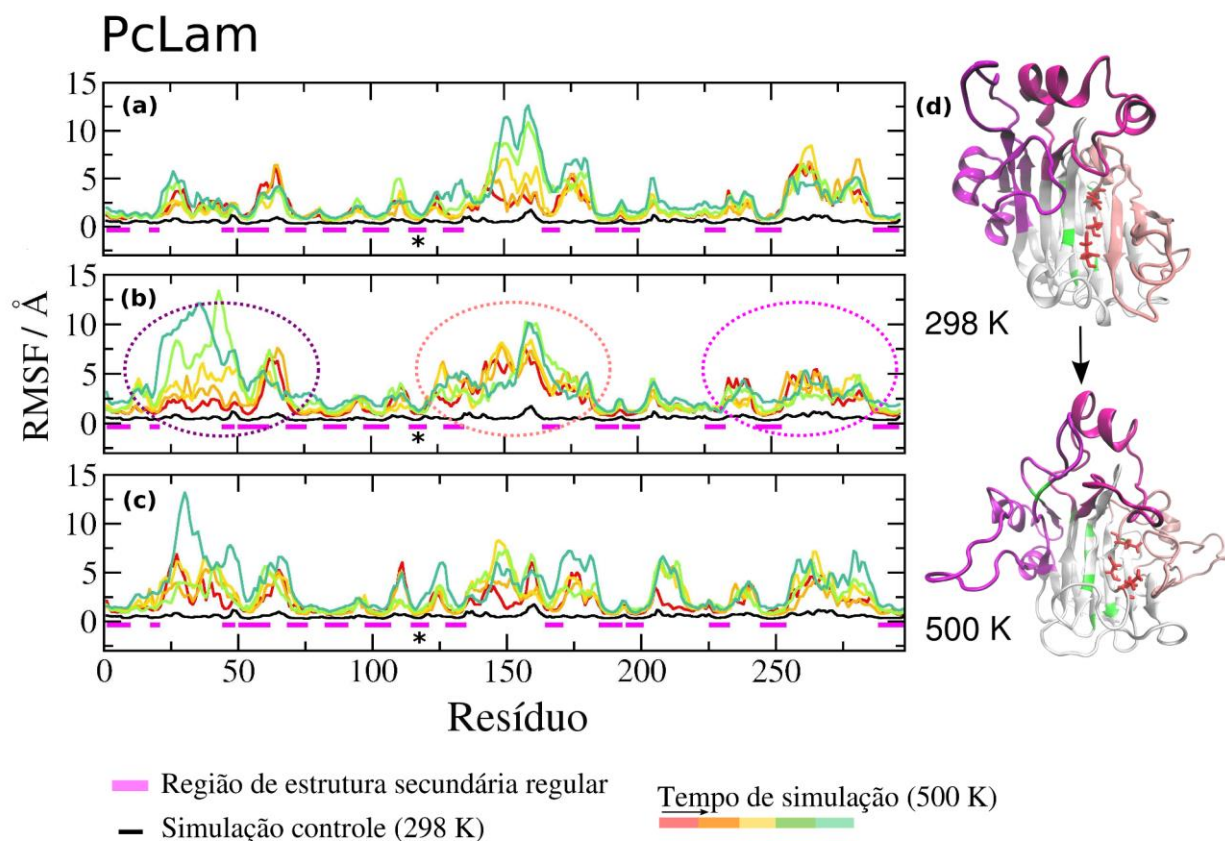


Figura 3.11. Flutuações dos resíduos em PcLam em relação à estrutura média em cinco trechos de 7 ns de simulação de 500 K. A linha em preto refere-se à simulação a 298 K. (a), (b) e (c) correspondem às três simulações independentes. As regiões circuladas em (b) são as que mais sofreram alterações na mobilidade com o aumento na temperatura. Estas regiões estão destacadas em tons de rosa nas estruturas tridimensionais em (d), onde também se apresenta a tríade catalítica em cor vermelha e os resíduos que interagem via ponte salina nas duas temperaturas, em verde. As estruturas das simulações a 500 K foram extraídas do final das simulações.

Outra característica importante observada nas simulações de desnaturação é que a

maior parte das pontes salinas, especialmente as localizadas nas fitas, é preservada mesmo à temperatura elevada, como pode ser observado nos resíduos coloridos em verde nas estruturas enzimáticas nas Figuras 3.9-11. Em RmLamR, o *cluster* de pontes salinas de seis ligações permanece “amarrando” as fitas e estabilizando a estrutura desta enzima. A persistência das pontes salinas, na maioria dos casos, coincide com as regiões que permanecem por mais tempo estruturadas.

3.6 Conclusões

Neste trabalho, foram estudadas as bases estruturais da termoestabilidade e termofilicidade de laminarinases através de simulações de MD. Foram feitas simulações comparativas de três laminarinases de diferentes níveis de termoestabilidade, sendo o foco principal, a laminarinase de *Rodothermus marinus*, cuja estrutura foi resolvida recentemente por nossos colegas cristalógrafos do IFSC-USP. Esta enzima hipertermoestável e hipertermofílica é formada majoritariamente de folhas- β que compõem um tipo de enovelamento conhecido como *jelly roll*, conservado entre as glicosídeo hidrolases da família 16. Esta enzima contém um alto número de pontes salinas, apenas comparável com outras GHs que são resistentes à condições ambientais extremas.

Neste estudo, verifica-se que a termoestabilidade das estruturas de laminarinases relaciona-se com o número de pontes salinas (quanto mais pontes salinas, maior a estabilidade), mas sua distribuição na estrutura pode, porém, desempenhar um papel ainda mais importante. Ocorre estabilização térmica adicional com a formação de *clusters* de PS e quando estas interações não permeiam o centro hidrofóbico destas proteínas, como nas termoestáveis aNLam e principalmente RmLamR. As pontes salinas que atravessam o centro hidrofóbico de uma estrutura consistem, na verdade, em um

fator de desestabilização, por facilitar a entrada de moléculas de água na região, como no caso da PcLam. Além disso, verificou-se que a mobilidade dos resíduos no sítio ativo não depende da temperatura de uma forma trivial. Há um compromisso da tríade estrutura-função-dinâmica, que torna a enzima mais eficiente em determinada temperatura, isto é, em sua temperatura ótima de atividade.

Simulações de desnaturação, em que o processo foi acelerado por uma temperatura bastante alta, complementaram o estudo sobre a termoestabilidade destas enzimas. A velocidade e a forma com que as enzimas foram desestruturadas ao longo das simulações refletiram os níveis de termoestabilidade. A enzima hipertermofílica RmLamR perde os contatos nativos lentamente e gradualmente na temperatura de simulação. Por outro lado, PcLam, a enzima menos estável termicamente, desestruturase com o rápido ganho de mobilidade dos *loops* e hélices nas regiões periféricas, tornando a cavidade catalítica inacessível logo nos primeiros nanossegundos de simulação.

Deste modo, embora alguns aspectos estruturais relacionados à termoestabilidade e termofilicidade pudessem ser identificados logo na estrutura tridimensional estática, as simulações de dinâmica molecular revelaram-se uma ferramenta valiosa na avaliação de como estes aspectos variam e o que causam na estrutura quando se alteram as condições térmicas.

Grande parte deste trabalho, referente às simulações comparativas a 25 e 90°C, já foi publicada, com o título “Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations”, na revista *The Journal of Physical Chemistry B* (Bleicher & Prates *et al.*, 2011). A porção restante do trabalho também deve ser submetida para publicação em breve.

Capítulo 4

Endoglucanase 3 de *Trichoderma harzianum*:

Organização estrutural e mecanismos de reconhecimento do substrato por endoglucanases que não possuem o módulo de ligação à celulose

4.1 Introdução

Celulases do fungo *Trichoderma reesei* estão entre as enzimas mais estudadas para a sacarificação da celulose. O sistema celulolítico de *T. reesei* expressa uma variedade de celobiohidrolases e endoglucanases que são, com poucas exceções, compostas de dois domínios, um grande centro catalítico (CCD, *catalytic core domain*) conectado a um pequeno módulo de ligação à celulose (CBM, *cellulose binding module*) através de uma longa cadeia peptídica rica em glicosilações.

Propõe-se que a função do CBM seja reconhecer e ligar-se à superfície celulósica, sendo o processo catalítico precedido pelo desprendimento de uma única cadeia de polissacarídeo da fibrila com subsequente inserção da cadeia no sítio ativo da enzima em seu CCD, através de um mecanismo ainda muito pouco compreendido. Sugere-se que a interação do CBM com as fibrilas levem à movimentação das celulases sobre a superfície celulósica (Liu *et al.*, 2009; Stahlberg *et al.*, 1991; Stahlberg *et al.*, 1996; Teeri, 1997). Ainda, propõe-se que o CBM cumpra o importante papel de aumentar a afinidade da enzima pela celulose e romper seus domínios cristalinos. A retirada do domínio CBM em enzimas pode impossibilitar o processo enzimático (Arantes, 2010). Por isso, então, a importância de compreender a função das celulases que não contêm o

domínio CBM e ainda assim são capazes de catalisar a hidrólise celulósica.

As endoglucanases 3 (também conhecidas como Cel12A (Henrissat, 1998)) de *T. reesei* e *T. harzianum* são exemplos de enzimas que carecem do domínio CBM. Sabe-se que estas enzimas são particularmente ativas em regiões amorfas da celulase e na hidrólise de oligossacarídeos solúveis (Karlsson *et al.*, 2002), mas pouco se sabe, a nível molecular, sobre como as enzimas Cel12A são capazes de reconhecer, ligar-se e hidrolisar as cadeias celulósicas na ausência de um CBM.

Neste trabalho, investigaram-se as interações enzima-substrato, visando elucidar como ocorre o reconhecimento e ligação do substrato na ausência do CBM em endoglucanase 3 de *Trichoderma harzianum*, cuja estrutura foi recentemente resolvida. Determinaram-se as principais interações desta enzima com oligossacarídeos (celotetraose e celopentaose) e foram feitas comparações estruturais desta enzima com o CBM de endoglucanase C de *Cellulomonas fimi*. Foram identificados aminoácidos na estrutura de Cel12A que têm papel fundamental na ligação com o substrato, fornecendo, então, bases estruturais para o fato de Cel12A dispensar um CBM. As informações obtidas poderão ser usadas para otimizar a eficiência enzimática através de técnicas de engenharia de proteínas.

4.2 Descrição Estrutural

A estrutura tridimensional da endoglucanase 3 de *Trichoderma harzianum* (ThEG3, PDB id: 4H7M) foi determinada por cristalografia de raios X pelo grupo de cristalografia do IFSC-USP, com resolução de 2,07 Å (PDB id: 4H7M). A estrutura, apresentada na Figura 4.1, compõe-se de duas camadas de folhas- β antiparalelas, em que as partes convexas e côncavas são formadas por seis (A1-A6) e nove (B1-B9) fitas, respectivamente, conforme mostrado nas Figuras 4.1B e 4.1C. As fitas- β são conectadas por vários *loops* e três α -hélices (H1, H2 e H3; Figura 4.1A, 4.1C). A parte côncava constitui a fenda catalítica, a qual liga-se às cadeias de celulose durante a

hidrólise. A Figura 4.1D mostra resíduos catalíticos Asp100, Glu117, e Glu201, rigorosamente conservados na família 12 de glicosídeo hidrolases (Sandgren *et al.*, 2001; Divne *et al.*, 1994).

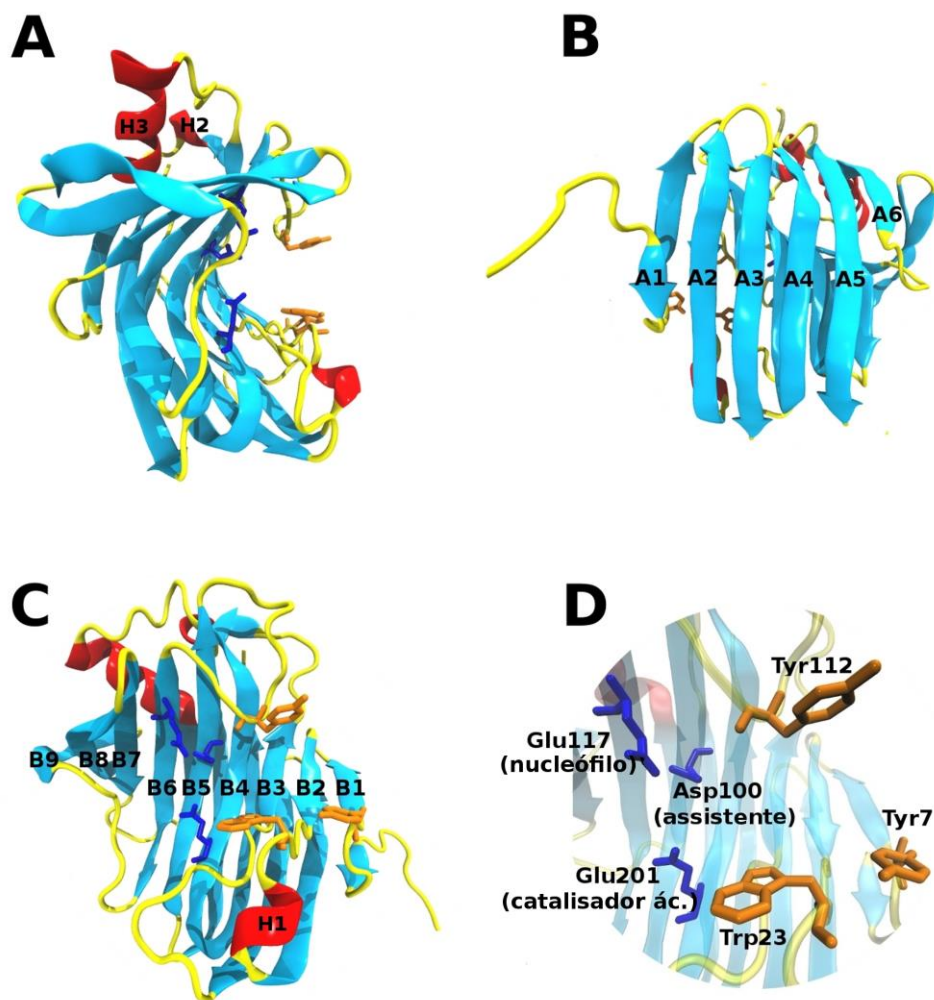


Figura 4.1. (A) A estrutura de β -jelly roll da endoglucanase 3 de *Trichoderma harzianum*. (B, C) O mesmo que A, após uma rotação de 90° ao redor do eixo y. (D) Imagem aproximada mostrando a tríade catalítica em azul e os três resíduos aromáticos que estão localizados na entrada da fenda de ligação ao substrato, em laranja.

A distância de 5,6 Å entre Glu117 e Glu201 é típica de pares nucleófilo/catalisador ácido envolvidos no mecanismo de hidrólise que leva ao mecanismo de retenção de configuração anomérica do produto (Davies & Henrissat,

1995). A Figura 4.1D também mostra os três resíduos aromáticos localizados nos *loops* que formam “polegar” e “dedos” da fenda catalítica da enzima (Törrönen *et al.*, 1994). As simulações de MD deste trabalho sugerem que estes resíduos devam exercer um papel fundamental na eficiência da hidrólise catalítica.

4.3 Metodologia

4.3.1 Construção dos modelos de complexo ThEG3-substrato

Simulações do complexo ThEG3-oligossacarídeo são centrais neste trabalho. Como ainda não há a estrutura cristalográfica de ThEG3 ligada ao substrato, diferentes oligossacarídeos foram modelados no canal de ligação com base na superposição de ThEG3 a estruturas sobreponíveis disponíveis de complexos proteína-substrato.

Inicialmente, então, foram construídos três diferentes sistemas compostos de pequenos oligossacarídeos ligados ao canal catalítico de ThEG3. Após o alinhamento de ThEG3 usando a ferramenta Multiseq do VMD (Humphrey *et al.*, 1996; Roberts *et al.*, 2006), as coordenadas dos ligantes celotetraose e da celopentaose foram copiadas das estruturas cristalográficas de Cel12A de *Thermotoga maritima* (TmEG3) (PDB id: 3AMM, Cheng *et al.*, 2011) e de endoglucanase C de *Cellulomonas fimi* (PDB id:1GU3, CfCBM) (Boraston *et al.*, 2002), respectivamente, de forma a posicionar estes ligantes no interior da fenda catalítica de ThEG3. Além destes dois modelos, um terceiro foi criado pelo prolongamento da celotetraose ligada à ThEG3 em uma unidade glicosídica, formando o ligante que passa a ser chamado celopentaose*. As construções de ThEG3-substrato contendo celotetraose, celopentaose, e celopentaose* serão, daqui para frente, denominadas ThEG3-tt, ThEG3-pt e ThEG3-pt*, respectivamente.

A fim de se avaliar a confiabilidade dos modelos de complexos propostos, foi utilizado um método alternativo para gerar a configuração inicial do substrato na

enzima. Foi, então, feito o *docking* molecular de celopentaose em ThEG3, utilizando o programa Autodock Vina (Trott and Olson, 2010). A escolha do aplicativo empregado como ferramenta para o *docking* vem do seu amplo uso para combinação proteína-ligante e dos bons resultados apresentados, além de ser um dos poucos programas de *docking* completamente gratuitos disponíveis para o uso acadêmico (Chang *et al.*, 2010; Magalhães *et al.*, 2007). Além disso, o programa foi testado através do *docking* da celotetraose na estrutura TmEG3 cristalográfica. O procedimento resultou em uma excelente sobreposição com o complexo cristalográfico, verificável visualmente e no valor de RMSD em relação à estrutura cristalográfica, de 2.2 Å.

A interface gráfica Autodocktools (Morris *et al.*, 2009) foi usada para selecionar as porções rígidas e flexíveis das moléculas e para gerar os arquivos PDBQT requeridos de proteína e ligante. Foram feitas rodadas usando duas conformações diferentes da enzima: I) a estrutura cristalográfica, e II) a enzima no momento de menor energia de interação com celopentaose*, selecionada das simulações de MD. Para fazer o *docking* da celopentaose na fenda de ThEG3, todas as ligações no ligante foram definidas como flexíveis, exceto aquelas nos anéis de glicose. A partir dos cálculos, empregando parâmetro de exaustividade igual à 20, três estruturas foram escolhidas a partir da análise visual da probabilidade de uma reação de hidrólise ser efetiva e afinidade de ligação correspondente.

Tabela 4.1. Modelos de ThEG3 complexada à celopentaose, obtidos através do método de *docking*.

	Afinidade / kcal . mol ⁻¹	RMSD / Å
Autodock-1	-8.1	2.0
Autodock-2	-8.7	5.6
Autodock-3	-8.2	3.5

Os modelos com menor energia de ligação são resultantes da situação II, o que é explicado pelo fato de que a enzima deve passar por sucessivas modificações

conformacionais de modo a acomodar da melhor forma o substrato para que a reação ocorra. Na Tabela 4.1, os valores de energia de afinidade das estruturas escolhidas estão presentes junto com o desvio quadrático médio em relação à estrutura original (a conformação de ThEG3-pt* de menor energia nas simulações de MD).

4.3.2 Construção dos sistemas e simulações

Foram conduzidas simulações de *apo*-ThEG3 e *apo*-TrEG3 (PDB id: 1H8V; Sandgren *et al.*, 2001), bem como da ThEG3 ligada aos diferentes oligossacarídeos. Foram também rodadas simulações auxiliares dos complexos de ThEG3 com a configuração inicial do ligante determinada pela aplicação de um método alternativo de *docking* e, finalmente, foram conduzidas simulações controle para as estruturas originárias dos três modelos de alinhamento, o complexo cristalográfico CfCBM-celopentaose e TmEG3-celotetraose. Os procedimentos empregados para preparar os sistemas para as simulações são descritos abaixo.

Na extremidade N-terminal da estrutura cristalográfica de ThEG3, há um piroglutamato (PCA), originado da ciclização e condensação de um resíduo de glutamina. Esta reação frequentemente ocorre em enzimas extracelulares de fungos e, atribui-se a ela, o papel de aumentar a resistência da enzima a proteases (Sandgren *et al.*, 2001). Além disso, TrEG3 encontra-se glicosilada na posição Asn164, com uma *N*-acetil-*D*-glucosamina (NAG) ligada covalentemente à cadeia lateral da Asn164, interagindo com o resíduo Asn91 de outra molécula na unidade cristalográfica assimétrica. Dado que não há evidências acerca da dimerização de TrEG3 (Sandgren *et al.*, 2001) e que TrEG3 exibe atividade catalítica mesmo na ausência de PCA (Xiao *et al.*, 2001), o resíduo de ácido glutâmico na posição N-terminal foi reconstruído e a NAG, removida. Foram removidos também os seis primeiros resíduos (N-terminal) da estrutura ThEG3 (EAEAEF), os quais foram artificialmente adicionados à sequência como *linker* entre a enzima recombinante e a cauda de histidina durante a purificação da

proteína (Liberato *et al.*, 2012).

Um dos objetivos deste trabalho era compreender as interações Cel12A-substrato que possibilitam o reconhecimento da celulose na ausência de um domínio CBM. A estratégia aplicada foi comparar as simulações de MD de ThEG3 e de CfCBM complexados à celopentaose, de forma a verificar se há resíduos que assumam funções similares em ambas as estruturas. Para cumprir este propósito, foram rodadas as simulações de CfCBM-celopentaose. Como um típico CBM da família 4, CfCBM liga-se preferencialmente a oligossacarídeos e celulose amorfa (<http://www.cazy.org/CBM4.html>, último acesso em abril de 2013) (Boraston *et al.*, 2004), similarmente à TrEG3, que interage fracamente com o substrato cristalino (Karlsson *et al.*, 2002). Os sete primeiros resíduos do N-terminal de CfCBM estão ausentes na estrutura cristalográfica, presumivelmente, devido à alta mobilidade que apresentam. Para modelar estes resíduos, foram usadas as coordenadas dos átomos da cadeia principal dos primeiros cinco resíduos de uma estrutura razoavelmente sobreponível (CBM de laminarinase de *Thermotoga maritima*, PDB id:1GUI). Os outros dois resíduos faltantes foram modelados através do programa Molden (Schaffenaar & Noordik, 2000).

As estruturas foram centralizadas em uma caixa cúbica de cerca de 80 Å de lado e foram hidratadas por 15.000 moléculas de água, usando o Packmol (Martínez & Martínez, 2003; Martínez *et al.*, 2009). Deste modo, em todos os sistemas, a camada de hidratação consistia em, pelo menos, 18 Å de espessura. Em todos os sistemas simulados, foram adicionados a quantidade basal de 50 íons cloreto e 50 íons sódio, com um excesso de um ou de outro de modo a manter o sistema eletricamente neutro a uma concentração salina de aproximadamente 0,16 M.

Tabela 4.2. Resumo dos sistemas simulados. Todas as "caixas de simulação" consistuem-se da enzima ou domínio imerso em água e estão eletricamente neutros pela presença de íons dissolvidos. Todos os sistemas apresentam pH 7,0.

Sistemas	Componentes principais	Condições	Simulações
TrEG3	Cel12A de <i>T. Reesei</i>	298 K, 1 bar	3 simulações x 40 ns
ThEG3	Cel12A de <i>T. harzianum</i>	298 K, 1 bar	3 simulações x 40 ns
ThEG3-tt	Complexo Cel12A de <i>T. harzianum</i> - celotetraose ¹	298 K, 1 bar	3 simulações x 40 ns
ThEG3-pt	Complexo Cel12A de <i>T. harzianum</i> - celopentaose ²	298 K, 1 bar	3 simulações x 40 ns
ThEG3-pt*	Complexo Cel12A de <i>T. harzianum</i> - celopentaose ³	298 K, 1 bar	3 simulações x 40 ns
CfCBM-pt	Complexo CBM de <i>C. Fimi</i> - celopentaose ⁴	298 K, 1 bar	3 simulações x 40 ns
Autodock-(1, 2, 3)	Complexo Cel12A de <i>T. harzianum</i> - celopentaose	298 K, 1 bar	1 simulação x 40 ns (cada modelo)

1. Coordenadas da celotetraose obtidas da estrutura cristalográfica de *Thermotoga maritima*.

2. Coordenadas da celopentaose obtidas da estrutura cristalográfica do CBM de *C. Fimi*.

3. Celopentaose construída a partir do prolongamento da extremidade não-redutora da celotetraose advinda de *T. Maritima*.

4. Celopentaose cristalográfica.

Todos os estados de ionização dos resíduos ionizáveis também foram determinados neste trabalho através do servidor H++ (<http://biophysics.cs.vt.edu/H++>, último acesso em janeiro de 2013; Anandakrishnan & Onufriev, 2012; Myers *et al.*, 2006; Gordon *et al.*, 2005). Com especial atenção aos resíduos da tríade catalítica, considerando a função na reação catalítica e as interações com o substrato,

consideraram-se protonados o catalisador ácido (Glu201 em ThEG3 e Glu200 em TrEG3) e o resíduo auxiliar (Asp100 em ThEG3 e Asp99 em TrEG3), ao passo que o nucleófilo (Glu117 em ThEG3 e Glu116 em TrEG3) foram mantidos na forma carregada. Similarmente à TrEG3 (Sandgren *et al.*, 2001), a estrutura cristalográfica de ThEG3 sugere que há ligações de hidrogênio entre as cadeias laterais de Glu117 e Asp100 e entre Glu201 e seu resíduo vizinho, Glu96. Os estados de protonação escolhidos são consistentes com estas interações e favorecem a reação de hidrólise (Davies & Henrissat, 1995).

Para rodar as simulações, foram adotados os parâmetros de simulação e os protocolos de minimização de energia e equilíbrio descritos no Capítulo 2. A partir dos sistemas pré-equilibrados, foram conduzidas três simulações independentes das estruturas *apo*-ThEG3 e TrEG3 por, pelo menos, 40 ns, totalizando 130 ns de simulação para cada estrutura. Foram rodadas também três simulações independentes de 40 ns para cada complexo proteína-substrato: ThEG3-tt, ThEG3-pt, ThEG3-pt*, TmEG3-tt e CfCBM-pt, totalizando 120 ns de simulação para cada um destes sistemas. Além disso, foram rodadas simulações de 40 ns para cada modelo de *docking* (Autodock-1, Autodock-2 e Autodock-3). Uma breve descrição dos sistemas e das simulações é apresentada na Tabela 4.2. As simulações não são suficientemente longas para amostrar movimentos de grande amplitude das proteínas, mas é possível captar, através delas, as flutuações da estrutura e acessar as interações proteína-substrato adequadamente.

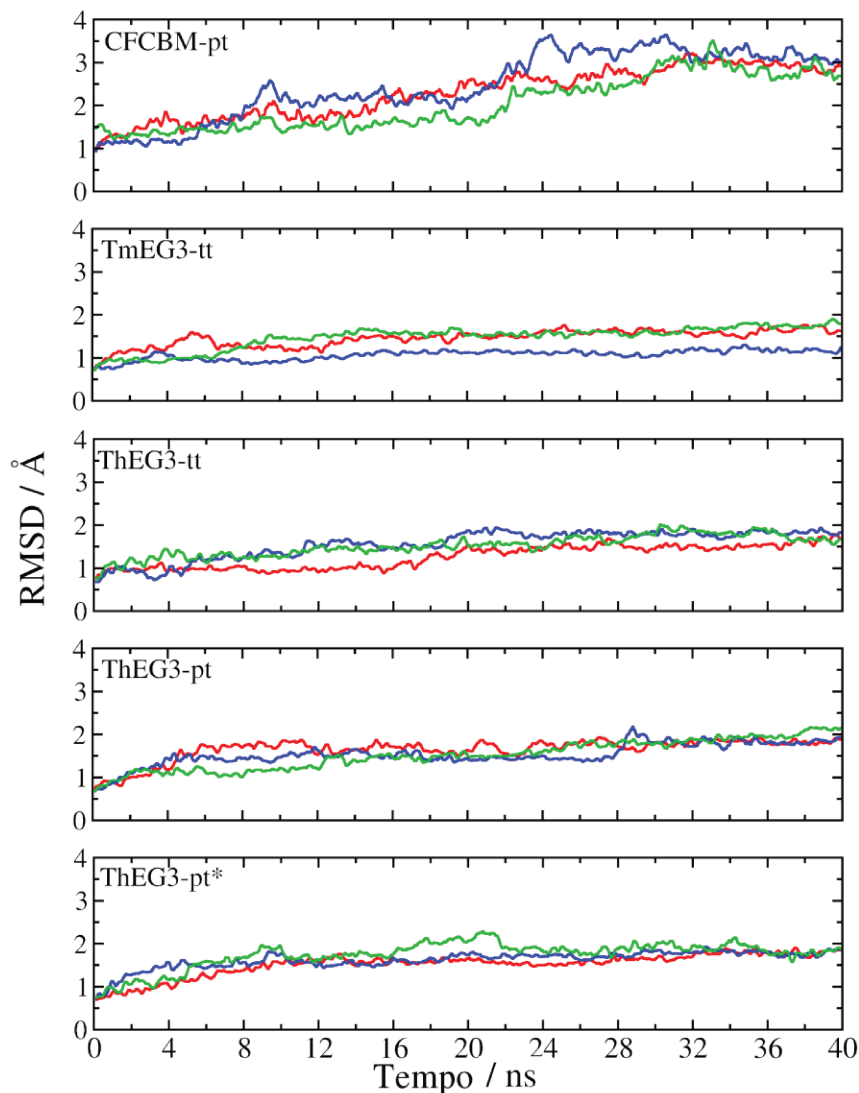


Figura 4.2. Evolução no tempo de simulação dos desvios quadráticos médios dos átomos da cadeia principal dos modelos ThEG3-substrato, além dos complexos cristalográficos CfCBM-pt e TmEG3-tt. As estruturas cristalinas foram usadas como referência. As linhas de diferentes cores correspondem às simulações independentes de cada sistema. Para melhor mostrar a estabilização nas simulações, os oito primeiros resíduos no *loop* N-terminal de ThEG3 não foram considerados, uma vez que esta região apresenta mobilidade muito elevada. Pela mesma razão, os 14 primeiros resíduos de CfCBM não foram considerados na análise. Os valores de RMSD de CfCBM são significativamente maiores do que de outras proteínas, uma vez que este CBM apresenta muitos *loops* longos e móveis, sendo um deles constituído de 11 resíduos. Abreviações: tt-celotetraose, pt-celopentaose, pt*-celopentaose*.

A estabilidade e a relaxação estrutural das enzimas foram monitoradas através da evolução temporal da raiz do desvio quadrático médio (RMSD) dos átomos C_{α} da

proteína ao longo das simulações (Figura 4.2). Verificou-se que, após alguns nanossegundos, o RMSD da cadeia principal tornou-se estável. Para todos os sistemas, os primeiros 6 ns de simulação não foram considerados para o cálculo de propriedades médias. Os valores de RMSD convergem para valores entre 1 e 2 Å. Os valores de RMSD para os modelos criados a partir de *docking* são apresentados posteriormente, na Seção 4.4.2, junto a outras análises.

4.4 Resultados

4.4.1 ThEG3 vs TrEG3: estrutura e dinâmica

Por TrEG3 ser uma enzima mais estudada, com vários trabalhos publicados sobre sua descrição estrutural e atividade enzimática (Sandgren *et al.*, 2001; Karlsson *et al.*, 2002; Macarrón *et al.*, 1993; Sandgren *et al.*, 2005), foi feito um estudo preliminar comparativo entre ThEG3 e TrEG3.

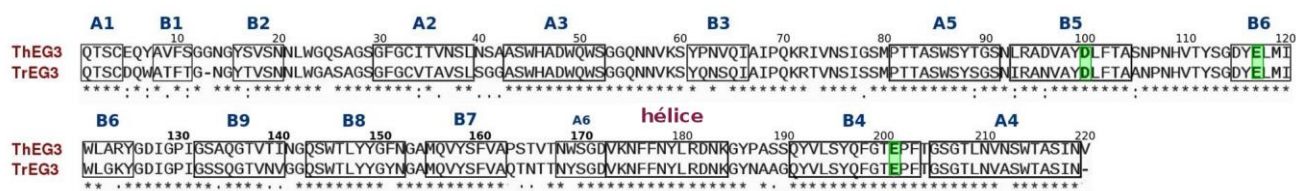


Figura 4.3. Alinhamento das sequências primárias de ThEG3 e TrEG3 com os resíduos da tríade catalítica marcados em verde. Os símbolos abaixo das sequências são adotados pela ferramenta ClustalW (<http://www.clustal.org/clustal2/>) (Larkin *et al.*, 2007), em que o asterisco indica a conservação da identidade do resíduo, o ponto-e-vírgula indica resíduos com propriedades muito similares, e o ponto, resíduos com propriedades menos similares.

A ThEG3 compõe-se de 220 resíduos e apresenta uma alta identidade de sequência primária com TrEG3, com apenas dois resíduos adicionais (Figura 4.3). Estes resíduos são Val220, localizada na porção C-terminal da enzima, e Gly13, que faz do *loop* que conecta as fitas B1 e B2 um pouco mais longo do que o correspondente em

TrEG3. A estrutura TrEG3, descrita em detalhes por Sandgren e colaboradores (Sandgren *et al.*, 2001) é muito parecida com ThEG3 na maioria de seus atributos estruturais. No entanto, cabe apontar algumas poucas diferenças na sequência primária que podem ser relevantes para a sua função enzimática. A Figura 4.4 mostra resíduos expostos ao solvente no canal de ligação ao substrato de ThEG3. Similarmente à TrEG3, pode-se verificar com facilidade, que há uma fileira de resíduos hidrofóbicos em uma das bordas do canal, formada por Tyr7, Trp23, Val58, Phe203 e Ile128. É interessante notar que, na família GH-12, é mais frequente encontrar um resíduo de triptofano do que uma tirosina na posição referente a Tyr7. É o caso da TrEG3, por exemplo. Os efeitos da substituição Tyr/Trp nesta posição ainda não são completamente claros. É provável que Tyr7 apresente interações um pouco mais fracas com o substrato devido à sua menor área para contato hidrofóbico, em relação ao triptofano. Isto é consistente com a cinética de Michaelis-Menten discutida recentemente, em que se relata $K_M \sim 21,4$ g/L para ThEG3 (Generoso *et al.*, 2012), sugerindo que a afinidade de ligação do substrato à ThEG3 é consideravelmente mais baixa, isto é, 14 vezes menor do que em TrEG3 ($K_M \sim 1,5$ g/L) (Nakazawa *et al.*, 2009).

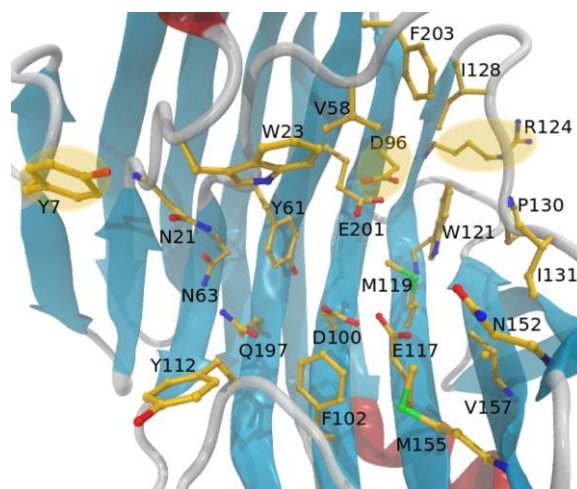


Figura 4.4. Fenda de ligação ao substrato da enzima ThEG3. Alguns resíduos importantes foram explicitados. Os resíduos em destaque por uma sombra amarela diferem na homóloga TrEG3. Os resíduos Tyr7, Asp96, Arg124 em ThEG3 são substituídos, respectivamente, por Trp7, Asn 95, e Lys123 em TrEG3.

A superfície mais funda da fenda, a “palma”, expõe ao solvente, aminoácidos polares como Asn21, Asn63, e Gln197. Próximo ao catalisador ácido Glu201, ThEG3 apresenta um outro resíduo ácido, Asp96, enquanto TrEG3 tem, em seu lugar, Asn95. Outra diferença entre as homólogas é o resíduo Arg124 em ThEG3, ocupando o lugar de Lys123, em TrEG3. De acordo com as simulações de MD, esta substituição leva a importantes diferenças na flexibilidade das duas enzimas nas vizinhanças da fita B9, conforme descrito abaixo.

Foram analisados os perfis de flexibilidade de ThEG3 e TrEG3 nas simulações de MD. As flutuações quadráticas médias (RMSF) das posições dos átomos C_α em relação às estruturas médias, as quais foram obtidas da divisão das três simulações independentes de cada sistema em blocos de 300 ps, são mostradas na Figura 4.5A. O perfil de mobilidade calculado apresenta boa concordância com os valores experimentais, obtidos através da conversão do fator de temperatura (fator-B) cristalográfico dos C_α em valor de RMSF (Phillips, 1990; Kuzmanic & Zagrovic, 2010), de acordo com:

$$RMSF_i^2 = \frac{3B_i}{8\pi^2}.$$

Cabe dizer que as poucas diferenças entre o fator-B e o RMSF não são inesperadas, uma vez que os dados foram obtidos em diferentes condições. Simulações de dinâmica molecular foram conduzidas em condições ambiente, próximas da condição de funcionamento enzimático ótimo, enquanto que a estrutura cristalográfica foi obtida em temperatura muito mais baixa e no estado cristalino.

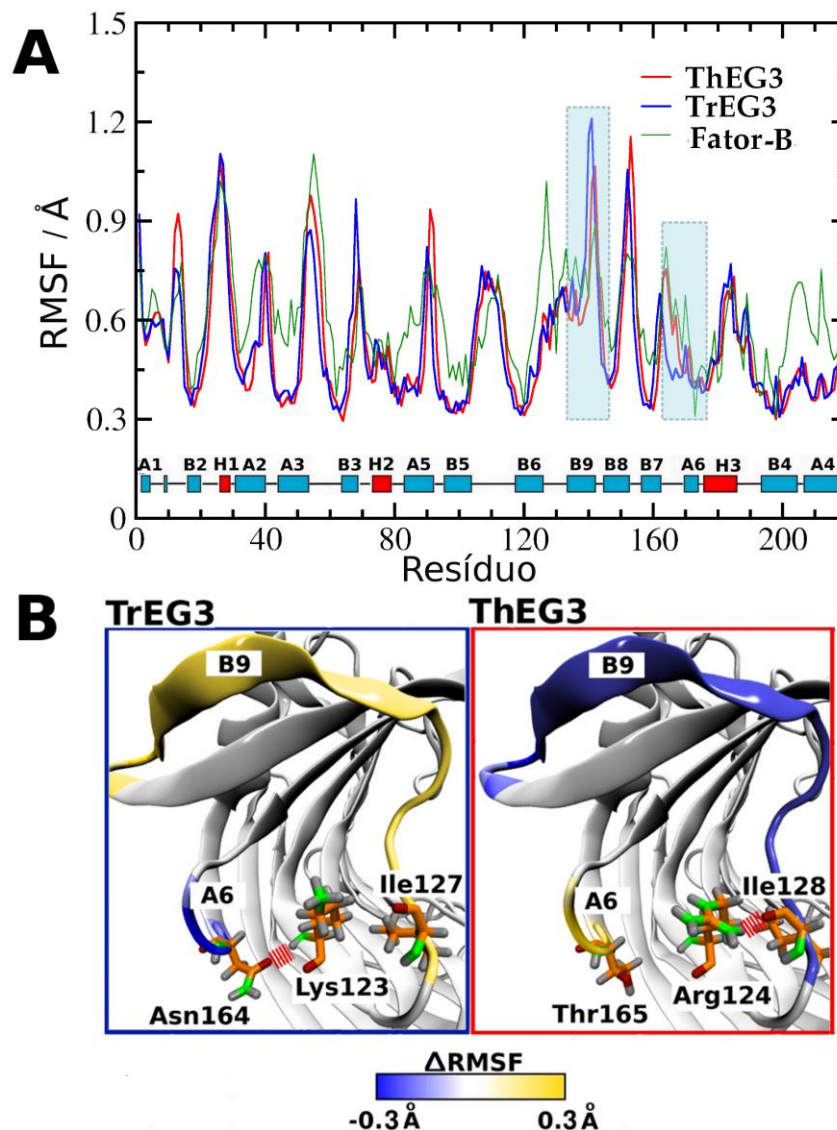


Figura 4.5. Perfil de mobilidade ao longo da sequência primária de ThEG3 e TrEG3, expresso através das flutuações quadráticas médias dos resíduos em relação às suas posições médias calculadas a cada bloco de 300 ps das simulações. Para comparação, os valores de RMSF dos resíduos de ThEG3 calculados através dos fatores-B cristalográficos também são mostrados (linha verde). (B) Regiões que apresentam mobilidade relativa maior (em amarelo) e menor (em azul) em ThEG3 e TrEG3 (tomando ThEG3 como referência para TrEG3 e vice-versa). Também mostram-se os resíduos envolvidos nas interações responsáveis pelas diferenças de mobilidade entre as duas homólogas.

A grande semelhança estrutural entre as duas enzimas, ThEG3 e TrEG3, reflete-se na dinâmica também muito similar entre elas. Poucas regiões apresentam diferenças significativas em mobilidade, mais notavelmente, os *loops* entre fitas- β . Por exemplo,

os resíduos na fita B9 de ThEG3 exibem menor mobilidade enquanto que resíduos no *loop* que conecta as fitas B7 e A6 apresentam maior mobilidade em relação à TrEG3. Estes efeitos locais são atribuídos à poucas diferenças na sequência primária das duas enzimas, a qual altera as interações entre resíduos específicos, conforme mostrado na Figura 4.5B. Em TrEG3, o resíduo Lys123 interage com Asn164 através de ligações de hidrogênio, restringindo a mobilidade do *loop* entre B7 e A6. Em ThEG3, Arg124 e Thr165 ocupam estas posições e interagem fracamente entre si. Em contrapartida, o resíduo Arg124 interage com Ile128, levando à estabilização da fita B9. Esta região é a extremidade redutora do sítio ativo, região em que o produto deve ligar-se, contribuindo para o mecanismo bem conhecido de inibição enzimática promovida pelo produto de reação. Sendo assim, supõe-se que esta diferença bastante localizada possa influenciar diretamente na inibição enzimática.

4.4.2 Ligação a oligossacarídeos

Estruturas cristalográficas disponíveis de enzimas da família GH-12 complexadas a um inibidor sugerem que o polímero deve ligar-se com a extremidade não-redutora voltada para a fita B1 (Sandgren *et al.*, 2012). A estrutura cristalográfica de Cel12A de *Thermotoga maritima* (Cheng *et al.*, 2011), recentemente disponibilizada, encontra-se complexada a uma molécula de celotetraose ligada a seu sítio ativo de forma que suporta esta hipótese. Os quatro resíduos de β -glicose nesta estrutura ocupam os subsítios -2, -1, +1, e +2 da fenda catalítica (Davies *et al.*, 1997).

Neste trabalho, foram construídos modelos de ThEG3 ligados à celotetraose e celopentaose* (ThEG3-tt e ThEG3-pt*), em que os substratos foram posicionados ao longo dos subsítios -2 a +2, como mostrado na Figura 4.6. Esta posição possibilita o contato favorável dos carboidratos com a tríade catalítica, o que poderia resultar na hidrólise que produziria duas moléculas de celobiose ou uma celobiose e uma celotriose como produtos de reação, o que seria concordante com a observação experimental de

que celobiose, não glicose, é o produto principal da hidrólise de TrEG3 (Karlsson *et al.*, 2002).

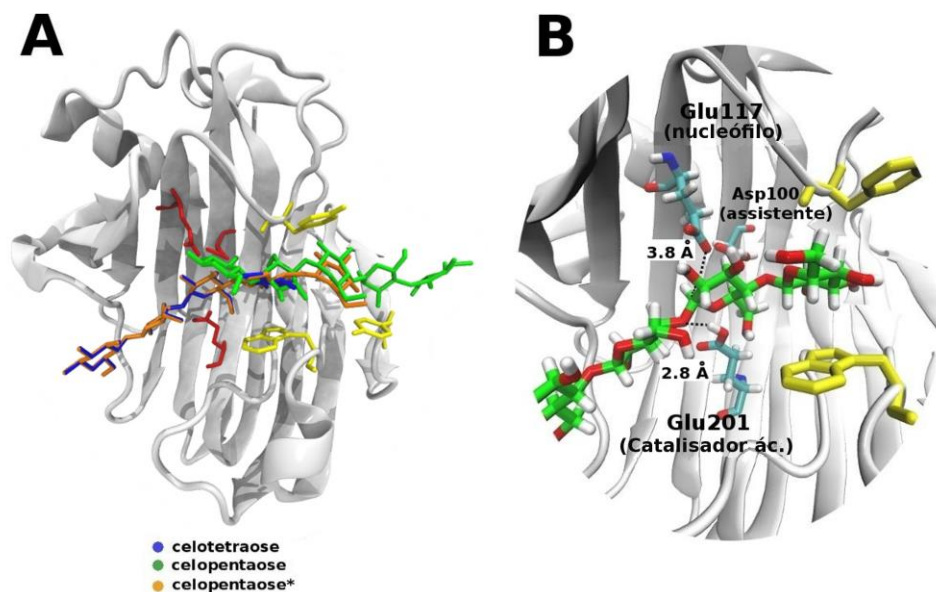


Figura 4.6. (A) Posições iniciais dos substratos em ThEG3 e (B) as interações dos resíduos catalíticos com a celotetraose.

Outro modelo gerado neste trabalho para o complexo ThEG3-celopentaose foi obtido a partir da estrutura cristalográfica do complexo CfCBM-celopentaose. Este modelo favorece o contato entre o carboidrato e os três resíduos aromáticos na entrada do canal de ligação ao substrato, em detrimento de um melhor contato com os resíduos catalíticos. Os resíduos de β -glicose neste modelo ocupam apenas os subsítios negativos -1 a -3. Neste caso, a distância entre os oxigênios do grupo carboxílico do nucleófilo (Glu117) e o carbono anomérico na extremidade redutora do oligossacarídeo é bastante grande (7,6 Å).

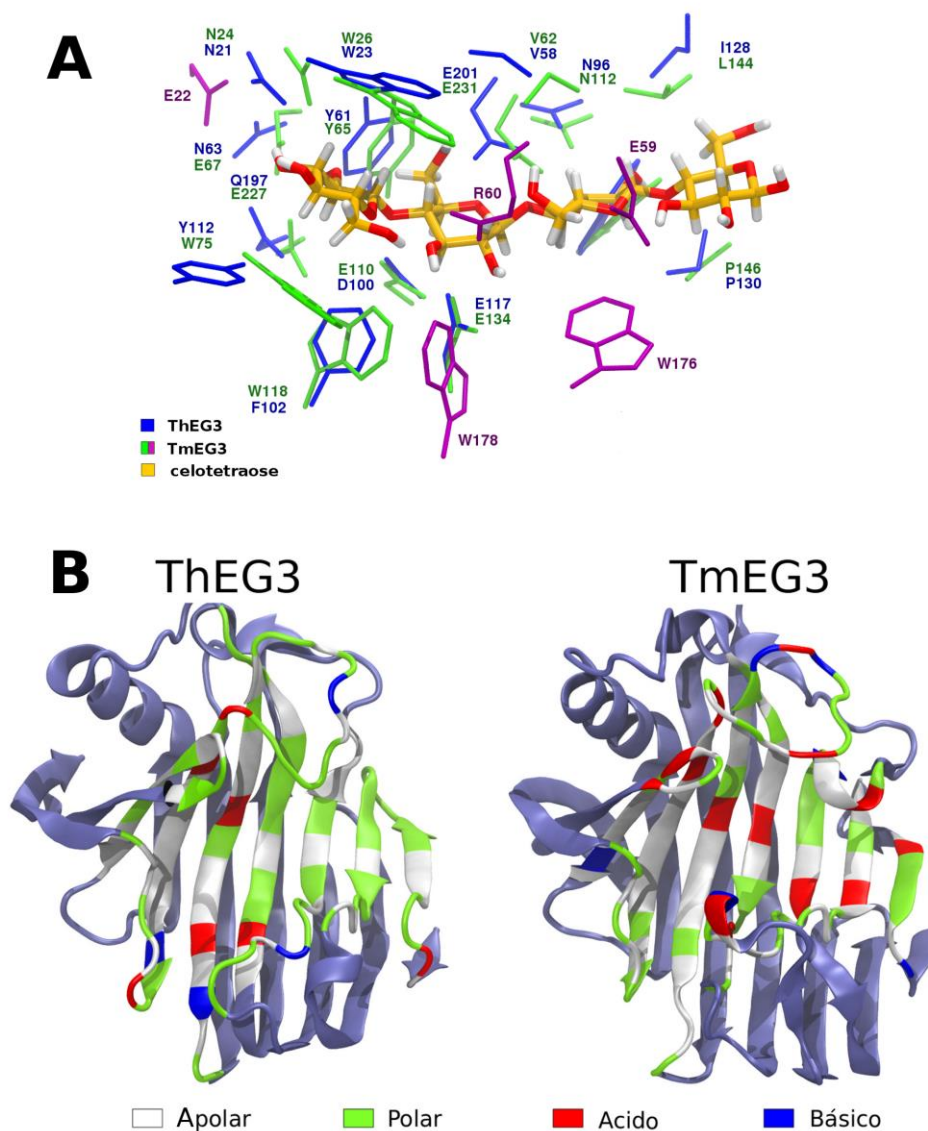


Figura 4.7. (A) Detalhes das fendas de ligação ao substrato de ThEG3 e TmEG3 e suas interações com o substrato após o alinhamento estrutural de ThEG3 com o complexo TmEG3-celopentaose. Os resíduos de ThEG3 são apresentados em azul e os de TmEG3, em verde ou violeta. A cor violeta é usada para mostrar os resíduos que não apresentam correspondentes em ThEG3. (B) Distribuição de resíduos de diferentes naturezas na fenda catalítica de ambas as enzimas.

Conforme mencionado em Metodologia (Seção 4.3.1), as coordenadas do ligante necessárias para construir os modelos ThEG3-tt e ThEG3pt* foram geradas através do alinhamento dos átomos da cadeia principal das estruturas cristalográficas ThEG3 e TmEG3-tt. Apesar de notavelmente diferentes (identidade 8%), a comparação restrita às

fendas catalíticas de ThEG3 e TmEG3, apresentada na Figura 4.7, indica que, nestas enzimas, a maioria dos resíduos importantes na ligação ao substrato são equivalentes ou similares em natureza. Isto é, alguns dos resíduos não são idênticos, mas provavelmente compartilham a mesma função na ligação ao substrato. Por exemplo, os resíduos polares Asn63 e Gln197 na fenda de ThEG3 são substituídos pelos resíduos ácidos Glu67 e Glu227 em TmEG3. A substituição dos resíduos polares por resíduos carregados é recorrente em enzimas termoestáveis e hipertermoestáveis, como a TmEG3 (Cheng *et al.*, 2011). A enzima TmEG3 é também muito mais rica em resíduos aromáticos em contato com o substrato. Por exemplo, há cinco triptofanos próximos à celotetraose na TmEG3 (Trp26, Trp75, Trp118, Trp176 e Trp178), enquanto que, em ThEG3, há apenas um (Trp23). Os resíduos Glu22, Glu59, Arg60, Trp176 e Trp178 de TmEG3 não apresentam correspondência em ThEG3 (Figura 4.7A, em violeta). A natureza dos resíduos na fenda catalítica é comparada na Figura 4.7B.

Os complexos modelados foram utilizados buscando compreender o comportamento dos substratos na fenda catalítica da enzima. Durante as simulações, os substratos nas estruturas modeladas ThEG3-tt e ThEG3-pt* tenderam a manter a posição inicial em relação aos resíduos catalíticos, conforme mostra a Figura 4.8. Apenas em uma das simulações de ThEG3-tt, a celotetraose não manteve sua posição inicial e os sítios alvo do substrato distanciaram-se dos resíduos catalíticos Glu201 e Glu117 após 12 ns de simulação. Um evento similar foi observado para a celopentaose em uma das simulações de ThEG3-pt*. No entanto, neste caso, o substrato chegou a reposicionar-se no interior da fenda, mas não de forma tão eficiente que não se deslocasse para fora novamente.

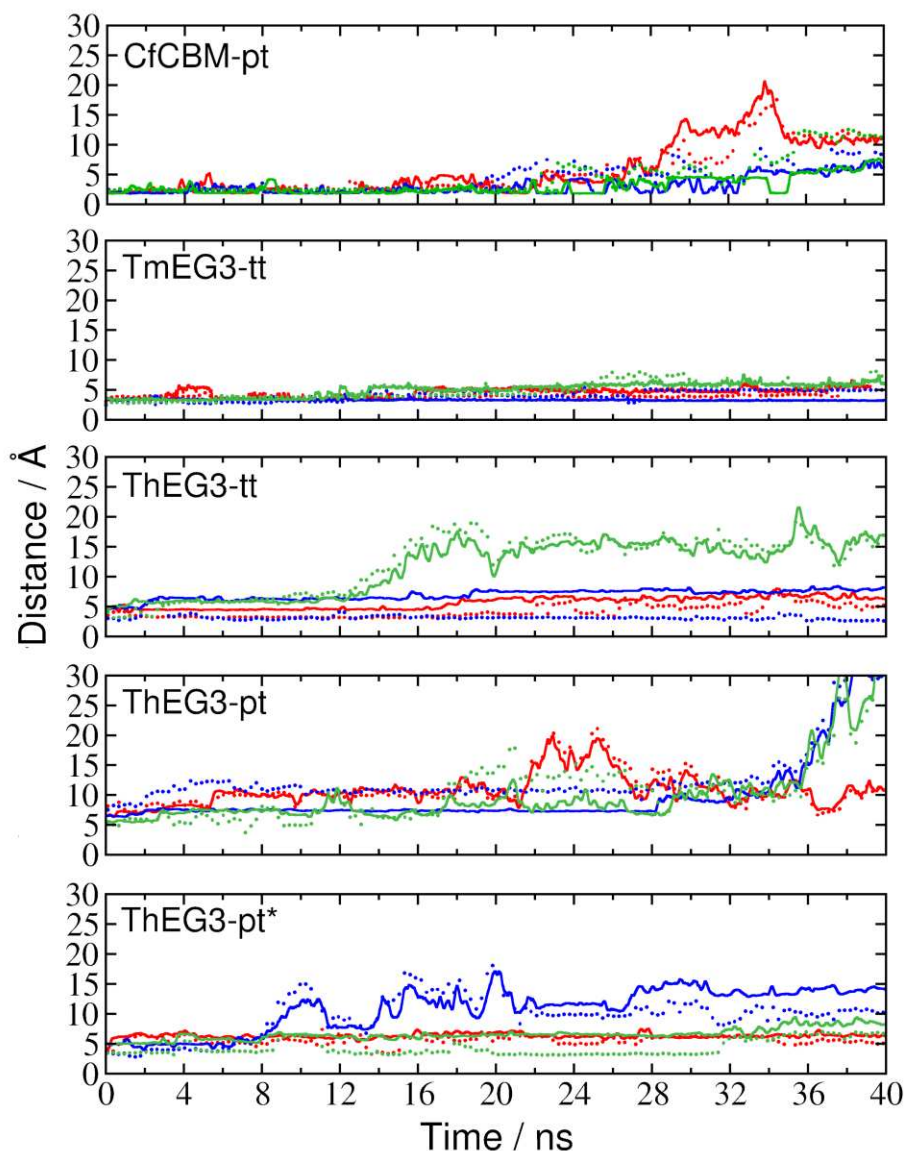


Figura 4.8. Evolução temporal das distâncias entre os resíduos catalíticos e átomos do substrato que são alvos na reação de hidrólise catalítica. Foram computadas as distâncias entre o catalisador ácido (Glu201 em ThEG3 e Glu134 em TmEG3), e um oxigênio glicosídico (linhas pontilhadas), e entre o nucleófilo (Glu117 em ThEG3 e Glu123 em TmEG3), e o átomo C1 na unidade de glicose (linhas cheias). Foram selecionados os átomos O e C1 nas ligações glicosídicas mais próximas (a segunda ligação glicosídica na celotetraose e celopentaose* e a primeira na celopentaose, contando a partir da extremidade redutora na cadeia do sacarídeo). E foram considerados o hidrogênio ácido no catalisador ácido e os átomos de oxigênio carboxílico no nucleófilo. Para o CfCBM, foram computadas as distâncias entre os resíduos Gln124 (oxigênio de amida na cadeia lateral) e um hidrogênio de hidroxila (ligada ao C6 da segunda unidade de glicose), e entre Gln128 (hidrogênio de amida na cadeia lateral) e um oxigênio de hidroxila (ligada ao C3 da primeira unidade de glicose). As curvas de diferentes cores correspondem às três simulações independentes de cada sistema.

As distâncias ligante-proteína também foram monitoradas ao longo das simulações das estruturas cristalográficas de TmEG3-tt e CfCBM-pt. Durante as simulações da estrutura cristalográfica TmEG-tt, o substrato oscila muito pouco além de sua posição inicial na fenda. Isto é consistente com o fato de a fenda catalítica de TmEG3 ser mais rica em resíduos carregados e aromáticos, os quais estabelecem contatos mais persistentes com o substrato. Por outro lado, as simulações do complexo cristalográfico CfCBM-pt e de ThEG3-pt mostram a saída do ligante da fenda, similarmente aos eventos observados em algumas das simulações de ThEG3-tt e ThEG3-pt mencionadas anteriormente.

As simulações de MD também forneceram informações valiosas sobre a interação dos substratos com os três resíduos aromáticos na entrada da fenda catalítica (Tyr7, Trp23 e Tyr112). Resíduos aromáticos, essencialmente triptofano e tirosina, são frequentemente encontrados na fenda catalítica das glicosidases. Tais resíduos podem emparelhar-se às faces dos anéis nas unidades glicosídicas, formando interações carboidrato- π , que exercem um papel central nos mecanismos de reconhecimento proteína-carboidrato e são essenciais em muitos processos biológicos (Boraston *et al.*, 2004; Laughrey *et al.*, 2008; Payne *et al.*, 2011). A importância dos resíduos aromáticos nos domínios de ligação à celulose de celulasas para a seletividade do substrato principal é amplamente reconhecida. Uma característica comum das famílias 4, 6, 9 e 22 de CBMs é a configuração dos aminoácidos aromáticos que podem emparelhar-se simultaneamente, como um “sanduíche”, às duas faces dos anéis de piranoses de oligossacarídeos solúveis ou cadeias polissacarídicas individuais em regiões amorfas da celulose (Boraston *et al.*, 2004). Os CBMs membros das famílias 1, 2, 3, 5 e 10, por sua vez, apresentam os anéis aromáticos em uma configuração planar, o que torna muito eficiente a ligação destes domínios à superfície de celulose cristalina. A Figura 4.9 mostra um CBM de fenda aberta, o CfCBM, que pertence à família 4, e um CBM de arranjo planar, o CBM da família 1, da celobiohidrolase I de *T. Reesei*.

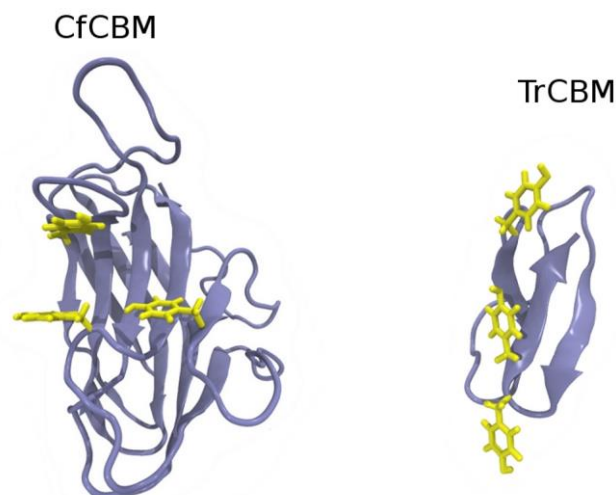


Figura 4.9. Estruturas dos domínios de ligação à celulose (CBMs) de endoglucanase C de *Cellulomonas fimi* (esquerda), um CBM da família 4, e de celobiohidrolase I de *T. reesei* (direita), um CBM da família 1. Os resíduos hidrofóbicos (amarelo) apresentam arranjos diferentes, dependendo do tipo de substrato ao qual o módulo se liga preferencialmente.

As principais posições dos substratos em relação aos anéis aromáticos e aos resíduos catalíticos são apresentadas na Figura 4.10, que mostra a sobreposição de diferentes instantes das trajetórias. Na ausência de um ligante, isto é, na *apo*-ThEG3, os três resíduos aromáticos encontram-se livres para fazerem movimentos de grande amplitude. A exposição frequente da superfície do anel à água pode contribuir favoravelmente para promover as primeiras interações entre a proteína e a cadeia de açúcar que poderiam dirigir o substrato para o interior da fenda catalítica.

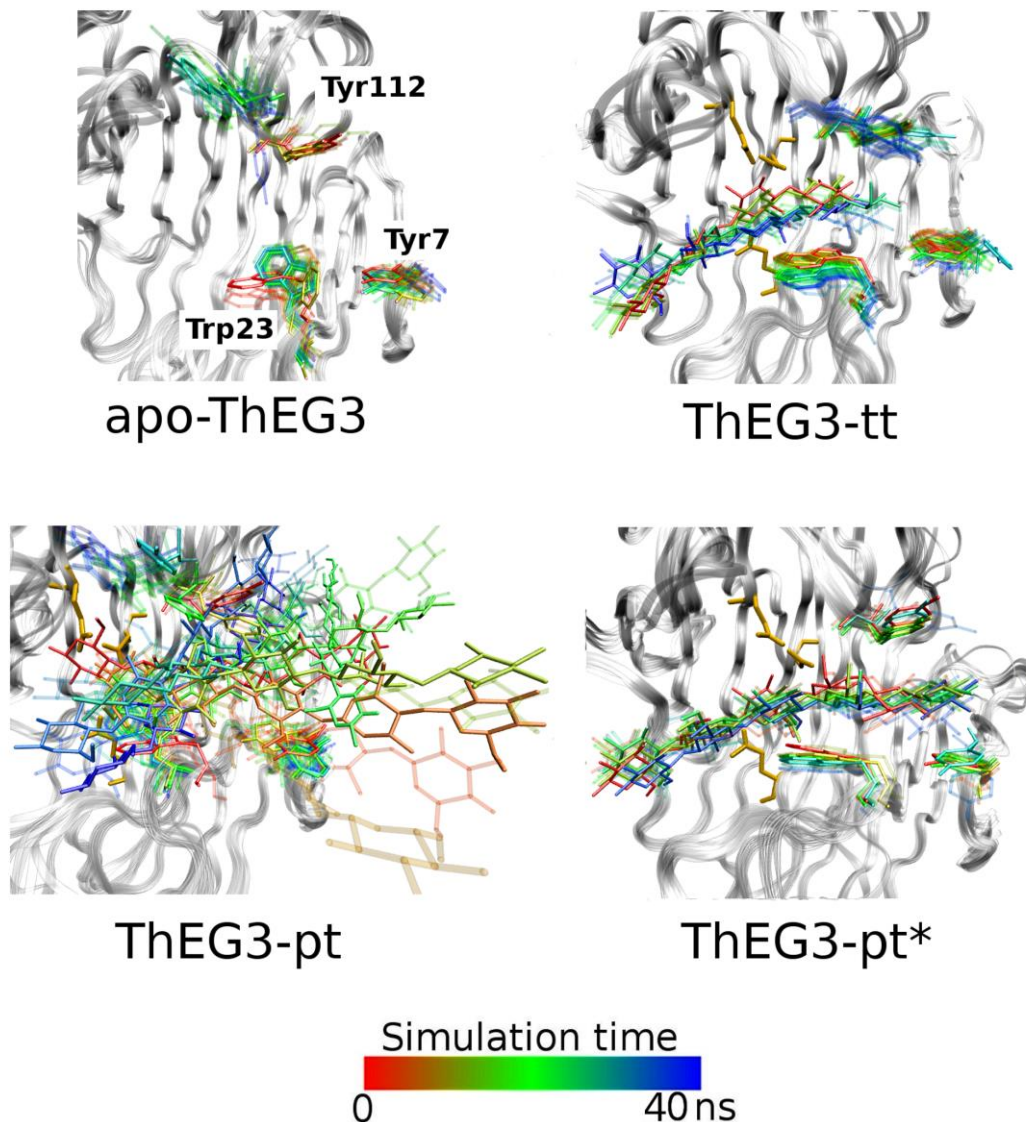


Figura 4.10. Sobreposição de *snapshots* das simulações que mostram como variam as posições dos resíduos aromáticos Tyr7, Trp23 e Tyr112, e dos oligossacarídeos celotetraose, celopentaose e celopentaose*. As cores dos resíduos aromáticos e dos oligossacarídeos variam na escala vermelho-verde-azul, de acordo com o tempo ao longo da trajetória de simulação.

No complexo ThEG3-tt, a molécula de celotetraose posiciona-se adequadamente em relação à tríade catalítica, mas não é longa o suficiente para interagir simultaneamente com os três resíduos aromáticos na entrada da fenda. No modelo de complexo ThEG3-pt, a posição do substrato é bastante flutuante e porções do ligante frequentemente deixam a fenda. Embora as unidades de glicose da cadeia de

celopentaose interajam com os três resíduos aromáticos na entrada da fenda, nenhuma ligação glicosídica fica acessível à tríade catalítica durante as simulações. Nas simulações do complexo ThEG3-pt*, por sua vez, a posição inicial e o comprimento do substrato formam uma combinação favorável que possibilita a interação da celopentaose tanto com os resíduos aromáticos quanto com os resíduos catalíticos. Estas características correlacionam-se bem com a observação experimental de inatividade ou baixa eficiência catalítica de TrEG3 na hidrólise de celotriose e celotetraose, ao contrário de oligossacarídeos mais longos como celopentaose (Karlsson *et al.*, 2002). Deste modo, as simulações sugerem que a eficiência catalítica de ThEG3 depende de um comprimento mínimo do substrato para posicioná-lo de forma produtiva na fenda catalítica. Isto deve também ser aplicável a outras enzimas com o mesmo enovelamento e mecanismo catalítico.

Embora sejam observados eventos de quase total desprendimento do substrato da fenda nos três modelos, ThEG3-tt, ThEG3-pt e ThEG3-pt*, além de CfCBM-pt, a importância dos anéis aromáticos em manter a ligação enzima-substrato é evidenciada pelo fato de que, na grande maioria do tempo, permanecem interações com, pelo menos, um ou dois anéis aromáticos. Como mostrado na Figura 4.11, as interações entre o substrato e os três anéis aromáticos, Tyr7, Trp23 e Ty112 permanecem quase sempre abaixo de -10 kcal/mol durante as simulações. Apenas em simulações de ThEG3-pt é que se observa o desprendimento completo do substrato em relação aos anéis aromáticos. Mas isto só chega a ocorrer nos últimos 4 ns de simulação. Exceto no caso da estrutura TmEG3-tt, em que o ligante permanece ligado à enzima por razões já discutidas, todos os outros sistemas simulados exibem instantes em que o módulo da energia de interação cai para menos que 5 kcal/mol.

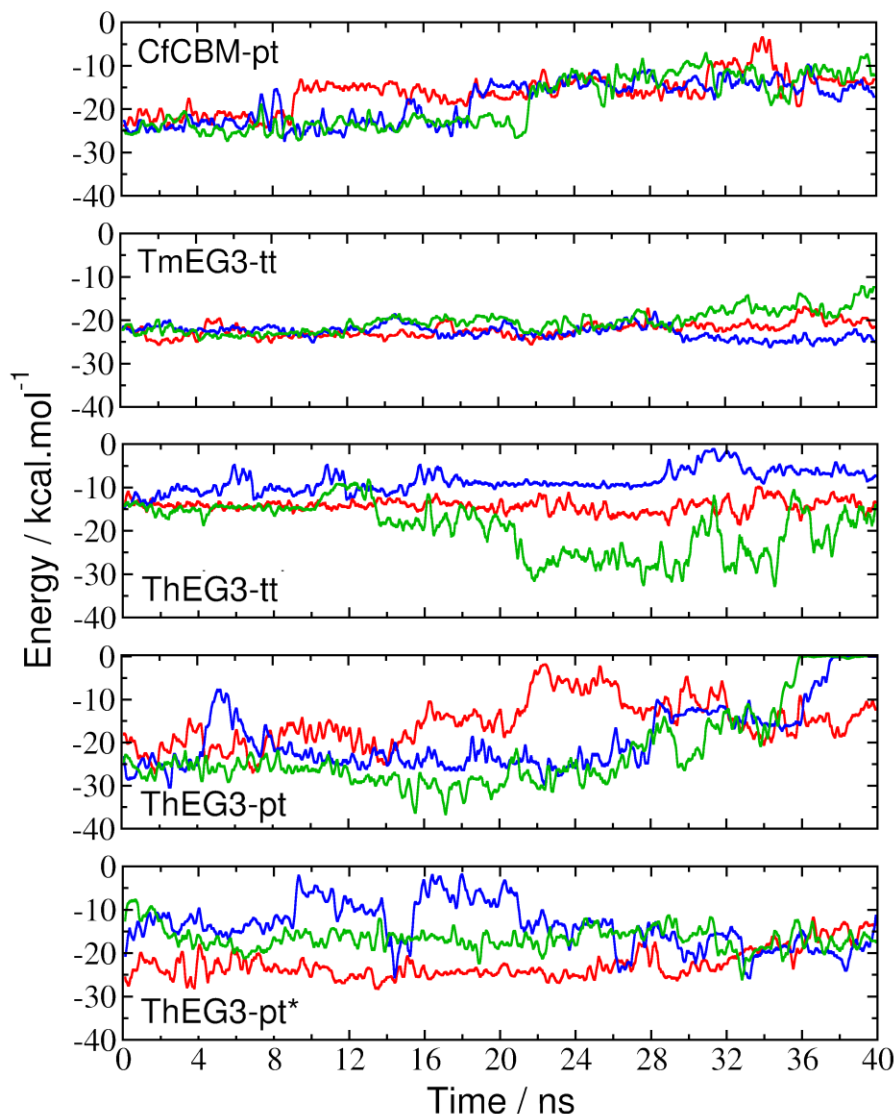


Figura 4.11. Evolução no tempo das energias de interação entre os resíduos aromáticos Tyr7, Trp23 e Tyr112 e os substratos obtidos de uma simulação de modelos de ThEG3-oligossacarídeo. Também são mostrados os gráficos correspondentes às energias de interação entre CfCBM (resíduos Tyr19, Tyr43 e Tyr85) e TmEG3 (resíduos Trp26, Trp75 e Trp176) com seus respectivos ligantes. As curvas de diferentes cores correspondem à três simulações independentes de cada sistema.

Uma análise cuidadosa das simulações revelam eventos interessantes. Em uma das simulações de ThEG3-tt, a energia de interação com os resíduos aromáticos passa a diminuir progressivamente após cerca de 12 ns de simulação (Figura 4.11, curva verde no painel correspondente à ThEG3-tt). A análise visual da trajetória referente a esta curva revela que, a partir deste momento de simulação, o substrato passa a deslizar-se

paralelamente à fenda de ligação na enzima, passando a ocupar os subsítios negativos, -1 e -2, em uma conformação muito semelhante à de ThEG3-pt, como mostrado na Figura 14.12A). Eventos semelhantes são reproduzidos em outros sistemas simulados. Em uma das simulações de ThEG3-pt*, verificou-se que, em um instante próximo a 10 ns, o substrato distancia-se da fenda e dos resíduos Tyr7 e Trp23, mas permanece conectado à enzima por contato hidrofóbico com Tyr112 (Figura 14.12B). Por volta de 15 ns, o carboidrato e os resíduos aromáticos encontram-se novamente interagindo fortemente. No entanto, neste momento, o substrato não está mais orientado ao longo da fenda. Apenas após alguns instantes, após 20-30 ns, é que o substrato volta a encaixar-se na fenda em uma conformação que se assemelha à de ThEG3-pt. Vale comentar que um encaixe similar de substrato é encontrado na estrutura cristalográfica da homóloga CelB2 de *Streptomyces lividans* (PDB id: 2NLR), em que um análogo da celotriose encontra-se ligado à mesma região da enzima, nos sítios -1 e -2 (Sulzenbacher *et al.*, 1999).

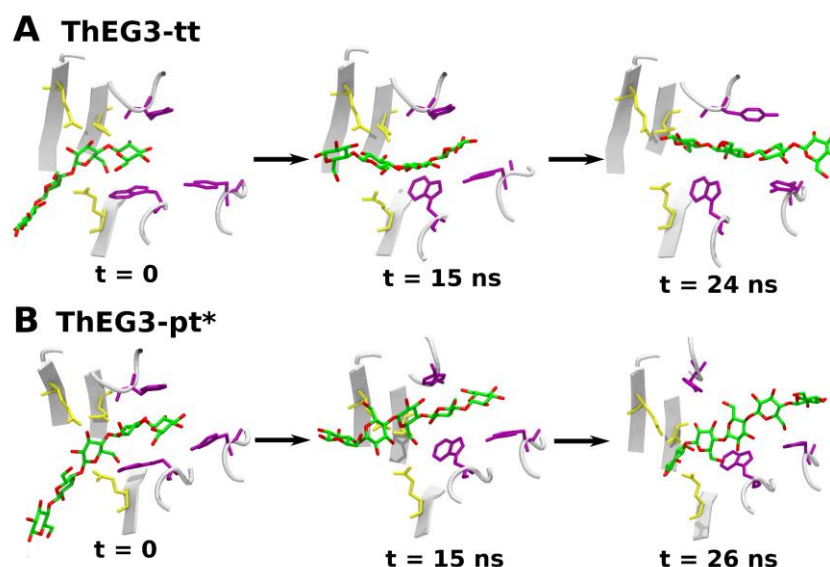


Figura 4.12. *Snapshots* de diferentes estágios ao longo das simulações de ThEG3-tt (A) e ThEG3-pt* (B). Estes eventos são representativos de outros que ocorrem ao longo das simulações dos três sistemas modelados, em que o substrato permanece temporariamente ligado à enzima por apenas um ou dois anéis aromáticos e depois volta a encaixar-se na fenda. Os resíduos catalíticos (Asp100, Glu117 e Glu201) são representados em amarelo e os resíduos aromáticos na entrada da fenda catalítica (Tyr7, Trp23 e Tyr112), em violeta.

Os eventos descritos há pouco sugerem que a cadeia oligossacarídica pode deslizar ao longo da fenda catalítica sem deixar completamente a fenda devido ao arranjo dos resíduos aromáticos na fenda de ThEG3. Isto é, quando o substrato volta a se posicionar paralelamente à fenda nos estágios finais da simulação, ele encontra-se deslocado em dois subsítios de ligação em relação à sua posição inicial. A liberação processiva de celobiose de substratos celulósicos é frequentemente observada na atividade de exoglucanases (Kipper *et al.*, 2005; Wilson, 2012). Este mecanismo é frequentemente relacionado à conformação tubular da fenda catalítica, em que o substrato é progressivamente clivado sem se desprender da fenda. As endoglucanases, por outro lado, exibem fendas catalíticas abertas e clivam as cadeias celulósicas em posições aleatórias (Sandgren *et al.*, 2005). Excepcionalmente, poucas celulases foram identificadas como endoglucanases processivas (Watson *et al.*, 2009; Sakon *et al.*, 1997) e sua processividade mostra-se independente do módulo de ligação à celulose. As simulações deste trabalho podem proporcionar um ponto de partida interessante para explorar as bases moleculares da processividade de endoglucanases, que ainda é pouco estudada. Neste sentido, porém, estas simulações são muito preliminares e investigações muito mais detalhadas seriam necessárias para elucidar tais mecanismos.

Experimentos de *docking* combinados às simulações de MD

Conforme descrito na Metodologia (Seção 4.3.1), também foram gerados modelos de complexo ThEG3-oligossacarídeo através da aplicação de *docking*, como um método alternativo para comparação com os três modelos propostos a partir de alinhamento com complexos cristalográficos.

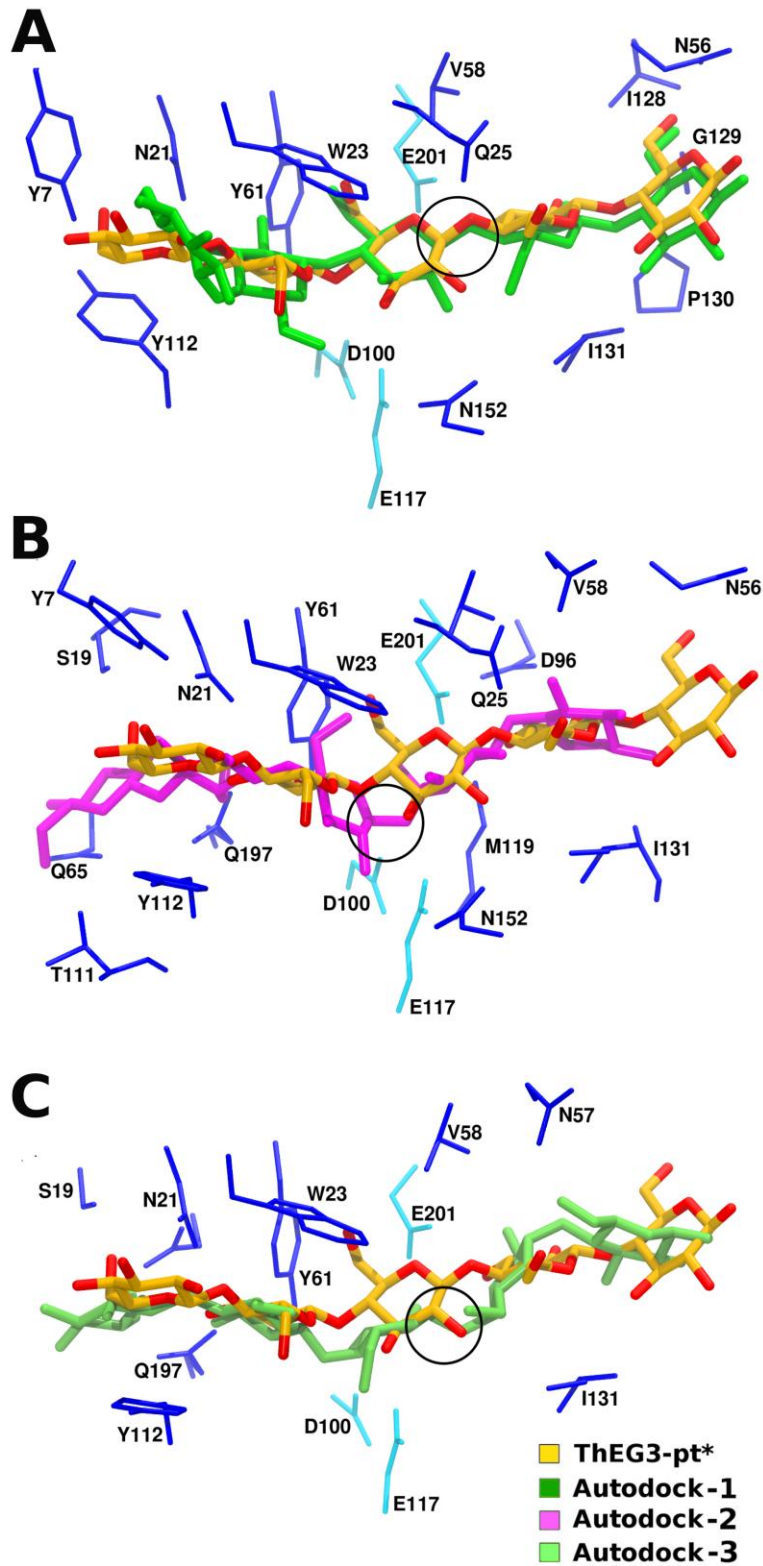


Figura 4.13. Modelos Autodock-1 (A), Autodock-2 (B) e Autodock-3 (C) e resíduos próximos. Os círculos pretos demarcam os supostos sítios de reação nos oligossacarídeos e a conformação desfavorável de Autodock-2 e Autodock-3.

Na Figura 4.13, são apresentados os três modelos selecionados resultantes do cálculo de *docking*. O complexo Autodock-1 apresenta-se como a estrutura mais próxima da original, enquanto que Autodock-2 encontra-se deslocado da estrutura original na direção dos subsítios de aminoácidos aromáticos. Autodock-2 e Autodock-3 apresentam um problema específico, envolvido diretamente na catálise, adotando uma conformação que não possibilita o ataque nucleofílico. Na Figura 4.13B e 4.13C, o círculo preto marca a região de provável reação. Nesta conformação, os elétrons livres do oxigênio glicosídico não estão orientados na direção do catalisador ácido E201. Além disso, este oxigênio glicosídico consiste no grupo de saída do ataque nucleofílico em C1 marcado com o círculo preto. O grupo de saída deve ser facilmente alinhado com os elétrons de ataque no nucleófilo (E117). Isto requereria uma custosa rotação de cerca de 180° ao redor da ligação marcada. Mesmo assim, estes sistemas também foram simulados para comparação com os outros modelos gerados.

Assim como os modelos anteriores, as simulações dos modelos de *docking* tornaram-se estáveis rapidamente, em poucos nanossegundos, como se verifica na Figura 4.14A. Destas simulações, algumas das análises feitas para os modelos de alinhamento foram repetidas aqui. É interessante notar que, na simulação do modelo mais próximo de ThEG3-pt*, Autodock-1, o substrato permanece estável no interior da fenda de ligação, mantendo a proximidade com os resíduos catalíticos. Similarmente, o substrato no modelo Autodock-2 manteve sua posição pouco alterada durante a simulação. Não houve, entretanto, a modificação na conformação do oxigênio glicosídico que favoreceria a reação de hidrólise, segundo descrito. Tampouco na simulação de Autodock-3.

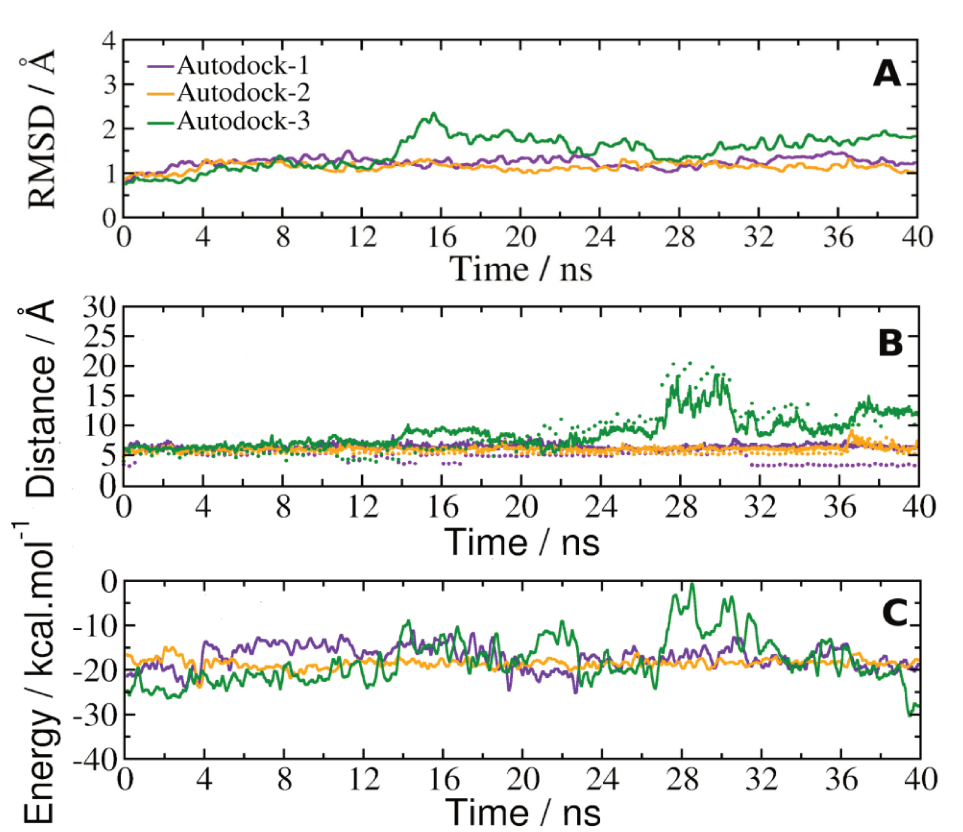


Figura 4.14. (A) Gráficos de desvio quadrático médio ao longo das simulações dos átomos da cadeia principal de ThEG3 nos modelos de *docking*, tomando a estrutura cristalográfica como referência. (B) Evolução no tempo das distâncias entre os resíduos catalíticos e os prováveis alvos de reação de hidrólise no substrato. Foram computadas as distâncias entre o catalisador ácido (Glu201 em ThEG3 e Glu134 em TmEG3), e um oxigênio glicosídico (linhas pontilhadas), e entre o nucleófilo (Glu117 em ThEG3 e Glu123 em TmEG3), e o átomo C1 na unidade de glicose (linhas cheias). (C) Evolução temporal nas energias de interação entre os resíduos aromáticos Tyr7, Trp23 e Tyr112 e os substratos obtidos de uma simulação de modelos de ThEG3-oligossacarídeo. As curvas de diferentes cores correspondem à três simulações independentes de cada modelo de *docking*.

Na simulação do complexo Autodock-3, por sua vez, o substrato distancia-se de sua configuração inicial ao longo da trajetória. É interessante que, neste caso, ocorre novamente o evento de deslocamento lateral do substrato descrito para as simulações de ThEG3-tt e ThEG3-pt* (Figura 4.15).

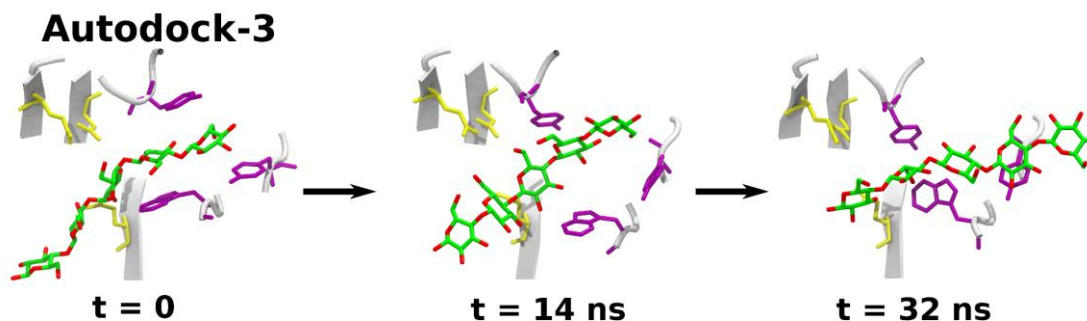


Figura 4.15. Evento de deslocamento lateral observado na simulação do modelo Autodock-3. Os resíduos catalíticos (Asp100, Glu117 e Glu201) são representados em amarelo e os resíduos aromáticos na entrada da fenda catalítica (Tyr7, Trp23 e Tyr112), em violeta.

Estes resultados podem não estar completamente desvinculados dos resultados de simulação para os modelos de alinhamento (ThEG3-tt, ThEG3-pt e ThEG3-pt*), uma vez que os modelos de *docking* foram obtidos com a enzima em uma configuração retirada de uma das simulações dos modelos de alinhamento. Vale comentar que a melhor pose da celopentaose obtida a partir da estrutura cristalográfica de ThEG3, que é completamente desvinculada dos experimentos de MD, apresentou valor de energia de afinidade -8.2 kcal/mol, similar a dos outros modelos apresentados. Este modelo, assim como Autodock-2 e Autodock-3, também não apresentou conformação e distância favoráveis das ligações glicosídicas para a hidrólise. Durante os últimos nanossegundos de uma simulação de 40 ns, este modelo se despreendeu completamente da fenda catalítica.

4.4.3 Motivo de CBM no centro catalítico

Finalmente, análises estruturais e simulações comparativas de MD de ThEG3 e CfCBM complexadas com celopentaose revelam que vários resíduos desempenham papéis similares na interação com o substrato, o que sugere fortemente que a ThEG3 incorpora a função CBM em sua estrutura. A Figura 4.16 mostra os resíduos que se mantêm em persistente contato hidrofóbico com o substrato, além de resíduos que formam ligações de hidrogênio com a celopentaose por, pelo menos, 10% do tempo de simulação no complexo ThEG3-pt* e CfCBM.

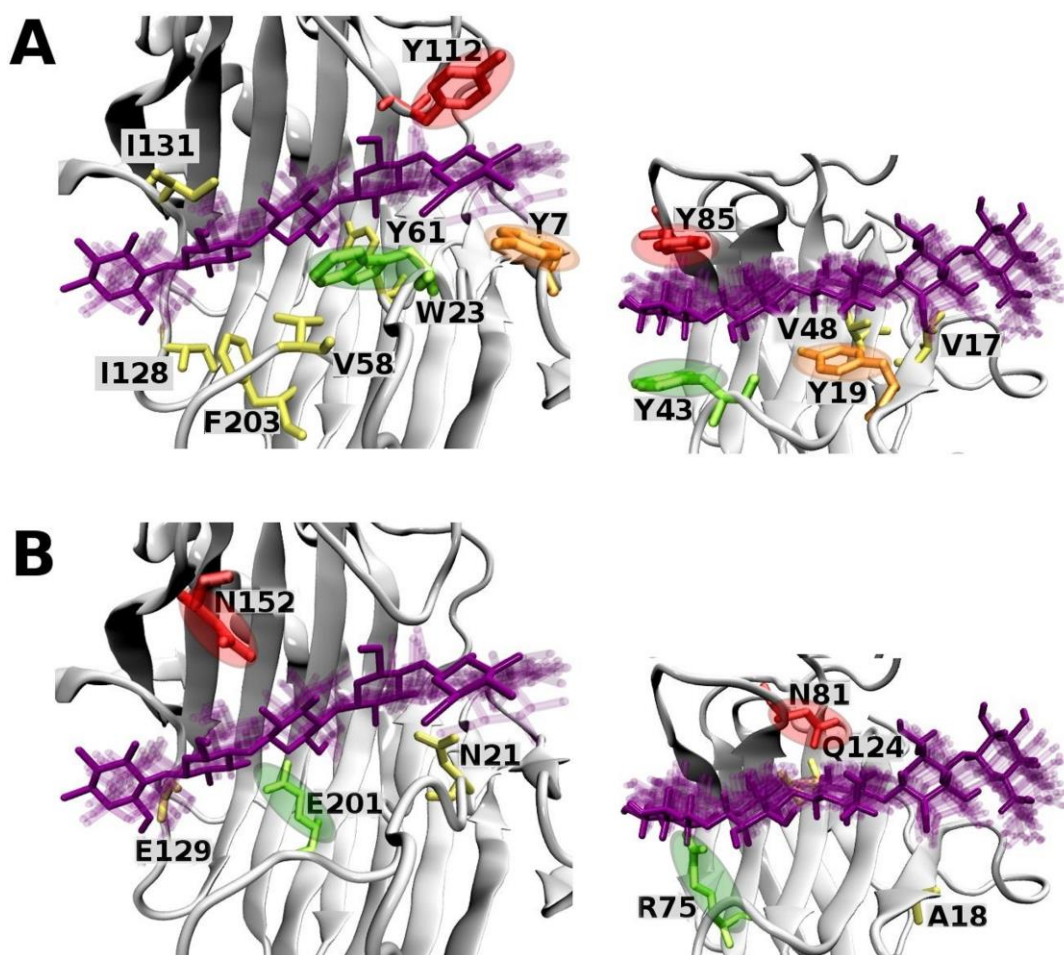


Figura 4.16. (A) Os resíduos de ThEG3 (esquerda) e CfCBM (direita) em contato hidrofóbico com a celopentaose e (B) resíduos que interagem via ligações de hidrogênio por pelo menos 10% do tempo de simulação. Círculos de cores iguais apontam os resíduos que são correspondentes em posição e função nas interações proteína-substrato nas estruturas da enzima (esquerda) e do CBM (direita).

Tabela 4.3. Frequências de contatos hidrofóbicos e ligações de hidrogênio nas simulações de ThEG3-pt* e CfCBM-pt. São mostrados apenas os resíduos que apresentam valores superiores a 10%.

ThEG3-pt*		Contatos hidrofóbicos (%)				
	Trp23	Ile131	Tyr112	Phe203	Val58	
	100	100	99	99	94	
	Tyr61	Ile128	Tyr7	Phe102	Trp121	
	94	72	67	42	19	
		Ligações de hidrogênio (%)				
	Gly129	Glu201	Asn21	Asn152		
	41	36	26	17		
CfCBM-pt		Contatos hidrofóbicos (%)				
	Tyr19	Val17	Val48	Tyr43	Tyr85	
	100	100	98	100	98	
		Ligações de hidrogênio (%)				
	Gln124	Asn81	Ala18	Arg75		
	37	33	34	29		

Estes resultados sugerem que a ausência de um CBM em Cel12A é parcialmente compensada pela presença de um grupo de resíduos similares a resíduos que são peça-chave na interação com o substrato em alguns tipos de CBM. Uma breve análise na estrutura do CBM de outra enzima, a laminarinase de *Thermotoga maritima*, mostrou, através do alinhamento às estruturas de ThEG3 e CfCBM, a presença de um conjunto de resíduos semelhantes (Boraston *et al.*, 2002). Nesta estrutura, há três resíduos de triptofano (Trp27, Trp61 e Trp102) em posições muito semelhantes aos resíduos aromáticos de CfCBM apresentados na Figura 4.15. Os resíduos Arg75 e Asn81 de CfCBM também têm correspondentes no CBM desta laminarinase, a saber, Glu138 e Asn98.

4.5 Conclusões

A partir da estrutura da endoglucanase 3 de *Trichoderma harzianum*, resolvida recentemente por cristalografia de raios-X, foram conduzidas simulações de dinâmica molecular buscando maior clareza sobre a dinâmica de enzimas Cel12A e seu mecanismo de interação com o substrato. Estruturas modelo desta enzima ligadas a celotetraose e celopentaose foram geradas baseando-se em estruturas cristalográficas de proteínas de enovelamento similar ligadas a oligossacarídeos. As simulações sugerem que há um arranjo de resíduos típico de CBM4 (CBM da família 4) presente na estrutura da Cel12A de *T. harzianum* e *T. reesei*, o qual é responsável pelo reconhecimento e ligação do substrato polimérico. Parte deste *cluster* de resíduos é formada por resíduos aromáticos que se emparelham com os anéis das unidades de glicose. Tais resíduos encontram-se espacialmente distantes dos resíduos catalíticos no sítio ativo. O sucesso da ligação produtiva do substrato e eficiência catalítica, portanto, requer cadeias de oligossacarídeo de um comprimento mínimo, de forma que os resíduos na tríade catalítica e os resíduos aromáticos típicos de CBM4 sejam simultaneamente alcançados. Estes resultados fornecem bases moleculares para a observação experimental de que Cel12A não hidrolisa eficientemente oligossacarídeos curtos como celotriose.

Em conjunto com os resultados de cristalografia de raios-X sobre a estrutura de Cel12A *T. harzianum*, o trabalho de simulações computacionais resultou na publicação do artigo: “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”, na revista PlosOne (Prates *et al.*, 2013).

Capítulo 5

β -Glicosidase de *Aspergillus niger*:

A função do domínio fibronectina do tipo 3

5.1 Introdução

No sistema celulolítico, as β -glicosidases são elementos cruciais na sacarificação completa do material celulósico por serem responsáveis pela hidrólise de pequenos oligossacarídeos e da celobiose em unidades de glicose. Também é destacada a importância destas enzimas na mistura reacional por contribuírem com a otimização da conversão enzimática da biomassa, através da diminuição da concentração de celobiose, um forte inibidor do complexo celulolítico, especialmente de celobiohidrolases (Korotkova *et al.*, 2009).

As β -glicosidases são isoladas de diversos fungos, sendo o fungo filamentosso *Aspergillus niger* o mais utilizado para a produção comercial. Estas enzimas têm sido muito estudadas no âmbito de aplicações tecnológicas e otimização da produção. Todavia, os estudos acerca da atividade e da relação estrutura-função ainda são escassos. Avançar na compreensão de seus mecanismos de ação é fundamental para a otimização de seu potencial como biocatalisador industrial, em processos como a produção de bioetanol celulósico.

Na literatura, é amplamente propagado que as β -glicosidases difundem livremente na solução e que se associam ao seu alvo reacional através de encontros ocasionais com moléculas de celobiose. Visando aumentar a produtividade (velocidade) enzimática nesta etapa final de sacarificação, algumas estratégias de ancoramento de β -glicosidases

ao substrato celulósico têm sido apresentadas. O Projeto Temático da FAPESP 08/55914-9, sob coordenação do Prof. Sandro R. Marana, IQ/USP, por exemplo, propõe a construção de enzimas quiméricas pela fusão de β -glicosidases a domínios de ligação a celulose (CBM) para este fim.

Bastante recentemente, o grupo de cristalografia do IFSC-USP aplicou a técnica de espalhamento de raios-X (SAXS) aliada a modelagem por homologia e *ab initio* para caracterizar a forma e enovelamento da β -glicosidase de *Aspergillus niger* (AnBgl1), enzima pertencente à GH-3 (Dan *et al.*, 2000), cuja estrutura tridimensional não havia sido estudada até então. Inesperadamente, identificou-se que AnBgl1 apresenta uma estrutura estendida, lembrando um girino (comumente, refere-se a este tipo de estrutura como *tadpole-like structure*), com o domínio catalítico preso por um longo *linker* a um domínio menor na extremidade C-terminal. Este tipo de estrutura é comum a enzimas que atuam em carboidratos insolúveis como a celulose, a saber, exoglucanases e endoglucanases. Conforme discutido em outros momentos nesta tese, estas enzimas frequentemente apresentam um grande domínio catalítico (CCD) ligado a um pequeno módulo de ligação à celulose (CBM) por um *linker* peptídico flexível e estendido. No entanto, é incomum encontrar este padrão entre enzimas que atuam em pequenos substratos solúveis, como as β -glicosidases. Outro fato inesperado é que o domínio menor não foi reconhecido como membro de uma das famílias de CBMs, mas sim como um domínio cujo enovelamento é de fibronectina do tipo 3 (FnIII).

As enzimas de maior homologia e identidade encontradas, β -glicosidase de *Thermotoga neapolitana* (TnBgl3B, PDB id 2X40) (Pozzo *et al.*, 2010) e β -glicosidase I de *Kluyveromyces marxianus* NBRC1777 (KmBgl1, PDB id 2ABZ) (Yoshida *et al.*, 2010), são também modulares e apresentam o domínio FnIII na ponta C-terminal, porém apresentam-se como estruturas compactas.

Ensaio biofísico de adsorção em diferentes substratos poliméricos feitos pelo

mesmo grupo de cristalógrafos (IFSC-USP) mostram que o complexo β -glicosidase – FnIII tem pouca ou nenhuma afinidade pelas superfícies de celulose, xilano ou amido (abaixo de 20% de associação), mas associa-se à lignina muito eficientemente. Sugere-se, assim, que as β -glicosidases possam, na verdade, estar ancoradas ao substrato ligno-celulósico ao invés de se difundirem livremente pelo meio reacional e indaga-se qual a contribuição do domínio FnIII na associação de β -glicosidase à lignina.

O domínio fibronectina de tipo III é um dos tipos de enovelamento mais presentes em proteínas modulares. Registram-se centenas de estruturas similares no *Protein Data Bank*, sendo a grande maioria componente de proteínas animais (Alahuhta *et al.*, 2010). Em animais, domínios fibronectina, de um modo geral, são conhecidos pela capacidade de ligação a membranas celulares, bem como a diversos compostos macromoleculares como colágenos, fibrinas e DNA. Em bactérias e fungos, este domínio é encontrado essencialmente em enzimas celulolíticas (Jee *et al.*, 2002). Pela multiplicidade de proteínas e espécies em que é encontrado, há muitos trabalhos que buscam desvendar os mecanismos genéticos que levaram a esta propagação não usual, bem como o seu funcionamento específico em cada sistema.

Hipóteses diversificadas acerca de funções destes domínios foram formuladas. Para cada tipo de organismo, este domínio pode cumprir um papel diferente. Mas deve haver uma vantagem natural para a especial abundância de FnIII em tantas glicosídeo hidrolases, especialmente extracelulares. Há trabalhos que sugerem que o domínio atue rompendo ligações entre fibras de celulose, facilitando o acesso das enzimas hidrolíticas (Kataeva *et al.*, 2002). Outra hipótese é a de que eles sejam como uma forma compactada do *linker* peptídico que conecta outros módulos da enzima, podendo ser estendida quando necessário, como um elástico. Ou, ainda, que eles funcionem, simplesmente, como espaçadores entre outros domínios (Jee *et al.*, 2002).

Neste trabalho, visando lançar luz acerca da razão funcional da arquitetura molecular de AnBg11, foca-se no domínio FnIII, empregando a técnica de dinâmica

molecular na análise das principais interações entre o domínio e pequenos fragmentos de lignina de parede celular de plantas. Novamente, as simulações partiram da estrutura fornecida pelo grupo de cristalografia do IFSC-USP, cujo enovelamento foi descrito a partir de experimentos de SAXS.

5.2 Construção dos sistemas e simulações

Uma vez que a lignina é um polímero notavelmente complexo, composto de arranjos variáveis de unidades de monolignóis reticulados (Vanholme *et al.*, 2010), seria impraticável obter amostragem satisfatória de sua variação conformacional e interações com o domínio FnIII nas simulações de dinâmica molecular. Deste modo, ao invés de um grande polímero, optou-se por simular pequenos fragmentos, o que favoreceria amostrar a interação com diferentes partes da molécula FnIII. Foram, então, construídos pequenos fragmentos de lignina que constituem homodímeros de guaiacil (LGG) e siringil (LSS), as duas unidades mais abundantes encontradas nos polímeros de lignina de plantas em geral (Vanholme *et al.*, 2010). O tipo de ligação mais recorrente entre as unidades, β -O-4', foi adotada para ambos os homodímeros (Sangha *et al.*, 2012). A Figura 5.1 ilustra a complexidade de um polímero de lignina encontrado em choupo e mostra os compostos desenhados para as simulações.

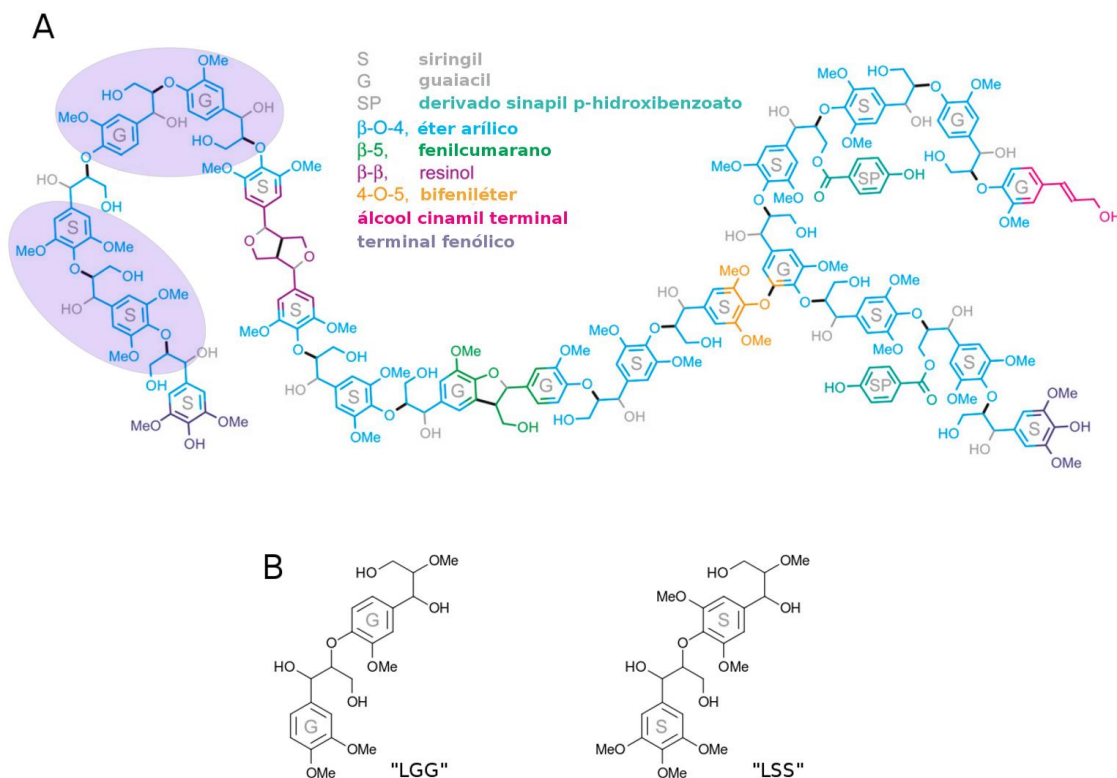


Figura 5.1. Representação de lignina de choupo (adaptada de Stewart *et al.*, 2009). Com base em um pequeno segmento deste polímero (regiões hachuradas), os dímeros de guaiacil e siringil foram modelados para as simulações de MD.

Simulações independentes de dinâmica molecular foram conduzidas a partir de diferentes configurações iniciais dos sistemas, geradas segundo quatro abordagens distintas, as quais são descritas na Seção seguinte.

5.2.1 Construção dos sistemas

- **Sistema 1:** Cinco fragmentos LGG foram posicionados ao redor da superfície do domínio FnIII conforme os resultados da aplicação do método de *docking* com o programa Autodock Vina (Trott and Olson, 2010);
- **Sistema 2:** Dez fragmentos de LGG foram aleatoriamente distribuídos ao redor do domínio FnIII;
- **Sistema 3:** Dez fragmentos de LGG foram posicionados em sítios específicos de

FnIII, escolhidos “intuitivamente” com base na possibilidade de formação de ligações de hidrogênio, empilhamento com anéis aromáticos, contatos hidrofóbicos e complementariedade de forma;

- **Sistema 4:** Dois fragmentos de lignina foram aleatoriamente posicionados na solução, distantes entre si e do domínio FnIII, de forma que as interações lignina-lignina fossem insignificantes. Neste caso, foram comparados os comportamentos dos fragmentos LGG (**sistema 4a**) e LSS (**sistema 4b**) com relação às interações com o domínio.

A Tabela 5.1 resume estas quatro abordagens.

Tabela 5.1. Breve descrição dos cinco sistemas gerados para o estudo de dinâmica molecular das interações de lignina com o domínio FnIII. Apresenta-se, também, na terceira coluna, o número de simulações e o tamanho de cada simulação.

Sistema	Composição	Abordagem usada para posicionar os dímeros	Simulações / ns
1	FnIII, 5LGG	Autodock Vina	60
2	FnIII, 10LGG	Aleatório	2 x 60
3	FnIII, 10LGG	Inspeção visual	3 x 60
4a	FnIII, 2LGG	Aleatório	3 x 60
4b	FnIII, 2LSS	Aleatório	3 x 60

5.2.2 Docking

O programa AutoDock Vina foi utilizado para propor os sítios mais favoráveis para o encaixe do fragmentos de lignina no domínio FnIII, no **sistema 1**. Nos arquivos de entrada, gerados com o programa AutoDockTools (Morris *et al.*, 2009), todos os ângulos diedrais de LGG, exceto os de ligações nos anéis aromáticos, foram definidos como flexíveis. Aplicou-se parâmetro de exaustividade igual à 20.

Tabela 5.2. Energias de afinidade dos nove modos de ligação gerados pela aplicação do método de *docking* e seus respectivos sítios ocupados. As poses escolhidas estão destacadas em vermelho.

Modo de ligação	Afinidade / kcal . mol ⁻¹	Sítio de ligação
M_a	-5,0	S2
M_b	-5,0	S5
M _c	-5,0	S5
M_d	-4,9	S3
M_e	-4,9	S1
M _f	-4,9	S1
M_g	-4,8	S4
M _h	-4,8	S4
M_i	-4,8	S1

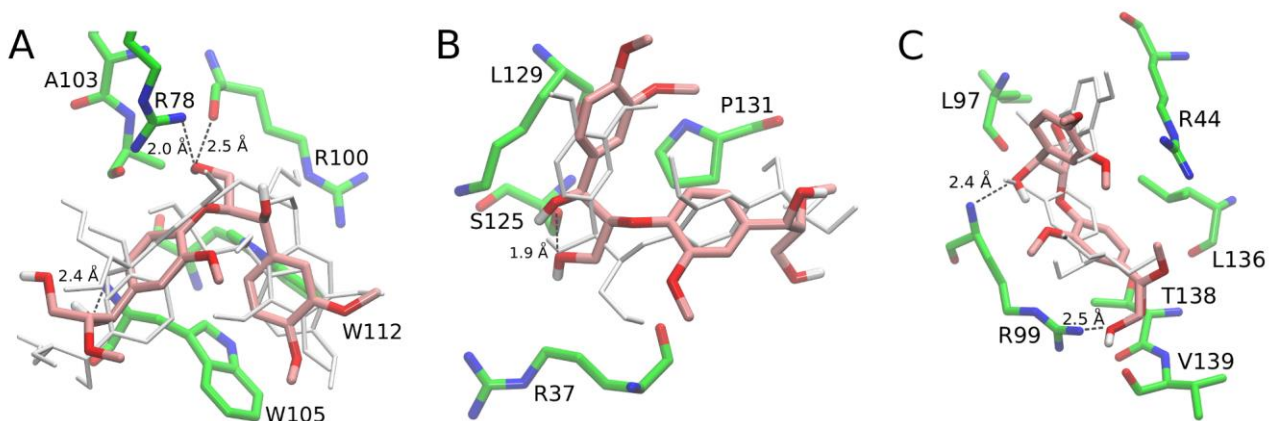


Figura 5.2. Modos de ligação de LGG sugeridos por *docking* nos sítios S1 (A), S4 (B) e S5 (C). O modo escolhido para MD é representado em rosa e as ligações de hidrogênio adicionais são destacadas com linhas tracejadas.

O cálculo resultou em nove modos de ligação, distribuídos em cinco sítios na superfície de FnIII. Os valores de afinidade são apresentados na Tabela 5.2. Os cinco sítios apontados como favoráveis ao encaixe dos dímeros foram nomeados, conforme é mostrado na Tabela 5.2, e serão mencionados algumas vezes no trabalho. Foram

selecionadas as cinco poses marcadas em vermelho na Tabela para conduzir simulações de dinâmica molecular, de forma que fosse possível estudar os cinco sítios simultaneamente. A escolha entre os modos de ligação que ocupam um mesmo sítio baseou-se na energia de afinidade como critério principal. O critério de desempate foi o número de interações favoráveis, como ligações de hidrogênio e contatos hidrofóbicos. A Figura 5.2 mostra os três sítios (S1, S4 e S5) em que coincidem dois ou três modos de ligação e indica as ligações de hidrogênio adicionais formadas com os modelos selecionados.

5.2.3. Simulações de MD

As caixas de simulação foram geradas usando o Packmol (Martínez & Martínez, 2003; Martínez *et al.*, 2009), antes de qualquer manipulação dos dímeros. O domínio FnIII foi posicionado no centro de uma caixa cúbica de cerca de 80 Å de lado em cada direção e hidratado por 15.000 moléculas de água. Foram adicionados 24 íons cloreto e 20 íons sódio para manter o sistema eletricamente neutro na concentração salina de 0,16 M. Os estados de ionização dos resíduos ionizáveis (Lys, Arg, His, Asp e Glu) foram determinados de acordo com os respectivos valores de pK_a em pH 5,0 e com o ambiente molecular (alta constante dielétrica na superfície da proteína e baixa constante dielétrica em seu interior), usando o servidor H++ (<http://biophysics.cs.vt.edu/H++>, último acesso em janeiro de 2013;; Anandakrishnan & Onufriev, 2012; Myers *et al.*, 2006; Gordon *et al.*, 2005). Como resultado, todos os resíduos ionizáveis foram considerados carregados.

As etapas de minimização e equilíbrio, bem como os parâmetros de simulação, foram descritos no Capítulo 2. A partir dos sistemas equilibrados, foram rodadas até três simulações independentes de cada sistema. A simulação do **sistema 1** durou 60 ns. Um total de 180 ns de tempo de simulação, divididos em três corridas independentes de 60 ns, foi gerado para cada um dos **sistemas 2, 3, 4a e 4b**, conforme é mostrado na Tabela 5.1. A estabilidade global e a relaxação estrutural das enzimas foi monitorada através da

evolução temporal do desvio quadrático médio dos átomos de C_{α} da proteína ao longo das simulações. Observou-se que, após cerca de 30 ns, os valores de RMSD tornaram-se estáveis para todos os sistemas. Este intervalo de tempo, então, não foi considerado para o cálculo de propriedades médias.

5.3 Estrutura do domínio fibronectina III de β -glicosidase de *Aspergillus niger*.

O enovelamento FnIIID consistiu-se de camadas paralelas de folhas- β , lembrando um "sanduíche" (em inglês, este tipo de estrutura é mesmo chamado de *beta sandwich structure*). O domínio apresenta fitas A, B e E em uma das faces e fitas C, D, F e G na outra. Uma pequena hélice e duas pequenas fitas- β intermedeiam as fitas E e F, conforme mostrado na Figura 5.3A. Há poucas estruturas destes domínios resolvidas, especialmente de enzimas de microorganismos. A estrutura resolvida de maior semelhança estrutural e de sequência primária encontrada a partir da busca pelo servidor Dali (Holm & Rosenström, 2010) é a do domínio FnIIID de β -glicosidase de *Thermotoga neopolitana* (PDB id: 2X40) (Pozzo *et al.*, 2010). Mesmo tratando-se de uma enzima de bactéria, são identificados muitos resíduos idênticos (identidade de 32%) e outros de natureza similar.

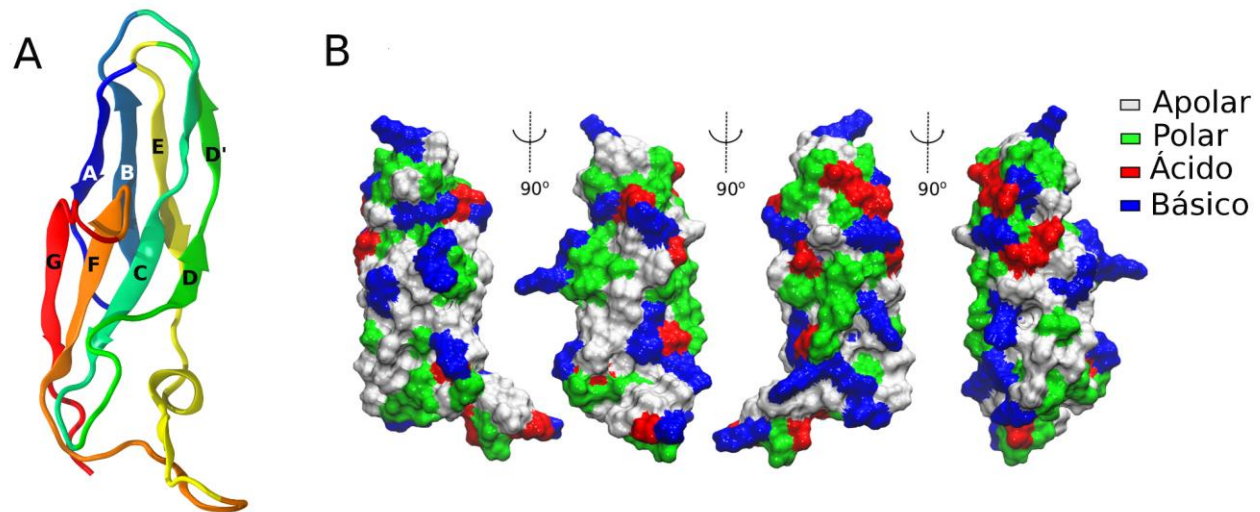


Figura 5.3. A) Estrutura de "sanduíche" de folhas- β do domínio fibronectina do tipo III. B) Superfície de van der Waals mostrando, de diferentes ângulos, a exposição de resíduos de diferentes natureza (vide legenda interna).

O alinhamento múltiplo de sequências de FnIIID de bactérias indica que resíduos que exercem papéis importantes na formação da estrutura principal (o "esqueleto") do domínio são totalmente conservados (Jee *et al.*, 2002). O alinhamento e comparação da natureza dos resíduos de FnIIID de *A. niger* e *T. Neapolitana* mostra que há grande semelhança tanto entre resíduos no centro hidrofóbico quanto na superfície do domínio. Além disso, verifica-se que estas enzimas são ricas em resíduos carregados e exibem grandes porções de resíduos apolares expostos ao solvente (Figura 5.3B).

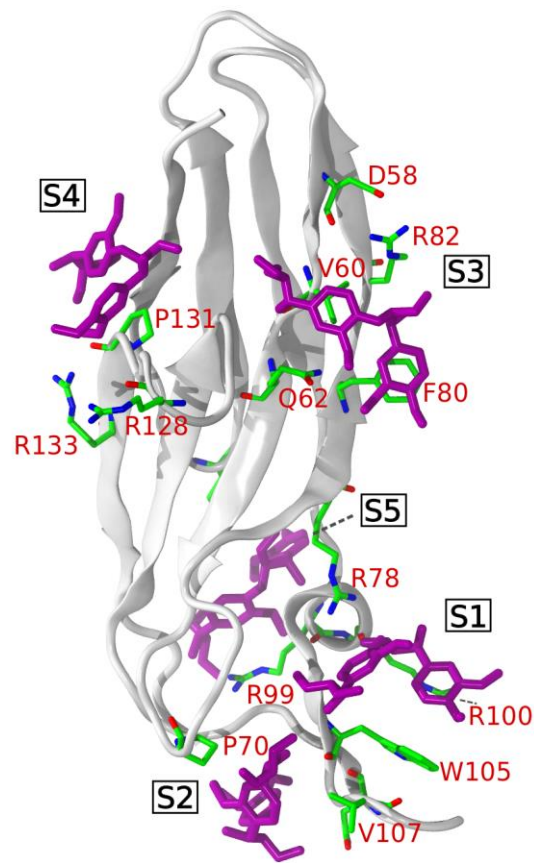


Figura 5.4. Os cinco sítios de ligação apontados pelos cálculos com o programa Autodock Vina.

A estrutura não exibe um sítio óbvio de ligação ao substrato, como são as típicas fendas encontradas em centros catalíticos e *clusters* de resíduos aromáticos em módulos de ligação à celulose de celulasas. O resultado do cálculo de *docking* pelo AutoDock Vina (**sistema 1**) é consistente com esta observação, uma vez que a multiplicidade de sítios de interação resultantes, alguns com diferentes modos de ligação, sugere que pode não existir um único sítio de interação na superfície do FnIII, mas vários. Uma outra característica marcante do domínio FnIII é que a maioria de seus resíduos encontra-se muito exposta ao solvente. Cabe adiantar o resultado de uma das análises das simulações: verificou-se que, em média, 89% de seus resíduos permanecem hidratados por, pelo menos, cinco moléculas de água (média calculada sobre todas as simulações).

5.4. Resultados das simulações

A energia de interação entre o domínio FnIII e os fragmentos de lignina LGG e LSS foi monitorada para todo o conjunto de doze simulações independentes aqui descritas. Os resultados são apresentados individualmente para os sistemas 1-4 na Figura 5.5.

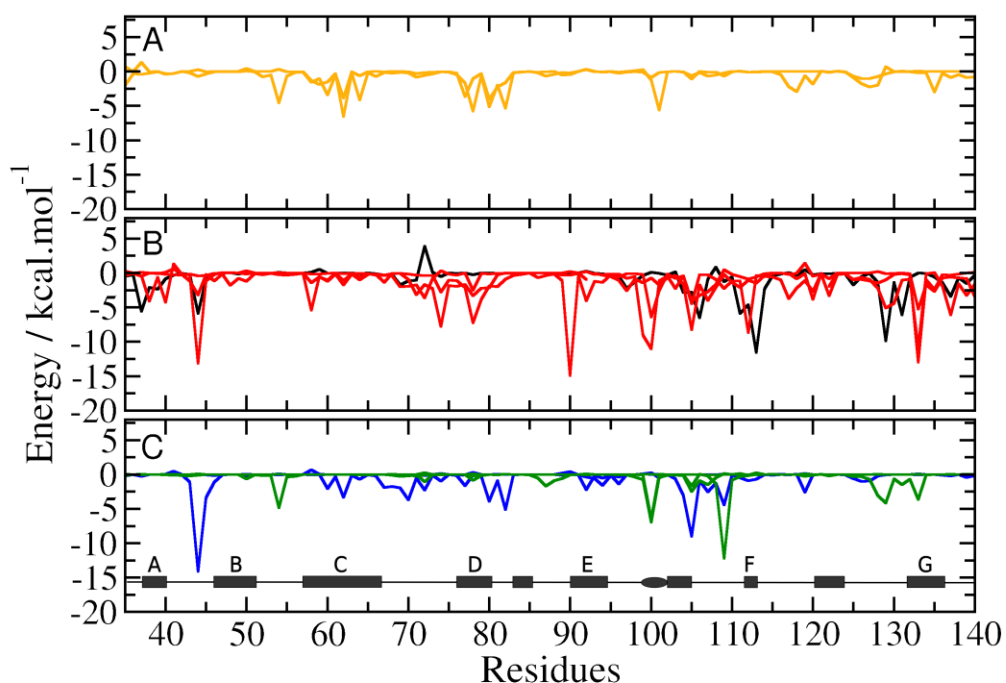


Figura 5.5. Energia de interação entre os dímeros LGG e LSS com os resíduos de FnIII. Os três painéis correspondem às diferentes abordagens adotadas para construir o arranjo inicial dos dímeros e rodar as simulações de MD. O painel (A) refere-se às simulações com distribuição inicial aleatória de 10 LGGs (**sistema 2**), o painel (B), às simulações de abordagens com encaixe guiado do substrato, isto é, utilizando o método de *docking* ou a simples inspeção visual (**sistema 1**, curva preta, e **sistema 3**, curvas vermelhas). Finalmente, o painel (C) refere-se às simulações dos **sistemas 4a** e **4b**, em que foi variada a unidade de monolignol.

Todas as diferentes abordagens para a construção da configuração inicial do sistema de simulação revelaram que há múltiplas regiões favoráveis para interações entre o domínio FnIII e os fragmentos de lignina. Nas simulações com 10 dímeros, especialmente no caso em que estes foram distribuídos aleatoriamente (**sistema 2**), a estatística de interações com o FnIII foi comprometida devido às frequentes interações entre os próprios fragmentos de lignina. Esta tendência já havia sido prevista nas simulações de Petridis e colaboradores (Petridis *et al.*, 2011), em que o colapso do polímero de lignina foi identificado a 300 K. Por outro lado, a análise das simulações de sistemas contendo apenas duas moléculas de LGG ou LSS, mostram que uma única simulação de 60 ns destes sistemas não é suficiente para explorar interações favoráveis em diferentes regiões do domínio. Se todas as simulações, geradas a partir de diferentes abordagens, forem analisadas em conjunto, muito mais informações sobre este sistema podem ser exploradas através de eventos repetidos.

Reunindo, então, em um único gráfico, os perfis de energia média de interação entre os resíduos de FnIII e os fragmentos de lignina das simulações dos sistemas 1-4 (Figura 5.6), claramente observa-se que há alguns resíduos que emergem com frequência como interagentes à lignina e são, portanto, promissores na análise dos prováveis sítios de ligação. Chama-se a atenção ao importante papel dos resíduos de arginina na interação com os fragmentos de lignina com o domínio FnIII. Os picos assinalados com 'X' na Figura 5.6A, revelam a alta ocorrência destas interações. Um análise mais detalhada mostra que as interações dispersivas e os contatos hidrofóbicos são predominantes em relação às ligações de hidrogênio entre resíduos de arginina em FnIII e LGG ou LSS. Isto é, os anéis aromáticos de lignina preferencialmente emparelham-se com o plano do cátion guanidíneo das argininas. Para demonstrar este tipo de interação, foi computada a evolução temporal do ângulo θ entre os vetores normais ao plano do cátion guanidíneo de um resíduo de arginina e ao plano de um dos anéis aromáticos de LGG (o anel que mais se aproxima do cátion). A Figura 5.7 mostra esta análise para três

casos representativos, simultaneamente à evolução temporal da energia de interação do respectivo par arginina-LGG. Com a aproximação destes grupos expressa pelo aumento no módulo de energia, as flutuações no valor de $\cos \theta$ tornam-se menores e este aproxima-se de 1 ou -1, indicando o emparelhamento.

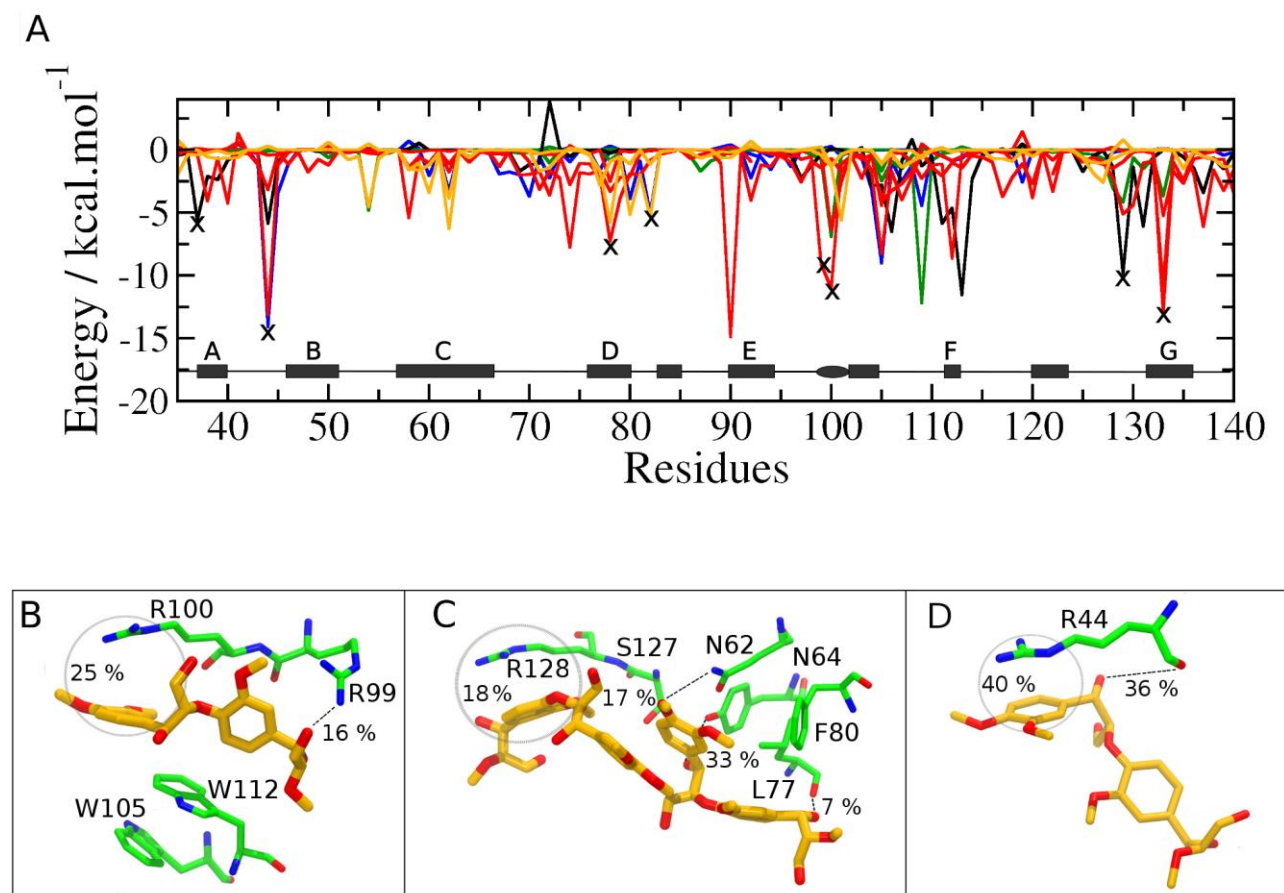


Figura 5.6. Interações entre fragmentos de lignina e o domínio FnIII verificadas nas simulações. (A) Perfis de energia de interação entre os dímeros LGG e LSS e resíduos do domínio FnIII correspondentes às diferentes abordagens adotadas para contruir a disposição inicial dos dímeros. Linha preta: **sistema 1**; Laranja: **sistema 2**; Vermelha: **sistema 3**; Azul e verde: **sistemas 4a/4b**. Os símbolos 'X' indicam os resíduos de arginina, que estão localizados em regiões do FnIII que emergem recorrentemente como prováveis sítios de ligação, independentemente da abordagem empregada para a disposição inicial de lignina. Os painéis (B), (C) e (D) são *snapshots* das simulações representativos dos modos de ligação lignina – FnIII. A frequência de interações do tipo cátion- π e via ligações de hidrogênio ao longo das simulações são indicadas.

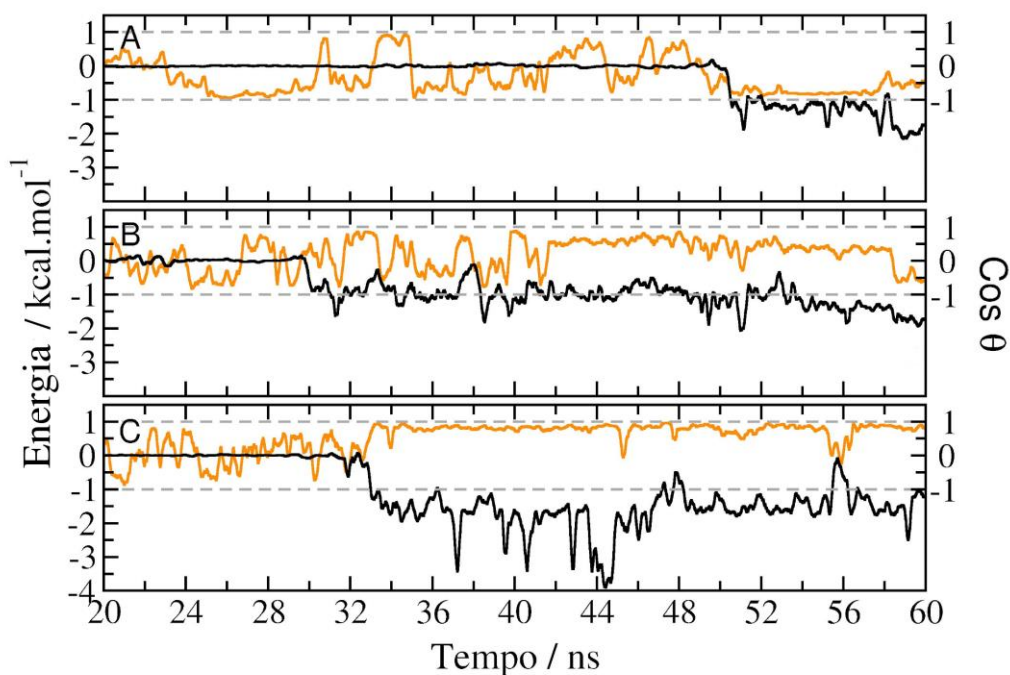


Figura 5.7. Gráficos que mostram simultaneamente a evolução temporal da energia de interação arginina-LGG (em preto) e do ângulo formado entre os vetores normais aos planos dos respectivos cátion guanidíneo de R e anel aromático de LGG (em laranja). Os painéis A, B e C referem-se às interações com R78 no **sistema 2**, R100 no **sistema 3** e R44 no **sistema 4a**.

Considerando a frequência em que ocorre cada tipo de interação nas simulações, três subdomínios de FnIII chamam a atenção como regiões importantes. O primeiro subdomínio, retratado na Figura 5.6B, corresponde à região dos longos *loops* entre as fitas C e D, e entre E e F (vide Figura 5.4, sítio S1). Sugere-se que esta região é atrativa para os análogos de lignina por estar parcialmente isolada da água, por ser cercada por cátions guanidíneos planares e grupos indol de argininas (R99 e R100) e triptofanos (W105 e W112), respectivamente. Também verifica-se a aglomeração de outros dímeros LGG nesta região, devido, principalmente, ao empilhamento π - π entre os fragmentos de lignina e os anéis aromáticos de triptofano.

Outro importante sítio de interação com lignina em FnIII é apresentado na Figura 5.6C e corresponde à região S3 que aparece no complexo gerado por *docking* (vide

Figura 5.4). Nesta região, um ou dois dímeros são frequentemente encontrados interagindo via ligações de hidrogênio com a tirosina Y64 e a leucina L77, além de formar contatos hidrofóbicos com a arginina R128. A Figura 5.6D, por sua vez, mostra um modo de interação frequentemente observado entre LGG ou LSS e a arginina R44. Este modo de ligação condiz com o sítio de ligação S5 obtido através dos experimentos de *docking*. Frequentemente observa-se uma ligação de hidrogênio uma das hidroxilas de lignina e a cadeia principal de R44.

5.5. Discussão sobre o papel do domínio FnIII de AnBgl1

A função do domínio FnIII, encontrado em proteínas diversas, vem sendo discutida em outros trabalhos. Atribui-se frequentemente a ele, o papel de mediar interações proteína-proteína. Não raro, este domínio aparece em cópias múltiplas entre as diferentes famílias de glicohidrolases, servindo também como espaçadores entre módulos catalítico e de ligação ao substrato (Campbell & Spitzfaden, 1994).

Kaeteva e colaboradores (Kaeteva *et al.*, 2002) demonstraram os efeitos do módulo FnIII de celobiohidrolase CbhA de *Clostridium thermocellum* em celulose. Os autores mostraram que o domínio FnIII é capaz de esfoliar e separar cadeias de celulose, expondo sítios adicionais da celulose para hidrólise pelo domínio catalítico. Os mesmos autores também demonstraram que efeitos mais fortes de FnIII são observados quando este está ligado ao domínio catalítico. A ligação covalente dos dois domínios aumenta a afinidade de FnIII pela superfície celulósica. Neste caso, no entanto, o enovelamento não é estendido, e Schubot e colaboradores (Schubot *et al.*, 2004) sugerem que este módulo tenham o importante papel de estabilizar o sítio ativo no domínio catalítico. Zhou e colaboradores (Zhou *et al.*, 2004) mostraram que a exclusão o domínio FnIII entre GH9-CBM3c e CBM2 em Cel9A de *T. Fusca* diminuiu a atividade em carboximetil celulase e em celulose microcristalina.

O domínio FnIII de β -glicosidase de *T. neapolitana* é a estrutura de maior

similaridade, em relação à sequência primária, com o domínio estudado neste trabalho. Nesse caso, porém, as superfícies do domínio FnIII e do domínio catalítico encontram-se em contato, em um arranjo que exclui a possibilidade de participação direta de FnIII no reconhecimento de pequenos substratos solúveis, muito embora ele possa estar envolvido na ancoragem da enzima a amplas superfícies de substratos poliméricos e na estabilização indireta do sítio ativo. Estudos funcionais de enzimas híbridas construídas com um domínio FnIII, de fato, confirmam seu papel na adesão à parede celular (Marín-Navarro *et al.*, 2008).

O domínio FnIII de AnBglu e seu homólogo de *T. Neapolitana* apresentam superfícies bastante expostas ao solvente, com um grande número de resíduos carregados e extensas porções apolares. Ambas estruturas são ricas em resíduos de arginina, com oito e nove resíduos Arg expostos em FnIII de AnBgl1 e TnBgl3B, respectivamente. Os experimentos de dinâmica molecular aqui apresentados sugerem que estes resíduos positivamente carregados possam aderir à matriz polimérica de lignina, principalmente através do emparelhamento cátion- π entre o cátion guanidíneo e os anéis aromáticos de LGG ou LSS. Este tipo de interação tem importância reconhecida na literatura por ser um fator auxiliar na estabilidade de proteínas e em interações proteína-ligante (Martis *et al.*, 2008; Dougherty, 1996). Além disso, sugere-se que a estrutura molecular de argininas é responsável por sua capacidade de aumentar a solubilidade de pequenas drogas orgânicas e proteínas em solução (Arakawa *et al.*, 2007; Arakawa *et al.*, 2008; Tsumoto *et al.*, 2007). Sugere-se que a arginina em solução seja responsável pela solubilização de moléculas de fulereno preferencialmente através de contatos hidrofóbicos e interações dispersivas com o grupo guanidíneo, enquanto que a porção polar da arginina aponta em direção ao solvente e cria uma cobertura hidrofílica, funcionando como uma molécula anfifílica (Li *et al.*, 2010).

Estes resultados de simulação, em conjunto com os ensaios experimentais de adsorção de AnBglu em materiais poliméricos (celulose, xilano, amido e lignina),

indicam que o domínio FnIII em AnBglu deve desempenhar um papel importante na ancoragem desta enzima à matriz lignocelulósica através da interação do domínio com lignina.

5.6 Conclusões

Neste trabalho, foi investigada a função do domínio do tipo fibronectina III de AnBgl1, a única β -glicosidase bimodular, que se conhece até o momento, em que o domínio catalítico é separado do domínio FnIII por uma longa cadeia peptídica estendida. Ensaio experimentais mostraram que a enzima tem alta afinidade pela superfície de lignina, mas não se liga a outros materiais poliméricos componentes da biomassa, como celulose. As simulações de MD deste trabalho indicam que o domínio FnIII é importante para direcionar a enzima a ancorar-se na porção de lignina da parede celular.

As simulações sugerem que o domínio FnIII liga-se à lignina de forma não específica em múltiplos sítios preferenciais de ligação, nos quais a interação cátion- π entre resíduos de arginina e os anéis aromáticos de lignina e ligações de hidrogênio com resíduos diversos simultaneamente contribuem para a estabilização da adesão lignina-FnIII. Além disso, os contatos hidrofóbicos entre os grupos indol de resíduos de triptofano e grupos aromáticos de lignina também se mostraram importantes fatores de estabilização da interação lignina-FnIII.

Sugere-se que a ancoragem da β -glicosidase seja especialmente útil para facilitar a imobilização da AnBgl1 na parede celular, uma região onde a celobiose é produzida pelo processo de despolimerização da biomassa através da ação concertada de exo- e endoglucanases. Uma vez que todas as enzimas variadas são extracelulares, produzidas simultaneamente no fungo, esta sinergia bioquímica na sacarificação da biomassa deve ser uma vantagem evolucionária. Claramente, é preciso avançar mais nos estudos bioquímicos para verificar esta hipótese.

Capítulo 6

Conclusões Gerais e Perspectivas

As enzimas glicosídeo hidrolases já ocupam amplo espaço no setor produtivo, sendo aplicadas nas indústrias papeleira e têxtil e como aditivos em alimentos de animais, por exemplo. A demanda e o incentivo à pesquisa sobre estas enzimas crescem ainda mais para que sejam usadas como catalisadores da sacarificação de material celulósico para obtenção de etanol combustível. No entanto, atualmente, os custos destas enzimas são relativamente altos e sua eficiência baixa, o que torna esta via economicamente pouco viável. Assim, muitos estudos têm sido direcionados para aprimorar o desempenho destas enzimas. O desempenho das enzimas esbarra em alguns desafios principais: 1) trata-se de uma catálise na interface líquido-sólido, a qual requer que o catalisador tenha uma afinidade ótima pela superfície do substrato, de forma que o catalisador prenda-se facilmente à superfície, mas que continue relativamente livre para a evolução da catálise ao longo de todo o substrato. Em comparação com a hidrólise ácida, a hidrólise enzimática é muito mais lenta; (2) a recalcitrância do material, isto é, a dificuldade de degradá-lo. A associação entre cadeias de celulose, lignina e hemicelulose é uma barreira natural à degradação; (3) as cadeias de celulose formam um arranjo cristalino resistente à hidrólise. Como são várias e distintas as enzimas necessárias para se completar o processo de sacarificação, é preciso combiná-las de maneira otimizada criando coquetéis enzimáticos eficientes.

São conhecidos três grupos principais de enzimas hidrolíticas que atuam na matriz celulósica, diferenciados pelo tipo de substrato e modo de ataque. As endoglucanases, o primeiro grupo, atuam especialmente clivando ligações glicosídicas nas regiões amorfas

da celulose; as exoglucanases, o segundo grupo, atuam hidrolisando as extremidades redutoras e não redutoras da matriz cristalina liberando celobiose; e, finalmente, as glicosidases clivam as moléculas de celobiose liberadas à solução, liberando moléculas de glicose. Estas enzimas atuam essencialmente de forma sinérgica para que o material celulósico possa ser degradado completamente. Conforme descrito ao longo desta tese, as propriedades estruturais (eg. formato da fenda e presença/ausência de domínios de ligação à celulose) destas enzimas estão intimamente relacionadas aos papéis que desempenham no processo hidrolítico. No entanto, os mecanismos de funcionamento de enzimas ainda não são completamente compreendidos e muitas questões estão sendo apenas agora levantadas. Uma delas, refere-se aos fatores que conferem termoestabilidade a celulasas, pois um aspecto importante a ser considerado para aprimorar o desempenho catalítico das enzimas para aplicações práticas é o desenvolvimento de enzimas estáveis em temperaturas elevadas (acima de 50 °C). Neste contexto, nossa contribuição deu-se nos estudos da laminarinase hipertermofílica de *Rhodothermus marinus* e outros microorganismos que produzem laminarinases de menor estabilidade térmica, descritos no **Capítulo 3**.

Foram conduzidas simulações comparativas de laminarinases que apresentam estruturas tridimensionais semelhantes, mas com diferentes níveis de estabilidade térmica. Verificou-se que o arranjo de pontes salinas, e não só sua abundância, é um fator importante na estabilização. As pontes salinas na enzima hipertermoestável (laminarinase de *R. marinus*) formam *clusters* que interconectam as folhas- β do domínio catalítico. Este tipo de ligação foi referida na tese como pontes salinas intra-folhas. Por outro lado, nem todas as pontes salinas contribuem para a estabilidade da proteína: a existência destas interações permeando o centro hidrofóbico, isto é, pontes salinas inter-folhas, na verdade, contribuem para desestabilizar a estrutura por facilitar a entrada de água em cavidades hidrofóbicas com o aumento da temperatura, como pôde ser visto

através de simulações em temperaturas elevadas de uma laminarinase mesofílica. Sugere-se que mutações que gerem pontes salinas favoráveis ou que extingam as desfavoráveis possam incrementar a estabilidade de glicosídeo hidrolases de estrutura conhecida. As simulações de desnaturação mostraram que os ganhos progressivos de mobilidade das cadeias laterais ocorrem de forma mais rápida nas enzimas menos estáveis. Especialmente nestas enzimas, o desenovelamento ocorre de fora para dentro, isto é, com as perdas de estrutura nas regiões periféricas ocorrendo inicialmente.

Um segundo aspecto importante nesta área do conhecimento, diz respeito à afinidade da enzima pelo substrato celulósico. A maioria das celulases apresenta um ou mais domínios CBM, responsáveis especificamente por aumentarem a superfície favorável ao contato com as cadeias celulósicas. A retirada deste domínio implica em uma diminuição severa na atividade enzimática. No **Capítulo 4**, foi apresentado um estudo sobre uma exceção dentre as GHs, as enzimas Cel12A (endoglucanase 3), que não possuem um CBM, mas ainda assim apresentam considerável atividade em regiões amorfas da celulose. Como resultado, verificou-se que a enzima apresenta um conjunto de resíduos arranjados similarmente a resíduos em CBMs que também atuam neste tipo de substrato. Trata-se de um conjunto de resíduos aromáticos que servem como âncoras hidrofóbicas ao substrato celulósico, e outros que interagem com ele via ligações de hidrogênio. Além disso, as simulações possibilitaram explicar porquê oligossacarídeos que incluem menos do que quatro unidades glicosídicas não são clivados por esta enzima. Sugere-se que possa ser profícuo o prosseguimento deste estudo através de mutações em enzimas com um enovelamento similar pela inclusão de resíduos em arranjo típico de CBM descrito neste trabalho.

Já no **Capítulo 5**, estudamos uma questão até então insuspeita e de certa forma inusitada, que é a interação de um subdomínio encontrado na β -glicosidase de *Aspergillus niger* (AnBgl1) com a matriz de lignina das paredes celulares de plantas.

Diferente do que normalmente se encontra entre as β -glicosidases, formadas apenas por um centro catalítico, constatou-se que a AnBgl1 é uma enzima bimodular, com um domínio fibronectina do tipo III além do centro catalítico. Provou-se, experimentalmente, que este domínio não tem a função de se prender a polissacarídeos, mas à lignina. Através das simulações, foi possível mostrar que a interação entre FnIII e lignina ocorrem especialmente pela participação de resíduos de arginina, através de interações cátion- π entre o grupo guanidíneo e os anéis aromáticos presentes em abundância na lignina. É interessante refletir que talvez não haja uma forma mais apropriada para interações com um substrato aromático tridimensional tão heterogêneo quanto a lignina. O domínio FnIII apresenta resíduos de arginina espalhados por toda a sua superfície, o que pode possibilitar que a lignina “abraçe” o domínio em várias regiões simultaneamente. Esta construção seria muito mais difícil com resíduos aromáticos, os quais geralmente não se encontram tão expostos ao solvente. Acredita-se que seja importante dar continuidade a este estudo através de testes de alteração no desempenho enzimático de AnBgl1 sem o domínio e através da substituição dos resíduos de arginina. Estes testes poderiam ser estendidos para outras β -glicosidases.

Até agora, três trabalhos foram concluídos. O estudo das bases moleculares da termofilicidade e termoestabilidade de laminarinases foi publicado na revista *Journal of Physical Chemistry B* (Bleicher & Prates *et al.*, 2011), o trabalho sobre os mecanismos de reconhecimento do substrato por Cel12A foi publicado na revista *Plos One* (Prates *et al.*, 2013) e o estudo da função do domínio fibronectina III (o conjunto de resultados experimentais e de MD) encontra-se em análise pela revista *The Journal of Biological Chemistry* (“*Aspergillus niger* beta-glucosidase has a cellulose-like tadpole molecular shape: insights into GH3 beta-glucosidases structure and function”). O artigo de estudo da desnaturação térmica das laminarinases também já está escrito e acredita-se que possa ser submetido em breve para publicação.

Referências Bibliográficas

Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* 106: 1589–1615.

Alahuhta M, Xu Q, Brunecky R, Adney WS, Ding S-Y, Himmel ME, Lunin, VV (2010) Structure of a fibronectin type-III like module from *Clostridium thermocellum*. *Struct Commun* F66: 878-880.

Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating *pK* prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. *Nucleic Acids Res* 40(W1):W537-541.

Aqvist J, Luzhkov VB, Brandsal BO (2002) Ligand binding affinities from MD simulations. *Acc Chem Res* 35: 358–365.

Arakawa T, Tsumoto K, Kita Y, Chang B, Ejima D (2007) Biotechnology applications of amino acids in protein purification and formulations. *Amino Acids* 33: 587-605.

Arakawa T, Kita Y, Koyama AH (2008) Solubility enhancement of gluten and organic compounds by arginine. *Int J Pharm* 355: 220-223.

Arantes V, Saddler JN (2010) Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol Biofuels* 3: 4-15.

Argos P, Rossmann GM, Grau UM, Zuber H, Frank G, Tratschin JD (1979) Thermal stability and protein structure. *Biochemistry* 18: 5698–5703.

Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58:521–54.

Beechem JM, Brand L (1985) Time-resolved fluorescence of proteins. *Ann Rev Biochem* 54: 43-71.

Berner R (2003) The long-term carbon cycle, fossil fuels and atmospheric composition. *Nature* 426:323-326.

Bhat MK (2000) Cellulases and related enzymes in biotechnology. *Biotechnol Adv* 18:355-383.

Bjornsdottir SH, Blondal T, Hreggvidsson GO, Eggertsson G, Petursdottir S, Hjorleifsdottir S, Thorbjarnardottir SH, Kristjansson JK (2006) *Rhodothermus marinus*: physiology and molecular biology. *Extremophiles* 10:1–16.

Bleicher L., Prates ET, Gomes TC, Silveira RL, Nascimento AS, Rojas AL, Golubev A, Martinez L, Skaf MS, Polikarpov I (2011) Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structures of the hyperthermostable laminarinase from *Rodothermus marinus* and molecular dynamics simulations. *J Phys Chem B* 115:7940-7949.

Bon EPS, Ferrara MA (2007) “Bioethanol production via hydrolysis of cellulosic biomass.” *Fao Symposium - “The role of agricultural biotechnologies for production of bioenergy in developing countries”*, disponível em <http://www.academia.edu/1441021/>, acessado em 28 de abril de 2013.

Bon EPS, Gírio F, Pereira Jr N (2008) “Enzimas na produção de etanol”. Em: *Enzimas em Biotecnologia*, cap. 10, pp. 241-271; Eds. Elba P. S. Bon, Maria Antonieta Ferrara, Maria Luísa Corvo.

Boraston AB, Nurizzo D, Notenboom V, Ducros V, Rose DR, Kilburn DG, Davies GJ (2002) Differential oligosaccharide recognition by evolutionary-related β -1,4 and β -1,3 glucan-binding modules. *J Molec Bio* 319:1143-1156.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382:769-781.

Borris R, Krah M, Brumer 3rd H, Kerzhner MA, Elyakova LA, Ivanen DR, Eneyskaya EV, Shishlyannikov SM, Shabalin KA, Neustroev KN (2003) Enzymatic synthesis of 4-methylumbelliferyl beta-(1,3)-D-glucooligosaccharides- new substrates for 1,3(4)-beta-glucanase. *Carbohydr Res* 338:1455-1457.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187-217.

Bryngelson FD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc. Nat. Acad. Sci* 84:7524-7528.

- Bumpus JA, Aust SD (1987) Biodegradation of DDT [1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane] by the white rot fungus *Phanerochaete chrysosporium*. Appl Environ Microbiol 53: 2001–2008.
- Campbell ID, Spitzfaden C (1994) Building proteins with fibronectin type III modules. Structure 2: 333-337.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233-238.
- Chang MW, Ayeni C, Breuer S, Torbett BE (2010) Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. Plos One 5: e11955.
- Cheng YS, Ko TP, Wu TH, Ma Y, Huang CH, Lai HL, Wang AH, Liu JR, Guo RT (2011) Crystal structure and substrate-binding mode of cellulase 12 Å from *Thermotoga maritima*. Prot-Struc Func Bioinf 79:1193-1204.
- Chipot C, Pohorille A, eds. (2007) Free energy calculations. Springer-Verlag: Berlin.
- Chundawat SPS, Beckham GT, Himmel ME, Dale BE (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. Ann Rev Chem Bio Eng 2: 121-145.
- Crowley PB, Golovin A (2005) Cation- π interactions in protein-protein interfaces. Proteins 58: 231-239.
- Daggett V (2002) Molecular dynamics simulations of the protein unfolding/folding reaction. Acc Chem Res 35: 422-429.
- Daggett V, Fersht AR (2003) Is there a unifying mechanism for protein folding? Trends Biochem Sci 28: 18-25.
- Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. Nat Rev Mol Cell Biol, 4: 497-502.
- Daggett V (2006) Protein folding-simulation. Chem Rev 106: 1898-1916.
- D'Almeida MO (1998) Celulose e Papel - Tecnologia de Fabricação de Pasta Celulósica. Ed. Departamento de Divulgação do IPT, v. 1 e 2, 2ª Edição.

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98: 10089-10092.

Darve E, Rodríguez-Gómez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys* 128: 144120.

Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3: 853-859.

Davies GJ, Wilson KS, Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321: 557-559.

Day R., Daggett V (2005) Ensemble versus single-molecule protein unfolding. *Proc Natl Acad Sci* 102: 13445-13450.

Din N, Damude H, Gilkes N, Miller R, Warren R, Kilburn D (1994) C1–Cx revisited: intramolecular synergism in a cellulase. *Proc Natl Acad Sci USA* 91: 11383– 11387.

Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 25: 331-339.

Divne C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, *et al* (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265: 524-527.

Dominy BN, Minoux H, Brooks III CL (2004) An electrostatic basis for the stability of thermophilic proteins. *Proteins* 57: 128–141.

Dougherty DA (1996) Cation- π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp. *Science* 271: 163-168.

Elcock AH (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J Mol Biol* 284: 489–502.

Fagerstrom A, Liljefors T, Stahlberg J, Sandgren M, Berg U, Isaksson R (2009) A study of the chiral recognition mechanisms of cellobiohydrolase Cel7A for ligands based on the beta-blocker motif: crystal structures, microcalorimetry and computational modelling of Cel7A-inhibitor complexes. To be published.

- Fersht AR, Daggett V (2002) Protein Folding and Unfolding at Atomic Resolution. *Cell* 108: 573-582.
- Feynman RP, Leighton RB, Sands M (1963). *The Feynman Lectures on Physics*, Vol. 1 Addison-Wesley.
- Fibriansah G, Masuda S, Koizumi N, Nakamura S, Kumasaka T (2007) The 1.3 Å crystal structure of a novel endo-beta-1,3-glucanase of glycoside hydrolase family 16 from alkaliphilic *Nocardiosis sp.* strain F96. *Proteins* 69: 683–690.
- Field MJ (2001) Simulating enzyme reactions: challenges and perspectives. *Comp Chem*, 23: 48-58.
- Figueira AC, Saidenberg DM, Souza PCT, Martinez L, Scanalan TS, Baxter JD, Skaf MS, Palma MS, Webb P, Polikarpov I (2010) Analysis of agonist and antagonist effects on thyroid hormone receptor conformation by hydrogen/deuterium exchange. *Molec Endoc* 25: 15-31.
- Fletcher R (2000). *Practical Methods of Optimization*, 2nd Ed.; Wiley: Padstow.
- Fletcher R, Reeves CM (1964) Function minimization by conjugate gradients. *Comput J* 7: 149-154.
- Foster MP, McElroy CA, Amero CD (2008) Solution NMR of large molecules and assemblies. *Biochem* 46: 331-340.
- Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254: 1598-1602.
- Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. *Prot Struc Func Gen* 23: 566-579.
- Generoso WC, Malagó Jr W, Pereira Jr N, Henrique-Silva F (2012) Recombinant expression and characterization of an endoglucanase III (cel12a) from *Trichoderma harzianum* (*Hypocreaceae*) in the yeast *Pichia pastoris*. *Gen Mol Res* 11: 1544-1557.
- Ghose TK (1987) Measurement of cellulase activities. *Pure Appl Chem* 59: 257-268.

Goldstein RA (2007) Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: Insights from the quasi-chemical approximation. *Protein Sci* 16:1887-1895.

Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pK_a s and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33:W368–W371.

Guvench O, Hatcher ER, Venable RM, Pastor RW, Mackerell Jr AD (2009) CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J Chem Theory Comput* 5: 2353-2370.

Guvench O, Mallajosyula SS, Raman EP, Hatcher ER, Vanommeslaeghe K, Foster TJ, Jamison FW, Mackerell AD (2011) CHARMM additive all-atom force field for carbohydrates derivatives and its utility in polysaccharides and carbohydrate-protein modeling *J Chem Theory Comput* 7: 3162-3180.

Hatcher ER, Guvench O, Mackerell Jr AD (2009) CHARMM additive all-atom force-field for acyclic polyalcohols, acyclic carbohydrates, and inositol. *J Chem Theory Comput* 5: 1315-1327.

Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 280: 309-361.

Henrissat B (1994) Cellulases and their interaction with cellulose. *Cellulose* 1: 169-196.

Henrissat B, Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Op Struct Biol*, 7: 637-644.

Henrissat B (1998) Glycosidase families. *Biochem Soc Trans* 26: 153-156.

Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* 450: 964-972.

Hirano A, Arakawa T, Shiraki K (2008) Arginine increases the solubility of coumarin: Comparison with salting-in and salting-out additives. *J Biochem* 144: 363-369.

Holm L, Rosentröm P (2010) Dali server: conservation mapping in 3D. *Nucl Acids Res* 38: W545-549.

- Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Molec Graphics* 14: 33-38.
- Isralewitz B, Gao M, Schulten K (2001) Review Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* 11: 224-230.
- IUBMB: Enzyme Nomenclature. Recommendations. Sand Diego: Academic Press; 1992.
- Jaenicke R. (1991) Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem* 202: 715–728.
- Jaenicke R, Bohm G (1998) The stability of proteins in extreme environments. *Curr Opin Struct Biol*, 8: 738–748.
- Jameson DM, Vetromile CM, James NG (2013) Investigations of protein-protein ninteractions using time-resolved fluorescence and phasors. *Methods* 59: 278-286.
- Jee JG, Ilegami T, Hashimoto M, Kawabata T, Ikeguchi M, Watanabe T, Shirakawa M (2002) Solution structure of the fibronectin type III domain from *Bacillus circulans* WL-12 chitanase A1. *J Biol Chem* 277: 1388-1397.
- Jorgensen WL, Chandrasekhar J, Madura JD (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926-935.
- Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105: 6474-6487.
- Karlsson J, Siika-aho M, Tenkanen M, Tjerneld F (2002) Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. *J Biotechnol* 99: 63-78.
- Kataeva IA, Seidel III RD, Shah A, West LT, Li X-L, Ljungdahl LG (2002) The fibronectin type-3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *App Env Microb* 68: 4292-4300.

Keitel T, Meldgaard M, Heinemann U (1994) Cation binding to a *Bacillus* 1,3-1,4-beta-glucanase. Geometry, affinity and effect on protein stability. *Eur J Biochem* 222: 203-214.

Kipper K, Våljamäe P, Johansson G (2005) Processive action of cellobiohydrolase Cel7A from *Trichoderma reesei* is revealed as 'burst' kinetics on fluorescent polymeric model substrates. *Biochem J* 385: 527-535.

Korotkova OG, Semenova MV, Morozova VV, Zorov IN, Sokolova LM, Bubnova TM, Okunev ON, Sinitsyn AP (2009) Isolation and properties of fungal beta-glucosidases. *Biochem* 74: 569-577.

Krah M, Misselwitz R, Politz O, Thomsen KK, Welfle H, Borriss R (1998) The laminarinase from thermophilic eubacterium *Rhodothermus marinus*. *Eur J Biochem* 257: 101-111.

Kumar S, Nussinov R (2002) Close-range Electrostatic Interactions in Proteins. *ChemBioChem* 3: 604-617.

Kumar R, Singh S, Singh OV (2008) Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J Ind Microbiol Biotechnol* 35:377-391.

Kuzmanic A, Zagrovic B (2010) Determination of ensemble-average pairwise root-mean square deviation from experimental b-factors. *Biophys J* 98: 861-871.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. *Bioinformatics* 23: 2947-2948.

Larsson AM, Bergfors T, Dultz E, Irwin DC, Roos A, Driguez H, Wilson DB, Jones TA (2005) Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: the role of tyrosine Y73 in substrate ring distortion. *Biochem* 44: 12915-12922.

Laughrey ZR, Kiehna SE, Riemen AJ, Waters ML (2008) Carbohydrate- interactions: what are they worth? *J Am Chem Soc* 130: 14625-14633.

Liberato MV, Generoso WC, Malagó Jr W, Henrique-Silva F, Polikarpov I (2012) Crystallization and preliminary X-ray diffraction analysis of endoglucanase III from *Trichoderma harzianum*. *Acta Crystallogr F* 68: 306-309.

- Leach AR (2001) *Molecular Modelling: Principles and applications*. 2^a Ed. Pearson Education, England.
- Little E, Bork P, Doolittle RF (1994) Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *J Molec Evol* 39: 631-643.
- Li J, Garg M, Shah D, Rajagopalan R (2010) Solubilization of aromatic and hydrophobic moieties by arginine in aqueous solutions. *J Chem Phys* 133: 054902.
- Liu YS, Zeng Y, Luo Y, Xu Q, Himmel ME *et al* (2009) Does the cellulose-binding module move on the cellulose surface? *Cellulose* 16: 587-597.
- Lu, N, Kofke DA (2001) Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *J Chem Phys* 114: 7303–7312.
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66: 506–77.
- Macarrón R, Acebal C, Castellón MP, Domínguez JM, Mata I, Petterson G, Tomme P, Claeysens M (1993) Mode of action of endoglucanase III from *Trichoderma reesei*. *Biochem J* 289: 867-873.
- Mackerell Jr. AD, Bashford D, Bellott M, Dunbrack Jr. RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher III WE, Roux B, Scenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102: 3586-3616.
- MacKerell Jr AD, Feig M, Brooks III CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25: 1400–1415.
- Magalhães CS, Barbosa HJC, Dardenne LE (2007) “Métodos de docking receptor-ligante para o desenho racional de compostos bioativos”, Em: *Métodos de química teórica e modelagem molecular*”, cap. 14, pp. 489-527; Eds Nelson H Morgon, Kaline Coutinho. Livraria da Física, São Paulo.

Marín-Navarro J, Gurgu L, Alamar S, Polaina J (2011) Structural and functional analysis of hybrid enzymes generated by domain shuffling between *Saccharomyces cerevisiae* (var. *diastaticus*) Stal glucoamylase and *Saccharomycopsis fibuligera* Bgl1 β -glucosidase. *App Micr Biotec* 89: 121-130.

Martínez JM, Martínez L (2003) Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *J Comput Chem* 24: 819–825.

Martínez L, Andrade R, Birgin EG, Martínez JM (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* 30: 2157–2164.

Martínez L, Sonoda MT, Webb P, Baxter JD, Skaf MS, Polikarpov I (2005) Molecular dynamics simulations reveal multiple pathways of ligand dissociation from thyroid hormone receptors. *Biophys J* 89: 2011-2023.

Martínez L, Webb P, Polikarpov I, Skaf MS (2006) Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: evidence of the likeliest escape pathway and its implications for the design of novel ligands. *J Med Chem* 49: 23-36.

Martínez L, Polikarpov I, Skaf MS (2008) Only subtle protein conformational adaptations are required for ligand binding thyroid hormone receptors: simulations using a novel multipoint steered molecular dynamics approach. *J Phys Chem B* 112: 10741-10751.

Martínez L, Nascimento AS, Nunes FN, Philips K, Aparício R, Dias SMG, Figueira ACM, Lin JH, Nguyen P, Apriletti JW, Neves FAR, Baxter JD, Webb P, Skaf MS, Polikarpov I (2009). Gaining ligand selectivity in thyroid hormone receptors via entropy. *Proc Nat Acad Sci* 106: 20717-20722.

Martínez L, Souza PCT, Garcia W, Batista FAH, Portugal RV, Nascimento AS, Nakahira M, Lima LMTR, Polikarpov I, Skaf MS (2010) On the denaturation mechanisms of the ligand binding domains of thyroid hormone receptors. *J Phys Chem B* 114: 1529-1540.

Martins RL, Singh SK, Gromiha MM, Santhosh C (2008) Role of cation- π interactions in single chain 'all- α ' proteins. *J Theor Biol* 250: 655-662.

Martyna GJ (1994) Remarks on constant-temperature molecular dynamics with momentum conservation. *Phys Rev* 50: 3234-3236.

- Mason PE, Neilson GW, Enderby JE, Saboungi ML, Dempsey CE, MacKerell AD, Brady JW (2004) The structure of aqueous guanidinium chloride solutions. *J Am Chem Soc* 126: 11462-11470.
- Masuda S, Endo K, Koizumi N, Hayami T, Fukazawa T, Yatsunami R, Fukui T, Nakamura S (2006) Molecular identification of a novel β -1,3-glucanase from alkaliphilic *Nocardiopsis* sp. Strain F96. *Extremophiles* 10: 251-255.
- McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267: 585-590.
- McQuarrie DA (2000) *Statistical Mechanics*. University Science Books, Sausalito.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comp Chem* 16: 2785-2791.
- Myers J, Grothaus G, Narayanan S, Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* 63: 928-938.
- Nakazawa H, Okada K, Onodera T, Ogasawara W, Okada H, *et al* (2009) Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. *Appl Microbiol Biotechnol* 83: 649-657.
- Nelson DL, Cox MM (2005) *Lehninger. Principles of Biochemistry*. 4^a Ed. WH Freeman and Company, Nova Iorque.
- Nookaraju A, Pandey SK, Bae H-J, Joshi CP (2013) Designing cell walls for improved bioenergy production. *Molecular Plant* 6: 8-10.
- Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Ann Rev Phys Chem* 48: 545-600.
- Pacheco TF (2011) "Produção de etanol: primeira ou segunda geração?". Artigos técnicos da Embrapa, disponível em <http://www.embrapa.gov.br/imprensa/artigos/2011/>, acessado em 28 de abril de 2013.
- Palmer AG (2001) NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* 30: 129-155.

Papoian GA (2008) Proteins with weakly funneled energy landscapes challenge the classical structure-function paradigm. PNAS 105: 14237-14238.

Payne CM, Bomble YJ, Taylor CB, McCabe C, Himmel ME *et al* (2011) Multiple functions of aromatic-carbohydrate interactions in a processive cellulose examines with molecular simulation. J Biol Chem 286: 41028-41035.

Petit L, Gibert M, Popoff MR (1999) *Clostridium perfringens*: toxinotype and genotype. Trends Microbiol 7: 104–110.

Petridis L, Schulz R, Smith JC (2011) Simulation analysis of the temperature dependence of lignin structure and dynamics. JACS, 133: 20277-20287.

Petridis L, Smith JC (2008) A molecular mechanics force field for lignin. J Comp Chem 30: 457-467.

Phillips Jr GN (1990) Comparison of the dynamics of myoglobin in different crystal forms. Biophys J 57: 381-383.

Philips JC, Braun R, Wangf W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J. Comput Chem 26: 1781-1802.

Phitswan P, Laohakunjit N, Kerdchoechuen O, Kyu KL, Ratanakhanokchai K (2013) Present and potential applications of cellulases in agriculture, biotechnology, and bioenergy. Folia Microbiol 58: 163-176.

Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscapes theory. Part I: Basic concepts. Quarterly Reviews of Biophysics 35: 205-286.

Pozzo T, Pasten JL, Karlsson EN, Logan DT (2010) Structural and functional analyses of b-glucosidase 3B from *Thermotoga neapolitana*: a thermostable three-domain representative of glycoside hydrolase 3. J. Mol. Biol. 397: 724-739.

Prates ET, Stankovic I, Silveira RL, Liberato MV, Henrique-Silva F, Pereira Jr N, Polikarpov I, Skaf MS (2013) X-ray structure and molecular dynamics simulations of endoglucanase 3 from *Trichoderma harzianum*: structural organization and substrate recognition by endoglucanases that lack cellulose binding module. Plos One 8: E59069.

- Priyakumar UD, Ramakrishna S, Nagarjuna KR, Reddy KS (2010) Structural and energetic determinants of thermal stability and hierarchical unfolding pathways of hyperthermophilic proteins, Sac7d and Sso7d. *J Phys Chem B*, 114: 1707–1718.
- Puhl AC, Cvorovic A, Campos JLO, Reinach PS, Saidenberg DM, Yuan J, Poikarpov I, Bernardes A, Palma MS, Webb P, Silveira RL, Ayers SD, Skaf MS (2012) Mode of PPAR Activation by Luteolin. *Molec Pharm* 76216.
- Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) 'MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7: 382.
- Roitberg A, Elber R (1991) Modeling of side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J Chem Phys* 95: 9277–87.
- Rubin E (2008) Genomics of cellulosic biofuels. *Nature* 454: 841-845.
- Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of cartesian equations of motion of a system with constraints – molecular dynamics of n-alkanes. *J Comp Phys*, 23: 327-341.
- Sakon J, Irwin D, Wilson DB, Karplus PA (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat Struct Biol* 4: 810-818.
- Sandgren M, Shaw A, Ropp TH, Bott SWR, Cameron AD, *et al* (2001) The X-ray Crystal Structure of the *Trichoderma reesei*, Family 12 Endoglucanase 3, Cel12A, at 1.9 Å Resolution. *J Mol Biol* 308: 295-310.
- Sandgren M, Ståhlberg J, Mitchinson C (2005) Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes. *Prog Biophys Molec Biol* 89: 246-291.
- Sangha AK, Petridis L, Smith JC, Ziebell A, Parks JM (2012) Molecular simulation as a tool for studying lignin. *Env Prog Sust Energy* 31: 47-54.
- Sarko A (1997) “Cellulose – How much do we know about its structure?” Em: Wood and Cellulosics: Industrial Utilization. Biotechnology, Structure and Properties. John & Wiley & Sons, Nova Iorque.

Schaftenaar G, Noordik JH (2000) Molden: a pre- and post-processing program for molecular and electronic structures. *J Comput Aided Mol Design* 14: 123-134.

Schlick T (2009) Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *F1000 Biol Reports* 1:51.

Schotte F, Somanb J, Olsonb JS, Wulff M and Anfinruda PA (2004) Picosecond time-resolved X-ray crystallography: probing protein function in real time. *Journal of Structural Biology* 147: 235–246.

Schneider T, Stoll E (1978) Molecular dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys Rev B* 17: 1302-1322.

Schubot FD, Kataeva IA, Chang J, Shah AK, Ljungdahl LG, Rose JP, Wang B-C (2004) Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. *Biochem* 43: 1163-1170.

Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 48:1198-1229.

Silveira RL, Martínez L, Skaf MS, Martínez J (2012) Enzyme Microheterogeneous Hydration and Stabilization in Supercritical Carbon Dioxide. *Journal of Physical Chemistry. B*, 116: 5671-5678.

Sims R, Taylor M, Saddler J, Mabee W (2008) “From 1st- to 2nd-Generation Biofuels Technologies: An Overview of Current Industry and RD&D Activities”. International Energy Agency Bioenergy, pp124; OECD/IEA, Paris.

Sonoda MT, Martínez L, Webb P, Skaf MS, Polikarpov I (2008) Ligand dissociation from estrogen receptor is mediated by receptor dimerization: evidence from molecular dynamics simulations. *Molec Endoc* 22: 1565-1578.

Souza PCT, Barra GB, Velasco LFR, Ribeiro ICJ, Simeoni LA, Togachi M, Webb P, Neves FAR, Skaf MS, Martinez L, Polikarpov I (2011) Helix 12 dynamics and thyroid hormone receptor activity: experimental and molecular dynamics studies of Ile280 mutants. *J Molec Biol* 412: 882-893.

Stahlberg J, Divne C, Koivula A, Piens K, Claeysens M *et al* (1996) Activity Studies and Crystal Structures of Catalytically Deficient Mutants of Cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol* 264: 337–349.

- Stahlberg J, Johansson G, Pettersson G (1991) A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. *Biotechnol Biofuels* 9: 286–290.
- Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Am Soc Plant Biol* 150: 621-635.
- Strop P, Mayo SL (2000) Contribution of surface salt bridges to protein stability. *Biochemistry* 39: 1251–1255.
- Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314: 141-151.
- Sulzenbacher G, Mackenzie LF, Wilson KS, Withers SG, Dupont C *et al* (1999) The crystal structure of a 2-fluorocellotriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 Å resolution. *Biochemistry* 38: 4826-4833.
- Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules – application to small water clusters. *J Chem Phys* 76: 537-649.
- Teeri T (1997) Crystalline cellulose degradation: new insights into the function of cellobiohydrolases. *Trends Biotechnol* 15: 160-167.
- Textor LC, Colussi F, Silveira RL, Serpa V; de Mello BL; Muniz JRC; Squina FM; Pereira N, Skaf MS (2013) Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from *Trichoderma harzianum*: deciphering the structural features of cellobiohydrolase catalytic activity. *FEBS* 280: 56-69.
- Thompson MJ, Eisenberg D (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol* 290: 595–604.
- Tomme P, Kwan E, Gilkes NR, Kilburn DG, Warren RAJ (1996) Characterization of CenC, an enzyme from *Cellulomonas fimi* with both endo-and exoglucanase activities. *J Bacteriol* 178: 4216-4223.
- Törrönen A, Harkki A, Rouvinen J (1994) Three-dimensional structure of endo-1,4-xylanase II from *Trichoderma reesei*: two conformational states in the active site. *EMBO J.* 13: 2493-2501.

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comp Chem* 31: 455-461

Tuka K, Zverlov V, Velikodvorskaya G (1992) Synergism between *Clostridium thermocellum* cellulases cloned in *Escherichia coli*. *Appl Biochem Biotechnol* 37: 201–207.

Tsai LC, Shyur LF, Lee SH, Lin SS, Yuan H (2003) Crystal structure of a natural circularly permuted jelly-roll protein: 1,3-1,4-beta-D-glucanase from *Fibrobacter succinogenes*. *J Mol Biol* 330: 607-620.

Tsou C-L (1988) Folding of the nascent peptide chain into a biologically active protein. *Biochem* 27: 1809–1812.

Tsou C-L (1988) The role of active site flexibility in enzyme catalysis. *Biochem* 63: 253–258.

Tsumoto K, Ejima D, Nagase K, Arakawa T (2007) Arginine improves protein elution in hydrophobic interaction chromatography - The cases of human interleukin-6 and activin-A. *J Chromatogr A*, 1154, 81-86.

Tsumoto K, Umetsu M, Kumagai I, Ejima D, Philo JS, Arakawa T (2004) Role of arginine in protein refolding, solubilization, and purification. *Biotechnol Prog* 20: 1301-1308.

Tsumoto K, Ejima D, Kita Y, Arakawa T (2005) Review: Why is arginine effective in suppressing aggregation? *Protein Pept Lett* 12: 613-619.

Tuckerman ME, Martyna GJ, Berne BJ (1990) Molecular dynamics algorithm for condensed systems with multiple time scales. *J Chem Phys* 93: 1287-1291.

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. *Plant Physiol* 153: 895-905.

Vendruscolo M, Paci E (2003) Protein folding: bringing theory and experiment closer together. *Curr Opin Struct Biol* 13: 82-87.

Verlet L (1967) Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys Rev* 159: 98-103.

- Vieille C, Zeikus G (1996) Thermozyms: Identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol* 14: 183-190.
- Vogt G, Woell S, Argos P (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* 269: 631–643.
- Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, MacKerell AD (2007) Additive and Classical Drude Polarizable Force Fields for Linear and Cyclic Ethers *J Chem Theory Comput* 3: 1120-1133.
- Walker L, Wilson D, Irwin D (1992) Fragmentation of cellulose by the major *Thermomonospora fusca* cellulase cellulases, *Trichoderma reesei* CBH1 and their mixtures. *Biotechnol Bioeng* 40: 1019– 1026.
- Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Molec Biol* 103: 227–249.
- Watson BJ, Zhang H, Longmire AG, Moon YH, Hutcheson SW (2009) Processive endoglucanases mediate degradation of cellulose by *Saccharophagus degradans*. *J Bacteriol* 191: 5697-5705.
- Wayman M, Parekh S (1990) “Biotechnology of biomass conversion: fuels and chemicals from renewable sources”. Em: *Biotechnology Series*, cap. 30, pp. 40-71. Open University Press, UK.
- Wefle K, Misselwitz R, Wefle H, Simon O, Politz O, Borris R (1994) Microcalorimetric determination of the thermostability of three hybrid 1,3-1,4-beta-glucanases. *J Biomol Struct Dyn* 11: 1417-1424.
- Wolfenden R, Snider M (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* 34: 938-945.
- White GW, Gianni S, Grossmann JG, Jemth P, Fersht AR, Daggett V (2005) Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J Mol Biol* 350: 757-775.
- Wilmouth RC, Clifton IJ and Neutze R (2000) Recent successes in time-resolved protein crystallography. *Nat Prod Rep Art* 17: 527–533.

Wilson DB (2009) Cellulases and biofuels. *Curr Opin Biotechnol* 20: 295-299.

Wilson DB (2012) Processive and nonprocessive cellulases for biofuel production – lessons from bacterial genomes and structural analysis. *Appl Microbiol Biotechnol* 93: 497-502.

Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. *Quarterly Rev Biophys* 38: 504-410.

Wood TM, Bhat KM (1988) Methods for measuring cellulase activities. *Methods Enzymol* 160: 87-117.

Xiao Z, Gao P, Qu Y, Wang T (2001) Cellulose-binding domain of endoglucanase III from *Trichoderma reesei* disrupting the structure of cellulose. *Biotech Lett* 23: 711-715.

Xiao L, Honig B (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol* 289: 1435–1444.

Yoshida E, Hidaka M, Fushinobu S, Koyanagi T, Minami H, Tamaki H, Kitaoka M, Katayama T, Kumagai H (2010) Structural and biochemical analyses of *Kluyveromyces marxianus* β -glucosidase: an intracellular GH3 enzyme with PA14 domain insertion. *Biochem J* 431: 39-49.

Zhang Y-HP, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: Screening and selection strategies. *Biotech Adv* 24: 452-481.

Zhang, Y-HP, Lynd LR (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol Bioeng* 88: 797-824.

Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB (2004) Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. *Biochem* 43: 9655-9663.

Zhu G, Youlin X, Nicholson LK, Sze KH (2000) Protein dynamics measurements by TROSY-based experiments. *J Magn Reson* 143: 423-426.

Anexo I

Figuras A1-A7: Parâmetros estruturais apresentados na Tabela 3.3 como função do tempo para cada simulação independente

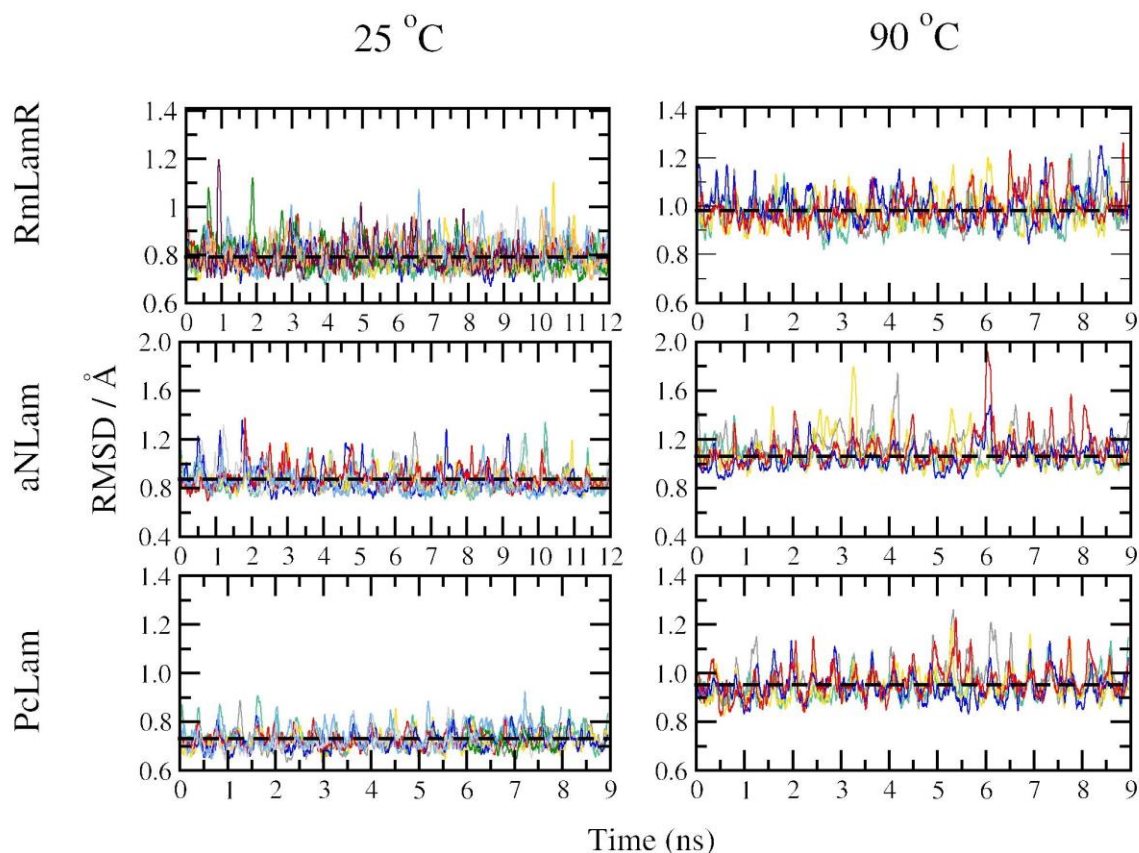


Figura A1. Evolução temporal dos desvios médios quadráticos dos átomos da cadeia principal considerando as estruturas médias calculadas a cada 4 ns de simulação como referência. As cores das linhas correspondem às simulações independentes e as linhas tracejadas correspondem aos valores médios (calculados para os últimos 6 ns de simulação).

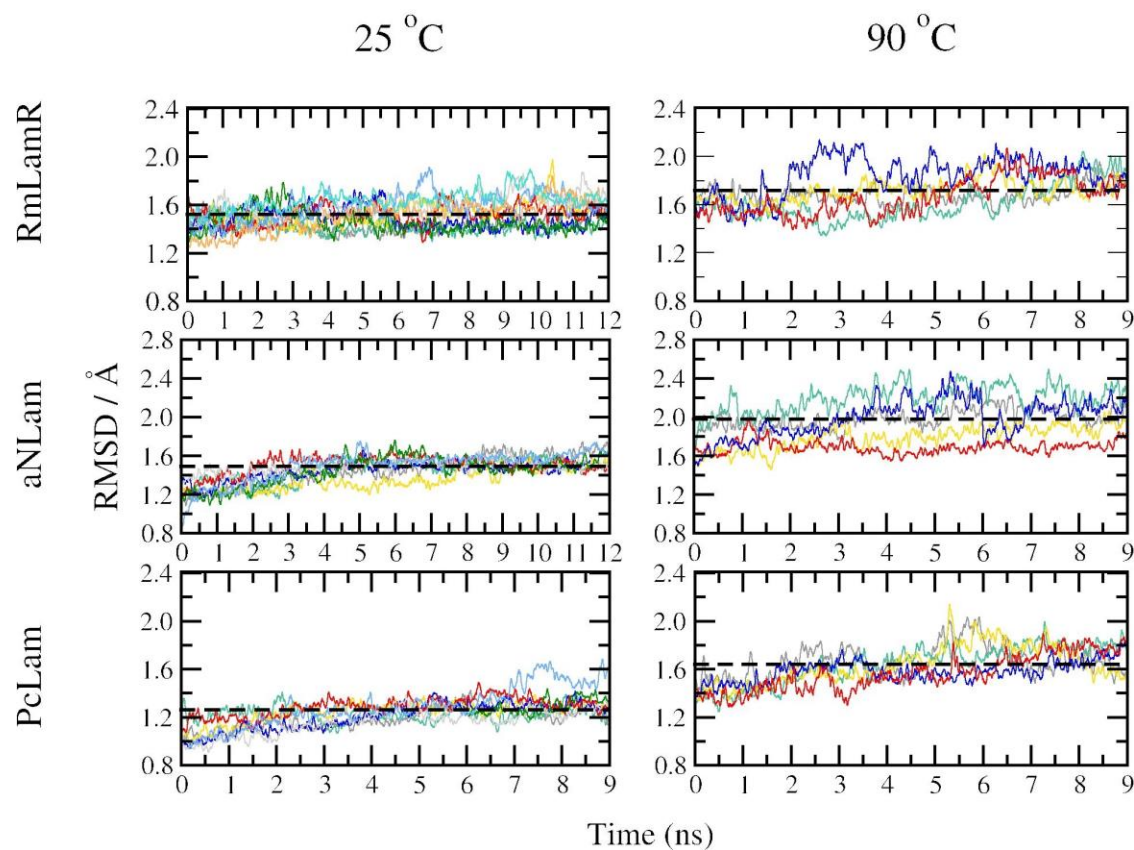


Figura A2. Evolução temporal dos desvios médios quadráticos dos átomos da cadeia principal considerando a estrutura cristalográfica como referência. As cores das linhas correspondem às simulações independentes e as linhas tracejadas correspondem aos valores médios (calculados para os últimos 6 ns de simulação).

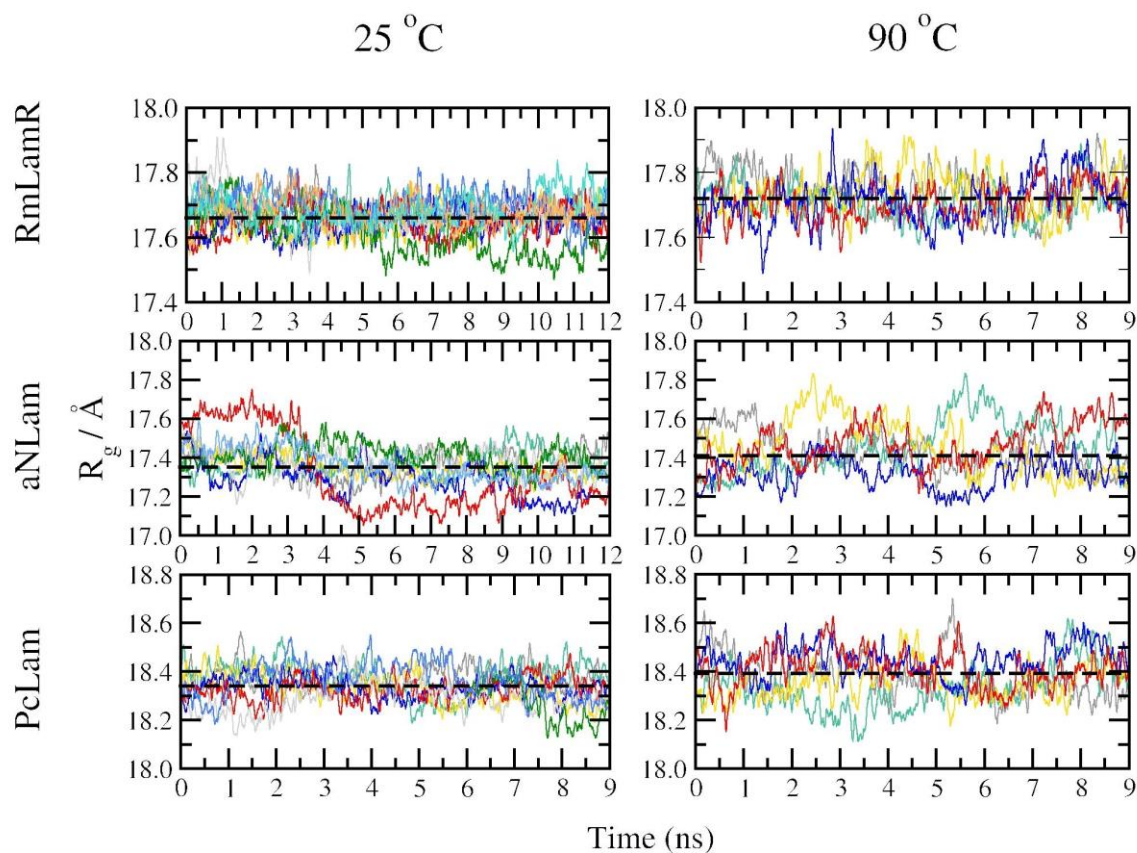


Figura A3. Evolução no tempo do raio de giração de laminarinases. As linhas tracejadas correspondem à médias finais (calculadas para os últimos 6 ns de simulação) sobre as simulações independentes de cada sistema (linhas coloridas). Os valores médios e desvios padrão são listados na Tabela 3.3 do texto principal.

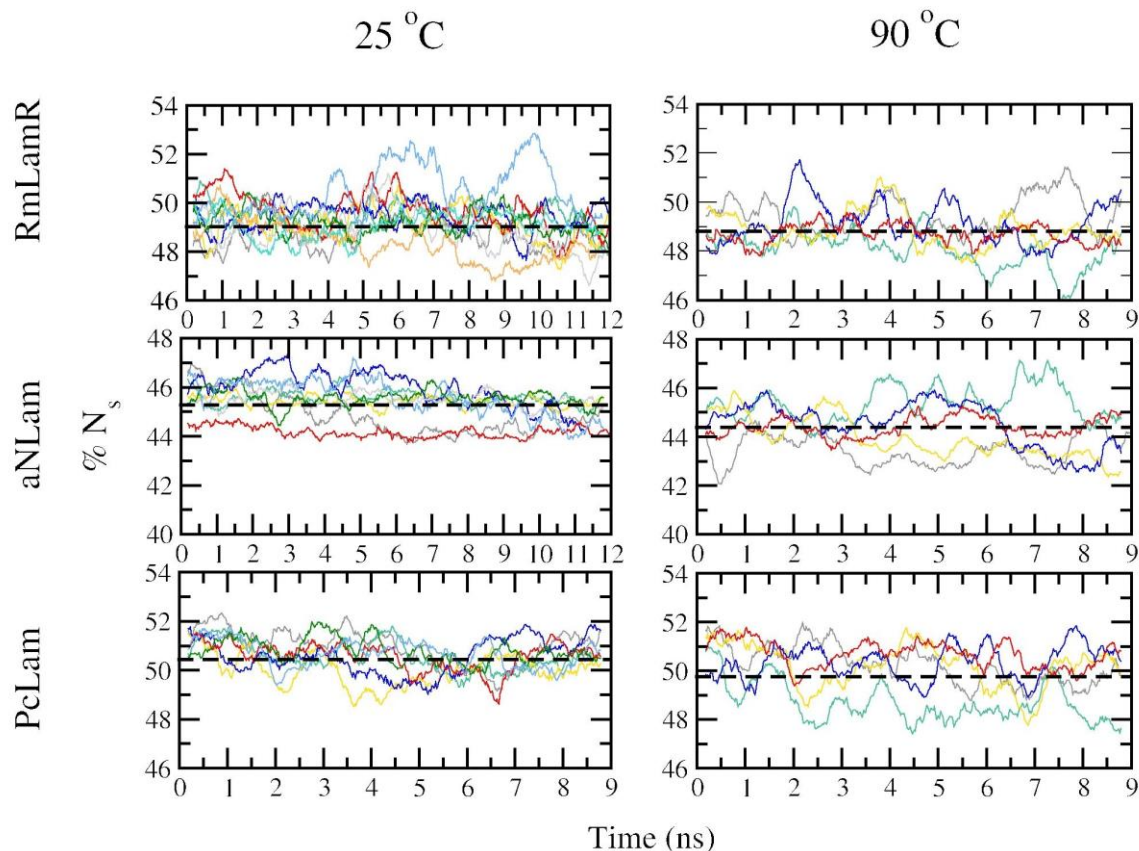


Figura A4. Evolução no tempo da porcentagem de estrutura secundária de laminarinases. As linhas tracejadas correspondem às médias finais (calculados para os últimos 6 ns de simulação) sobre as simulações independentes de cada sistema (linhas coloridas). Os valores médios e desvios padrão estão listados na Tabela 3.3 do texto principal.

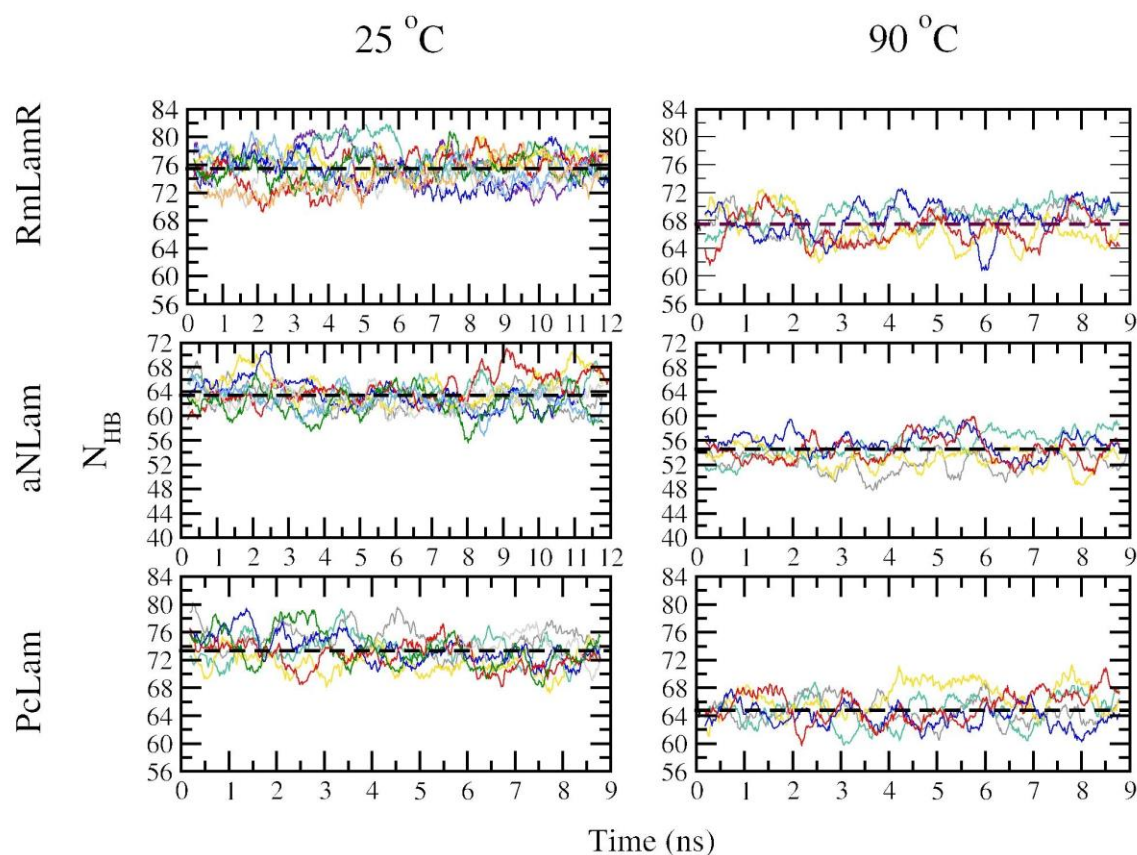


Figura A5. Evolução no tempo do número de ligações de hidrogênio em laminarinases. As linhas tracejadas correspondem aos valores de média final (calculados para os últimos 6 ns de simulação) sobre as simulações independentes para cada sistema (linhas coloridas).

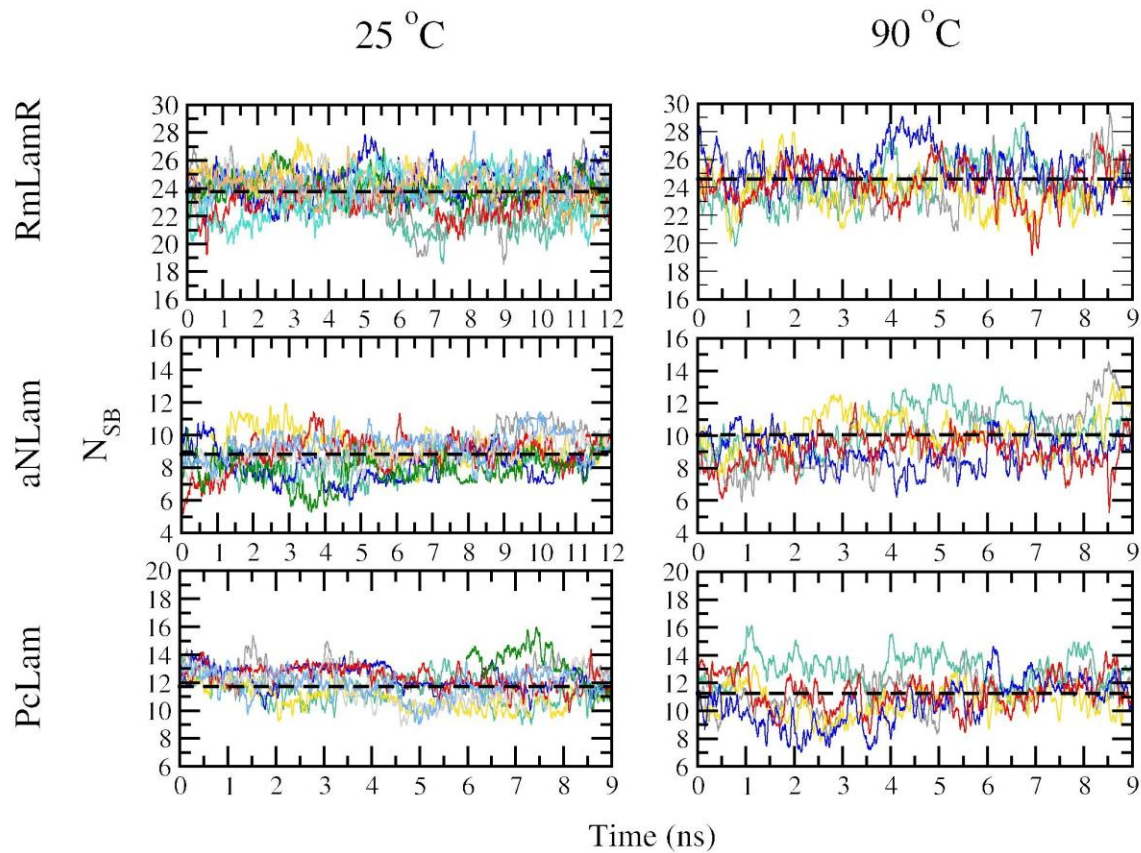


Figura A6. Evolução no tempo no número de pontes salinas nas laminarinases. As linhas tracejadas correspondem aos valores de média final (calculados para os últimos 6 ns de simulação) sobre as simulações independentes para cada sistema (linhas coloridas).

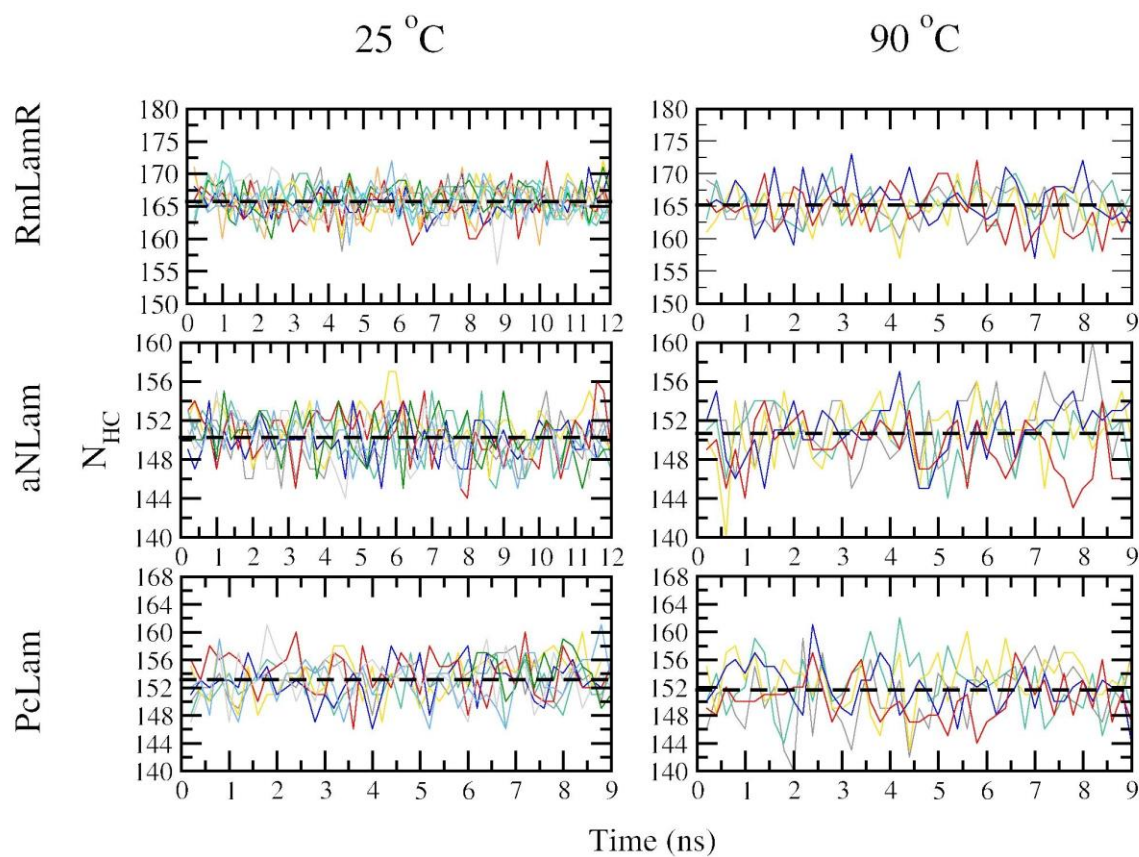


Figura A7. Evolução no tempo no número de contatos hidrofóbico nas laminarinases. As linhas tracejadas correspondem aos valores de média final (calculados para os últimos 6 ns de simulação) sobre as simulações independentes para cada sistema (linhas coloridas).