



IVANA STANKOVIC

**DINÂMICA MOLECULAR DE CELULASES:
ESTUDOS DE RECONHECIMENTO DE SUBSTRATO
E PROPRIEDADES CONFORMACIONAIS**

CAMPINAS

2014



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**

IVANA STANKOVIC

**DINÂMICA MOLECULAR DE CELULASES: ESTUDOS DE
RECONHECIMENTO DE SUBSTRATO E PROPRIEDADES
CONFORMACIONAIS**

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 IVANA STANKOVIC, E
ORIENTADA PELO PROF. DR. MUNIR SALOMÃO SKAF.**

ASSINATURA DO ORIENTADOR

**CAMPINAS
2014**

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

St24d Stankovic, Ivana, 1986-
Dinâmica molecular de celulasas : estudos de reconhecimento de substrato e propriedades conformacionais / Ivana Stankovic. – Campinas, SP : [s.n.], 2013.

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

1. Dinâmica molecular. 2. Celulasas. 3. Seletividade. 4. Propriedades conformacionais. I. Skaf, Munir Salomão. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Molecular dynamics of cellulases : study of substrate recognition and conformational properties

Palavras-chave em inglês:

Molecular dynamics

Cellulases

Selectivity

Conformational properties

Área de concentração: Físico-Química

Titulação: Doutora em Ciências

Banca examinadora:

Munir Salomão Skaf [Orientador]

Hubert Karl Stassen

Fabio Marcio Squina

Rogério Custodio

Nelson Henrique Morgon

Data de defesa: 20-02-2014

Programa de Pós-Graduação: Química

Agradecimentos

Quero aqui expressar de coração os meus agradecimentos às seguintes pessoas e instituições:

- No primeiro lugar ao meu orientador Prof. Dr. Munir Salomão Skaf por ter me aceitado como aluna. Pela orientação e oportunidade de aprender uma área completamente nova para mim e pela grande vontade de providenciar sempre recursos para os projetos de pesquisa.
- À Universidade por ter me aceitado no doutorado, não me separando dos alunos brasileiros em nenhum sentido.
- Ao meu grupo de pesquisa que me ensinou toda a base de dinâmica molecular e também de como ser uma pesquisadora.
- Ao CNPq pelo apoio financeiro durante o doutorado.
- Ao grupo colaborador do Instituto de Física de São Carlos da Universidade de São Paulo.
- Aos professores do Instituto de Química pelas discussões do trabalho.
- Ao Prof. Dr. Edvaldo Sabadini pela minha primeira pesquisa brasileira.
- A toda a comunidade científica de química computacional pela distribuição gratuita de informação e software.

Curriculum Vitae

• Informações Pessoais

Nome: Ivana Stankovic

Data de nascimento: 27/10/1986

• Formação Universitária

Físico-Química, Período: 10/2005 à 04/2010.

Faculdade de Físico-Química, Universidade de Belgrado

Monografia: *A abordagem geral para a derivação da função de partição e funções termodinâmicas com base na teoria de ensemble de Gibbs.*

Orientadora: Profa. Dra. Ljiljana Kolar-Anić.

Período: 02/2010 à 04/2010.

• Doutorado

Dinâmica Molecular de Celulases: Estudos de Reconhecimento de Substrato e Propriedades Conformacionais.

Orientador: Prof. Dr. Munir S. Skaf, Instituto de Química, UNICAMP.

Ingresso: 03/08/2010, Defesa: 20/02/2014, Bolsa CNPq

• Prêmios e Distinções

1. *Best Poster of Congress*, para o trabalho “*Substrate Length Selectivity of the Endoglucanase from Xanthomonas campestris pv. campestris*”, concedido por *XXXVIII Congress of Theoretical Chemists of Latin Expression*, 2012.

• Artigos publicados

1. PRATES, E. T. ; STANKOVIC, I. ; SILVEIRA, R. L. ; LIBERATO, M. V. ; HENRIQUE-SILVA, F. ; NEI PEREIRA, Jr. ; POLIKARPOV, I. ; SKAF, M. S . 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*, v. 8, p. e59069, 2013.
2. STANKOVIC, I. ; MARKOVIC, V. M. ; KOLAR-ANIC, L. Z. System with Variable Energy, Volume and Number of Particles: Evaluation of Partition Function and Thermodynamic Quantities. *Russian Journal of Physical Chemistry*, v. 85, p. 2257-2263, 2011.

• Artigos em preparação

1. Substrate Length Selectivity of the Endoglucanase from *Xanthomonas campestris pv. campestris*: a Molecular Dynamics study
2. Molecular Dynamics of Pro-Thr linker of the Endoglucanase from *Xanthomonas campestris pv. campestris*
3. Application of the Molecular Dynamics Flexible Fitting on the SAXS structure of the Endoglucanase from *Xanthomonas campestris pv. campestris*

• Resumos publicados em anais de congressos

1. STANKOVIC, I. ; PRATES, E. T. ; LIBERATO, M. V. ; POLIKARPOV, I. ; SKAF, M. S. . Oligossaccharide Interactions With Endoglucanases That Lack the Cellulose Binding Domain: Molecular Dynamics Simulations of Endoglucanase III from *Trichoderma Harzianum*. In: São Paulo School of Advanced Science, 2012, Campinas. Agrochemical & Drug Design, 2012. p. 40-40.

- **Formação na pos-graduação**

1. 2011/2S - Exame de Qualificação Geral
Tema: Mecanismos de Hidrolise Enzimática de Polissacarídeos,
Banca: Prof. Dr. Rogério Custódio, Prof. Dr. Fábio Gozzo e Prof. Dr. Nelson Morgon, Conceito: Aprovado.
2. 2011/1S - Física Estatística I,
Ministrado por: Prof. Dr. Alex Antonelli, Conceito: B
3. 2010/2S - Introdução à Química Quântica e Espectroscopia,
Ministrado por: Prof. Dr. Pedro Vazquez, Conceito: A

- **Formação Complementar**

1. Introdução ao FORTRAN 90, Centro Nacional de Processamento de Alto Desempenho - SP, Carga horária: 15h, 2010
2. Introdução ao C, Centro Nacional de Processamento de Alto Desempenho - SP, Carga horária: 15h, 2010

DINÂMICA MOLECULAR DE CELULASES: ESTUDOS DE RECONHECIMENTO DE SUBSTRATO E PROPRIEDADES CONFORMACIONAIS

A degradação enzimática da celulose para a produção de bioetanol é realizada por um conjunto de proteínas denominadas celulasas, as quais são produzidas por vários fungos e bactérias. O mecanismo molecular das interações físicas entre as celulasas e a celulose é pouco conhecido. Para investigar estas interações, bem como as propriedades conformacionais e dinâmicas, nesta Tese usamos simulações por dinâmica molecular (MD). Abordamos duas celulasas: a endoglucanase 3 de *Trichoderma harzianum* (ThEG3) e a endoglucanase de *Xanthomas campestris pv. campestris* (XccEG). O estudo dos mecanismos de reconhecimento de substrato por ThEG3 em que falta o módulo de ligação à celulose (CBM), revela que a ausência de um CBM nesta estrutura é compensada pela presença de resíduos similares aos resíduos típicos de um CBM. O segundo estudo, sobre seletividade da XccEG, explica por que esta enzima catalisa a hidrólise somente de oligossacarídeos de quatro ou mais unidades. Nossas simulações indicam que os quatro subsítios característicos da fenda de ligação ao substrato da XccEG precisam ser simultaneamente estabilizados pelas interações com substrato para promover uma ligação efetiva substrato-enzima. No terceiro estudo, investigamos as propriedades mecânicas do linker entre o domínio catalítico (CCD) e o CBM desta mesma enzima, composto por blocos de Thr e Pro, utilizando a técnica de troca de réplicas (REMD). Nossos resultados sugerem as bases moleculares da maior rigidez deste linker em comparação ao linker de celobiohidrolase 2 de *Trichoderma reesei*. Por fim, adaptamos o método de ajuste flexível por dinâmica molecular (MDFF) para gerar modelos de estrutura terciária a partir de dados de SAXS e o aplicamos para criar um modelo da estrutura intacta CCD-linker-CBM de XccEG.

MOLECULAR DYNAMICS OF CELLULASES: STUDY OF SUBSTRATE RECOGNITION AND CONFORMATIONAL PROPERTIES

The enzymatic degradation of the cellulose for the production of bioethanol is performed by a group of proteins called cellulases which are produced by various fungi and bacteria. The molecular mechanism of the physical interactions between the cellulases and a cellulose is not well understood yet. In this work we use molecular dynamics simulations (MD) to investigate these interactions as well as conformational and dynamic properties. We have studied two cellulases: the endoglucanase 3 from *Trichoderma harzianum* (ThEG3) and the endoglucanase from *Xanthomas campestris pv. campestris* (XccEG). The study of the mechanisms of substrate recognition by ThEG3 which lacks the cellulose binding module (CBM) shows that the absence of a CBM in this structure is compensated by the presence of residues similar to typical CBM residues. The second study, on the selectivity of XccEG, explains why this enzyme catalyses only the hydrolysis of oligosaccharides with four or more units. Our simulations indicate that the four characteristic subsites of the substrate binding cleft of XccEG need to be simultaneously stabilized by the interactions with the substrate to provide effective enzyme-substrate binding. In the third study, we investigated the mechanical properties of the linker between the catalytic domain (CCD) and CBM of the same enzyme, composed of Thr-Pro blocks, using the replica exchange molecular dynamics (REMD). Our results suggest the molecular basis of higher rigidity of this linker in comparison to the linker of cellobiohydrolase 2 from *Trichoderma reesei*. Finally, we have adapted the method of Molecular Dynamics Flexible Fitting (MDFF) to generate tertiary structure models from SAXS data and applied it to create a model of the intact structure CCD-linker-CBM of the XccEG.

Sumário

Lista de Abreviaturas	xxi
Lista de Figuras	xxiii
Lista de Tabelas	xxxii
1 Introdução	1
1.1 Celulases: Função e importância	2
1.2 Mecanismo químico: a hidrólise	2
1.3 Mecanismo físico: reconhecimento de substrato	7
1.4 Estrutura modular	8
1.4.1 Sinergismo	10
1.4.2 Estrutura intacta	11
2 Objetivos	15
3 Metodologia: Dinâmica Molecular	17
3.1 Fundamentos teóricos	18
3.2 Aproximações e vínculos	23
3.3 Análise de trajetória	26
3.4 Amostragem conformacional	27
3.5 Técnica de Troca de Réplicas	28
3.6 Modelagem estrutural por Dinâmica Molecular	33

4	Endoglucanase sem CBM	37
4.1	Introdução	37
4.2	Simulações	41
4.3	Resultados e discussões	43
4.4	Conclusão	47
5	Seletividade da Endoglucanase de <i>Xanthomonas campestris pv. campestris</i>	49
5.1	Introdução	49
5.2	Simulações	52
5.3	Resultados e discussões	56
5.4	Conclusão	65
6	Propriedades mecânicas do <i>linker</i> da XccEG	67
6.1	Introdução	67
6.2	Simulações	69
6.3	Resultados e discussões	73
6.4	Conclusão	86
7	Modelagem com SAXS	89
7.1	Introdução	89
7.2	Preparação do sistema	90
7.3	Resultados e discussões	98
7.4	Conclusão	109
8	Conclusões gerais	111
8.1	Endoglucanase sem CBM	111
8.2	Seletividade da Endoglucanase Xcc	112
8.3	Propriedades mecânicas do <i>linker</i> da XccEG	112

8.4 Modelagem com SAXS 113

Referências Bibliográficas **129**

Lista de Abreviaturas

CBH2	Linker da Celobiohidrolase 2 de <i>Trichoderma reesei</i>
CBM	Módulo de ligação ao carboidrato
CCD	Domínio catalítico
CfCBM	CBM da Endoglucanase C de <i>Cellulomonas fimi</i>
PCA	Análise de Componentes Principais
(PT) ₁₂	Linker da Endoglucanase de <i>Xanthomonas campestris pv. campestris</i>
MD	Dinâmica Molecular
MDFF	Ajuste Flexível por Dinâmica Molecular
REMD	Técnica de Troca de Réplicas
RMSD	Raiz do desvio quadrático médio
SAXS	Espalhamento de raios X de baixo ângulo
ThEG3	Endoglucanase de <i>Trichoderma harzianum</i>
XccEG	Endoglucanase de <i>Xanthomonas campestris pv. campestris</i>

Lista de Figuras

1.1	Ação das celulasas na celulose.	2
1.2	Ligação β -1,4-glicosídica, entre os átomos C1 e O4'.	3
1.3	Duas configurações anoméricas da glicose.	4
1.4	Mecanismos de hidrólise glicosídica.	5
1.5	Três topologias das celulasas relacionadas a interação com substrato: fenda, túnel e bolsão. Sítio ativo está representado em vermelho.	8
1.6	Subsítios de ligação ao substrato de uma glicosidase ou unidades monoméricas de polissacarídeo são denominados por números pos- itivos ou negativos dependendo do lado em relação a ligação glicosídica.	9
1.7	Esquema da estrutura bimodal de celulase. O domínio onde ocorre a reação química (CCD) e o domínio menor que interage fisicamente com substrato (CBM). A terceira parte da molécula é um peptídeo sem estrutura terciária definida, o linker.	9
1.8	Estrutura cristalográfica de uma celulose resolvida separadamente para os domínios CCD e CBM.	12
1.9	Estrutura de SAXS de uma celulase bimodal.	12
3.1	Transição do estado α para o estado β trocando apenas a temper- atura. Temperatura é denominada por T e a energia total por E	29
3.2	Número de trânsito como função de probabilidade média de troca. Os números indicando as retas representam número de réplicas. Adaptado de [52].	33

4.1	Estruturas de ThEG3 e TrEG3 alinhadas. As diferenças entre os resíduos das fendas catalíticas são destacadas.	39
4.2	Estrutura de CBM da endoglucanase C de <i>Celulomonas fimi</i> <i>Cellulomonas fimi</i> (CfCBM), e da celobiohidrolase I de <i>Trichoderma reesei</i> (TrCBM). A celulose é representada em cinza como uma cadeia ou como cristalina. Resíduos aromáticos (em amarelo) são posicionados diferentemente, dependendo do substrato em que agem [62]. . .	41
4.3	Posições iniciais dos substratos inseridos na fenda da ThEG3. A tríade catalítica, Glu117, Glu201 e Asp100, estão representados em vermelho e os resíduos aromáticos na boca da fenda em amarelo. . .	42
4.4	Movimentos dos resíduos aromáticos e substratos estão apresentados por instantes sobrepostos das simulações coloridos em uma escala da vermelho à azul de acordo com o tempo de trajetória. . . .	45
4.5	A: Resíduos em contato hidrofóbico com substratos por mais do que 50 % do tempo de simulação. B: Resíduos que interagem via ligações de hidrogênio com o substrato por, pelo menos, 10 % do tempo de simulação. Círculos de mesmas cores destacam resíduos em posições e papéis correspondentes nas interações proteína-substrato nas duas proteínas, ThEG3 e CfCBM.	46
4.6	As estruturas alinhadas das endoglucanases da família 12. Os nomes se referem aos nomes dos organismos que produzem as enzimas representadas. Resíduos aromáticos representados como barras. A: As endoglucanases da família 12 com CBM. B: As endoglucanases da família 12 sem CBM.	48

5.1	Espectros de Eletroforese capilar de oligossacarídeos marcados com APTS: Hidrólise completa de oligossacarídeo formado por cinco (A) e seis (B) glicoses. Os dois oligossacarídeos estão marcados com APTS pela extremidade não-redutora e os símbolos C6, C5, C4, C3 e C2 indicam o grau de polimerização dos oligômeros de glicose. . . .	52
5.2	Esquema de seletividade de XccEG. A XccEG não cliva cadeias menores que quatro unidades glicosídicas e não cliva uma unidade da extremidade de cadeia.	53
5.3	Docagem da celotetraose. A: XccEG (cor de laranja) alinhada com a endoglucanase E1 de <i>Acidothermus cellulolyticus</i> (azul). O substrato cristalizado em complexo com a endoglucanase E1 de <i>Acidothermus cellulolyticus</i> B: Esquema da posição da celotetraose dentro da fenda da XccEG.	55
5.4	A evolução temporas da distância entre o oxigênio do grupo -OH de Glu182 e o oxigênio glicosídico de substrato para todos os sistemas simulados contendo substrato. Algumas curvas são interrompidas por falta de espaço no eixo <i>y</i>	56
5.5	RMSD para todos os sistemas simulados contendo substrato.	57

5.6	<p>A: A distância entre oxigênio do grupo -OH de Glu182 e oxigênio glicosídico de substrato durante as simulações, para sistemas com celobiose, celotriose-2-1+1, celotriose-1+1+2 e a celotetraose. B: Número de átomos de cada unidade de substrato a 4 Å de cada subsítio, em relação à estrutura inicial docada. Subsítios foram definidos pelo critério de energia de < -5kcal/mol, Figura 5.7. C: RMSD de cada substrato em relação à estrutura inicial. As linhas vermelhas foram obtidas como a média de blocos de 20 pontos da curva que representa as propriedades médias das três simulações. Algumas curvas são interrompidas por falta de espaço no eixo <i>y</i>.</p>	58
5.7	<p>Energia das interações EGXc-celotetraose para cada um dos subsítios. Estão destacados loops compartilhados entre subsítios diferentes. . .</p>	59
5.8	<p>Mobilidade dos loops da fenda. Os substratos são representados em núvens brancas. A- sem substrato, B- celobiose, C- celotriose-2-1+1, D- celotriose-1+1+2, E- celotetraose.</p>	61
5.9	<p>XccEG possui somente 4 subsítios para ligar à celulose.</p>	62
5.10	<p>Energia de interação XccEG-celotetraose para cada um dos subsítios. A primeira docagem (azul) e a docagem alternativa (vermelho), onde a distância entre o átomo de oxigênio do grupo-OH do catalisador ácido Glu182 e o oxigênio glicosídico da celotetraose são 3,61 Å e 3,05 Å, respectivamente.</p>	64

5.11	Enzimas da família 5 que possuem um análogo ao loop5 e que foram cristalizadas com substratos. Os nomes se referem aos nomes dos organismos que produzem as enzimas representadas. Substratos estão representados em barras e as fendas (o loop e a extremidade) em tubos. Todas possuem loop do tipo loop5 da XccEG e fenda de tamanho parecido. A extremidade das fendas de cada enzima foi determinado como conjunto de átomos de enzima a 7Å ao redor do último átomo de substrato do lado redutor, o átomo O1.	65
6.1	A função de autocorrelação da energia potencial normalizada, para ambos os linkers, (PT) ₁₂ e CBH2.	72
6.2	A: Espaço de conformação reduzido a duas componentes principais a 298K. Cada ponto é posição de sistema no espaço de conformação em um instante de simulação. B: Distribuição de parâmetro de ordem R não apresenta mudanças abruptas ao longo do tempo de simulação.	74
6.3	Perfil de energia livre em função da distância ponta-a-ponta R , para os 2 linkers, (PT) ₁₂ (vermelho) e CBH2(preto). As linhas finas representam o ajuste à lei de Hook.	75
6.4	Perfil de energia livre em relação ao raio de giro R_g , para os 2 linkers, (PT) ₁₂ (vermelho) e CBH2(preto).	77
6.5	Ângulos torcionais da cadeia principal, ϕ e ψ	78
6.6	A distribuição de probabilidade dos ângulos torcionais da cadeia principal, vermelho para (PT) ₁₂ , preto para CBH2. A função no eixo x é definida pela equação 6.8.	78
6.7	Estrutura da prolina.	78
6.8	Ligação de hidrogênio (amarelo) entre a cadeia lateral da Thr e os resíduos vizinhos.	79

6.9	Número total de ligações de hidrogênio em função de tempo.	81
6.10	Mobilidade estimada por RMSF por blocos de 100ps para dois modos R_1 e R_2 e a barreira. A ordem da mobilidade crescente é: R_1 , R_2 , barreira. As linhas horizontais representam a média de RMSF por resíduos.	82
6.11	Mudança relativa dos parâmetros do ajustamento da energia livre à lei de Hook, constante de força k e a distância ponta-a-ponta R_0 , com aumento de temperatura.	83
6.12	A distribuição espacial das ligações de hidrogênio em (PT) ₁₂ . Os pontos pretos representam existência das ligações de hidrogênio entre um par de resíduos. À temperatura mais alta, permanecem ligações entre resíduos vizinhos (as ligações perto à diagonal do gráfico). . .	84
6.13	A: Instantes na simulação do linker (PT) ₁₂ sobrepostos, cada 1ns. Alinhamento dos instantes foi feito em relação ao átomo central do linker, o nitrogênio do resíduo Thr15. Representação de cadeia principal em cores de escala de tempo. B: Mapa volumétrico com isovalor de ocupância igual a 4%. À 298K, o linker percorre maior espaço.	86
7.1	Envelope SAXS da XccEG. Representação de perspectiva.	90
7.2	A estrutura cristalográfica do dímero de chiquimato quinase I de <i>Escherichia coli</i> (esquerda) e o seu envelope SAXS (direita).	93
7.3	Valores de χ , porcentagem dos átomos pesados dentro do envelope e RMSD em relação à estrutura cristalográfica para cada uma das combinações de docagem. O eixo x apresenta número de combinação de docagens de dois monômeros, o primeiro dígito é de um monômero e o segundo é de outro monômero. As oito estruturas com o menor valor de χ estão representadas em barras vermelhas e o mínimo global está enquadrado.	94

7.4	Estrutura cristalográfica do dímero do chiquimato quinase I de <i>Escherichia coli</i> (PDBid: 1KAG) e oito estruturas modeladas de menor valor de χ . Os números indicam números de docagens da Tabela 7.1. Os modelos estão alinhados em relação a um dos monômeros (o monômero da parte de baixo do dímero). Bolinha vermelha representa o início da sequência de cada monômero.	95
7.5	Mapa volumétrico de densidades eletrônicas obtidas por SAXS. A: Envelope SAXS representado como isovalor 0,8 do mapa volumétrico. B: Fatia do mapa volumétrico que inclui imagem de esferas equidistantes. C: Fatia do mapa volumétrico com os valores do mapa truncados no 0,8, cada valor de menor que 0,8 foi substituído por 0,8. . . .	97
7.6	Modelos feitos pela docagem dos domínios separadamente. Resíduos catalíticos e resíduos aromáticos do CBM são representados em vermelho. A: O loop5 nao impede a saída do substrato. B: O loop5 impede a saída do substrato. C: Enzimas da família 5 de glicosidases possuem um análogo ao loop5 e substratos que não ultrapassam esta barreira.	100
7.7	Concordância com a curva SAXS experimental durante o tempo de simulação MDFF. As curvas vermelhas representam simulação à 1000K de temperatura em que somente o linker foi livre para movimentos. . . .	101
7.8	Durante o procedimento de modelagem por MDFF, estruturas internas dos domínios são preservadas, que mostra valor baixo de RMSD e a porcentagem dos átomos pesados dentro do envelope cresce até 82%.	102
7.9	Sem vínculos de envelope, o sistema é capaz de manter χ em torno de 3,3 enquanto o RMSD e a porcentagem de átomos pesados dentro do envelope exibiram patamar de valores razoáveis.	103

7.10	A estrutura inicial (à esquerda) e final (à direita) da modelagem da XccEG intacta em solução.	104
7.11	O modelo final. Os resíduos aromáticos estão em vermelho. As linhas vermelhas representam os eixos de CCD e CBM.	104
7.12	A: O modelo final da modelagem da XccEG posicionado em cima da celulose cristalina. B: O modelo de CBM da celobiohidrolase I de <i>Trichoderma reesei</i> [92] inicial (azul) e após simulação (vermelho). CBM desenhado como faixa e os resíduos aromáticos como barras. .	105
7.13	Posição do CBM da endoglucanase de <i>Xantomonas campestris pv. campestris</i> em cima da camada da celulose, inicial (azul) e após 40ns de simulação (vermelho). CBM desenhado como faixa e os resíduos aromáticos como barras.	106
7.14	Diferenças estruturais (RMSF) e dinâmicas (RMSFblocos) entre o modelo após ajuste com SAXS e cristalográfico do CCD da XccEG. A unidade simétrica cristalográfica (direito).	107
7.15	O instante inicial e final de uma simulação MDFF de 1ns, partindo do modelo do CBM paralelo à fenda do CCD. Eixo do CCD e resíduos aromáticos do CBM são representados em vermelho.	108

Lista de Tabelas

1.1	Distância entre resíduos catalíticos para quatro tipos de glicosidases, adaptado de [7].	7
5.1	Trechos das simulações usadas nos cálculos das propriedades médias do sistema, determinadas com as análises de RMSD e a distância entre o oxigênio do grupo -OH de Glu182 e o oxigênio glicosídico de substrato.	55
6.1	Sequência do linker da celobiohidrolase 2 de <i>Trichoderma reesei</i> . Aminoácidos da sequência são representados pelo código de uma letra.	69
6.2	Glicosilação do linker CBH2. Aminoácidos da sequência são representados pelo código de uma letra e unidades manosídicas pelo símbolo Ó.	70
6.3	Parâmetros das curvas de perfil de energia livre em função de distância ponta-a-ponta ajustados à lei de Hook, para os dois linkers, (PT) ₁₂ e CBH2.	76
6.4	Parâmetros das curvas de perfil de energia livre em função de raio de giro ajustadas à lei de Hook, para os dois linkers, (PT) ₁₂ e CBH2.	76

6.5	Frequência de ligações de hidrogênio no linker (PT) ₁₂ em função de distância entre resíduos na sequência. <i>l</i> e <i>p</i> denominam a cadeia lateral e principal, respectivamente. Por simplicidade, as frequências de baixo de 0,01% não estão mostradas na tabela. A maior frequência é entre Thr e primeiro(Pro) ou segundo resíduo vizinho (Thr).	80
6.6	Porcentagem das ligações de hidrogênio para os dois linkers, (PT) ₁₂ e CBH2, no total, e apenas da cadeia lateral da Thr.	80
7.1	Docagens escolhidas pelo valor mínimo de χ . A porcentagem dos átomos pesados dentro do envelope e RMSD em relação à estrutura cristalográfica estão representados para cada docagem.	93

Capítulo 1

Introdução

Com a preocupação de significativa parcela do mundo em desacelerar um possível aquecimento global e reduzir as emissões de gases resultantes da queima de combustíveis fósseis, aumentou consideravelmente o interesse na utilização do bioetanol como uma fonte energética menos poluente e produzida de uma fonte renovável. No Brasil, um aumento significativo e sustentável na produção de bioetanol poderia ser alcançado com o melhor aproveitamento do bagaço da cana-de-açúcar. A celulose, principal componente do bagaço, é um material insolúvel em água, composto de longas cadeias formadas por moléculas de glicose unidas por ligações β -1,4-glicosídicas. A principal dificuldade na utilização eficiente da celulose está na sua grande resistência em sofrer hidrólise [1].

1.1 Celulases: Função e importância

A degradação da celulose em açúcares solúveis pode ser alcançada através da ação de enzimas celulases produzidas por vários fungos e bactérias que catalisam a hidrólise das ligações glicosídicas. Diferentes tipos das glicosidases, endoglucanases, exoglucanases e β -glicosidases, atuam em conjunção no mecanismo de hidrólise. As exoglucanases clivam as extremidades das cadeias polissacarídicas da celulose, liberando glicose ou celobiose, um dissacarídeo de glicose. As endoglucanases hidrolisam aleatoriamente ligações intramoleculares β -1,4- glicosídicas da celulose que estão acessíveis, disponibilizando novas extremidades para o ataque das exoglucanases. E, por fim, as β -glicosidases realizam a hidrólise da celobiose em glicose, Figura 1.1 [2].

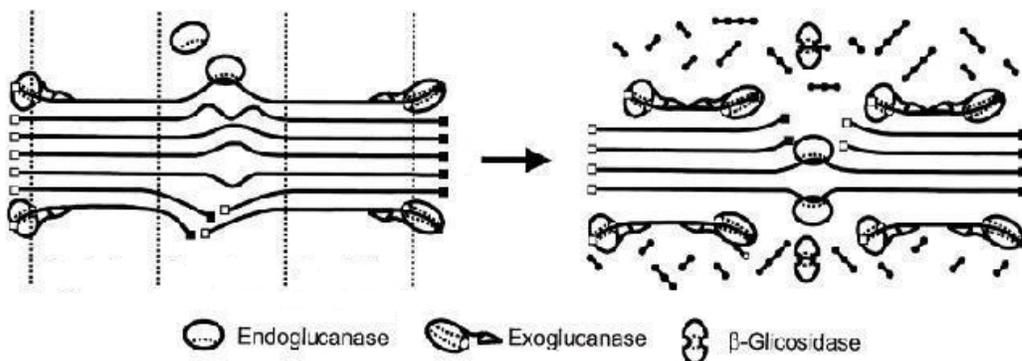


Figura 1.1. Ação das celulases na celulose.

1.2 Mecanismo químico: a hidrólise

As celulases catalisam a hidrólise da celulose ou, mais especificamente, a ligação glicosídica. São proteínas que mostram grande diversidade estrutural, mas apresentam similaridade na função de hidrólise. A celulose é um polissacarídeo constituído de monômeros de glicose unidos pelas ligações β -1,4-glicosídicas, Figura 1.2. A celulose pode apresentar regiões cristalinas e

amorfas [1].

Dependendo da configuração relativa entre a hidroxila no carbono anomérico (C1) e a hidroxila no carbono com mais alta numeração, um monossacarídeo é chamado de α ou β , Figura 1.3. A extremidade de polissacarídeo com carbono anomérico livre é chamada de extremidade redutora porque é capaz de formar um grupo aldeído em carbono anomérico através de abertura de anel e atuar como agente redutor. Enquanto isso, a extremidade oposta é chamada de não-redutora.

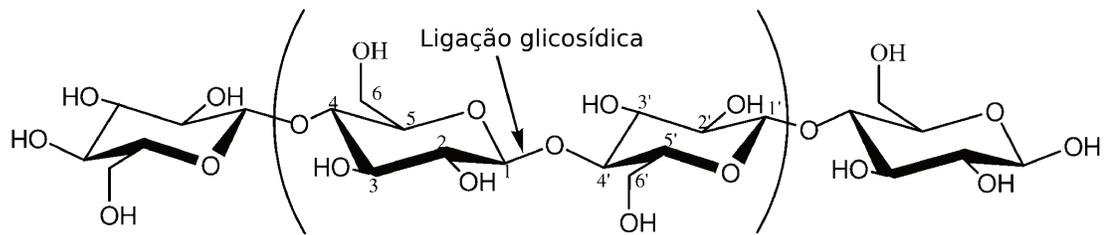


Figura 1.2. Ligação β -1,4-glicosídica, entre os átomos C1 e O4'.

A ligação β -1,4-glicosídica que ocorre em celulose é uma ligação entre a hidroxila do carbono anomérico de uma glicose e a hidroxila na posição 4 de outra glicose, com a remoção de uma molécula de água. Portanto, é uma ligação anídrica e as enzimas que a clivam são hidrolases, as enzimas que requerem água. Mais especificamente, é uma ligação covalente entre o carbono anomérico (C1) de uma unidade de glicose e oxigênio da hidroxila 4 da outra, agora chamado de oxigênio glicosídico, Figura 1.2. A liberdade de movimentação de uma cadeia de celulose é determinada por esta ligação que é mais flexível do que as ligações de anéis.

O mecanismo de hidrólise da ligação glicosídica foi proposta por Koshland em 1953 [8], Figura 1.4. Esta hidrólise ocorre através de dois mecanismos principais que dão origem à retenção ou inversão integral da configuração anomérica do substrato. Ela ocorre via catálise ácida e exige dois resíduos críticos: um

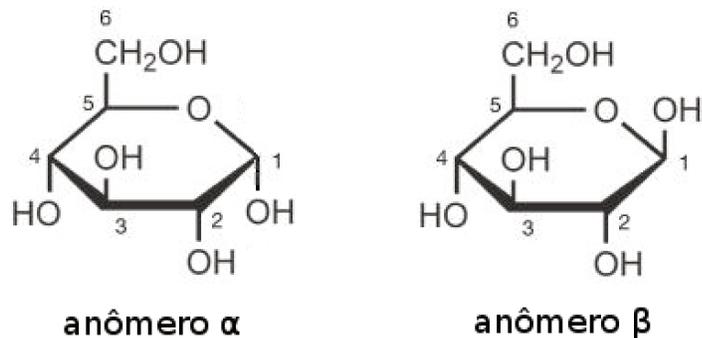


Figura 1.3. Duas configurações anoméricas da glicose.

catalisador ácido e um nucleófilo/base. O catalisador ácido doa hidrogênio ao oxigênio glicosídico e quebra a ligação glicosídica, liberando o produto da hidrólise. O carbono anomérico que formava a ligação glicosídica fica com a carga positiva, portanto sujeito ao ataque nucleofílico. Estes dois resíduos catalíticos junto com um resíduo auxiliador que é protonado e forma ligação de hidrogênio com o nucleófilo mantendo-o na posição certa, são chamados de sítio ativo.

No caso do mecanismo de retenção, o papel do nucleófilo pertence a um resíduo da enzima que se liga covalentemente ao carbono anomérico, assim formando um intermediário covalente com o carboidrato. No segundo passo da reação, este intermediário é sujeito a outro ataque nucleofílico de uma molécula de água da solução que, para ser capaz de atuar como nucleófilo, precisa ser desprotonada. E esta desprotonação agora é realizada por outro resíduo da enzima que tinha papel de catalisador ácido e agora tem papel de catalisador básico, estando desprotonado. Depois da hidrólise, a enzima permanece igual e o substrato, depois de duas inversões da configuração anomérica em seguida, permanece com a configuração inalterada. O primeiro passo, em que o intermediário covalente é formado, chama-se glicosilação pelo evento

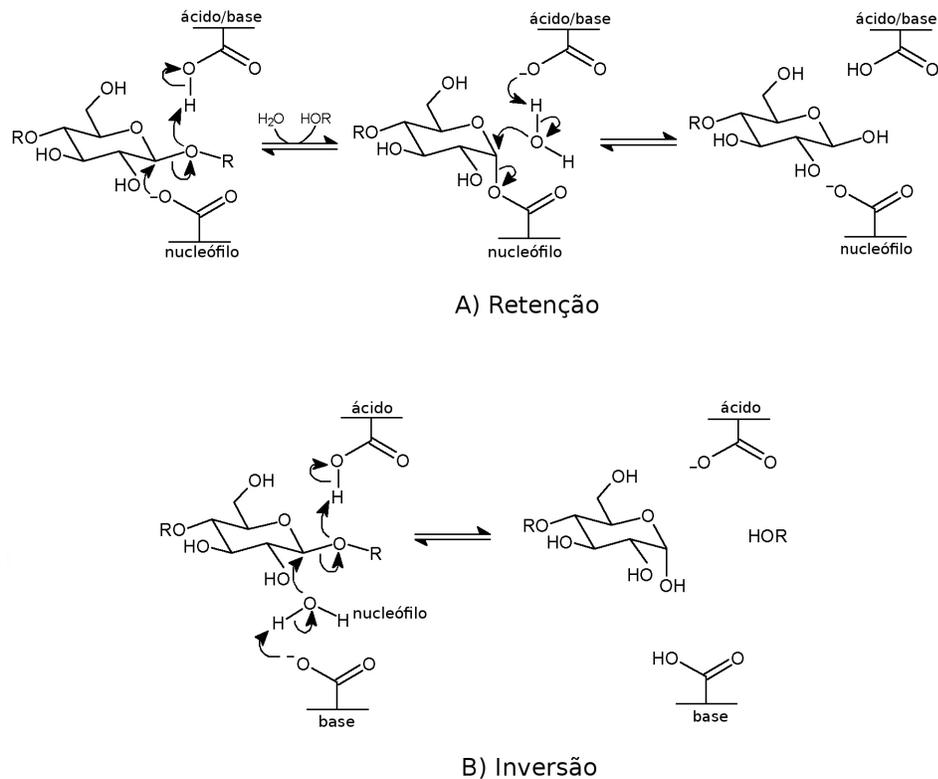


Figura 1.4. Mecanismos de hidrólise glicosídica.

de ligar o glicosídeo (polissacarídeo) à enzima, e o segundo passo chama-se desglicosilação porque este glicosídeo separa-se da enzima.

No caso do mecanismo da inversão, o papel do nucleófilo pertence a uma molécula de água que posiciona-se entre o substrato e a enzima. Ela está desprotonada por um resíduo da enzima, o catalisador básico, e pronta para o ataque nucleofílico ao carbono anomérico do substrato. A posição do catalisador básico impede a formação da configuração inalterada do substrato, mantendo a molécula de água nucleofílica no lugar certo. Depois da hidrólise, a enzima fica com um próton trocado entre os resíduos catalíticos e o substrato com a configuração anomérica invertida. Praticamente, é o mesmo mecanismo de protonação seguida por um ataque nucleofílico com a diferença que, no caso

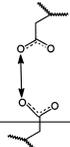
da retenção, este processo acontece duas vezes [4, 5].

Os resíduos catalíticos de enzima, como tem sido observado na maioria dos casos, são aspartato ou glutamato. Ambos possuem o grupo carboxila (-COOH) na cadeia lateral que pode atuar como um doador de próton na catálise ácida e também possuem oxigênios ricos em pares de elétrons livres para ataque nucleofílico.

Os grupos ionizáveis são relativamente raros no interior de proteínas, mas eles são essenciais na catálise. Em microambientes altamente polares ou polarizáveis, a forma carregada de um grupo ionizável irá predominar. Já em microambientes menos polares ou polarizáveis, a forma neutra será favorecida e os valores de pK_a serão deslocados em relação aos valores normais em água. Isto acontece de modo que, para os grupos ácidos, os valores de pK_a tendem a ser superiores aos valores normais de pK_a , e para os grupos básicos, os valores de pK_a tendem a ser inferiores aos valores normais [6].

O mecanismo de hidrólise (retenção ou inversão) depende da estrutura da enzima. As enzimas com o mecanismo de retenção e as enzimas com o mecanismo de inversão diferem em distância entre os resíduos catalíticos. Ambos os mecanismos apresentam a mesma distância entre o catalisador ácido e o oxigênio glicosídico situando-os perto o suficiente para eles formarem a ligação de hidrogênio. Os dois mecanismos começam a divergir quanto à distância entre o nucleófilo/base e o carbono anomérico. Como o mecanismo de inversão deve acomodar espaço suficiente para a molécula de água, o nucleófilo/base está situado mais longe do carbono anomérico. A tabela 1.1 mostra as distâncias médias entre resíduos catalíticos para os quatro tipos de glicosidases medidas em várias enzimas por vários métodos experimentais [7].

Tabela 1.1. Distância entre resíduos catalíticos para quatro tipos de glicosidases, adaptado de [7].

	<p>Inversão</p> <p>$(9.0 \pm 1.0) \text{ \AA}$</p>	<p>Retenção</p> <p>$(5.3 \pm 0.2) \text{ \AA}$</p>
---	---	---

1.3 Mecanismo físico: reconhecimento de substrato

Fora do mecanismo químico da hidrólise [8], pouco sabe-se a respeito dos mecanismos moleculares envolvidos na interação física enzima-substrato que é a primeira condição para a reação química ocorrer. Entender tais mecanismos em nível molecular é necessário para a compreensão de como as celulasas funcionam e, portanto, de extrema importância para o desenvolvimento de enzimas com maior atividade celulolítica. Há diferentes modos de atuação das celulasas. Estes dependem principalmente da forma da enzima e das interações hidrofóbicas e hidrofílicas entre enzimas e substratos e não são dependentes do mecanismo da reação de hidrólise. Para a hidrólise acontecer, não basta a existência do sítio ativo em uma enzima, é necessário também direcionar e manter a cadeia polissacarídica no sítio ativo, e aí as interações não covalentes têm papel importante. As interações não covalentes também podem ser envolvidas em reconhecimento e, às vezes, em bloquear substratos.

As enzimas de estruturas tridimensionais similares vão exibir modos de ligação similares [9]. A topologia geral que determina a atuação das celulasas existe em três classes gerais: 1) fenda, 2) túnel e 3) bolsão [10–13]. Celulasas com as três topologias estão representadas pela superfície molecular na Figura 1.5. As endoglucanases exibem a forma de fenda aberta que permite uma ligação a cadeia celulósica na posição aleatória [10, 12]. O sítio ativo das exoglucanases encontra-se em forma de túnel ou bolsão. O túnel permite que a

enzima libera o produto, permanecendo firmemente ligado à cadeia de polissacarídeo [12]. A forma do bolsão é relacionada aos polissacarídeos ramificados ou cristalinos com poucas extremidades livres. O tamanho do bolsão reflete o número de subsítios, que interagem com o substrato, e o comprimento do produto [10]. Além destas três topologias, existem as topologias intermediárias.

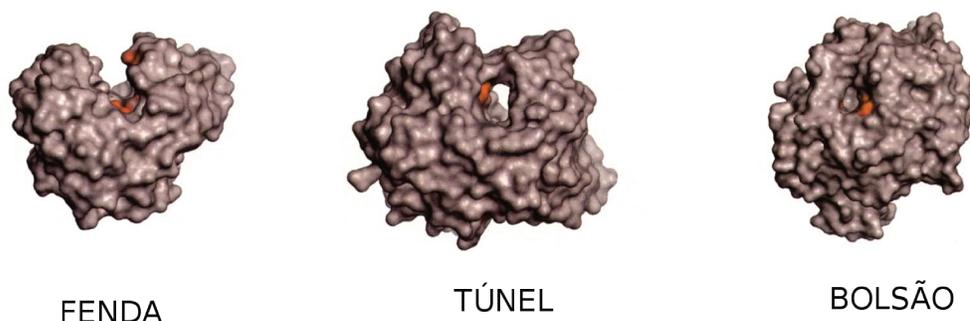


Figura 1.5. Três topologias das celulasas relacionadas a interação com substrato: fenda, túnel e bolsão. Sítio ativo está representado em vermelho.

Dentro da região de ligação à celulose de uma celulase, os subsítios de ligação são frequentemente determinados. O subsítio é a parte da enzima que interage com uma unidade glicosídica de substrato. Os subsítios das enzimas são enumeradas porcentagem inteiros positivos do lado redutor do ponto de clivagem de substrato e por inteiros negativos do lado não-redutor do ponto de clivagem. Da mesma maneira, enumeram-se as unidades do substrato. Os subsítios são mostrados na Figura 1.6. Esta determinação de subsítios poderia ajudar no entendimento de interação enzima-substrato e assim revelar aspectos moleculares de atividade de enzima.

1.4 Estrutura modular

A celulose pode apresentar regiões cristalinas. Como não é solúvel, o acesso ao substrato cristalino representa um desafio para as enzimas. Por isso

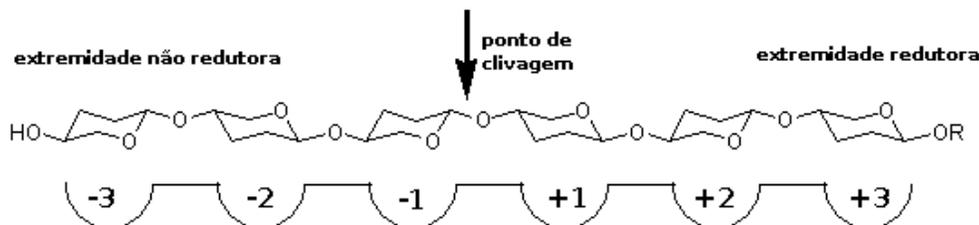


Figura 1.6. Subsítios de ligação ao substrato de uma glicosidase ou unidades monoméricas de polissacarídeo são denominados por números positivos ou negativos dependendo do lado em relação a ligação glicosídica.

as celulasas apresentam uma estrutura complexa e isto traz implicações para a sua função. Uma característica geral da maioria das celulasas é a estrutura modular [5], incluindo módulo catalítico (do inglês, catalytic core domain, CCD) e módulo de ligação ao carboidrato (do inglês, carbohydrate binding module, CBM), Figura 1.7.

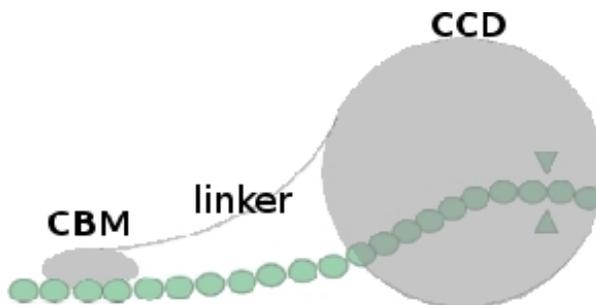


Figura 1.7. Esquema da estrutura bimodal de celulase. O domínio onde ocorre a reação química (CCD) e o domínio menor que interage fisicamente com substrato (CBM). A terceira parte da molécula é um peptídeo sem estrutura terciária definida, o linker.

A hidrólise ocorre no CCD. O CBM tem papel importante na afinidade da enzima pela celulose.

Os dois domínios, CCD e CBM, são conectados por um peptídeo sem estrutura secundária, chamado de linker. A função deste *linker* não é completamente conhecida e nesta tese serão exploradas algumas de suas propriedades físico-químicas.

1.4.1 Sinergismo

Através dos vários experimentos realizados principalmente em celulases, foi observado que a atividade coletiva das endo e exoglucanases é maior do que a das enzimas individuais. Este fenômeno chama-se sinergismo. Foram relatadas quatro formas de sinergismo [14]:

1) Sinergismo endo/exo entre as endoglucanases e as exoglucanases de modo que as endoglucanases providenciam novas extremidades do substrato para as exoglucanases atuarem.

2) Sinergismo exo/exo entre as exoglucanases atuando em extremidades redutora e não-redutora do substrato.

Uma observação de microscopia de força atômica junto com os ensaios de atividade enzimática [15] trata um substrato preparado de mistura de celulose cristalina e a celulose amorfa e as celulases de *Trichoderma reesei*. As celulases estudadas foram as exoglucanases: a celobiohidrolase I que atua da extremidade redutora de substrato e a celobiohidrolase II que atua da extremidade não-redutora; e uma endoglucanase. Foi mostrado que a endoglucanase junto com a celobiohidrolase I fornecem uma superfície da celulose de maior cristalinidade acessível para a celobiohidrolase II removendo as regiões amorfas. Depois, a celobiohidrolase atua liberando cadeias amorfas novas. Além disso, sem a ação prévia da endoglucanase, as exoglucanases foram pouco ativas na celulose. A soma das atividades de todas as três enzimas separadamente foi menor do que a atividade da mistura das enzimas.

3) Sinergismo entre exoglucanases e β -glicosidases que clivam a celobiose, o produto das exoglucanases. As β -glicosidases atuam nos substratos de duas unidades glicosídicas, as celobioses, ou seja, precisam de ação prévia das enzimas que liberam este produto. Também, as β -glicosidases consomem celo-

biose, que por uma vez, são inibidores da atividade enzimática das exoglucanases e endoglucanases.

4) Sinergismo intramolecular entre o domínio catalítico e o domínio de ligação ao substrato.

Sabe-se que o domínio CBM aumenta adsorção de enzima inteira na superfície da celulose e assim aumenta a atividade da enzima [16]. Por exemplo, a endoglucanase 2 e a endoglucanase 5 de *Thermomonospora fusca* expressas sem CBM exibem uma redução significativa tanto na atividade quanto na capacidade de interagir com a celulose cristalina [17]. A atividade foi determinada incubando enzimas com substrato e medindo a quantidade de açúcares redutores que são produto da reação da hidrólise. Para os ensaios da capacidade de interação, uma solução das enzimas foi misturada com a celulose cristalina e incubadas. Depois, a quantidade de enzima livre em solução foi medida com os ensaios de atividade.

1.4.2 Estrutura intacta

Para melhor entender o mecanismo físico da interação enzima-substrato no caso das enzimas modulares, é preciso modelar a estrutura da enzima intacta. Nesta tese uma modelagem será feita combinando a estrutura cristalográfica e estrutura de espalhamento de raios X de baixo ângulo (SAXS) da endoglucanase de *Xanthomonas campestris pv. campestris*.

A estrutura de enzima com resolução atômica é normalmente obtida pela cristalografia de raios X ou pela Ressonância magnética nuclear. Porém, a cristalografia é restringida a descrever conformações estáveis de macromoléculas em cristal. Normalmente os domínios muito flexíveis, como linkers, não são capturados pela cristalografia e existem estruturas dos domínios enzimáticos apenas separadamente. A Figura 1.8 representa a estrutura cristalográfica

dos domínios CCD e CBM da Endoglucanase de *Xanthomonas campestris* pv. *campestris*.



Figura 1.8. Estrutura cristalográfica de uma celulose resolvida separadamente para os domínios CCD e CBM.

O espalhamento de raios X de baixo ângulo (SAXS) oferece informação complementar sobre forma de biomolécula em solução e domínios flexíveis com uma resolução consideravelmente mais baixa. A Figura 1.9 representa estrutura SAXS da Endoglucanase de *Xanthomonas campestris* pv. *campestris*.



Figura 1.9. Estrutura de SAXS de uma celulase bimodal.

Juntas, estas técnicas podem fornecer modelo de multi-escalas para criar uma imagem completa da macromolécula e também descrever mecanismos alostéricos, supramoleculares etc. No banco de dados PDB [18] existem estruturas de domínios separados resolvidos em resolução atômica. O problema é a orientação relativa dos domínios no espaço e a conformação em solução no caso das estruturas cristalográficas. Para combinar resolução atômica com

os dados de SAXS e obter uma estrutura tridimensional para a enzima intacta usamos simulações de dinâmica molecular para gerar um grande conjunto de conformações e calcular espalhamento SAXS teórico que depois será comparado com os dados experimentais. Na literatura existem estudos de modelagem em solução com estrutura de baixa resolução como: programas para o cálculo de curva SAXS teórica baseada em múltiplas conformações [19, 20], Ajuste Flexível por Dinâmica Molecular (MDFF), que realiza o encaixe de estrutura cristalográfica em estrutura de baixa resolução [21], comparação de curva SAXS teórica com a experimental depois de obter conformações por MD [22, 23].

Capítulo 2

Objetivos

O objetivo desta tese foi a aplicação de Dinâmica Molecular para o estudo de proteínas envolvidas na conversão de biomassa celulósica em pequenas moléculas de açúcares. A técnica de MD é empregada para explorar mecanismos de interações não-covalentes que precedem a reação de hidrólise. O foco principal deste trabalho foram quatro problemas diferentes:

1) Endoglucanase sem CBM.

Entender como a endoglucanase 3 de *Trichoderma harzianum* é capaz de reconhecer, ligar-se e hidrolisar as cadeias celulósicas na ausência do domínio de ligação ao carboidrato.

2) Seletividade da Endoglucanase de *Xanthomonas campestris pv. campestris*.

Explicar a seletividade da Endoglucanase de *Xanthomonas campestris pv. campestris* quanto ao comprimento do substrato. Entender por que a Endoglucanase não cliva cadeias de celulose menores que quatro glicoses de comprimento.

3) Propriedades mecânicas do linker da XccEG.

Explorar a sequência específica do linker de XccEG, que consiste em doze blocos de prolina e treonina intercaladas, estudando propriedades mecânicas,

como a elasticidade.

4) Modelagem com SAXS.

Adaptar a técnica de Ajuste Flexível por Dinâmica Molecular às estruturas de SAXS para modelagem de estrutura intacta de biomoléculas em solução. Aplicar à endoglucanase de *Xanthomonas campestris pv. campestris* e comparar com os resultados experimentais.

Capítulo 3

Metodologia: Dinâmica Molecular

Para o estudo de proteínas envolvidas na conversão de biomassa celulósica em pequenas moléculas de açúcares e suas interações não-covalentes com substratos, que são investigados nesta tese, a técnica de dinâmica molecular é empregada.

A dinâmica molecular trata o sistema classicamente em que a estrutura eletrônica dos átomos é tratada implicitamente. Porém, este modelo simplificado consegue capturar bem as propriedades físicas porque o campo de força utilizado é parametrizado por cálculos quânticos e dados experimentais. Não é possível modelar reações químicas ou polarização (no caso dos modelos de átomos de carga constante), mas é possível modelar interações não-covalentes que são a base molecular de reconhecimento biomolecular [24].

A dinâmica molecular foi apresentada pela primeira vez na aplicação de esferas rígidas no ano 1957 [25]. Muitas informações importantes sobre o comportamento dos líquidos simples surgiu a partir destes estudos. A primeira simulação de proteína apareceu em 1977 no trabalho do inibidor de tripsina pancreática bovina [26]. Hoje, na literatura, existem simulações de Dinâmica Molecular de proteínas solvatadas, complexos DNA-proteína, bem como sis-

temas de lipídeos, abordando uma variedade de problemas incluindo a termodinâmica da ligação ao ligante e o enovelamento de proteínas pequenas. Além do aumento da complexidade dos sistemas abordados, o número de técnicas de simulação tem expandido. Nos anos setenta do século XX apareceu uma técnica mista mecânica quântica - mecânica clássica, que é empregada para estudar as reações enzimáticas no contexto da proteína completa [27]. O prêmio Nobel de Química de 2013 foi outorgado a três cientistas que desenvolveram esta técnica [28]. Agora existem muitas técnicas especializadas para problemas específicos, incluindo cálculos de energia livre, modelagem estrutural etc. Técnicas de simulação de dinâmica molecular são largamente utilizadas em processos de refinamento de resultados experimentais, tais como cristalografia de raios X e a determinação da estrutura de Ressonância Magnética Nuclear.

Uma simulação de dinâmica molecular exige cálculos muito extensos porque os processos que ela aborda envolvem um número grande de interações entre átomos e escalas de tempo diferentes. Os métodos atuais e o poder de computação permitem simulações de sistemas de um milhão de átomos (grande o suficiente para a maioria das proteínas de interesse) por centenas de nanossegundos, os limites que são constantemente superados pela evolução contínua de arquitetura de computadores. Um exemplo de alto custo computacional da dinâmica molecular é que uma simulação de um microssegundo de um sistema relativamente pequeno (aproximadamente 25000 átomos) realizado em 24 processadores demora vários meses para completar [30].

3.1 Fundamentos teóricos

Dinâmica molecular é uma técnica de simulação computacional que propaga sistema no tempo com equações baseadas na segunda lei de Newton, para

cada átomo [31], [32]:

$$\vec{F}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} \quad (3.1)$$

É impossível resolver as integrais devido ao número enorme de partículas e interações, então as equações de movimento são resolvidas discretamente, passo a passo. O algoritmo mais usado atualmente nas simulações computacionais das biomoléculas é *Velocity-Verlet* [31]. Este algoritmo é usado por causa de conveniência de calcular posições e velocidades ao mesmo passo de tempo, o que significa que todas as propriedades derivadas daqui serão sincronizadas. As posições de cada átomo em cada passo são calculadas da seguinte maneira:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{1}{2} \frac{\vec{F}(t)}{m} \Delta t^2 \quad (3.2)$$

Onde t é tempo, \vec{F} a força calculada a partir do potencial $V(\vec{r})$ e m a massa da partícula. A força do próximo passo, $\vec{F}(t + \Delta t)$, é obtida a partir da posição do passo $t + \Delta t$:

$$\vec{F}(t + \Delta t) = -\vec{\nabla}V(\vec{r}(t + \Delta t)) \quad (3.3)$$

e em seguida as velocidades são calculadas da seguinte maneira:

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{F}(t) + \vec{F}(t + \Delta t)}{2} \frac{\Delta t}{m} \quad (3.4)$$

Para resolver estas equações, posições e velocidades iniciais são necessárias. As velocidades iniciais de cada átomo são distribuídas aleatoriamente assim que satisfaçam a distribuição de Maxwell-Boltzmann para uma temperatura desejada. As posições iniciais de átomos da proteína e do substrato são determinadas experimentalmente, normalmente pela cristalografia de raios X. A cristalografia de raios X não é suficientemente sensível para detectar raios es-

palhados de átomos de hidrogênio, por isso as posições de hidrogênios são determinadas computacionalmente em relação ao meio onde se encontram. Os estados da protonação de aminoácidos ou outras moléculas são determinados calculando os valores de pK_a a um pH desejado.

O campo de força é a energia potencial que descreve interações entre diferentes átomos através das interações entre átomos não covalentemente ligados, representadas pelos potenciais de Lennard-Jones e de Coulomb, e das interações entre átomos ligados, descritas pelos estiramentos e deformações angulares das ligações covalentes e a torção de diedros como na forma aqui representada [38]:

$$\begin{aligned}
 V(\vec{R}) = & \sum_i \sum_{j \neq i} (4\epsilon_{ij} [(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}) + \\
 & + \sum_{lig} k_b (b - b_0)^2 + \sum_{ang} k_\theta (\theta - \theta_0)^2 + \sum_{died} k_{\phi n} [1 + \cos(n\phi - \delta_n)] + \\
 & + \sum_{died-impr} k_\omega (\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u (u - u_0)^2
 \end{aligned} \tag{3.5}$$

O primeiro termo da primeira soma, que é o potencial de Lennard-Jones, descreve as interações de van der Waals entre dois átomos i e j , separados por uma distância r_{ij} , onde σ é a distância em que o potencial inter partículas é zero e ϵ_{ij} é a profundidade do potencial. O segundo termo da primeira soma descreve as interações eletrostáticas ou Coulômbicas entre dois átomos i e j , com cargas parciais q_i e q_j , respectivamente, separados por uma distância r_{ij} , onde ϵ_0 é a constante de permissividade do vácuo. Na segunda soma da equação, que representa o estiramento de uma ligação química, k_b é a constante de força, b a distância entre dois átomos e b_0 a distância de equilíbrio entre esses átomos covalentemente ligados. Na terceira soma, k_θ representa a constante de força associada à deformação angular, θ é o ângulo normal entre os três átomos e

θ_0 é o ângulo de equilíbrio. Já a quarta soma, de torções diedrais, é descrita por uma soma de cossenos em que k_{ϕ_n} é a constante relacionada à altura da barreira potencial; n é a multiplicidade, que determina o número de pontos mínimos na curva ao rotacionar a ligação completamente; ϕ é o valor do ângulo diedro; δ_n é o valor do ângulo de fase, que determina quando o ângulo torsional passa por um ponto de mínimo. As torções representam a interação entre quatro átomos ligados sucessivamente. A quinta soma, responsável pelas torções impróprias é usada para tratar distorções fora do plano, como as que ocorrem nos hidrogênios aromáticos, por exemplo. k_{ω} é a constante de força e $\omega - \omega_0$ é o ângulo fora do plano. A última soma do potencial interno é o Urey-Bradley, que é um termo cruzado entre dois átomos separados por duas ligações químicas. Esse termo é usado para que os parâmetros melhor reproduzam o espectro vibracional experimental, dado que apenas os termos harmônicos de estiramento e deformação angular de alguns átomos podem não ser suficientes.

Parâmetros como σ , ϵ , k_b etc. da equação 3.5 para uma grande variedade de moléculas e átomos encontram-se determinados, incluindo solventes, íons, polipeptídeos, bases nucleotídicas, lipídeos, sacarídeos, etc. Uma característica geral de campos de força é a transferibilidade de parâmetros. Assim os parâmetros de aminoácidos são usados em todas as proteínas.

O sistema a ser simulado computacionalmente é construído a partir de estrutura cristalográfica das biomoléculas, moléculas de solvente distribuídas aleatoriamente em volta das biomoléculas e normalmente NaCl é adicionado para eletroneutralizar sistema e imitar o ambiente fisiológico. Muitas vezes esta distribuição espacial de partículas resulta em maus contatos entre elas. Antes de realizar simulações, é necessário minimizar a energia de sistema para eliminar a energia repulsiva excessiva entre a estrutura cristalográfica e átomos adicionados na caixa de simulação. A minimização de energia é um procedi-

mento de mudança de posições de partículas no sentido de diminuir a energia potencial total do sistema.

Em seguida, o sistema é equilibrado por MD. Isto é preciso porque as posições dos átomos do solvente de um sistema criado computacionalmente não refletem uma imagem real do sistema. Então, para relaxar o sistema de uma situação artificial, é preciso simular um tempo até equilíbrio de algumas grandezas como temperatura, energia e flutuação de posições. A trajetória depois deste ponto é considerada em análise de termodinâmica ou dinâmica do sistema. Uma pré-equilíbrio é normalmente realizada por partes, fixando partes do sistema de forma que relaxamento de uma parte não atrapalhe o relaxamento de uma outra parte do sistema. Por exemplo, sistema é equilibrado com os átomos da proteína fixos e em seguida com os átomos apenas de cadeia principal de proteína fixos, e finalmente com todas as partículas livres para se movimentar.

Nesta tese será usado o programa NAMD [40] que usa o algoritmo Velocity-Verlet [31] para a resolução das equações de movimento de átomos. O NAMD é usado tanto para a dinâmica quanto para minimização de energia. A minimização é feita através do método dos Gradientes Conjugados [42]. As posições dos hidrogênios são determinadas computacionalmente através da extensão Psfgen do programa VMD [34]. Os estados da protonação de aminoácidos são determinados calculando valores pK a um pH desejado e a aproximação de solvente contínuo usando o servidor H++ [35–37]. O campo de força usado é CHARMM22 para proteínas [38] e CHARMM36 para carboidratos [39]. O campo de força CHARMM foi desenvolvido tanto a partir de dados experimentais quanto de cálculos quânticos [40]. O modelo para água é TIP3P [38].

3.2 Aproximações e vínculos

A dinâmica molecular trata sistemas de muitas partículas que exige um número enorme de potenciais de pares. Como é impossível resolver analiticamente equações de movimento nesta condição e também para diminuir o custo computacional, várias aproximações são empregadas.

Para aproximar os resultados dos cálculos aos dados experimentais, é preciso simular um número representativo de partículas do estado termodinâmico. Na MD seria impossível tratar um número de partículas tão grande com os custos computacionais de hoje. Em vez disso, um número de partículas representativo é usado junto com a condição de contorno periódico de uma caixa de simulação. Quando uma partícula sai da caixa, ela aparece do lado oposto, preservando o número de partículas e o momento total do sistema. A condição periódica diminui os efeitos de superfície de líquido e assim mimetiza o seio da solução.

Interações não-covalentes exigem um custo computacional grande porque teoricamente cada um dos átomos interage com cada um, no espaço infinito. Interações não-covalentes de curto alcance como as de Van der Waals decaem no mínimo com r^{-6} . Nos algoritmos de simulações computacionais, estas interações são normalmente truncadas depois de um raio de corte esférico de magnitude da ordem de 10Å para proteínas [33, 42]. Para evitar a queda abrupta de interação de Van der Waals, são usadas funções suavizantes a partir de r igual ao raio de corte até um r onde o potencial é zero.

As interações eletrostáticas são de maior alcance, decaem com r^{-1} , então na borda da caixa de simulação ainda apresentam um valor significativo. Em vez de corte abrupto de interações, o método baseado em somas de Ewald (Particle Mesh Ewald) [41] é usado. Este contém um artifício matemático usado na resolução das equações. Um número infinito de réplicas da caixa em três di-

mensões é considerado. A soma que representa a energia total eletrostática no conjunto de infinitas réplicas da caixa diverge, ou converge muito lentamente. O artifício matemático consiste em multiplicar o membro da soma por uma função de forma que a soma convirja rapidamente e que a outra parte adicionada para anular convirja rapidamente no espaço recíproco. Espaço recíproco é a transformada de Fourier do espaço real.

Para poupar tempo de computação, o método de múltiplos passos de tempo é usado. Interações que mudam com velocidades diferentes são calculadas com frequências diferentes. Um passo de tempo maior corresponde a interação que muda mais lentamente. Estes passos diferentes têm de ser múltiplos inteiros um do outro. Por exemplo, interações Van der Waals são calculadas a cada passo e as interações eletrostáticas a cada dois passos. O programa NAMD usa o algoritmo RESPA (reference system propagator algorithm) [42, 43]. O algoritmo RESPA é baseado na solução numérica das equações do sistema de referência. A essência do algoritmo é de definir um sistema dinâmico de referência para os movimentos rápidos e derivar equações de movimento para o desvio das coordenadas rápidas a partir das coordenadas do sistema de referência. Estes desvios são acoplados às equações de movimento das coordenadas lentas. O sistema dinâmico rápido é integrado por n pequenos passos de tempo $n\delta t$ mantendo as coordenadas lentas fixas. As condições iniciais para cada passo de tempo grande $n\delta t$ são escolhidas para que este desvio seja sempre mantido pequeno.

Vários vínculos são usados para definir propriedades macroscópicas desejadas e assim imitar o sistema real que se quer estudar. Sempre são usados vínculos de conservação de momento e momento angular para sistemas fechados. Para definir a temperatura, a dinâmica de Langevin é usada, e para definir pressão usa-se método de Langevin-Hoover [33]. Na dinâmica de Langevin, a

temperatura é mantida através da modificação suave das equações do movimento de Newton. O acoplamento com o reservatório térmico é modelado pela adição de dois termos à força que age sobre átomo, uma força dissipativa e uma força flutuante relacionada com a temperatura desejada. Na dinâmica de Langevin-Hoover, as equações incluem o volume como uma variável dinâmica. Da forma similar à dinâmica de Langevin, uma força flutuante age como um barostato [44]. Estas dinâmicas são estocásticas, então cada nova trajetória gerada do mesmo ponto inicial será diferente. Isto é importante porque pode aumentar amostragem do espaço de fase.

Uma outra aproximação empregada para diminuir o custo computacional é fixar ligações covalentes envolvendo átomo de hidrogênio. A vibração desta ligação representa a maior frequência sendo o átomo de menor massa, e assim demanda um passo pequeno em resolução das equações de movimento. As ligações envolvendo átomos de hidrogênio serão consideradas rígidas usando o método SHAKE [45]. A cada passo de simulação, após aplicar o procedimento normal de integração, a posição do átomo de hidrogênio é ajustada para reproduzir o comprimento da ligação desejado. Para os grupos moleculares contendo um átomo de hidrogênio este procedimento é direto, mas no caso de vários hidrogênios existe um procedimento iterativo de ajuste de cada ligação hidrogênio-átomo pesado até que todas as ligações estejam dentro de uma certa tolerância do comprimento desejado. O algoritmo é um pouco diferente para molécula de água. Uma restrição é aplicada também à distância entre os dois átomos de hidrogênio e o procedimento de três ligações é seguido, de modo que o ângulo entre os átomos de hidrogênio é também fixo [46]. Depois de corrigir as posições atômicas, as velocidades calculadas também devem ser ajustadas para seguirem as posições alteradas. Sem a correção de velocidade, a energia cinética é calculada incorretamente. A correção de velocidade é ex-

ecutada pelo método RATTLE [45]. As velocidades calculadas pela integração normal são ajustadas através da remoção do componente paralelo à ligação, sem alterar a velocidade do centro de massa do par de átomos ligados.

3.3 Análise de trajetória

A trajetória gerada pela MD é uma sequência de velocidades e posições de todas as partículas durante o tempo de simulação. A trajetória revela muitas informações dinâmicas, mecânicas e termodinâmicas do sistema.

Partindo das posições é possível derivar todas as propriedades geométricas como: raio de giro, estrutura secundária definida para proteínas específicas, solvatação em volta de uma partícula, ligação de hidrogênio, raiz do deslocamento médio quadrático (RMSD) e energia de interação entre átomos, entre outros.

A energia potencial entre átomos é obtida diretamente a partir das posições atômicas porque depende apenas da posição relativa entre os átomos e do campo de força já definido nas simulações.

A ligação de hidrogênio é considerada existente se a distância entre o doador e o receptor de hidrogênio é no máximo 3Å e o ângulo doador-hidrogênio-receptor no mínimo 160°, como usado para proteínas pelo programa VMD.

O RMSD (raiz do desvio quadrático médio) é uma medida das diferenças estruturais em relação a uma referência específica:

$$RMSD(t) = \sqrt{\frac{\sum_i^N (\vec{r}_i(t) - \vec{r}_{ref,i})^2}{N}} \quad (3.6)$$

Esta grandeza mostra por exemplo, o afastamento estrutural da estrutura cristalográfica. Monitorando desvios da raiz quadrada média (RMSD) em relação a estrutura inicial é possível constatar uma equilibração estrutural do sistema.

RMSF é uma outra função frequentemente usada. É uma média temporal do desvio quadrático para cada resíduo de molécula.

$$RMSF(i) = \sqrt{\frac{\sum_t^{t_{tot}} (\vec{r}_t(i) - \vec{r}_{ref}(i))^2}{t_{tot}}}, \quad (3.7)$$

onde i representa o número de resíduo. Para estimar mobilidade de uma molécula, usa-se o RMSF em relação à estrutura média de blocos pequenos da simulação. Assim, a mobilidade não depende da estrutura inicial, mas sim da flutuação de posição.

3.4 Amostragem conformacional

Como as interações entre todos os átomos de um sistema são calculadas, o tempo de realização de simulação cresce consideravelmente com o número de átomos de sistema. Sistemas de proteínas em água podem conter dezenas de milhares de átomos.

Com o poder computacional atual obtém-se simulação de proteínas de até microssegundo, levando em muitos casos a uma amostragem inadequada dos estados conformacionais [30]. Para observar, por exemplo, o enovelamento de proteína partindo de uma conformação qualquer, é preciso simular o tempo na ordem de segundos. Isto é impossível com recursos atuais [47]. Uma solução é empregar técnicas de amostragem ampliada que de uma maneira levam sistema às regiões pouco acessíveis de espaço de fase. A razão da inacessibilidade de uma região é o perfil de energia livre. Para sistemas complexos como biomoléculas, ele é bastante rugoso incluindo barreiras energéticas altas entre regiões diferentes. Em princípio, a amostragem ampliada é conseguida por duas maneiras gerais: por mudança do campo de força ou por aumento de temperatura [48, 49]. O aumento de temperatura providencia uma energia ex-

tra para o sistema superar barreiras entre as regiões diferentes. A mudança do campo de força é feita através de atenuação de potencial de diedro ou atenuação das interações de van der Waals. Estas mudanças implicam em amostragem não-Boltzmanniana e requerem um reponderamento dos estados da trajetória resultante para obter médias canônicas das propriedades moleculares. Neste sentido, a técnica de aumento de temperatura é mais direta, mas requer cuidado na escolha da magnitude do aumento de temperatura, pois a estrutura e a função de algumas moléculas são altamente sensíveis ao aumento de temperatura.

3.5 Técnica de Troca de Réplicas

Uma maneira de estimular amostragem de espaço de fase é aumentar a temperatura do sistema. O aumento da temperatura aumenta o módulo de velocidade média das partículas do sistema e assim os átomos ganham energia para superar barreiras potenciais entre conformações diferentes, e conseguem amostrar mais o espaço de fase. Uma das técnicas que aumenta a amostragem por aumento de temperatura é Troca de Réplicas [50]. Réplicas são cópias do mesmo sistema a temperaturas diferentes que são simuladas simultaneamente, trocando temperaturas entre elas. Por exemplo, uma réplica a temperatura baixa pode encontrar-se em poço potencial e uma outra réplica a temperatura alta pode encontrar-se em um pico de energia livre. Elas trocam temperaturas e a réplica de poço ganha calor que a ajuda a sair do poço. A nova temperatura da réplica é determinada por reescalamiento de velocidade porque a energia cinética média $3/2Nk_B T$ é vinculada a temperatura. As tentativas de trocas entre réplicas são feitas periodicamente. Trocas são aceitas ou não de acordo com critério de Metropolis [50]. Este critério tem origem nas simulações de Monte Carlo onde energia de uma configuração gerada é comparada com a

energia da configuração anterior e a probabilidade de Boltzmann é comparada com um número aleatório gerado cada passo.

Observemos agora uma troca de temperaturas entre duas réplicas. Para isso combinaremos fórmula para a distribuição canônica (para as simulações executadas no ensemble NVT) e a fórmula para balanço detalhado para assegurar uma troca em equilíbrio. O estado total inicial de duas réplicas com as



Figura 3.1. Transição do estado α para o estado β trocando apenas a temperatura. Temperatura é denominada por T e a energia total por E .

temperaturas ainda não trocadas será chamado de estado α , e o estado de duas réplicas com as temperaturas trocadas será chamado de estado β , Figura 3.1. A probabilidade canônica do estado α e β , respectivamente, é:

$$P_{\alpha} = \frac{e^{\frac{-E_1}{k_B T_1}} e^{\frac{-E_2}{k_B T_2}}}{\sum_i e^{\frac{-E_i}{k_B T_1}} \sum_i e^{\frac{-E_i}{k_B T_2}}}, \quad P_{\beta} = \frac{e^{\frac{-E_1}{k_B T_2}} e^{\frac{-E_2}{k_B T_1}}}{\sum_i e^{\frac{-E_i}{k_B T_2}} \sum_i e^{\frac{-E_i}{k_B T_1}}} \quad (3.8)$$

onde a energia total é denominada E , a temperatura é denominada T , k_B é a constante de Boltzmann e os índices 1 e 2 se referem às duas réplicas. E a equação do balanço detalhado entre estes dois estados seria:

$$P_{\alpha} P_{\alpha \rightarrow \beta} = P_{\beta} P_{\beta \rightarrow \alpha} \quad (3.9)$$

Combinando estas três expressões, obtemos a fração:

$$\frac{P_{\alpha \rightarrow \beta}}{P_{\beta \rightarrow \alpha}} = \exp\left(-\frac{(E_2 - E_1)(T_2 - T_1)}{k_B T_1 T_2}\right) = e^{-\Delta} \quad (3.10)$$

Aqui é usado a convenção de que se uma destas duas probabilidades, $P_{\alpha \rightarrow \beta}$ e $P_{\beta \rightarrow \alpha}$, for maior que outra, ela será igual a 1:

$$se \quad P_{\alpha \rightarrow \beta} \geq P_{\beta \rightarrow \alpha}, \quad P_{\alpha \rightarrow \beta} = 1 \quad (3.11)$$

$$se \quad P_{\beta \rightarrow \alpha} > P_{\alpha \rightarrow \beta}, \quad P_{\beta \rightarrow \alpha} = 1 \quad (3.12)$$

Então, a expressão para a probabilidade de troca entre as réplicas 1 e 2, ou seja, a probabilidade de transição do estado α para o estado β será:

$$P_{\alpha \rightarrow \beta} = \begin{cases} 1, & se \quad e^{-\Delta} \geq 1, \\ e^{-\Delta}, & se \quad e^{-\Delta} < 1 \end{cases} \quad (3.13)$$

ou simplesmente

$$P_{\alpha \rightarrow \beta} = \min(1, e^{-\Delta}), \quad onde \quad \Delta = (E_2 - E_1)(T_2 - T_1)/k_B T_1 T_2 \quad (3.14)$$

Assim simulações andam trocando temperaturas entre pares de réplicas vizinhas. Mas antes de simular, é preciso ajustar quatro parâmetros para simulações:

1) Faixa de temperaturas. A faixa deveria abranger a temperatura de interesse até uma temperatura arbitrariamente alta que ajude na superação de barreiras energéticas. Aqui é preciso tomar cuidado na escolha da temperatura máxima porque as biomoléculas altamente estruturadas vão perder estrutura tridimensional e portanto a função com aumento de temperatura.

2) Número de réplicas. O número de réplicas determina quão densamente as temperaturas são distribuídas, dada a probabilidade média de troca, $\langle P_{\alpha \rightarrow \beta} \rangle$

e escopo de temperaturas desejados. Como as temperaturas são normalmente distribuídas exponencialmente para assegurar constância de probabilidade de trocas entre diferentes pares de réplicas vizinhas, a fração entre temperaturas vizinhas é constante.

$$\frac{T_2}{T_1} = c \quad (3.15)$$

A probabilidade média de troca pode ser estimada usando a expressão para energia total média de sistema proteína-água [51]:

$$\langle E \rangle = N_{atom} k_B T, \quad (3.16)$$

onde N_{atom} é número de átomos de sistema. A probabilidade média de troca seria a probabilidade de troca em função da energia média, em vez da energia atual:

$$\langle P_{\alpha \rightarrow \beta} \rangle = \exp\left(-\frac{(\langle E_2 \rangle - \langle E_1 \rangle)(T_2 - T_1)}{k_B T_1 T_2}\right) \quad (3.17)$$

$$\langle P_{\alpha \rightarrow \beta} \rangle = \exp\left(-\frac{(N_{atom} k_B T_2 - N_{atom} k_B T_1)(T_2 - T_1)}{k_B T_1 T_2}\right) \quad (3.18)$$

$$\langle P_{\alpha \rightarrow \beta} \rangle = \exp\left(-\frac{N_{atom}(c-1)^2}{c}\right) \quad (3.19)$$

Apos calcular a constante c , ela pode ser substituída na expressão que conecta faixa de temperaturas com o número de réplicas

$$\frac{T_{max}}{T_{min}} = c^{N_{rep}-1}, \quad (3.20)$$

e assim o número de réplicas pode ser calculado.

3) Tempo entre tentativas de troca. O alvo da técnica Troca de Réplicas é aumentar amostragem de equilíbrio. Por isso é importante o sistema se relaxar

entre mudanças de temperatura. O tempo entre tentativas de troca é considerado o tempo de correlação da energia potencial de sistema [52].

4) Tempo total de simulação. Normalmente, a temperatura mais baixa do conjunto de réplicas é a temperatura de interesse. As réplicas de alta temperatura tem como alvo ampliar a amostragem das partes menos acessíveis do espaço de conformações. No entanto, para estas réplicas de temperaturas altas serem aproveitadas para a amostragem ampliada, elas precisam trocar temperatura com a réplica de temperatura baixa. Então, um número grande de réplicas de temperaturas altas não faz sentido na ausência de uma boa mistura das réplicas porque aumenta o custo computacional sem que a temperatura máxima seja aproveitada para ajudar a saída do poço energético. Abraham e Gready [52] estimaram o número mínimo de tentativas de troca necessário para satisfazer eficiência de mistura das réplicas. A figura 3.2 mostra este número (denominado número de trânsito) em função de número de réplicas e a probabilidade média de troca. Dada uma estimativa de taxa de trocas média entre as réplicas, eles calcularam o número mínimo de tentativas de troca de tal maneira que providencie a probabilidade de 95% de pelo menos uma réplica tenha visitado a temperatura máxima antes de visitar a temperatura mínima. Abraham e Gready recomendaram que o número de tentativas de troca seja de duas ordens de grandeza maior que o número de trânsito.

Para analisar o sistema a uma temperatura de interesse, é preciso juntar passos de simulações a essa temperatura. Está trajetória concatenada não representa dinâmica do sistema. Representa um conjunto ampliado de conformações a essa temperatura. Nesta tese, iremos aplicar a técnica de Troca de Réplicas para estudar as propriedades elásticas e conformacionais do linker que conecta o CCD ao CBM.

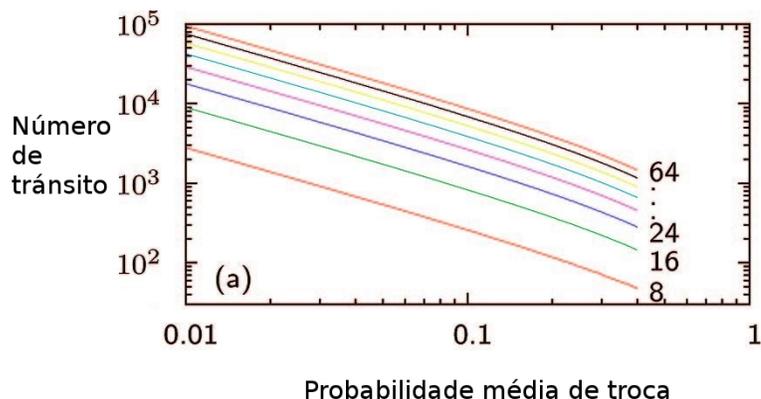


Figura 3.2. Número de trânsito como função de probabilidade média de troca. Os números indicando as retas representam número de réplicas. Adaptado de [52].

3.6 Modelagem estrutural por Dinâmica Molecular

Como foi mencionado na Introdução, Seção 1.4.2, a dinâmica molecular é usada para combinar uma estrutura cristalográfica com uma estrutura de SAXS porque ela gera conjunto de conformações assim como uma estrutura de SAXS representa média temporal das várias conformações de uma molécula. Isto porque se sabe que as biomoléculas são flexíveis e representam uma dinâmica no tempo.

Neste trabalho será feita uma combinação de métodos da literatura, MDFF [21] e o cálculo de curva SAXS teórica obtida ao longo de uma trajetória inteira que é depois comparada com a curva experimental. A técnica de MDFF, que foi originalmente desenvolvida para estruturas de crio-eletromicroscopia, será primeiro adaptada ao uso de estrutura de SAXS. Esta metodologia será aplicada à endoglucanase de *Xanthomonas campestris pv. campestris*. Aqui será representado brevemente o princípio da técnica de MDFF e o cálculo do espalhamento SAXS teórico. Os detalhes do procedimento inteiro da modelagem da endoglucanase de *Xanthomonas campestris pv. campestris* estão no Capítulo 7.

O princípio da técnica de MDFF. Uma estrutura tridimensional de SAXS é obtida pela modelagem computacional a partir da curva unidimensional de SAXS. A estrutura tridimensional representa apenas a superfície de molécula sem estrutura interna. Esta superfície é denominada frequentemente envelope de SAXS. O alvo da modelagem é sobrepor uma estrutura atômica com o envelope e mudá-la minimamente assim que ela entrasse no espaço de envelope. Isto é conseguido através das simulações. A técnica MDFF foi desenvolvida para combinar a estrutura atômica com a estrutura de crio-eletromicroscopia [21], mas aqui será adaptada para estruturas de SAXS. Durante a simulação, cada átomo de biomolécula modelada é restringido ao espaço do envelope usando uma força proporcional ao gradiente da densidade eletrônica do mapa volumétrico criado a partir do envelope. Praticamente um termo a mais é adicionado ao campo de força para o átomo i [21]:

$$U_i = U_{CFi} + m_i \xi \frac{\phi(r)}{\phi_{max}}, \quad (3.21)$$

onde U_{CFi} é a energia potencial que representa um campo de força convencional, ξ é fator de escala arbitrário e $\phi(r)$ é a densidade eletrônica de um ponto r no mapa volumétrico. A força que cada átomo sente é ponderada por sua massa (m_i). Ao mesmo tempo, cada átomo é submetido ao campo de força convencional que define a relação com outros átomos. A modelagem é resultado de compromisso entre estas duas forças.

Curva SAXS teórica. A curva SAXS é obtida experimentalmente a partir da intensidade dos raios espalhados e do ângulo de observação. Nesta tese, o programa CRY SOL foi usado para calcular a curva SAXS teórica a partir do conjunto de conformações gerado nas simulações. O espalhamento de uma molécula é calculado como média espacial de espalhamento proveniente de

todos os átomos [53]:

$$I(q) = \langle f(q) \rangle_{\Omega}, \quad (3.22)$$

onde $q = 4\pi \sin\theta/\lambda$, $f(q)$ é a intensidade de raios espalhados da molécula e $I(q)$ é a média espacial da intensidade de raios espalhados. O símbolo $\langle \rangle_{\Omega}$ denota a média de todas as orientações da molécula no espaço recíproco. A intensidade $f(q)$ é calculada da seguinte forma:

$$f(q) = A(\vec{q})^* A(\vec{q}), \quad (3.23)$$

$$A(q) = \sum_n b_n(q) \exp(i\vec{q} \cdot \vec{r}_n), \quad (3.24)$$

onde $b_n(q)$ é o fator de estrutura, n é o índice de átomo e \vec{r}_n é a posição no espaço real do átomo n . O fator de estrutura para o espalhamento de raios X encontra-se determinado para cada tipo de átomo [54].

Após a determinação da curva teórica, ela é comparada com a curva experimental. Para avaliar a concordância entre as duas curvas, aqui será usada a grandeza χ [53]:

$$\chi = \sqrt{\frac{1}{N_P} \sum_q \left(\frac{cI_{teor,q} - I_{exp,q}}{\sigma_q} \right)^2}, \quad (3.25)$$

onde $\sigma(q)$ é erro da curva experimental, N_P é número de pontos experimentais e c é fator de normalização da intensidade teórica I_{teor} porque a intensidade de espalhamento tem valor relativo, não absoluto.

Capítulo 4

Endoglucanase sem CBM

4.1 Introdução

Uma característica geral da maioria das celulasas é a estrutura modular, incluindo módulo catalítico CCD e módulo de ligação ao carboidrato CBM. A função e o mecanismo de CBM são ainda muito pouco compreendidos. Sabe-se que os CBMs têm papel importante na afinidade da enzima pela celulose [16]. Por exemplo, a endoglucanase 2 e a endoglucanase 5 de *Thermomonospora fusca* expressa sem CBM exibe uma redução significativa tanto na atividade quanto na capacidade de interagir com a celulose cristalina [17]. A atividade foi determinada incubando enzimas com o substrato e medindo a quantidade de açúcares redutores que são produto da reação da hidrólise. Para os ensaios da capacidade de interação, solução das enzimas foi misturada com a celulose cristalina e incubadas. Depois, a quantidade de enzima livre em solução foi medida com os ensaios de atividade.

Acredita-se que CBM reconhece e liga-se à superfície celulósica, separa uma cadeia da celulose disponibilizando para o CCD. Simulações de MD realizadas somente com CBM da celobiohidrolase I do fungo *Trichoderma reesei* mostram que o CBM movimenta-se livremente na superfície da celulose e muda de conformação quando encontra a extremidade redutora da celulose. Três resíduos de tirosina formam interações hidrofóbicas com a superfície da

celulose possibilitando movimentos em ambas as direções, perpendicular e paralela, mas ao mesmo tempo sendo firmemente ligados à superfície. Depois de uma torção da extremidade redutora do substrato, o CBM interage através da quarta tirosina com a extremidade redutora. Supostamente, a mudança conformacional do substrato é necessária para reconhecimento pela enzima [55].

A vantagem de CBM verificou-se diminuída pela redução da quantidade de água no sistema hidrolítico [56]. Em uma solução mais concentrada, a probabilidade de contato enzima-substrato aumenta. Em baixas concentrações de substrato (1% w/w), CBMs foram mais importantes no desempenho catalítico das celobiohidrolases I e II de *Trichoderma reesei*, em comparação com as endoglucanases Cel5A e Cel7B. O aumento da concentração do substrato, mantendo a proporção de enzima para substrato, aumenta a adsorção de Cel7A de *Trichoderma reesei*, independente da presença do CBM. No caso da concentração do substrato de 20% (w/w), o desempenho hidrolítico de celulases sem CBM alcançou o das celulases com CBM. Este fenômeno é provavelmente devido à adsorção não-produtiva de enzimas.

Há celulases que não contêm CBM e não se conhece como elas são capazes de reconhecer, ligar-se e hidrolisar as cadeias celulósicas na ausência de um CBM. Neste trabalho feito em colaboração com outros alunos do grupo, Érica Teixeira Prates e Rodrigo Leandro Silveira, foi investigada uma dessas enzimas, a endoglucanase 3 de *Trichoderma harzianum* [PDB id: 4H7M, [57]] (aqui chamada de ThEG3) e aqui estão mostrados apenas os resultados mais importantes e da maior participação da autora da tese. O alvo é entender como a ThEG3 interage com o substrato celulósico, que parte da enzima é responsável para a função de ligação à celulose que é normalmente função do domínio CBM.

A enzima ThEG3 [58] apresenta alta similaridade sequencial (83 %) em

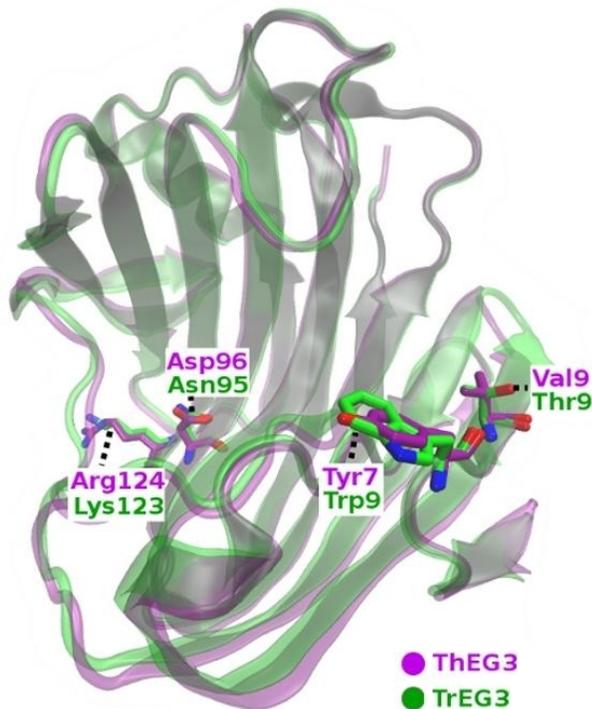


Figura 4.1. Estruturas de ThEG3 e TrEG3 alinhadas. As diferenças entre os resíduos das fendas catalíticas são destacadas.

relação a endoglucanase 3 de *Trichoderma reesei* (TrEG3) [PDB id: 1H8V, [59]] e apresenta o mesmo tipo da estrutura terciária, o sanduíche de folhas β , e por isso foi escolhida como referência de dados experimentais faltantes para a ThEG3, Figura 4.1. São poucas as diferenças entre as estruturas da ThEG3 e TrEG3. Os resíduos da fenda catalítica que diferem são destacados na Figura 4.1. A ThEG3 possui um resíduo carregado, Asp96, na posição de um resíduo polar na TrEG3, Asn95. Os outros resíduos que diferem pertencem ao mesmo tipo de aminoácidos. O alinhamento estrutural das duas estruturas foi feito por minimização do valor de RMSD entre as duas estruturas, usando o programa VMD [34].

Para comparar as interações com a celulose, foi escolhida a endoglucanase C de *Cellulomonas fimi* (CfCBM) [PDBid:1GU3, [60]], Figura 4.2, porque esta

age principalmente em oligossacarídeos e celulose amorfa, similarmente à TrEG3, que interage muito fracamente com substratos cristalinos [61]. A estrutura do CfCBM apresenta um arranjo de resíduos aromáticos similar ao da ThEG3, e esta é a outra razão para usar CfCBM nos estudos comparativos da interação com substrato.

Sabe-se que os resíduos aromáticos são importantes para reconhecimento e ligação ao substrato celulósico. Cadeias laterais de tirosina, triptofano ou fenilalanina emparelham-se aos anéis glicosídicos de carboidrato [62]. Um estudo de simulações de MD e análise de modos normais foram usados para examinar o papel funcional de resíduos aromáticos no túnel da celobiohidrolase 2 de *Trichoderma reesei* [63]. A enzima contém três triptofanos que interagem com substrato e cada um deles foi mutado computacionalmente para alanina. Os cálculos de energia livre de interação com a celulose foram feitos para a enzima nativa e para três proteínas mutantes. A mudança de energia livre foi positiva nos casos das três proteínas mutantes em relação à proteína nativa. Então, os resíduos aromáticos associados à aquisição de substrato e estabilização de produto têm maior influência sobre a energia livre de ligação e, portanto, são propensos a ter um grande impacto sobre a atividade da enzima. Os resultados consistentes foram obtidos com ensaios experimentais de NMR que mostraram proximidade espacial maior entre substrato e resíduos aromáticos do que entre substrato e outro tipo de resíduo no caso da proteína mutada [64].

A estrutura do CfCBM apresenta três resíduos aromáticos característicos posicionados na boca da fenda, o que é uma característica geral das celulasas que clivam preferencialmente a celulose amorfa ou oligossacarídeos solúveis enquanto existe outro tipo de CBM com resíduos aromáticos alinhados em paralelo que age preferencialmente na celulose cristalina [62]. Um exemplo é o CBM da celobiohidrolase I de *Trichoderma reesei* (TrCBM), Figura 4.2. A

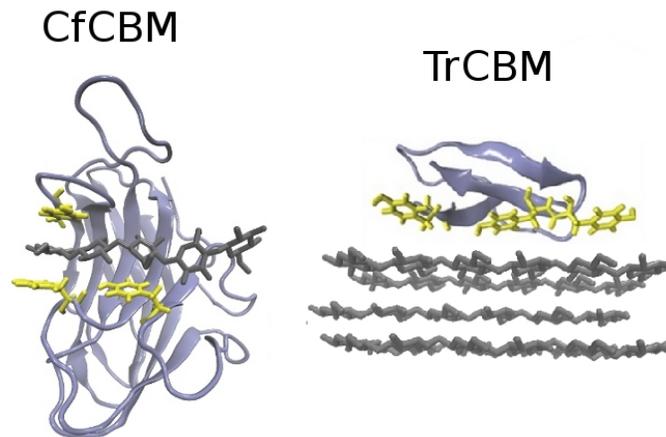


Figura 4.2. Estrutura de CBM da endoglucanase C de *Cellulomonas fimi* (*CfCBM*), e da celobiohidrolase I de *Trichoderma reesei* (*TrCBM*). A celulose é representada em cinza como uma cadeia ou como cristalina. Resíduos aromáticos (em amarelo) são posicionados diferentemente, dependendo do substrato em que agem [62].

disposição dos resíduos na região da ligação facilita acesso ao substrato. A forma de fenda interage com uma cadeia da celulose, enquanto a região plana contendo resíduos aromáticos alinhados interage mais facilmente com a superfície da celulose cristalina.

O arranjo dos resíduos aromáticos da ThEG3 está representada em amarelo na Figura 4.3. Da mesma forma que os resíduos aromáticos na estrutura do *CfCBM*, estes resíduos formam a entrada da fenda.

4.2 Simulações

As duas proteínas, ThEG3 e *CfCBM*, foram simuladas com substratos em sua fenda por MD, a fim de identificar similaridades nas interações com substrato. A *CfCBM* foi simulada com a celopentaose, o oligossacarídeo com o qual foi cristalizada. E como a ThEG3 não foi cristalizada em complexo com substrato, três substratos diferentes foram docados na sua fenda. A celopen-

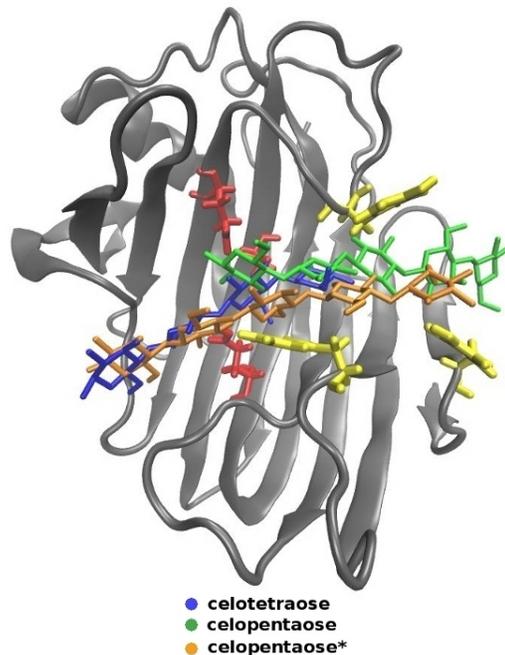


Figura 4.3. Posições iniciais dos substratos inseridos na fenda da ThEG3. A tríade catalítica, Glu117, Glu201 e Asp100, estão representados em vermelho e os resíduos aromáticos na boca da fenda em amarelo.

taose da estrutura cristalográfica da CfCBM foi inserida na fenda da ThEG3 pelo alinhamento estrutural das proteínas; a celotetraose da estrutura cristalográfica da endoglucanase Cel12A de *Thermotoga maritima* [65] também foi inserida pelo alinhamento das proteínas e o terceiro substrato foi construído pelo prolongamento da celotetraose em uma unidade glicosídica (o substrato denominado celopentaose*), Figura 4.3. A ThEG3 também foi simulada sem substrato na sua fenda ao fim de eliminar os possíveis mecanismos não relacionados à interação com substrato. A tríade catalítica, o nucleófilo Glu117, o catalisador ácido Glu201 e o resíduo auxiliador Asp100, estão representados em vermelho. (Ver o apêndice com os códigos de aminoácidos, 8.4.) Os três resíduos aromáticos na boca da fenda são Tyr7, Trp23 e Tyr112 e estão representados em amarelo.

Os três substratos apresentam diferenças de posição em relação a estes dois grupos de resíduos, o sítio ativo e os resíduos aromáticos na boca da fenda. A celotetraose é posicionada perto do sítio ativo, mas não representa uma boa posição para interagir com os resíduos aromáticos, ao contrário da celopentaose que alcança os resíduos aromáticos, mas não o sítio ativo. A celopentaose* é bem posicionada para interagir com os dois grupos de resíduos característicos.

Caixas de simulação de lado de 80 Å foram construídas adicionando 53 Cl⁻ e 50 Na⁺ ions, deixando o sistema eletricamente neutro a uma concentração salina de aproximadamente 0,16M. O estado de protonação da tríade catalítica foi determinada de acordo com a sua função na hidrólise [5]. O catalisador ácido, Glu201, e o resíduo auxiliador, Asp100, foram considerados protonados, enquanto o nucleófilo, Glu117, foi mantido na sua forma carregada. O estado de protonação dos outros resíduos foi determinado pelo procedimento do programa *H++* [36, 37] considerando o valor de *pH* 7. Todos os sistemas foram simulados 3 vezes 40ns à temperatura de 298K e pressão de 1atm. O raio de corte foi de 12 Å com a função suavizante de 10Å para as interações de van de Waals.

4.3 Resultados e discussões

Posição e comprimento do substrato. Após as simulações realizadas com os quatro sistemas da ThEG3, a saber: sem substrato, com a celotetraose, com a celopentaose e com a celopentaose*, o comportamento dos substratos durante o tempo de simulação foi monitorado. A Figura 4.4 mostra a posição dos três substratos em relação aos resíduos catalíticos e anéis aromáticos da ThEG3, Tyr7, Trp23 e Tyr112, como sobreposição de instantes da simulação. Na ausência do substrato (apo-ThEG3) os três anéis aromáticos são livres para

executarem movimentos de larga amplitude.

No complexo com a celotetraose (ThEG3-tt), a celotetraose está adequadamente posicionada em relação à tríade catalítica (representada em amarelo), mas não é longa o suficiente para interagir simultaneamente com os três resíduos aromáticos. Neste sistema, dois dos três resíduos aromáticos que não interagem com a celotetraose executam movimentos de amplitude maior ainda do que no sistema apo-ThEG3.

No complexo com a celopentaose (ThEG3-pt), por outro lado, o substrato interage com os três resíduos aromáticos na entrada da fenda, mas a ligação glicosídica mais próxima ainda está longe do sítio ativo. Na simulação deste complexo, os resíduos aromáticos permanecem ligados ao substrato, mesmo quando o substrato afasta-se da tríade catalítica e quase sai da fenda da enzima. Assim a importância dos anéis aromáticos na ligação ao substrato está confirmada.

No complexo ThEG3-pt*, entretanto, a posição inicial e comprimento do substrato permite tanto a interação com os resíduos aromáticos quanto com os resíduos catalíticos. Percebemos que, para uma ligação enzima-substrato efetiva, a posição e o comprimento do substrato têm papel determinante. O comprimento mínimo de substrato é de cinco unidades glicosídicas. Tomando os dados experimentais da atividade catalítica da TrEG3 como os dados faltantes da ThEG3, este comportamento concorda com a atividade hidrolítica de TrEG3 - ela age apenas em oligossacarídeos maiores que quatro unidades glicosídicas [61]. Deste modo, as simulações sugerem que a eficiência catalítica da ThEG3 depende de um tamanho mínimo do substrato que permita uma posição adequada na fenda.

Função do CBM. Análises comparativas de estrutura e de simulações de MD dos complexos ThEG3-pt* e CfCBM-celopentaose foram realizadas na procura

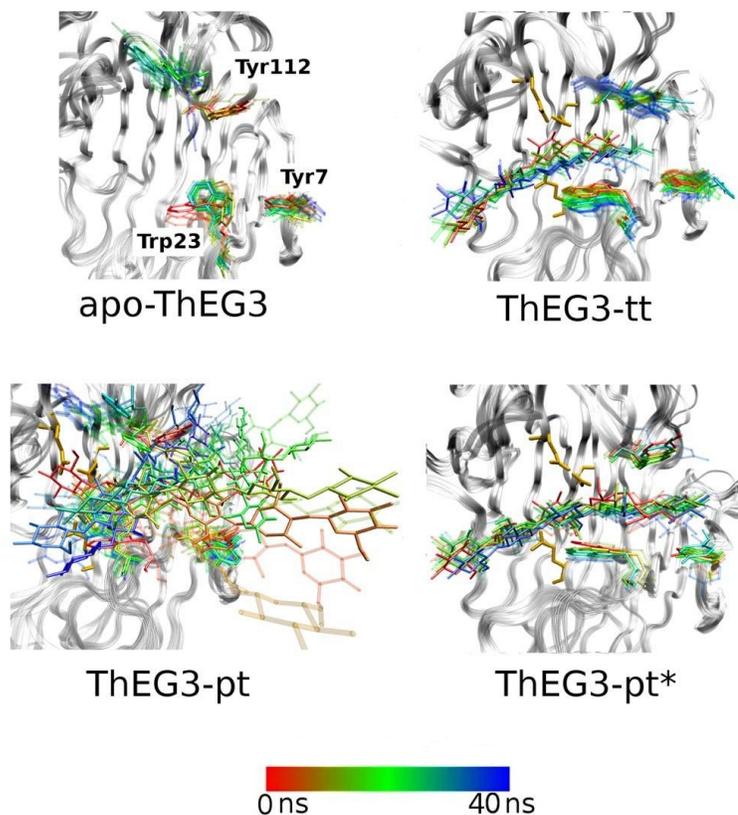


Figura 4.4. Movimentos dos resíduos aromáticos e substratos estão apresentados por instantes sobrepostos das simulações coloridos em uma escala da vermelho à azul de acordo com o tempo de trajetória.

de similaridades na interação com substrato e possível explicação da atividade da ThEG3 sem o domínio CBM. Tanto a inspeção da estrutura, quanto as análises das simulações revelam que vários resíduos da ThEG3 e CfCBM exibem papéis similares na interação com o substrato. Uma análise foi feita das interações não-covalentes com substrato, considerando contatos hidrofóbicos e ligações de hidrogênio. A Figura 4.5 mostra os resíduos das duas proteínas que interagem hidrofobicamente (A) e hidrofílicamente (B) com o substrato com frequência no tempo acima de 50% e 10%, respectivamente. O critério de ligação de hidrogênio é geométrico, como descrito na Seção 3.3 da Metodologia e os contatos hidrofóbicos foram considerados existentes se qualquer átomo

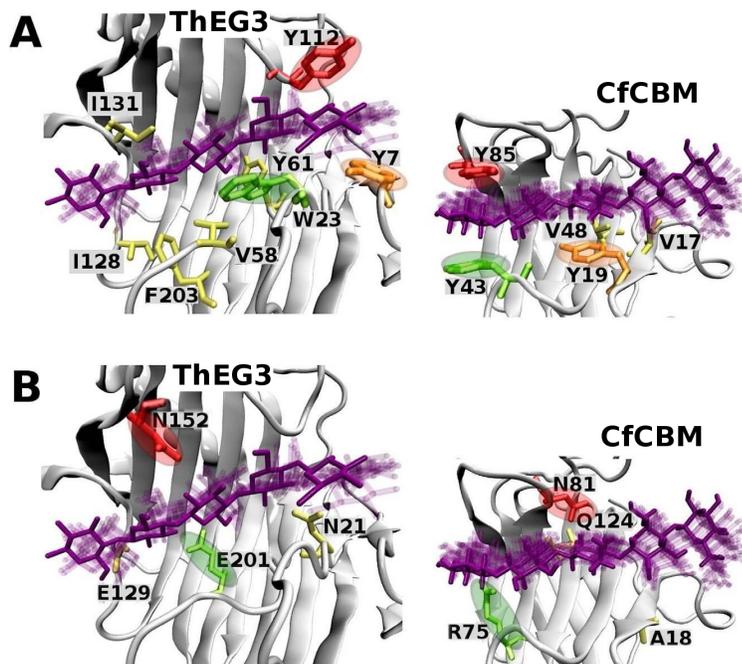


Figura 4.5. **A:** Resíduos em contato hidrofóbico com substratos por mais do que 50 % do tempo de simulação. **B:** Resíduos que interagem via ligações de hidrogênio com o substrato por, pelo menos, 10 % do tempo de simulação. Círculos de mesmas cores destacam resíduos em posições e papéis correspondentes nas interações proteína-substrato nas duas proteínas, ThEG3 e CfCBM.

pesado de um resíduo de proteína encontra-se a distância de 5Å de qualquer átomo pesado do substrato. Os três resíduos aromáticos da ThEG3 identificados como importantes, Y7, W23, Y112, mimetizam os resíduos Y19, Y43 e Y85 em termos de posição e função na CfCBM e os resíduos N152 e E201 mimetizam N81 e R75.

Outras enzimas sem CBM. Este mecanismo de interação com substrato na ausência do domínio CBM pode ser o mesmo para outras glicosidases da estrutura terciária similar à estrutura da ThEG3, as glicosidases da família 12, [66]. As estruturas de todas as glicosidases desta família que têm sua estrutura resolvida com a resolução atômica, [18, 66] foram comparadas. Estas estruturas são separadas em dois grupos: que contém CBM e que não contém CBM.

Trata-se de endoglucanases em todos os casos, exceto a xilo-endoglucanase XEG de *Aspergillus aculeatus*, Figura 4.6B, que atua em substratos ramificados como xiloglucano, além na celulose. Após alinhamento estrutural, foi procurado um padrão de resíduos aromáticos na mesma posição dos resíduos aromáticos que formam contatos com substrato com a maior frequência na ThEG3, os resíduos Y112, Y23 e Y7. A Figura 4.6 mostra que as endoglucanases contendo CBM podem ou não conter também os três resíduos aromáticos característicos, mas no caso das endoglucanases sem CBM, elas sempre contêm os três resíduos aromáticos. A exceção é a xilo-endoglucanase XEG de *Aspergillus aculeatus*, Figura 4.6B, que contém dois dos três resíduos aromáticos característicos. Ela interage com substratos ramificados [67] e pode apresentar um mecanismo de ligação diferente. Para ligar aos substratos ramificados deve possuir mais sítios de ligação, e assim conter vários outros resíduos de ligação ao substrato além dos dois resíduos aromáticos.

4.4 Conclusão

A endoglucanase 3 de *Trichoderma harzianum* (ThEG3) apresenta uma estrutura sem domínio de ligação ao carboidrato. Sabendo que este domínio é importante na hidrólise da celulose, o alvo deste trabalho foi entender que parte da estrutura é responsável pela função de um CBM. As simulações de dinâmica molecular foram realizadas com ThEG3 e o CBM da endoglucanase C de *Cellulomonas fimi* (CfCBM).

Os modelos da estrutura da ThEG3 complexada com a celotetraose e a celopentaose foram gerados usando estruturas cristalográficas de proteínas disponíveis de enovelamento semelhante ligadas a estes substratos. As simulações sugerem que a ausência de um CBM na estrutura da ThEG3 é compensada pela presença de resíduos similares a resíduos de um CBM que reconhecem o

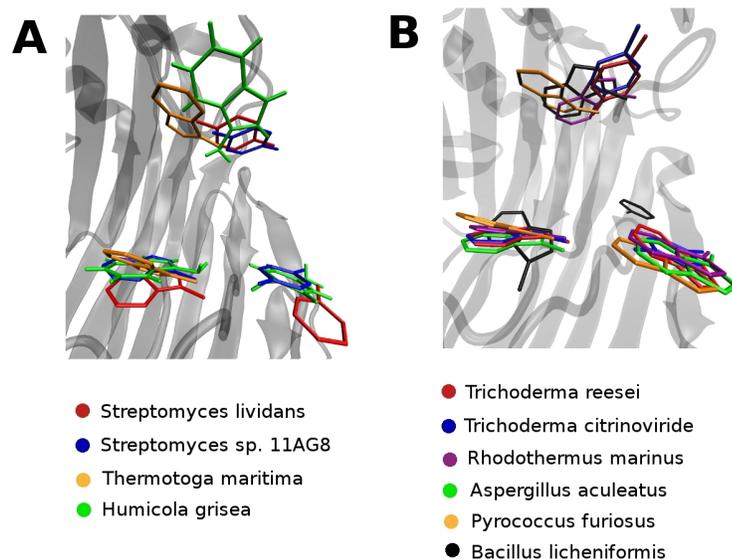


Figura 4.6. As estruturas alinhadas das endoglucanases da família 12. Os nomes se referem aos nomes dos organismos que produzem as enzimas representadas. Resíduos aromáticos representados como barras. **A:** As endoglucanases da família 12 com CBM. **B:** As endoglucanases da família 12 sem CBM.

substrato.

Além da explicação da atividade da ThEG3 na ausência de CBM, aqui foi explicado porque esta enzima age em substratos de no mínimo cinco unidades glicosídicas. A tríade catalítica é espacialmente longe dos resíduos aromáticos que mostram-se importantes na interação enzima-substrato, e por isso o substrato precisa ter um comprimento mínimo. Este resultado fornece uma base molecular para a observação experimental de que a enzima não hidrolisa eficientemente oligossacarídeos curtos como a celotriose.

Capítulo 5

Seletividade da Endoglucanase de *Xanthomonas campestris* pv. *campestris*

5.1 Introdução

A seletividade das celulasas é um ponto importante no mecanismo de hidrólise destas enzimas. Ela pode ser exibida em relação ao tipo de substrato, ramificado ou não, cristalino ou solúvel, etc. Quando se trata de substrato solúvel, oligossacarídeo, a seletividade pode ser exibida em relação ao seu comprimento. As celulasas que clivam uma faixa larga de comprimentos de substrato poderiam ser mais eficazes que as seletivas porque são capazes de promover hidrólise de substrato e, em seguida, de seus produtos em vários passos consecutivos. Desta forma a mesma enzima é reaproveitada para a hidrólise de substratos de comprimentos diferentes. Por outro lado, uma especificidade pode ser a propriedade desejada na natureza ou nos processos industriais. Por exemplo, as β -glicosidases são seletivas para ação em celobiose, o oligossacarídeo de duas unidades, e atuam em sinergismo com as endo e exoglucanases concluindo o processo de produção de glicose na indústria de bioetanol.

Os estudos da base molecular da seletividade de comprimento do substrato referem-se principalmente à estrutura da enzima, ao arranjo dos aminoácidos dentro dos subsítios de ligação. Um exemplo foi apresentado no Capítulo 4 desta tese. A explicação da seletividade foi o arranjo espacial relativo entre os subsítios diferentes. O sítio ativo é espacialmente longe dos resíduos aromáticos que mostram-se importantes na interação enzima-substrato, e por isso o substrato precisa ter um comprimento mínimo de cinco unidades glicosídicas.

Um estudo de atividade e cristalografia da glicosidase de *Streptococcus mutans* mostra que esta enzima liga-se aos substratos mais longos além da celobiose e mostra as diferenças estruturais entre ela e outras glicosidases da mesma família [68]. A glicosidase de *Streptococcus mutans* não apresenta um impedimento estérico para a interação com substratos longos realizado por um loop longo na entrada do bolsão catalítico existente em outras enzimas da família. Ela também apresenta um resíduo aromático no subsítio +2, W238, que é relacionado aos substratos longos. Os ensaios de atividade dos mutantes da região do resíduo aromático e também da região do loop foram consistentes com as observações estruturais. Ambos os mutantes, o que não contém o resíduo aromático característico e o com o loop trocado por um maior exibiram atividade menor em substratos longos.

Alguns subsítios de ligação mostram-se mais importantes que outros, como no estudo combinado de cristalografia e atividade enzimática da endoglucanase D de *Clostridium cellulovorans* [69]. Estudos de cromatografia de produtos da hidrólise mostram que a enzima cliva preferencialmente substratos de quatro ou mais unidades glicosídicas. Na estrutura cristalográfica do complexo enzima-celotriose, três subsítios da enzima no lado não-redutor são ocupados e os do lado redutor desocupados pelo substrato. Nesta posição o substrato não é cli-

vado. Foi sugerido que estes três subsítios exibem alta afinidade por substrato e que, independentemente do comprimento do substrato, eles são sempre ocupados. Assim, um substrato precisa ter no mínimo quatro unidades de comprimento para alcançar os subsítios do lado redutor, ou seja, realizar contato com o sítio ativo e assim ser hidrolisado.

Neste capítulo, procuramos explicar o princípio da seletividade de uma endoglucanase da origem bacteriana, a Endoglucanase de *Xanthomonas campestris* pv. *campestris* (XccEG). Esta enzima cliva preferencialmente substrato de quatro ou mais unidades glicosídicas. Estudos utilizando eletroforese capilar [58] mostraram que a enzima é capaz de promover a hidrólise de dois oligossacarídeos formados por cinco e seis unidades de glicose. Inicialmente, os dois polissacarídeos marcados anteriormente com a sonda fluorescente 8-aminopireno-1,3,6-ácido trissulfônico (APTS) foram submetidos a hidrólise enzimática. Os seus produtos foram observados por espectroscopia de fluorescência através de excitação a 488 nm e as emissões coletadas a 520 nm. Podemos observar, Figura 5.1, que o polissacarídeo contendo cinco glicoses foi hidrolisado liberando uma molécula com duas unidades de glicose e a enzima não foi capaz de hidrolisar o substrato restante, com três glicoses (C3). No caso que tem substrato formado por seis unidades de glicose, observamos o mesmo padrão de hidrólise (liberando C2), só que neste caso o substrato restante de quatro unidades de glicose (C4) ainda foi reduzido a C2. Porém, a clivagem do C3, que também foi formado, não teve continuidade pela ação da enzima. Com o resultado da hidrólise destes oligossacarídeos, é possível concluir que o substrato precisa ter no mínimo quatro unidades de glicose para ser hidrolisado pela enzima, já que nos dois casos, quando o substrato foi reduzido a C3, este se manteve mesmo nos casos em que enzima e substrato ficaram incubados por tempos mais longos. Além disso, podemos observar que a XccEG cliva no

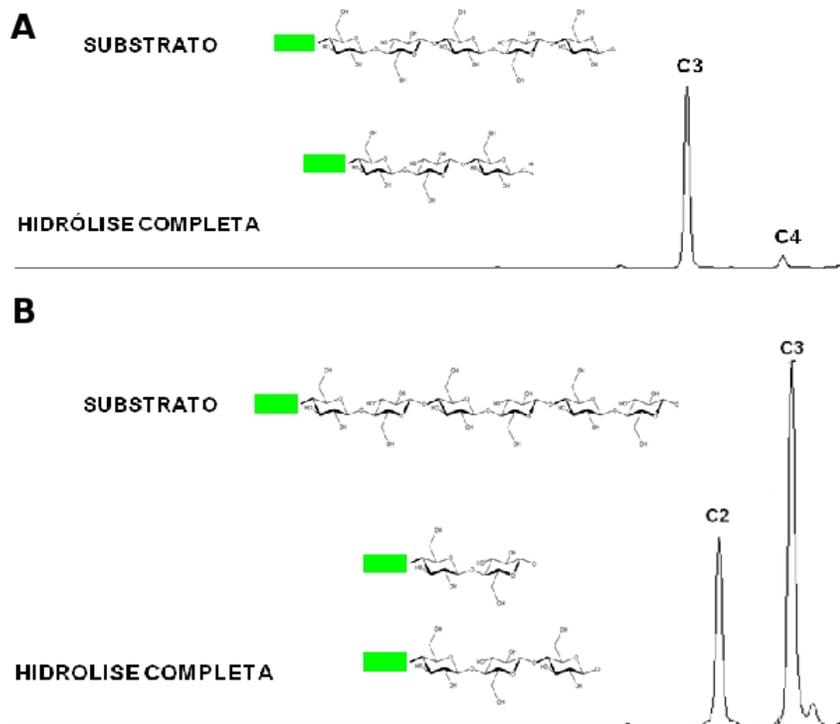


Figura 5.1. Espectros de Eletroforese capilar de oligossacarídeos marcados com APTS: Hidrólise completa de oligossacarídeo formado por cinco (A) e seis (B) glicoses. Os dois oligossacarídeos estão marcados com APTS pela extremidade não-redutora e os símbolos C6, C5, C4, C3 e C2 indicam o grau de polimerização dos oligômeros de glicose.

mínimo duas unidades da extremidade do substrato. O esquema desta seletividade é representado na Figura 5.2.

5.2 Simulações

Para investigar ao nível molecular a seletividade da Endoglucanase de *Xanthomonas campestris pv. campestris* (XccEG), simulações de Dinâmica Molecular foram feitas em complexos com substrato de comprimento mínimo para o qual a enzima é ativa, ou seja a celotetraose, e com substratos mais curtos, a celobiose e a celotriose. A estrutura cristalográfica da XccEG isenta de sub-

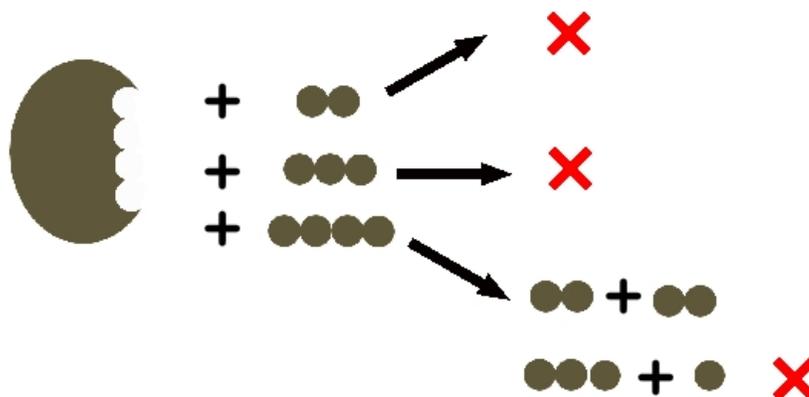


Figura 5.2. Esquema de seletividade de XccEG. A XccEG não cliva cadeias menores que quatro unidades glicosídicas e não cliva uma unidade da extremidade de cadeia.

strato foi utilizada como ponto de partida para as simulações. Modelos para a XccEG com substratos foram gerados por meio da docagem de celotetraose dentro da fenda através do alinhamento com a enzima da mesma família, a família 5 de glicosidases, a Endoglucanase E1 de *Acidothermus cellulolyticus* (pdb: 1ECE, [70]), Figura 5.3A, através do programa VMD [34]. O alinhamento estrutural das duas proteínas foi feito por minimização do valor de RMSD entre as duas estruturas, usando o programa VMD [34].

Como já mencionado anteriormente, os subsítios das enzimas onde se ligam as unidades glicosídicas do carboidrato são enumeradas por inteiros positivos do lado redutor do ponto de clivagem e de números inteiros negativos do lado não-redutor do ponto de clivagem. No caso da celotetraose os subsítios são: -2,-1,+1 e +2. Da mesma maneira são denominadas as unidades do substrato, Figura 5.3B. Os demais três sistemas foram obtidos cortando a celotetraose: uma celobiose, uma celotriose na posição -2-1+1 e uma celotriose na posição -1+1+2.

Três simulações independentes de 30ns para cada um dos cinco sistemas

foram realizadas: XccEG sem substrato, XccEG com celotetraose, XccEG com celotriose na posição -2-1+1, XccEG com celotriose na posição -1+1+2 e XccEG com celobiose. Os sistemas foram tratados no ensemble NPT à temperatura de 298K e pressão de 1atm. As interações de curta alcance, de van der Waals, não foram consideradas acima de raio de 12 Å e a função suavizante foi empregada a partir de 10 Å. De acordo com suas funções na catálise [5], o catalisador ácido Glu182 foi considerado protonado, o nucleófilo Glu303 desprotonado e o resíduo auxiliador Tyr263 protonado. Isso porque sabe-se que em sistema real existe uma ligação de hidrogênio entre o grupo -OH do catalisador ácido (glutamato ou aspartato) e o oxigênio glicosídico do carboidrato para garantir a doação de próton. O nucleófilo ataca o carbono anomérico do substrato e por isso é tratado como desprotonado, e o auxiliador forma ligação de hidrogênio com o nucleófilo mantendo-o na posição favorável para catálise [5]. Uma caixa de simulação cúbica, de lado de 80Å, foi construída, contendo uma molécula de XccEG, substrato em casos de simulações com substrato e 0.16M NaCl de densidade de 1g/cm³, usando o programa Packmol [71]. A distância entre o oxigênio do grupo -OH do catalisador ácido Glu182 e o oxigênio glicosídico da celotetraose é 3.61Å, Figura 5.3B, enquanto na estrutura original (1ECE.pdb) é 3.02Å.

Para os sistemas com substrato, as partes de simulação para cálculos de propriedades que se referem às interações com substrato, como mobilidade de loops e energia de interação, foram escolhidas assim que satisfaçam ambas as condições - de valor RMSD atingir patamar, e de substrato ainda estar presente na fenda da proteína. A Figura 5.4 mostra a distância entre o oxigênio do grupo -OH de Glu182 e o oxigênio glicosídico de substrato. A Figura 5.5 mostra o RMSD para todas as simulações de todos os sistemas contendo substrato. A tabela 5.1 mostra os intervalos de tempo escolhidos para análises com base

nos valores de RMSD e da distância entre o oxigênio do grupo -OH de Glu182 e o oxigênio glicosídico. A energia de interação enzima-substrato foi calculada

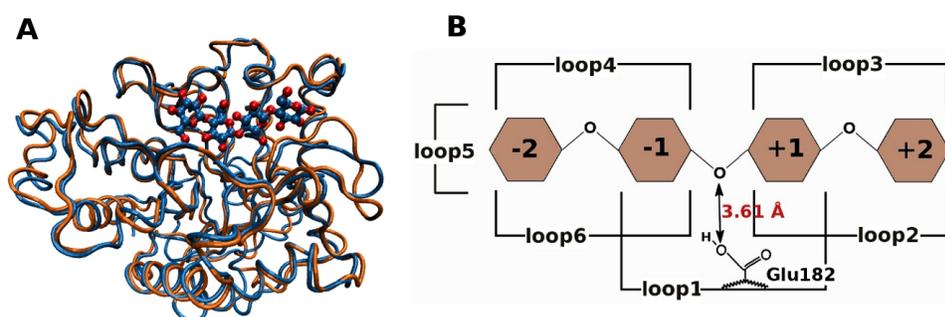


Figura 5.3. Docagem da celotetraose. **A:** XccEG (cor de laranja) alinhada com a endoglucanase E1 de *Acidothermus cellulolyticus* (azul). O substrato cristalizado em complexo com a endoglucanase E1 de *Acidothermus cellulolyticus* **B:** Esquema da posição da celotetraose dentro da fenda da XccEG.

como soma de energia potencial de van der Waals e a energia eletrostática entre cada resíduo do substrato e cada resíduo da enzima.

Tabela 5.1. Trechos das simulações usadas nos cálculos das propriedades médias do sistema, determinadas com as análises de RMSD e a distância entre o oxigênio do grupo -OH de Glu182 e o oxigênio glicosídico de substrato.

sistema/simulação	sim A	sim B	sim C
celobiose	(8-15)ns	(1-4)ns	(6-10)ns
celotriose-2-1+1	(1-3)ns	(8-13)ns	(1-5)ns
celotriose-1+1+2	(8-13)ns	(8-14)ns	(1-3)ns
celotetraose	(6-24)ns	(13-22)ns	(21-28)ns

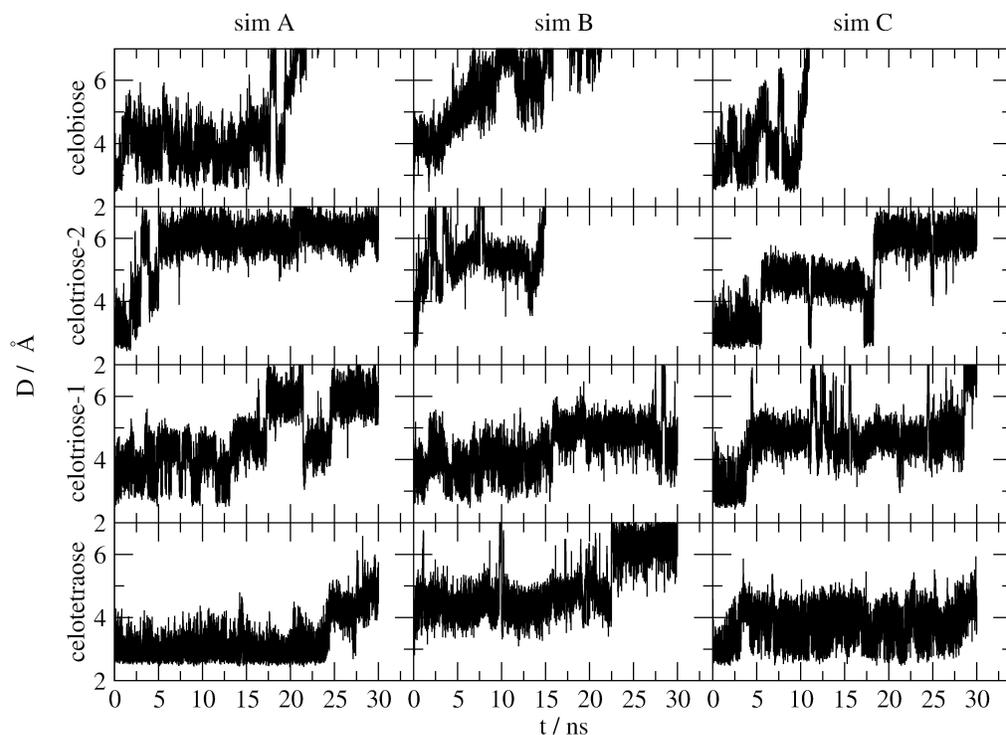


Figura 5.4. A evolução temporal da distância entre o oxigênio do grupo -OH de Glu182 e o oxigênio glicosídico de substrato para todos os sistemas simulados contendo substrato. Algumas curvas são interrompidas por falta de espaço no eixo y .

5.3 Resultados e discussões

Seletividade confirmada nas simulações. As simulações mostram que os subsítios curtos perdem contato com a fenda catalítica e conseqüentemente poderiam perder a chance de serem hidrolisados. O substrato foi considerado presente na fenda da proteína quando a distância entre o oxigênio do grupo -OH do catalisador ácido Glu182 e o oxigênio glicosídico de substrato fosse menor que 4 Å [8]. Esta distância foi escolhida como a soma da distância necessária para o catalisador formar ligação de hidrogênio com o oxigênio glicosídico (3 Å)

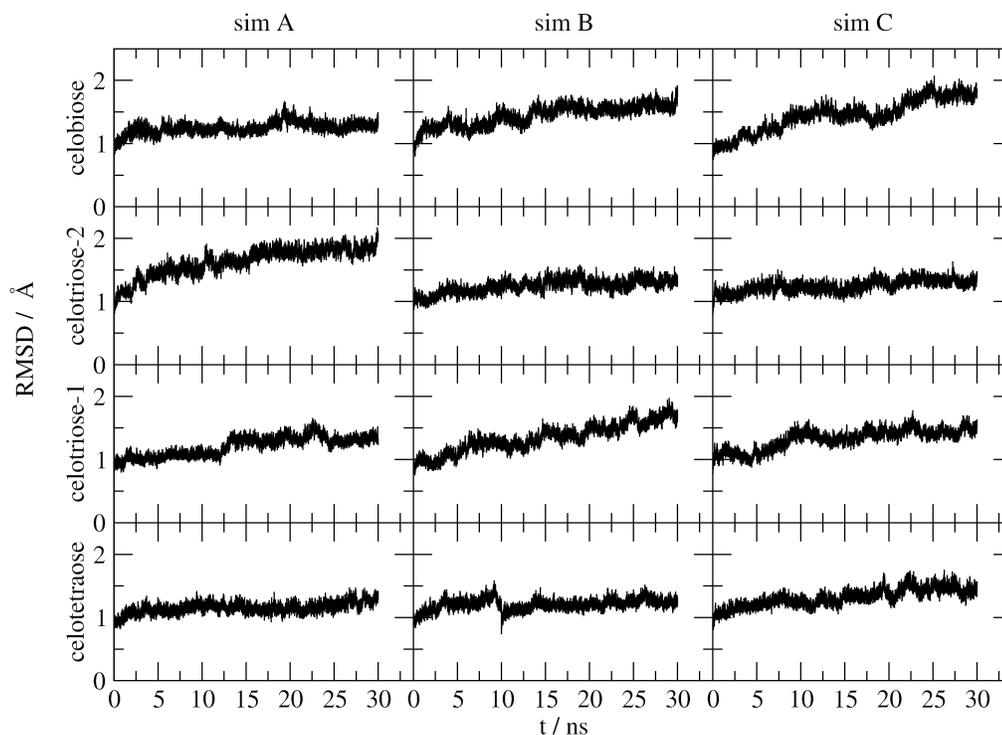


Figura 5.5. RMSD para todos os sistemas simulados contendo substrato.

e uma amplitude de flutuações temporais (1 Å). A docagem da celotetraose feita por alinhamento com a estrutura 1ECE.pdb, Figura 5.3B, fornece uma posição favorável para o ataque à ligação glicosídica pelo catalisador ácido [5]. Durante as simulações com a celotetraose, esta distância flutua ao redor do valor inicial, Figura 5.6A, o que mostra a presença da celotetraose dentro da fenda catalítica. Além disso, existe ligação de hidrogênio com a probabilidade de 28.3 % entre o catalisador ácido e o oxigênio glicosídico. Os esquemas dos substratos são mostrados de baixo dos gráficos na Figura 5.6. No caso dos substratos curtos, esta distância enzima-substrato aumenta em tempo menor.

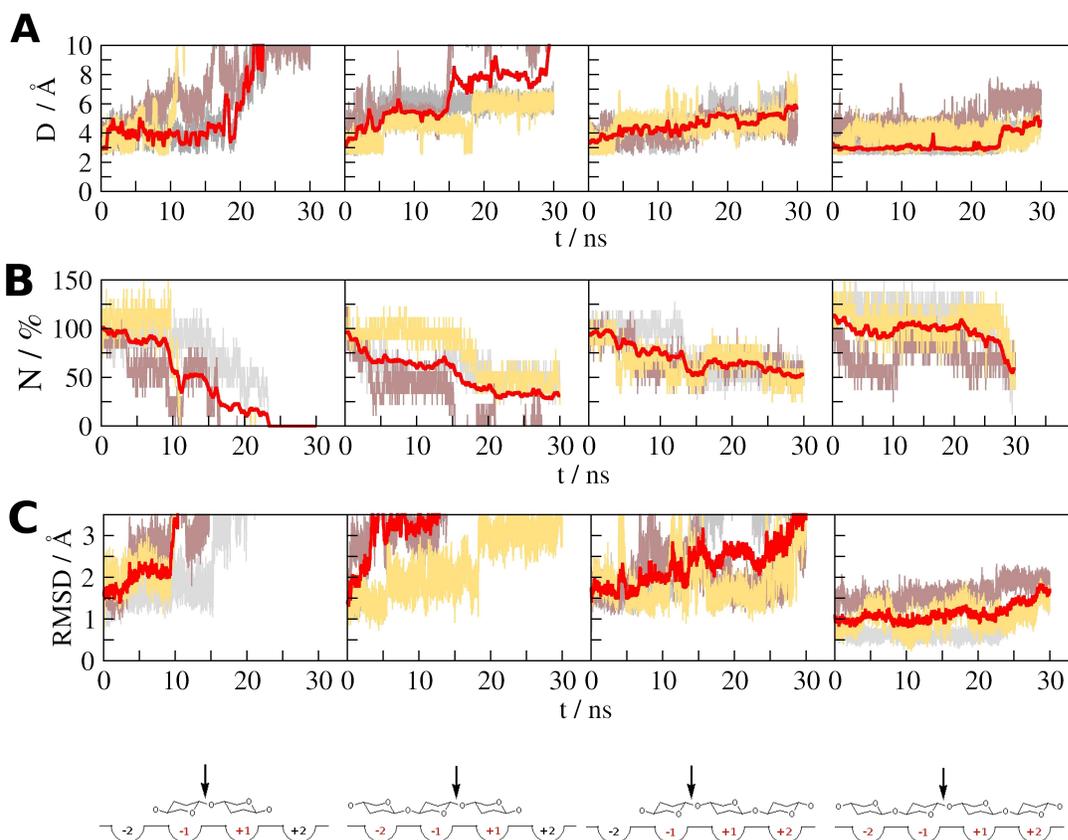


Figura 5.6. **A:** A distância entre oxigênio do grupo -OH de Glu182 e oxigênio glicosídico de substrato durante as simulações, para sistemas com celobiose, celotriose-2-1+1, celotriose-1+1+2 e a celotetraose. **B:** Número de átomos de cada unidade de substrato a 4 Å de cada subsítio, em relação à estrutura inicial docada. Subsítios foram definidos pelo critério de energia de $< -5\text{kcal/mol}$, Figura 5.7. **C:** RMSD de cada substrato em relação à estrutura inicial. As linhas vermelhas foram obtidas como a média de blocos de 20 pontos da curva que representa as propriedades médias das três simulações. Algumas curvas são interrompidas por falta de espaço no eixo y .

Também foi feita uma contagem dos átomos de cada unidade de substrato a 4 Å ao redor de cada subsítio ao longo do tempo, Figura 5.6B. A soma destes números para cada subsítio foi dividida pela soma inicial, referindo-se à estrutura inicial docada. Primeiro, os subsítios foram definidos pelo critério de energia de interação com substrato de $< -5\text{kcal/mol}$. A Figura 5.7 mostra

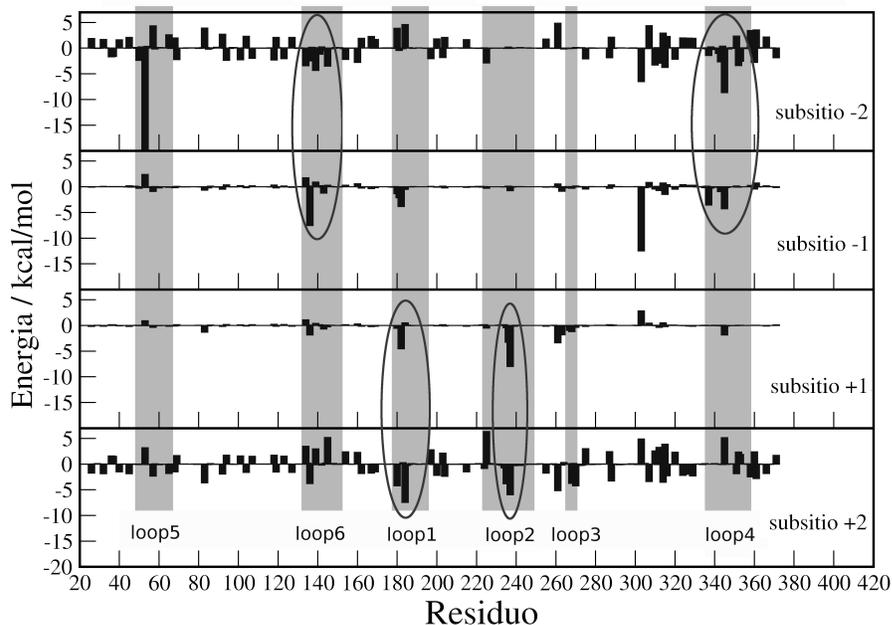


Figura 5.7. Energia das interações EGXc-celotetraose para cada um dos subsítios. Estão destacados loops compartilhados entre subsítios diferentes.

a energia de interação enzima-substrato entre a proteína e cada unidade da celotetraose. A energia da interação entre a XccEG e cada um dos subsítios da celotetraose foi calculada. Por exemplo, entre a unidade -2 da celotetraose e todos os resíduos da enzima. Esta análise foi feita somente para o sistema em equilíbrio contendo a celotetraose, Tabela 5.1. O critério de energia foi escolhido com base nas energias calculadas de modo que cada um dos subsítios contenha pelo menos um resíduo. A hipótese é que os subsítios ocupados pelos substratos curtos interagem da mesma maneira que os respectivos subsítios com a celotetraose. A energia calculada representa energia de interação não-covalente entre cada resíduo da enzima e cada unidade de substrato, como média no tempo de simulação. Os resíduos relevantes são aqueles cuja energia de interação com substrato é negativa. Escolhendo os resíduos da proteína que representam a energia de interação menor que -5kcal/mol com uma certa

unidade de substrato, o subsítio da proteína que refere-se a esta unidade de substrato foi localizado. Assim o número de átomos do substrato ao redor do subsítio identificado foi determinado. A celotetraose mantém inalterado o número de resíduos interagentes com os subsítios da enzima.

O RMSD de cada substrato foi calculado durante as simulações em relação às estruturas iniciais docadas. Em cada sistema, a proteína foi alinhada com ela mesma ao longo do tempo ao fim de calcular o RMSD do substrato, Figura 5.6C. Os substratos curtos, a celobiose e as duas celotrioses, mudam consideravelmente a posição, enquanto a celotetraose exibe RMSD de valor baixo, em torno de 1Å, durante a maior parte do tempo.

Com base nos resultados da Figura 5.6, a celotetraose mantém a distância da enzima mais curta, o número de átomos dentro da fenda maior e o valor de RMSD mais baixo em comparação com os substratos mais curtos. Aparentemente, a celotetraose exibe a menor tendência de perder contato com a enzima.

Os substratos mais curtos, independentemente da posição, afastam-se do sítio ativo. Aparentemente, a existência simultânea das interações nos dois subsítios -2 e +2 é importante para a fixação do substrato à fenda catalítica da XccEG. Aqui, os resultados teóricos concordam com os experimentais quanto a ligação ao substrato que é a condição necessária para a hidrólise. Os substratos mais curtos que quatro unidades perdem contato com a enzima. Segundo nossas simulações, a celotriose poderia ocupar igualmente a posição -2-1+1 e a posição -1+1+2. Em nenhum dos casos, a celotriose estaria em posição de ser clivada corroborando os ensaios enzimáticos.

Mobilidade da fenda catalítica. Os loops são regiões de enzimas bastante importantes para interação com substrato, sendo regiões mais móveis e capazes de mudar a conformação de enzima para receber ou soltar o substrato. Normalmente encontram-se na superfície da proteína onde há acesso fácil para

o substrato. A fenda de ligação ao substrato da XccEG consiste em seis loops. Calculando energia de interação com a celotetraose, vemos que 84% desta energia pertence aos seis loops, Figura 5.7. No resto, são incluídos resíduos catalíticos, Glu303 e Tyr263, que normalmente encontram-se em regiões menos móveis da proteína.

A Figura 5.8 mostra a mobilidade dos seis loops da fenda da XccEG durante o tempo de simulação e a escala de cores que se refere ao valor RMSF dos resíduos calculado por blocos de 300ps. Na medida que o número de

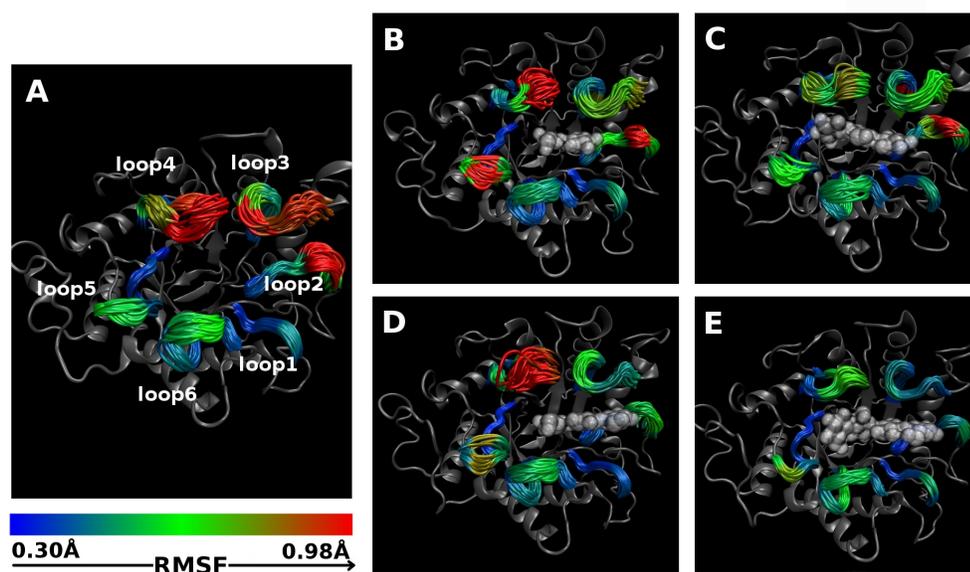


Figura 5.8. Mobilidade dos loops da fenda. Os substratos são representados em núvens brancas. A- sem substrato, B- celobiose, C- celotriose-2-1+1, D- celotriose-1+1+2, E- celotetraose.

subsítios ocupados por substrato aumenta, a mobilidade dos loops diminui. Em outras palavras, os loops são estabilizados pelas interações com substrato. É necessário estabilizar os loops dos quatro subsítios da fenda da enzima para o substrato permanecer interagindo com a enzima. Como a XccEG exibe atividade muito maior em celotetraose do que em substratos curtos, como mostram

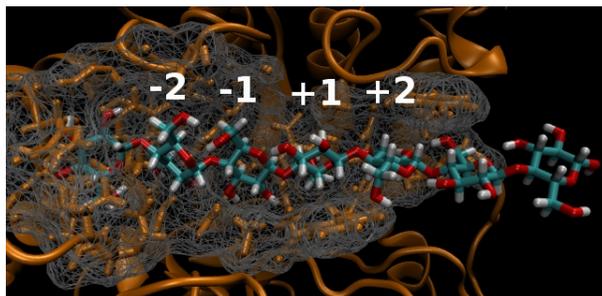


Figura 5.9. XccEG possui somente 4 subsítios para ligar à celulose.

os dados de eletroforese capilar, estes resultados teóricos sugerem que é preciso estabilizar os loops dos quatro subsítios da fenda para estabilizar o substrato em posição adequada para a hidrólise.

Se observarmos a estrutura terciária da XccEG, entenderemos porque é necessário estabilizar exatamente quatro e não mais subsítios da enzima, Figura 5.9: a enzima é composta por quatro subsítios de ligação à celulose. Em lugar do hipotético subsítio -3 encontra-se o loop5 que por meio da sua posição bloqueia entrada do substrato na fenda, ou seja, ele forma a extremidade da fenda. No outro lado, em lugar do hipotético subsítio +3 não existem interações com substrato, sendo este lugar o final da fenda da enzima. A XccEG poderia catalisar hidrólise de cadeias mais longas que quatro unidades glicosídicas. No lado positivo da fenda não existe impedimento espacial para substratos longos e no lado negativo um substrato longo poderia interagir com a enzima em uma conformação curvada de forma que evite a barreira espacial do loop5.

Para entender como a mobilidade dos subsítios desocupados por substrato afetam os subsítios ocupados, a identificação dos resíduos que formam os subsítios da enzima foi realizada, Figura 5.7. Os subsítios -2 e -1 compartilham os loops 4 e 6, e os subsítios +2 e +1 compartilham os loops 1 e 2, encirculado na Figura 5.7. Assim a mobilidade do subsítio -2 pode ser envolvida com a mobilidade do subsítio -1. Eventualmente o subsítio -1 perde contato enzima-

substrato, a ligação glicosídica afasta-se do sítio ativo e como resultado final a hidrólise não ocorre. Igual acontece nos subsítios de numeração positiva.

Docagem alternativa. Para assegurar a plausibilidade das interações entre a XccEG e a celotetraose, uma docagem alternativa foi feita. A partir da mesma configuração inicialmente docada, um procedimento de simulação foi feito usando uma restrição harmônica de constante de força de 10 kcal/mol Å² e uma distância de equilíbrio de 3 Å entre o átomo do oxigênio do grupo -OH do catalisador ácido Glu182 e o oxigênio glicosídico da celotetraose. Desta forma, a minimização de 1000 passos foi feita apenas com os átomos de proteína livres de se moverem, e em seguida, a equilibração de 10 ps apenas com as cadeias laterais da proteína livres para se moverem. A distância resultante foi 3,05 Å. Depois disso, o mesmo procedimento de 3 simulações de 30 ns tem sido realizado. Nas simulações, existe uma ligação de hidrogênio entre o grupo -OH do catalisador ácido Glu182 e o oxigênio glicosídico da celotetraose com uma probabilidade média de 20.5 %. A energia da interação proteína-substrato é muito semelhante para as duas docagens diferentes, Figura 5.10, então podemos considerar a análise da identificação dos subsítios confiável.

Seletividade das outras enzimas. A mesma base estrutural de seletividade pode ser procurada entre as endoglucanases da mesma família, a família 5 de glicosidases, [66]. A Figura 5.11 mostra todas as endoglucanases da família 5 de glicosidases cristalizadas com substrato e quais contêm um loop do tipo loop5 da XccEG. Os nomes na figura se referem aos nomes dos organismos que produzem as enzimas representadas. As enzimas foram alinhadas com a XccEG ao fim de comparar o tamanho das fendas catalíticas. Todas elas apresentam o loop na mesma posição, representado em cores diferentes para cada enzima e desenhado como tubo. O tamanho das fendas também é similar ao da fenda da XccEG. A extremidade das fendas do lado redutor foi determi-

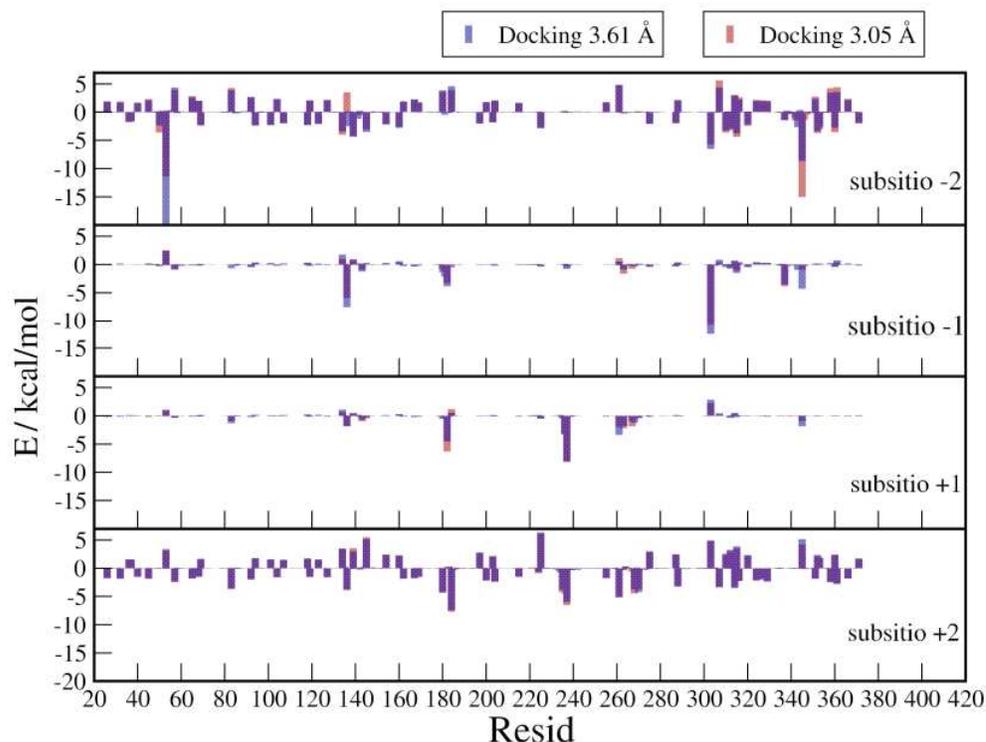


Figura 5.10. Energia de interação XccEG-celotetraose para cada um dos subsítios. A primeira docagem (azul) e a docagem alternativa (vermelho), onde a distância entre o átomo de oxigênio do grupo-OH do catalisador ácido Glu182 e o oxigênio glicosídico da celotetraose são 3,61 Å e 3,05 Å, respectivamente.

nado como conjunto de átomos de enzima a 7Å ao redor do último átomo de substrato do lado redutor, o átomo O1. A extremidade das fendas das enzimas alinhadas também está representado em cores diferentes para cada enzima e desenhado como tubo.

A endoglucanase E1 de *Acidothermus cellulolyticus* [70] e a endocelulase de *Pyrococcus horikoshii* [72] foram cristalizadas com a celotetraose e também com duas moléculas de celobiose que indica o produto da hidrólise. Então, estas duas enzimas atuam na celotetraose e têm duas celobioses como produto. O mesmo tipo de atividade é exibido pela XccEG.

A endoglucanase E5 de *Thermobifida Fusca* [PDBid: 2CKR] foi cristalizada

com a celopentaose que forma uma curvatura na região do loop e não foi cristalizada com substratos menores.

As três enzimas comparadas estruturalmente com a XccEG têm como substrato um oligossacarídeo de quatro ou mais unidades glicosídicas da mesma maneira que a XccEG. Sugere-se que, em base de tamanho da fenda e a mesma estrutura terciária (enzimas da mesma família) que poderia implicar em mobilidade similar, estas três enzimas poderiam exibir o mesmo tipo de seletividade: substrato de quatro ou mais unidades e produto de no mínimo duas unidades glicosídicas.

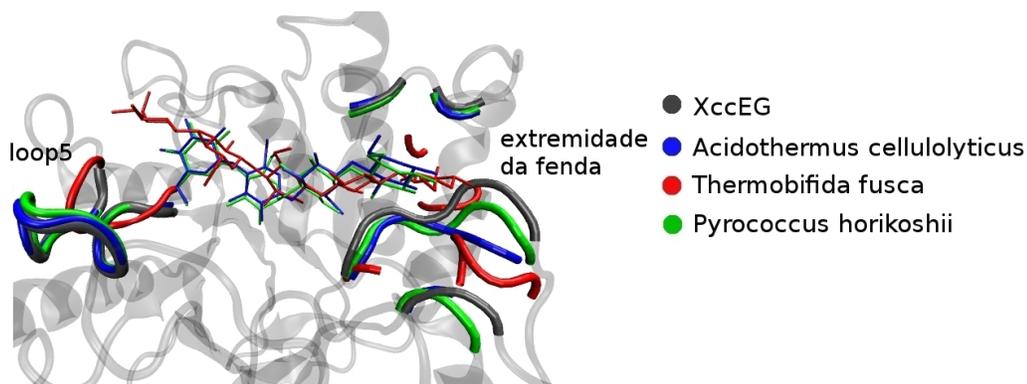


Figura 5.11. Enzimas da família 5 que possuem um análogo ao loop5 e que foram cristalizadas com substratos. Os nomes se referem aos nomes dos organismos que produzem as enzimas representadas. Substratos estão representados em barras e as fendas (o loop e a extremidade) em tubos. Todas possuem loop do tipo loop5 da XccEG e fenda de tamanho parecido. A extremidade das fendas de cada enzima foi determinado como conjunto de átomos de enzima a 7Å ao redor do último átomo de substrato do lado redutor, o átomo O1.

5.4 Conclusão

Entender a causa de seletividade das celulasas poderia ajudar no desenho de novas enzimas com as atividades desejadas. A endoglucanase de *Xanthomonas campestris pv. campestris* (XccEG) exibe uma seletividade em relação ao comprimento de substrato. Esta enzima cataliza somente a hidrólise de

cadeias da celulose de comprimento de quatro ou mais unidades glicosídicas e além disso cliva mais que uma unidade da extremidade de substrato.

Empregando a técnica de simulações de dinâmica molecular e analisando interações não covalentes com substratos, a base molecular da seletividade da XccEG foi proposta. A sua fenda catalítica é composta por quatro subsítios de ligação à celulose que são constituídos por loops altamente móveis. Para o substrato permanecer ligado dentro da fenda catalítica, e assim na posição adequada para a hidrólise ocorrer, todos os quatro subsítios precisam estar estabilizados pelas interações com substrato. Então, o comprimento mínimo de substrato que possa ser clivado é de quatro unidades glicosídicas.

Também foi procurada a mesma base estrutural de seletividade entre as endoglucanases da mesma família da XccEG. Sugere-se que, com base no tamanho da fenda e a mesma estrutura terciária (enzimas da mesma família) que poderia implicar em mobilidade similar, outras endoglucanases poderiam exibir o mesmo tipo de seletividade: substrato de quatro ou mais unidades e produto de no mínimo duas unidades glicosídicas.

Capítulo 6

Propriedades mecânicas do *linker* da XccEG

6.1 Introdução

O linker é um peptídeo entre dois domínios de uma celulase, que normalmente não forma estrutura secundária como hélice α ou fita β . Sabe-se que ele mantém CCD e CBM conectados ou poderia fixar a distância entre eles. Um linker rígido possui habilidade de manter mais ou menos fixa a distância entre as pontas e assim promoveria a separação dos domínios CCD e CBM, prevenindo contatos não favoráveis entre eles [73]. Mas, além disso, não se sabe se possui mais alguma função.

Existem também estudos sobre a elasticidade do linker das celobiohidrolases [76, 77]. As análises das simulações de dinâmica molecular mostraram que estes linkers apresentam um perfil de energia livre parecido com o perfil de energia potencial de um corpo elástico. Também, a influência da glicosilação ao comportamento elástico foi investigada. A distância de uma a outra ponta do linker, que refere-se às conformações de maior frequência nas simulações, é maior no caso do linker glicosilado do que com o mesmo linker simulado sem a glicosilação. O perfil geral da energia livre, por outro lado, não foi afetado, somente o valor de comprimento do linker é deslocado para os valores

Tabela 6.1. Sequência do linker da celobiohidrolase 2 de *Trichoderma reesei*. Aminoácidos da sequência são representados pelo código de uma letra.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
P	G	A	A	S	S	S	S	S	T	R	A	A	S	T	T	S	R	V	S	P	T	T	S	R	S	S	S	A	T	P	P	P	G	S	T	T	T	R	V	P	P	V	G	S	G

simulações Nos baseamos neste linker porque ele já foi abordado na literatura nos estudos de dinâmica molecular [77].

Propriedades mecânicas, como a rigidez e a elasticidade, foram estudadas para investigar semelhanças diferenças entre estes dois linkers. Segundo a lei de Hook para uma mola ideal, a força restitutiva providencia um comportamento elástico:

$$\vec{f} = -k\vec{x} \quad (6.1)$$

Quanto maior a constante de força k , ou em outras palavras, a rigidez da mola, maior é a elasticidade. Então, sob a hipótese de que um linker fixa a distância entre os domínios CCD e CBM, a elasticidade seria uma propriedade desejável. A proposta neste trabalho, é obter perfil de energia livre ao longo da distância ponta-a-ponta de linker e assim estimar a elasticidade. Para obter o maior número possível de conformações, é preciso amostrar bastante o espaço de conformações. Portanto, a técnica de Troca de Réplicas foi usada.

6.2 Simulações

As estruturas iniciais dos dois linkers, (PT)₁₂ e CBH2, foram preparadas por previsão de estrutura tridimensional em base das propriedades de aminoácidos, servidor I-TASSER, [82, 83]. O N-terminal dos linkers foi acetilado (adicionando o grupo $-COCH_3$) e C-terminal foi amidado (adicionando o grupo $-CONH_2$) para evitar efeito dos terminais carregados.

A celobiohidrolase 2 de *Trichoderma reesei* é de origem de um fungo, po-

Tabela 6.2. Glicosilação do linker CBH2. Aminoácidos da sequência são representados pelo código de uma letra e unidades manosídicas pelo símbolo Ò.

S5	S6	S7	S8	S9	T10	...	S14	T15	T16	S17	...	S20	...	T22
Ò	Ò	Ò	Ò	Ò	Ò		Ò	Ò	Ò	Ò		Ò		Ò
					Ò			Ò	Ò					Ò
					Ò			Ò	Ò					Ò
T23	S24	...	S26	S27	S28	...	T30	...	S35	T36	T37	T38	...	S45
Ò	Ò		Ò	Ò	Ò		Ò		Ò	Ò	Ò	Ò		Ò
Ò							Ò			Ò	Ò	Ò		
Ò							Ò			Ò	Ò	Ò		

dendo então ser glicosilada, por isso o seu linker (CBH2) foi glicosilado na fase de preparação para simulações. A glicosilação seguiu padrão da Celobiohidrolase 1 de *Trichoderma reesei* [76] onde uma unidade de manose é ligada a Ser e duas ou três unidades de manose a Thr. Resultados de Espectroscopia de Massas quanto à quantidade de O-ligadas $\alpha(1,2)$ -manoses [84] mostram que 39-46 unidades no total são ligadas ao linker. Então, as unidades manosídicas foram distribuídas da seguinte forma: uma unidade ligada a cada Ser e três unidades ligadas a cada Thr, 41 unidades no total, Tabela 6.2. Como queremos amostrar todas as conformações de linker, até as mais estendidas, a caixa de simulação teve que ser suficientemente grande. Por isso, para diminuir o tempo computacional, o solvente implícito [85] foi usado, o que dispensa o uso da condição periódica assim como o método de Ewald no tratamento de interações eletrostáticas. Para o tratamento das interações eletrostáticas e de Van der Waals foi usado o raio de corte de 18 Å junto com a função suavizante a partir de 10 Å.

Simulação para estimar autocorrelação de energia potencial foi realizada a 298K com o passo de tempo de 2fs. 17ns de simulação de equilíbrio foram extraídos para o linker (PT)₁₂ e 28ns para o linker CBH2. A energia potencial foi coletada a cada 100 passos de tempo de simulação, ou a cada 0,2ps. O critério para equilíbrio é convergência de energia potencial e de RMSD. A função de autocorrelação foi calculada como:

$$C(t) = \frac{1}{t_{max}} \sum_{t_0=1}^{t_{max}} U(t_0)U(t_0 + t) \quad (6.2)$$

onde t é intervalo de tempo e t_0 é passo de tempo para qual a energia potencial U foi calculada. Para cada valor de t , o valor de $t_0 + t$ não pode ultrapassar o número de U armazenado pela simulação, então t_{max} é o máximo passo disponível para o dado intervalo atual t . A função de autocorrelação obtida, Figura 6.1, foi ajustada à função exponencial:

$$C(t) \approx Ae^{-Bt} + Ce^{-Dt} \quad (6.3)$$

e o tempo de correlação foi calculado como área de baixo da curva normalizada:

$$\tau = \Delta t \sum_{t=1}^{t_{cor}} \frac{C(t)}{C(0)}, \quad (6.4)$$

onde Δt representa o intervalo de tempo para qual o valor da energia potencial foi salva ao longo de simulação e este intervalo é $\Delta t = 0,2ps$.

Depois de um tempo, o decaimento exponencial de $C(t)$ é dominado pelo ruído constante de simulação, então o $C(t)$ foi truncado depois de 100 passos de tempo. O tempo de correlação calculado é 2,8ps para o linker (PT)₁₂ e 4,3ps para o linker CBH2. Este tempo foi usado como o tempo entre as tentativas de troca de temperatura na técnica de troca de réplicas.

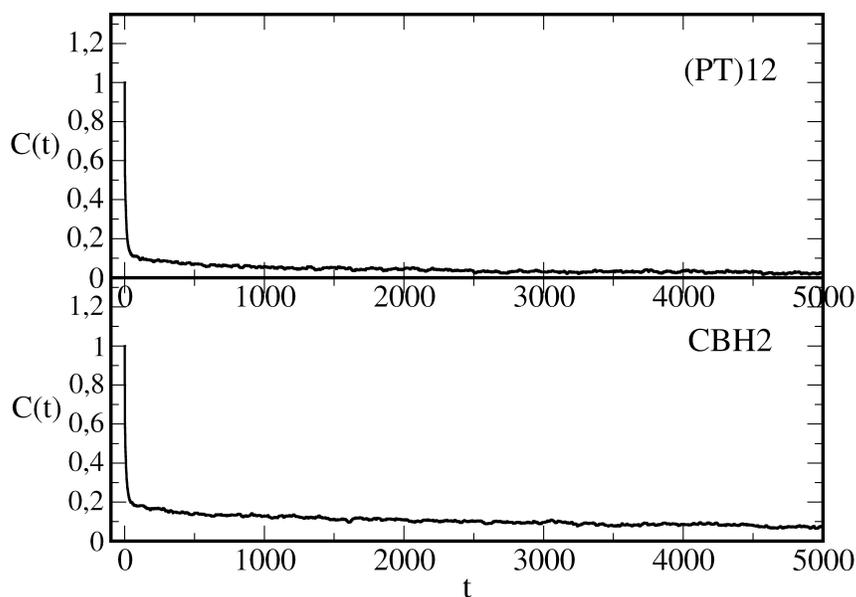


Figura 6.1. A função de autocorrelação da energia potencial normalizada, para ambos os linkers, $(PT)_{12}$ e CBH2.

O número de réplicas foi ajustado para atingir a probabilidade de troca igual a 0,5. Este número é verificado, durante as simulações, estimando a taxa de troca. Ela é igual a fração entre o número de trocas e o número total de tentativas de troca, calculado como a média entre todas as réplicas. Para o linker $(PT)_{12}$, o número de réplicas é 16 e para o CBH2 é 24. Cada sistema foi simulado em réplicas de temperatura mínima de 298K e máxima de 550K. A temperatura máxima foi escolhida com base no estudo dos linkers das celobiohidrolases [76, 77]. Os sistemas foram tratados no ensemble NVT e cada réplica primeiro foi equilibrada a temperatura própria por 35ps. O tempo de correlação da energia potencial para o sistema aquecido à temperatura máxima foi de 20ps. O número de trocas total para satisfazer a eficiência de mistura foi

estimado com base na recomendação de Abraham e Grady (vide Figura 1 da ref. [52]). Para (PT)₁₂ este número é cerca de 20 000 e para o CBH2 30 000.

As trajetórias a 298K foram usadas para todas as análises. Do tempo total de simulações das réplicas foi extraído 32ns para (PT)₁₂ e 38ns para CBH2 a 298K.

6.3 Resultados e discussões

Para testar se o sistema visita adequadamente o espaço de conformações, uma análise de PCA [86] foi feita. Foram mostrados instantes na simulação no espaço reduzido de projeção a duas componentes principais, ou seja, de maior autovalor, Figura 6.2A. Ambos os linkers percorrem o espaço de fase bastante uniformemente. A distribuição de parâmetro de ordem R , a distância ponta-a-ponta de linker, também é ampla ao longo do tempo de simulação, a Figura 6.2B. Se durante o tempo o linker volta a percorrer um subespaço do início de simulação, isto significa que a trajetória não contém um subespaço proibido e que pode ser usada para as análises. Portanto, as trajetórias inteiras foram usadas nas análises.

A diferença entre os dois linkers. O perfil de energia livre em função da distância ponta-a-ponta R para ambos os linkers foi estimada como

$$F = -k_B T \ln W(R) + C, \quad (6.5)$$

Figura 6.3, onde $W(R)$ é a distribuição da distância ponta-a-ponta R durante simulação, que foi calculada por meio do histograma de valores de R considerando janelas de 1Å de comprimento, e C é constante.

O erro do perfil de energia livre foi estimado como desvio padrão entre três

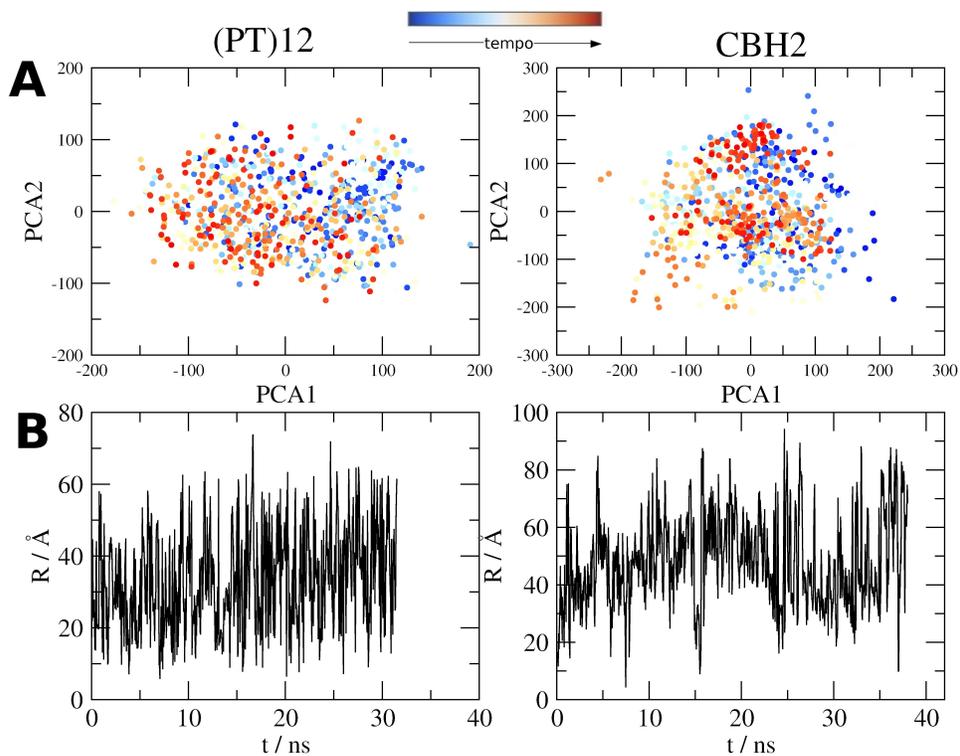


Figura 6.2. **A:** Espaço de conformação reduzido a duas componentes principais a 298K. Cada ponto é posição de sistema no espaço de conformação em um instante de simulação. **B:** Distribuição de parâmetro de ordem R não apresenta mudanças abruptas ao longo do tempo de simulação.

curvas obtidas a partir de três trechos subsequentes de simulação:

$$\Delta = \sqrt{\frac{\sum_{i=1}^3 (F_i - \langle F \rangle)^2}{3}} \quad (6.6)$$

Para comparar comportamento elástico dos 2 linkers, as curvas obtidas

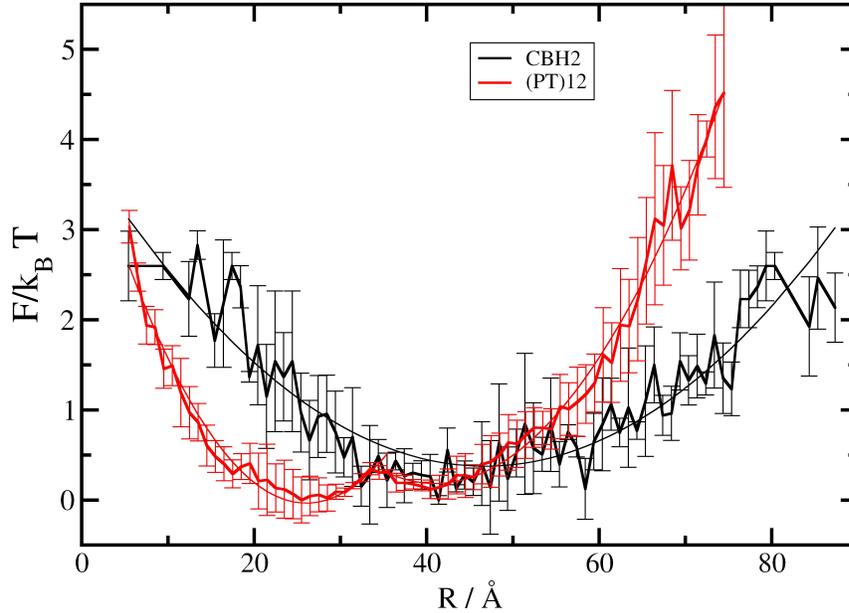


Figura 6.3. Perfil de energia livre em função da distância ponta-a-ponta R , para os 2 linkers, $(PT)_{12}$ (vermelho) e CBH2(preto). As linhas finas representam o ajuste à lei de Hook.

foram ajustadas à expressão da energia potencial da lei de Hook:

$$F \approx \frac{k}{2}(R - R_0)^2 + F_0, \quad (6.7)$$

Tabela 6.3. O parâmetro R_0/N_{res} representa o comprimento do linker do mínimo de energia dividido pelo número de resíduos. Isto porque os dois linker possuem números de resíduos diferentes.

O perfil de energia livre do linker $(PT)_{12}$ é levemente bimodal com o modo R_1 mais comprimido e o modo R_2 mais estendido que o linker CBH2. A diferença entre os dois comprimentos R_1 e R_2 é aproximadamente do comprimento de 3 glicoses na cadeia celulósica cristal. O linker $(PT)_{12}$ é mais rígido do que

	$k / \text{cal/mol}\text{\AA}^2$	R_0/N_{res}
(PT) ₁₂ , R1	12,47	0,93
(PT) ₁₂ , R2	7,74	1,46
CBH2	3,21	1,02

Tabela 6.3. Parâmetros das curvas de perfil de energia livre em função de distância ponta-a-ponta ajustados à lei de Hook, para os dois linkers, (PT)₁₂ e CBH2.

	$k / \text{cal/mol}\text{\AA}^2$	R_{g0}/N_{res}
(PT) ₁₂ , R1	330,17	0,42
(PT) ₁₂ , R2	139,11	0,57
CBH2	164,86	0,47

Tabela 6.4. Parâmetros das curvas de perfil de energia livre em função de raio de giro ajustadas à lei de Hook, para os dois linkers, (PT)₁₂ e CBH2.

o linker CBH2 porque exibe maior constante de força k . Também podemos observar que na faixa de flutuação térmica $\Delta F = k_B T$ o linker CBH2 exibe uma faixa de extensão de 60 Å e o linker (PT)₁₂ de 40 Å abrangendo os dois modos R_1 e R_2 , o que significa que o linker (PT)₁₂ é mais específico quanto a manter o comprimento.

Um outro parâmetro de ordem, o raio de giro, também foi usado na comparação dos linkers, Figura 6.4. A Tabela 6.4 mostra os parâmetros: a constante de força k e o raio de giro de mínimo de energia R_{g0} . De novo, o linker (PT)₁₂ exibe uma leve bimodalidade e a maior rigidez do que o linker CBH2.

A análise de RMSF por blocos também mostra que o linker (PT)₁₂ é mais rígido que o linker CBH2, o que é consistente com a maior elasticidade porque esta rigidez providencia uma força restitutiva necessária para comportamento

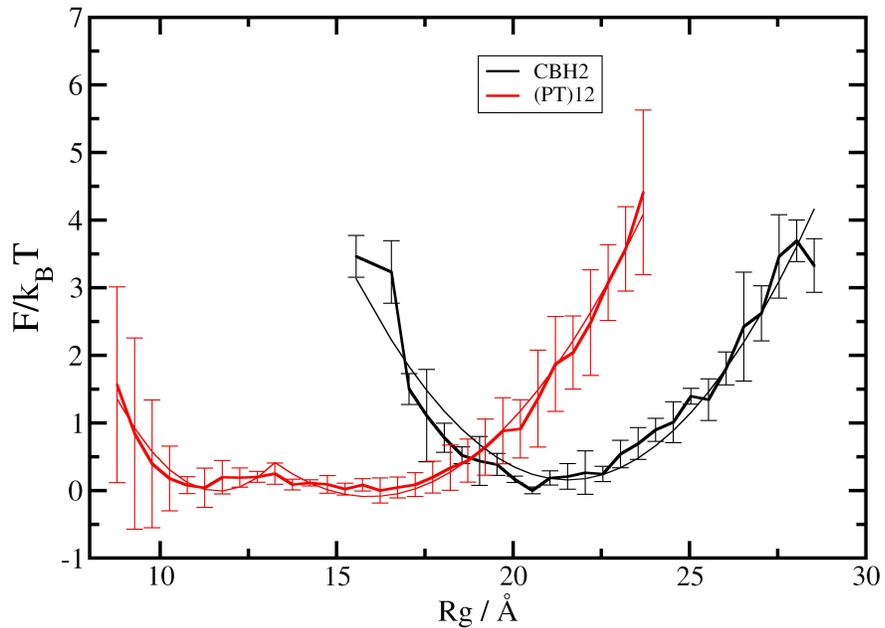


Figura 6.4. Perfil de energia livre em relação ao raio de giro R_g , para os 2 linkers, (PT)₁₂(vermelho) e CBH2(preto).

elástico. O valor de RMSF computado por blocos de 100ps como média de todos os resíduos é 8,79 Å para CBH2 e 6,15Å para (PT)₁₂. Este valor, bem como a distribuição de ângulos torcionais, Figura 6.6, mostra maior rigidez do linker (PT)₁₂. A coordenada no eixo x é a raiz quadrada da média sobre todos os resíduos da soma dos quadrados dos ângulos:

$$\Sigma_{\psi,\phi} = \sqrt{\frac{\sum_i(\psi_i^2 + \phi_i^2)}{N_{res}}} \quad (6.8)$$

A distribuição é mais estreita para o linker (PT)₁₂, o que significa que ele exibe variação menor dos ângulos. O esquema dos ângulos torcionais da cadeia principal da proteína é representada na Figura 6.5.

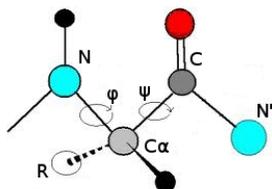


Figura 6.5. Ângulos torcionais da cadeia principal, ϕ e ψ .

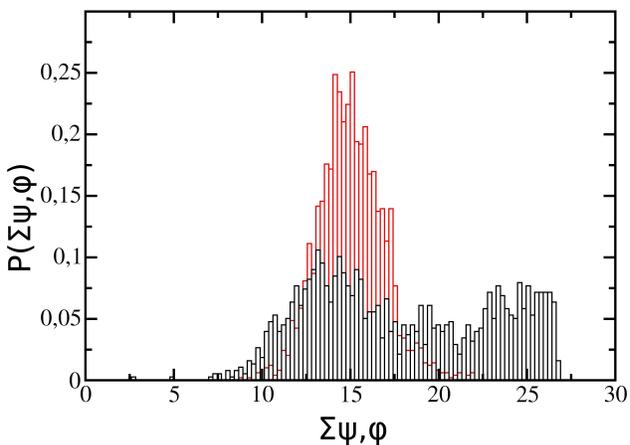


Figura 6.6. A distribuição de probabilidade dos ângulos torcionais da cadeia principal, vermelho para (PT)₁₂, preto para CBH2. A função no eixo x é definida pela equação 6.8.

A rigidez pode ser explicada pela própria sequência do linker (PT)₁₂. A prolina é o aminoácido mais rígido devido o anel e este anel envolve a cadeia principal de proteína o que influencia a rigidez de proteína, Figura 6.7.

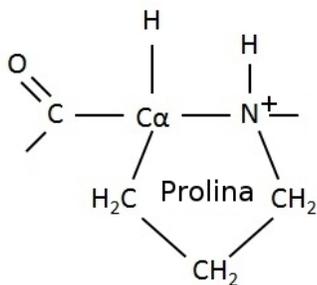


Figura 6.7. Estrutura da prolina.

A treonina, por sua vez, devido a posição do grupo -OH da sua cadeia lateral, possui a propriedade de formar ligação de hidrogênio com o resíduo vizinho ou segundo vizinho, formando assim um “anel” que lembra o da prolina, Figura 6.8. A rigidez exibida pelo linker (PT)₁₂ poderia ter origem na rigidez intra-

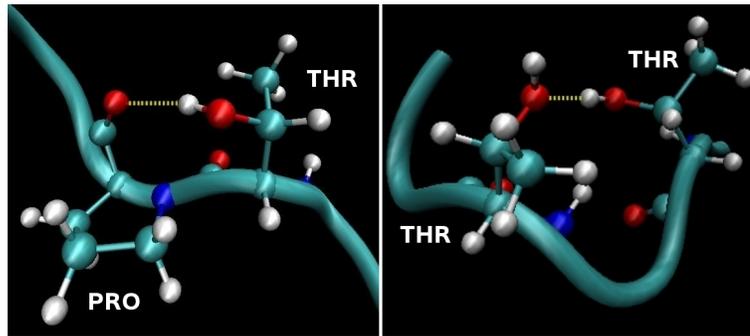


Figura 6.8. Ligação de hidrogênio (amarelo) entre a cadeia lateral da Thr e os resíduos vizinhos.

resíduo (proveniente da Pro) e rigidez inter-resíduo vizinhos (proveniente da Thr). As probabilidades de todos os tipos de ligações de hidrogênio presentes na simulação do linker $(PT)_{12}$ foram determinadas como função de distância entre resíduos na sequência, Δres , Tabela 6.5. Nesta análise os primeiros três resíduos que não são Pro nem Thr não foram considerados. O que se mostra mais provável são justamente as ligações de hidrogênio entre a cadeia lateral da Thr e o resíduo vizinho ou segundo vizinho: Thr(l)-Pro (a cadeia lateral de Thr com o primeiro vizinho que é Pro), Thr(l)-Thr(p) (a cadeia lateral de Thr com a cadeia principal da segunda vizinha, Thr) e Thr(l)-Thr(l) (a cadeia lateral de Thr com a cadeia lateral da segunda vizinha, Thr).

O número total de ligações de hidrogênio em função do tempo foi determinado para ambos os linkers, Figura 6.9. Todas as ligações de hidrogênio existentes na proteína foram contadas, mesmo se mais que uma ligação é formada entre um par de resíduos. O número médio de ligações de hidrogênio para $(PT)_{12}$ é 0,712, o que significa que em 71,2% do tempo existe ligação de hidrogênio, Tabela 6.6. Nas simulações de $(PT)_{12}$, 99% das todas as ligações de hidrogênio são da cadeia lateral da treonina (70,5% no tempo dos 71,2%), enquanto nas simulações de CBH2, mesmo com 9 treoninas presentes – Thr10, Thr15, Thr16, Thr22, Thr23, Thr30, Thr36, Thr37, Thr38– não existem ligações

Tabela 6.5. Frequência de ligações de hidrogênio no linker (PT)₁₂ em função de distância entre resíduos na sequência. *l* e *p* denominam a cadeia lateral e principal, respectivamente. Por simplicidade, as frequências de baixo de 0,01% não estão mostradas na tabela. A maior frequência é entre Thr e primeiro(Pro) ou segundo resíduo vizinho (Thr).

	%				
Ligação de H	Δ res				
	1	2	3	4	5
Thr(l)-Thr(p)		9,40		1,30	
Thr(l)-Pro	2,90		0,98		0,17
Thr(l)-Thr(l)		4,10		0,14	
Pro-Thr(p)	0,01		0,23		0,01
Thr(p)-Thr(p)		0,14			
Pro-Pro		0,02			

Tabela 6.6. Porcentagem das ligações de hidrogênio para os dois linkers, (PT)₁₂ e CBH2, no total, e apenas da cadeia lateral da Thr.

Ligação de H		
	total	Thr(l)
(PT) ₁₂	71,2 %	70,5 %
CBH2	42,0 %	0,0 %

envolvendo a cadeia lateral da treonina, Tabela 6.6. Então, podemos observar que as ligações de hidrogênio envolvendo a cadeia lateral da treonina são importantes na indução da rigidez do linker (PT)₁₂.

O linker CBH2 é glicosilado, enquanto o (PT)₁₂ não é glicosilado. Existem estudos computacionais que mostram que a glicosilação do linker CBH2 afeta o

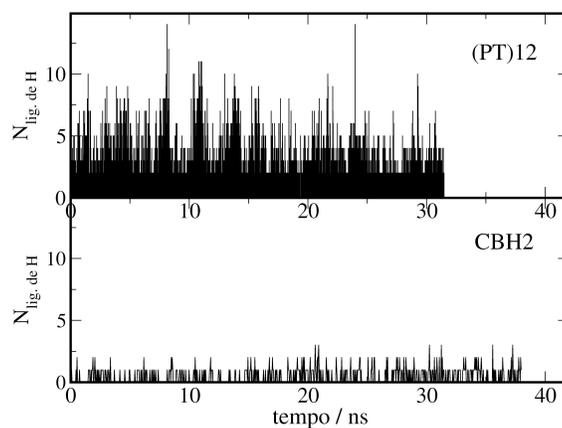


Figura 6.9. Número total de ligações de hidrogênio em função de tempo.

seu comprimento. Quando na forma glicosilada, este linker exibe maior comprimento [77]. A extensão do modo R_1 do linker $(PT)_{12}$ é 0,93 Å por resíduo e para o modo R_2 1,46 Å por resíduo, enquanto do linker CBH2 1,02 Å por resíduo. O linker $(PT)_{12}$ chega a ser estendido tanta quanto o linker CBH2 mesmo não contendo a glicosilação. A prolina e treonina são aminoácidos de natureza diferente, a prolina é apolar e a treonina polar. Prolina é capaz de formar interações hidrofóbicas com outras prolinas, e treonina de formar ligações de hidrogênio com cadeias laterais de outras treoninas na cadeia do linker $(PT)_{12}$. Aqui se sugere que estes dois resíduos, quando intercalados, impedem um agrupamento dos resíduos do mesmo tipo pelas interações hidrofílicas ou hidrofóbicas e, assim, impedem uma compressão do linker. Esta poderia ser a razão de porque o linker $(PT)_{12}$ é igualmente estendido quanto o linker CBH2, mesmo sem a glicosilação.

Bimodalidade do perfil de energia livre. Para explorar mais a bimodalidade do linker $(PT)_{12}$, conformações com a distância ponta-a-ponta $R = R_1 \pm 2\text{Å} = (25 \pm 2)\text{Å}$ e com $R = R_2 \pm 2\text{Å} = (41 \pm 2)\text{Å}$ foram separadas, bem

como as conformações relacionadas à barreira de $k_B T/2$, com comprimentos de $R = R_{barreira} \pm 2\text{Å} = (35 \pm 2)\text{Å}$. Na busca de diferença entre os dois modos da elasticidade, R_1 e R_2 num lado, e a barreira no outro lado, algumas análises foram realizadas nestes subespaços conformacionais (definidos pela faixa de 4 Å de distâncias em torno dos três comprimentos característicos do linker, R_1 , R_2 e $R_{barreira}$). A análise de RMSF por blocos de 100ps foi realizada em cima destes três conjuntos de conformações.

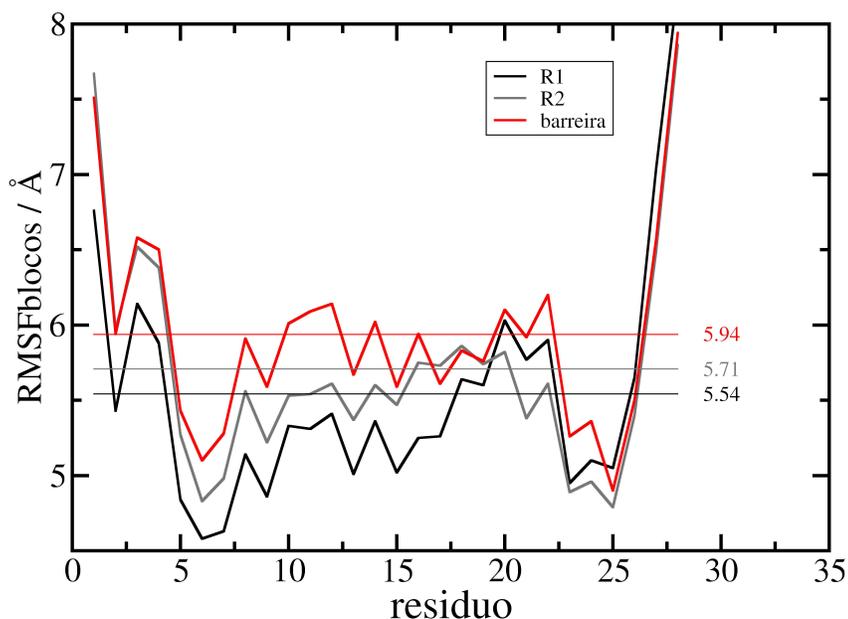


Figura 6.10. Mobilidade estimada por RMSF por blocos de 100ps para dois modos R_1 e R_2 e a barreira. A ordem da mobilidade crescente é: R_1 , R_2 , barreira. As linhas horizontais representam a média de RMSF por resíduos.

A mobilidade é a maior no caso da barreira, depois no caso do modo R_2 , e a menor no caso do modo R_1 , Figura 6.10. Vemos, na Figura 6.3, que a probabilidade de conformações diminui nesta mesma ordem. Como a barreira refere-se a energia mais alta no perfil de energia livre em relação aos mínimos, ela rep-

representa o número menor de conformações, ou representa subespaço de menor probabilidade. Em outras palavras, a mobilidade cresce com a diminuição da probabilidade. Isto significa que no caso do linker (PT)₁₂, o comportamento flexível é menos provável.

Dependência da temperatura. A técnica de troca de réplicas possibilita análise de trajetórias a várias temperaturas, dentro da faixa definida nas réplicas. Aqui será investigado comportamento dos linkers ao longo de temperatura. O linker (PT)₁₂ não perde a bimodalidade com aumento de temperatura, mas aqui será tratado como unimodal por simplicidade. Os parâmetros de elasticidade, k e R_0 , foram determinados ajustando perfis de energia livre para as temperaturas diferentes. Como esperado, a rigidez de um linker regular, CBH2, diminui com o aumento da temperatura. A constante de força do CBH2 diminui consideravelmente com o aumento da temperatura, até 40%.

O linker (PT)₁₂ exibe maior constante de força k ainda com temperatura e quase não muda a extensão (distância ponta-a-ponta R_0), Figura 6.11. Em out-

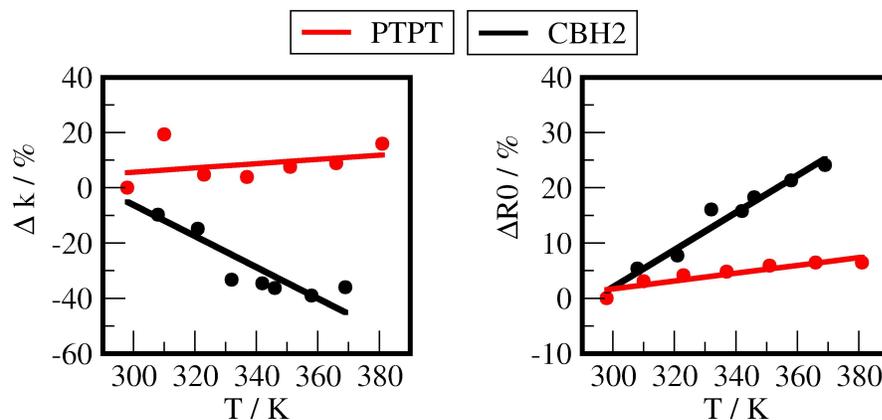


Figura 6.11. Mudança relativa dos parâmetros do ajustamento da energia livre à lei de Hook, constante de força k e a distância ponta-a-ponta R_0 , com aumento de temperatura.

ras palavras, este linker aumenta a sua rigidez com a temperatura, exibindo as-

sim o comportamento da mola mais forte. Este efeito é relativamente modesto como se pode ver pelos resultados da Figura 6.11. A faixa de temperatura representada na Figura 6.11 é a região de maior linearidade da curva.

A fim de investigar este aumento da rigidez do $(PT)_{12}$ com temperatura, a análise das todas as ligações de hidrogênio em $(PT)_{12}$ foi realizada como função da posição do resíduo na sequência, Figura 6.12. O número de resíduo na sequência do linker é plotado nos dois eixos do gráfico e cada ponto representa uma ligação de hidrogênio existente entre um par de resíduos. Como 99% das todas as ligações são da cadeia lateral da treonina, esta análise refere-se a ela e sua influência na rigidez. À temperatura elevada, 381K, as ligações de hidrogênio entre resíduos distantes na sequência são rompidas e aquelas entre vizinhos são preservadas. Como à temperatura mais alta o linker é mais rígido, podemos concluir que as ligações entre vizinhos são relacionadas com um processo que contribui à elasticidade, enquanto aquelas entre os resíduos afastados na sequência com um processo que tem o efeito oposto.

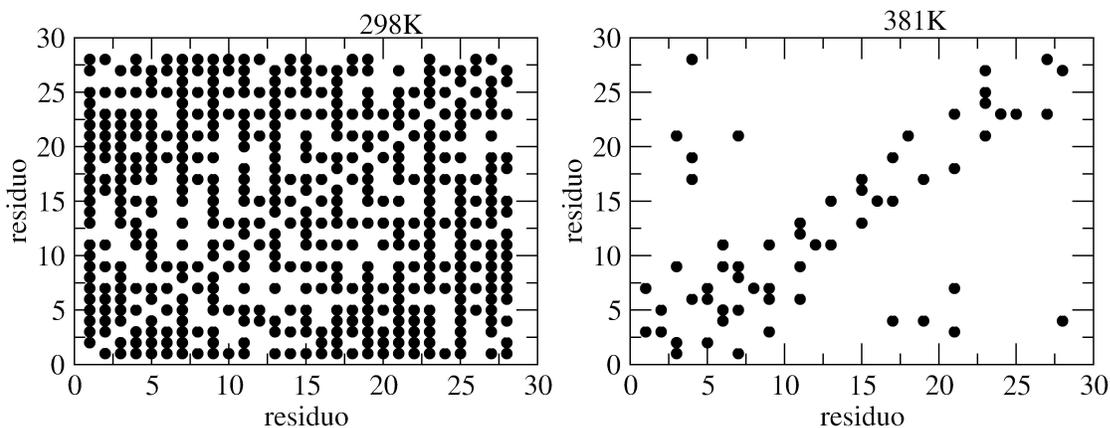


Figura 6.12. A distribuição espacial das ligações de hidrogênio em $(PT)_{12}$. Os pontos pretos representam existência das ligações de hidrogênio entre um par de resíduos. À temperatura mais alta, permanecem ligações entre resíduos vizinhos (as ligações perto à diagonal do gráfico).

O comportamento do linker nas duas temperaturas, 298K e 381K, foi visualizado através da representação das estruturas formadas a cada 1ns alinhadas em relação ao átomo central, o nitrogênio da Thr15. A 298K, o linker apresenta uma variedade maior das estruturas executando dobramentos em várias direções, Figura 6.13A. A diferença entre o tamanho de dobramentos nas duas temperaturas diferentes também foi representada pelo espaço varrido pelo linker. O volume que cada sistema ocupa durante o tempo de simulação foi calculado para a ocupância dos átomos igual ou maior de 4%, Figura 6.13B. No caso da temperatura de 298K o volume é 10608 \AA^3 , e no caso de 381K o volume é 9214 \AA^3 . O espaço varrido para a temperatura mais baixa é maior porque o linker executa mais dobramentos, em mais direções. Ambas as análises, a visualização das estruturas e o volume varrido, são consistentes com o resultado do cálculo do perfil de energia livre, e mostram a mudança estrutural menor, ou seja, uma rigidez maior no caso da temperatura mais alta.

À temperatura mais baixa, o linker executa dobramentos em várias direções, de várias maneiras diferentes de forma que exiba uma faixa maior de comprimento R , Figura 6.13A, o que implica em um perfil de energia livre mais largo, ou seja, a constante de força k menor comparado com as temperaturas altas. As ligações de hidrogênio entre os resíduos afastados na sequência são formadas durante estes dobramentos.

À temperatura mais alta, o linker exibe menos dobramentos e as ligações de hidrogênio entre os resíduos afastados na sequência são formadas em menor escala, restam apenas as ligações entre os vizinhos que contribuem à rigidez formando anel entre resíduos. Devido às flutuações térmicas maiores à temperatura mais alta, os resíduos afastados na sequência afastam-se e as ligações de hidrogênio entre eles são rompidas, enquanto os resíduos vizinhos sendo ligados covalentemente não se afastam e as ligações respectivas permanecem.

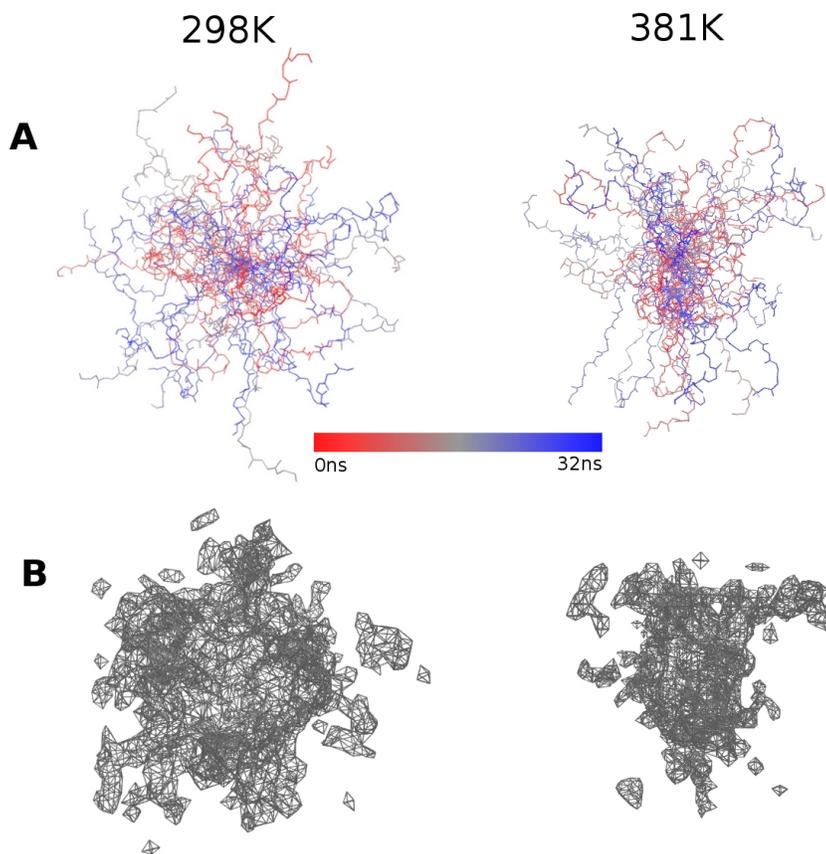


Figura 6.13. A: Instantes na simulação do linker $(PT)_{12}$ sobrepostos, cada 1 ns. Alinhamento dos instantes foi feito em relação ao átomo central do linker, o nitrogênio do resíduo Thr15. Representação de cadeia principal em cores de escala de tempo. **B:** Mapa volumétrico com isovalor de ocupância igual a 4%. À 298K, o linker percorre maior espaço.

6.4 Conclusão

Domínios funcionais das celulases que contêm mais que um domínio são conectados covalentemente por um peptídeo sem estrutura secundária definida, denominado *linker*. Além de papel de conectar os domínios, a função dos linkers é ainda pouco compreendida. Na hipótese do linker como fixador de distância entre os domínios de uma celulase, a rigidez poderia ser uma pro-

priedade desejável.

Aqui foi investigado o linker da endoglucanase de *Xanthomonas campestris* pv. *campestris* ((PT)₁₂) que apresenta uma sequência peculiar, formada por doze blocos de Pro-Thr.

Na busca de peculiaridades de comportamento do linker (PT)₁₂, o cálculo de propriedades relacionadas a rigidez foi empregado. Este linker foi comparado com um linker da sequência habitual, neste caso o linker da celobiohidrolase 2 de *Trichoderma reesei* (CBH2). O cálculo do perfil de energia livre ao longo de um parâmetro de ordem relacionado com a extensão do linker (distância ponta-a-ponta) foi feito. Para obter várias conformações de linker, uma técnica de amostragem ampliada foi usada nas simulações, a técnica de troca de réplicas.

O perfil de energia livre obtido foi comparado com as equações da lei de Hook e as informações sobre a rigidez e a extensão dos linkers foram obtidas. A rigidez foi estimada por outras análises também, como RMSF, a distribuição dos ângulos torcionais ou espaço varrido pelo linker. O linker (PT)₁₂ é mais rígido que o linker CBH2.

A sequência do linker (PT)₁₂ composta por um resíduo rígido, Pro, e um resíduo que induz rigidez, Thr, induz a rigidez do linker como um todo. A prolina apresenta um anel na sua estrutura, e este anel envolve a cadeia principal da proteína afetando assim a mobilidade da cadeia. A treonina poderia induzir rigidez pelo fato de que a sua cadeia lateral frequentemente forma ligação de hidrogênio com os resíduos vizinhos da cadeia de proteína. Desta maneira, a treonina forma um “anel” com outros resíduos, composto em parte por uma ligação de hidrogênio. A análise de ligações de hidrogênio mostra que estas ligações da cadeia lateral da treonina com o primeiro ou segundo vizinho da cadeia ocorrem nas simulações.

Já existem estudos que mostram que a glicosilação de linker aumenta a

extensão do linker CBH2 [76, 77] e aqui as simulações mostram que o linker (PT)₁₂ que não é glicosilado chega a ser estendido igualmente devido a sua sequência específica.

A mesma análise do perfil de energia livre foi realizada para várias temperaturas. A rigidez do linker (PT)₁₂ aumenta com o aumento da temperatura o que não é esperado e não ocorre no caso do linker com a sequência habitual, o CBH2. À temperatura mais baixa, as ligações de hidrogênio tanto entre resíduos vizinhos quanto entre os resíduos afastados na sequência ocorrem. Enquanto à temperatura mais alta, permanecem apenas as ligações entre os vizinhos que contribuem à rigidez formando anel entre resíduos.

Capítulo 7

Modelagem com SAXS

7.1 Introdução

A estrutura da enzima é fundamental para sua função. Para entender melhor o mecanismo físico das celulasas bimodais, ou aquelas que contém domínios muito móveis que não podem ser resolvidos pela cristalografia de raios X, é desejável obter um modelo estrutural para a enzima inteira. Neste capítulo apresentamos um modelo para a XccEG intacta, combinando dados de SAXS e modelagem molecular.

A técnica de Ajuste Flexível por Dinâmica Molecular (MDFF) foi desenvolvida para modelagem computacional a partir de uma estrutura atômica e a estrutura de crio-eletromicroscopia [21]. A crio-eletromicroscopia é um tipo de microscopia eletrônica de transmissão onde a amostra é estudada em temperaturas criogênicas (normalmente às temperaturas do nitrogênio líquido). Uma eventual vantagem desta técnica experimental é o fato de que ela pode fornecer a estrutura da macromolécula em ambiente nativo em oposição à cristalografia de raios X que requer a colocação de amostra em um ambiente não-fisiológico (cristal), o que pode ocasionalmente levar a modificações conformacionais funcionalmente irrelevantes. Recentemente, a estrutura do capsídeo do vírus da imunodeficiência humana (HIV) foi modelada com a técnica MDFF combinando a estrutura cristalográfica com a estrutura da crio-eletromicroscopia [87]. O es-

palhamento de raios X de baixo ângulo (SAXS) também fornece a estrutura da macromolécula em ambiente nativo, mas em baixa resolução e, ao contrário da crio-eletromicroscopia não captura a estrutura interna da proteína. O SAXS fornece apenas a superfície da molécula observada.

Aqui, a técnica MDFF será adaptada para ser usada com as estruturas de SAXS para modelagem de estrutura intacta de biomoléculas em solução. Depois disso, esta técnica será aplicada à modelagem da Endoglucanase de *Xanthomonas campestris pv. campestris* intacta (XccEG).

A sequência da XccEG é conhecida [79], mas somente a estrutura do CCD foi resolvida com resolução atômica por cristalografia de raios X [58]. O alvo é modelar a estrutura tridimensional intacta.

7.2 Preparação do sistema

Como foi explicado na Seção 3.6, para modelar a XccEG intacta, o método de Ajuste Flexível por Dinâmica Molecular foi adaptado [21]. Para isso, a curva SAXS obtida experimentalmente e o envelope SAXS modelado computacionalmente a partir desta curva [58] foram usados. O envelope de SAXS é uma su-



Figura 7.1. Envelope SAXS da XccEG. Representação de perspectiva.

perfície da enzima representada pelas esferas fictícias equidistantes que preenchem o volume definido pela superfície, Figura 7.1. No caso da estrutura SAXS da XccEG, a distância entre esferas é 5,8Å.

Como a cristalografia não é capaz de captar estruturas muito móveis, normalmente o que se tem é a estrutura atômica em partes, ou seja, o CCD e o CBM de uma celulase separadamente. A modelagem do CBM e do linker da XccEG foi feita por previsão de estrutura tridimensional com base nas sequências primárias no servidor I-TASSER, [82, 83]. De posse das estruturas em partes, partimos para modelar a orientação relativa dos domínios.

Docagem rígida: Antes da simulação, uma docagem rígida de baixa resolução do CCD e CBM foi realizada de forma independente dentro da estrutura de SAXS com o programa SITUS [88]. Para a docagem do CBM, o envelope de SAXS primeiro foi cortado manualmente na sua parte mais estreita e depois a estrutura atômica do CBM foi docada. A docagem do CCD foi realizada com o envelope SAXS inteiro.

O procedimento da docagem rígida pelo programa SITUS é o seguinte: um modelo de esferas é construído a partir da estrutura atômica e, em seguida, comparado com o modelo-alvo de esferas (o envelope de SAXS). Uma pesquisa exaustiva 6D, que é composta por três graus de liberdade translacionais e três rotacionais, foi feita até que o coeficiente de correlação mínimo entre os dois modelos de esferas fosse alcançado, equação 7.1, onde ρ é a densidade de pontos do modelo de esferas.

$$CCC = \int \rho_{alvo}(\vec{r})\rho(\vec{r})d^3\vec{r} \quad (7.1)$$

As estruturas resultantes de maior coeficiente de correlação são guardadas para a análise posterior.

Teste de docagem: A fim de testar o uso das estruturas SAXS na mode-

lagem de biomoléculas modulares, um dímero do chiquimato quinase I de *Escherichia coli* foi escolhido. Sua estrutura já foi resolvida pela cristalografia de raios X (PDBid: 1KAG [74]) e por SAXS (BIOISISid: 1KINHP [75]) o que possibilita comparação direta de resultado da modelagem com a estrutura cristalográfica.

O envelope SAXS foi cortado pela metade e cada monômero da estrutura cristalográfica foi docado rigidamente a cada metade do envelope independentemente. As docagens com maior coeficiente de correlação foram guardadas. Após a docagem, todas as combinações de dois monômeros docados independentemente foram comparadas com a estrutura cristalográfica tanto visualmente quanto calculando o valor de RMSD em relação à estrutura cristalográfica. Além disso, mais algumas análises foram realizadas para avaliar a docagem. Foram calculadas a grandeza χ , que apresenta a concordância com a curva SAXS experimental, a equação 3.25, e a porcentagem dos átomos pesados dentro do envelope. O critério para o cálculo da porcentagem dos átomos pesados dentro do envelope será explicado na Seção 7.3.

A Figura 7.2 mostra um arranjo dos monômeros em chiquimato quinase I de *Escherichia coli* diferente. É notável que o ângulo entre os eixos dos dois monômeros é diferente para a estrutura cristalográfica e a de SAXS. Isso provavelmente se deve ao fato de que proteína apresenta empacotamento diferente em cristal e em solução. Por isso, a estrutura cristalográfica não será necessariamente o alvo da modelagem.

O critério para escolher a docagem melhor será o valor mínimo de χ . A Figura 7.3 mostra os valores de χ , a porcentagem dos átomos pesados dentro do envelope e o RMSD em relação à estrutura cristalográfica para cada uma das combinações de docagem. O eixo x representa o índice de combinações de docagens dos dois monômeros, o primeiro dígito é de um monômero e o

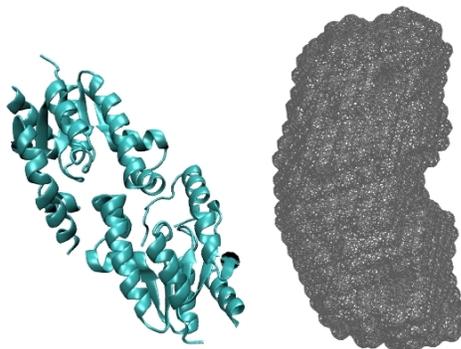


Figura 7.2. A estrutura cristalográfica do dímero de chiquimato quinase I de *Escherichia coli* (esquerda) e o seu envelope SAXS (direita).

Tabela 7.1. Docagens escolhidas pelo valor mínimo de χ . A porcentagem dos átomos pesados dentro do envelope e RMSD em relação à estrutura cristalográfica estão representados para cada docagem.

Docagem	48	18	49	46	11	19	68	59
χ	2,91	3,05	3,14	3,22	3,26	3,27	3,35	3,38
N / %	93,38	92,40	95,98	94,79	94,90	95,01	93,70	95,55
RMSD / Å	15,26	13,32	17,38	11,68	9,96	16,69	12,31	14,14

segundo é de outro monômero. Por exemplo: a docagem 43 é a combinação da quarta docagem independente do primeiro monômero e a terceira docagem do segundo monômero. As oito estruturas com o menor valor de χ estão representadas em barras vermelhas e o mínimo global está enquadrado. A Tabela 7.1 mostra valores de χ , a porcentagem dos átomos dentro do envelope e o RMSD para estas oito melhores docagens. Observamos que o menor valor de χ não corresponde à porcentagens máximas dos átomos dentro do envelope, nem aos menores valores de RMSD. Também, os oito menores valores de χ não correspondem às porcentagens máximas dos átomos pesados, nem aos

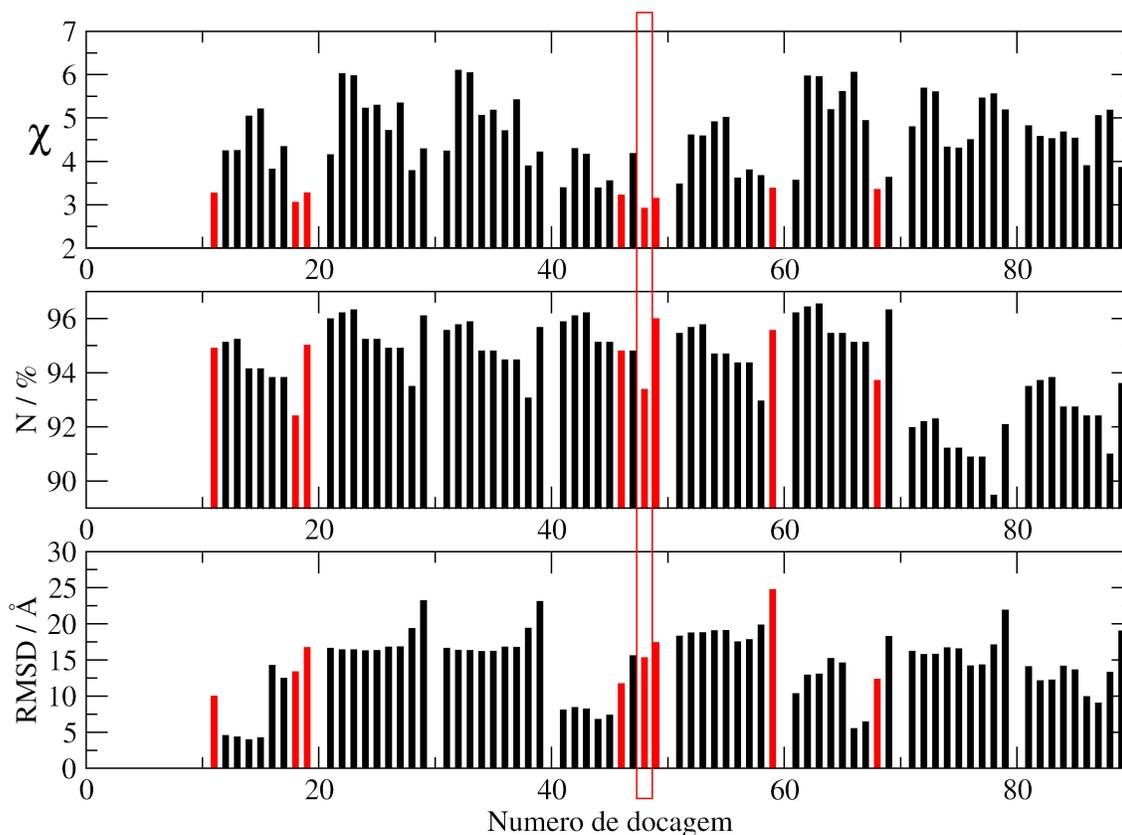


Figura 7.3. Valores de χ , porcentagem dos átomos pesados dentro do envelope e RMSD em relação à estrutura cristalográfica para cada uma das combinações de docagem. O eixo x apresenta número de combinação de docagens de dois monômeros, o primeiro dígito é de um monômero e o segundo é de outro monômero. As oito estruturas com o menor valor de χ estão representadas em barras vermelhas e o mínimo global está enquadrado.

valores de RMSD mínimos. Ou seja, no caso deste dímero, os modelos que mais concordam com a curva experimental não correspondem a sobreposição máxima com o envelope nem com a estrutura cristalográfica.

As oito melhores estruturas do dímero estão desenhadas na Figura 7.4. Os

números de baixo de cada estrutura indicam índices de docagens da Tabela 7.1. Para comparar a semelhança entre os arranjos dos dois monômeros, todas as estruturas foram alinhadas em relação a um dos monômeros (o monômero da parte de baixo do dímero) e o início da sequência de cada monômero foi marcado pela bolinha vermelha. Vemos que os melhores modelos apre-

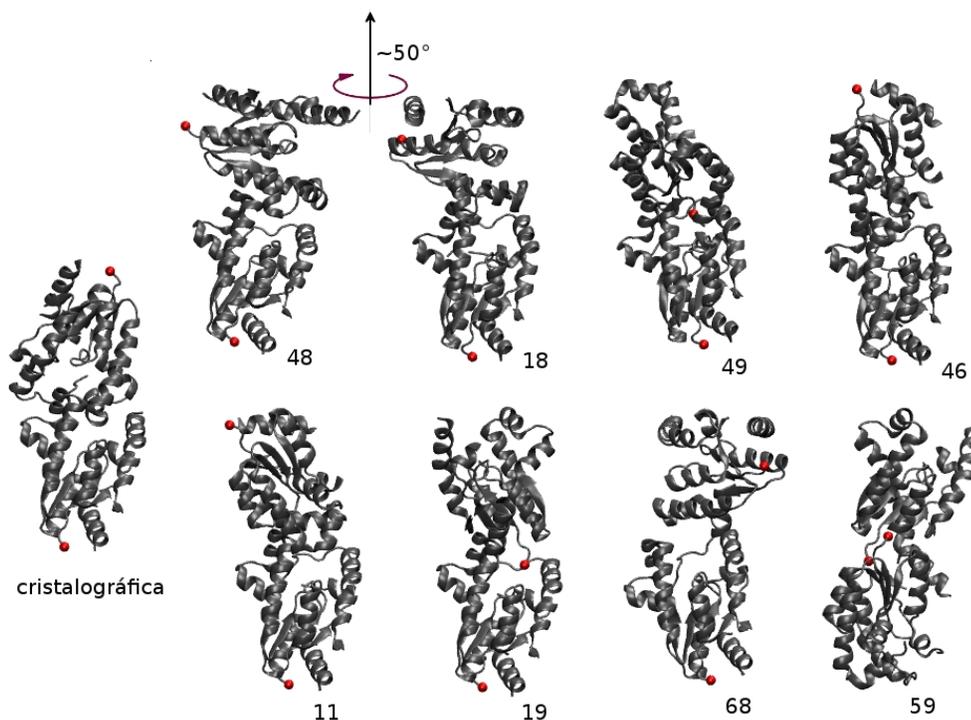


Figura 7.4. Estrutura cristalográfica do dímero do chiquimato quinase I de *Escherichia coli* (PDBid: 1KAG) e oito estruturas modeladas de menor valor de χ . Os números indicam números de docagens da Tabela 7.1. Os modelos estão alinhados em relação a um dos monômeros (o monômero da parte de baixo do dímero). Bolinha vermelha representa o início da sequência de cada monômero.

sentam orientação relativa entre os monômeros diferente da estrutura cristalográfica e também diferente entre eles. Os dois melhores modelos, 48 e 18, têm a orientação similar, os inícios das sequências apontando à extremidade da proteína, mas os seus monômeros fazem um ângulo de 50 graus aproxi-

madamente. No caso deste dímero, existem várias maneiras diferentes para docar no envelope com uma faixa estreita de valores de χ . A docagem não é específica suficientemente para este tipo de proteína.

Como o envelope de SAXS representa uma estrutura de resolução muito baixa, a especificidade da docagem no envelope depende do forma da molécula. Quanto mais irregularidades existem na forma de molécula, mais especificamente ela se encaixa dentro do envelope SAXS. Uma molécula de alto grau de simetria pode ser docada em várias maneiras diferentes ao envelope. Além disso um outro critério baseado nos resultados experimentais ou computacionais pode ser usado como critério de escolha do melhor modelo junto com o critério de maior concordância com a curva experimental, χ .

A proteína modelada neste capítulo, a Endoglucanase de *Xanthomonas campestris pv. campestris*, apresenta uma simetria muito mais baixa do que a do chiquimato quinase I de *Escherichia coli*, Figura 7.1. Portanto, existe esperança para especificidade maior da modelagem.

Mapa volumétrico: Para docagem e também para as simulações, um modelo de SAXS no formato de um mapa volumétrico é usado. Portanto, um modelo de esferas de SAXS tem de ser convertido em um mapa volumétrico. A partir do modelo de esferas da endoglucanase de *Xanthomonas campestris pv. campestris*, o mapa volumétrico foi construído adicionando uma função gaussiana com $\sigma = 2.9\text{Å}$ em cada ponto de uma rede de 1Å de espaçamento. O valor de σ foi escolhido como metade da distância entre duas esferas vizinhas que é constante para o modelo inteiro. O mapa volumétrico resultante representa uma soma de funções gaussianas em todos os pontos do espaço. Os valores do mapa volumétrico são normalizados de forma que tomem valores de 1 até 0. Zero representa a máxima densidade de elétrons do modelo de SAXS. Para definir o envelope, um isovalor específico do mapa volumétrico foi

determinado. Este valor deveria representar a superfície que abrange todas as esferas no modelo SAXS. Então, ele foi determinado de tal forma que coincide com a superfície das esferas sondada com uma esfera de raio igual a metade da distância entre duas esferas vizinhas. Este isovalor do mapa volumétrico foi igual a 0,8.

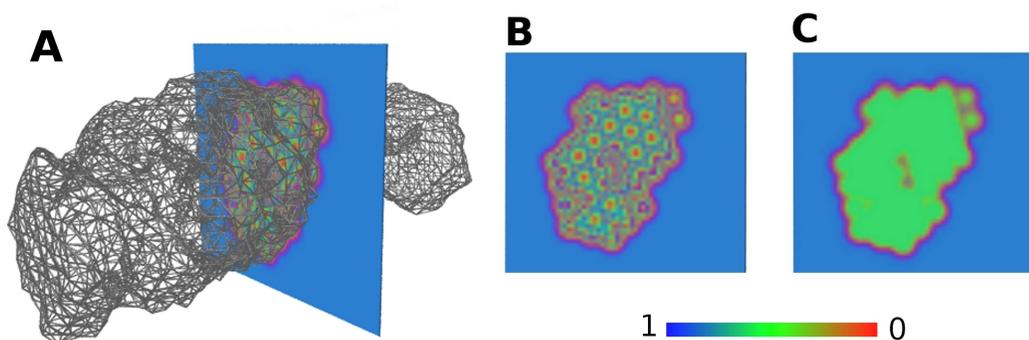


Figura 7.5. Mapa volumétrico de densidades eletrônicas obtidas por SAXS. **A:** Envelope SAXS representado como isovalor 0,8 do mapa volumétrico. **B:** Fatia do mapa volumétrico que inclui imagem de esferas equidistantes. **C:** Fatia do mapa volumétrico com os valores do mapa truncados no 0,8, cada valor de menor que 0,8 foi substituído por 0,8.

Parâmetros para simulações: Como as esferas do modelo SAXS são fictícias, elas foram retiradas após construção do mapa volumétrico, Figura 7.5, para uso nas simulações. Cada valor do mapa de menor que 0,8 foi substituído por 0,8. Após a docagem rígida, a estrutura atômica intacta foi hidratada e eletroneutralizada em uma caixa de NaCl-água de $1g/cm^3$ de densidade. Uma caixa de simulação com água explícita de tamanho $140\text{Å} \times 70\text{Å} \times 70\text{Å}$ foi construída. A camada de água em volta da enzima é de pelo menos 10Å de espessura. Simulações foram feitas tratando esta caixa de simulação no ensemble NPT a 298K e 1atm. No tratamento de interações não-covalentes de curto alcance foi usado raio de corte de 10Å e a função suavizante depois de 9Å . Na integração das equações de movimento foi usado um passo de tempo

mais curto, de 1fs, para evitar que a força externa do envelope SAXS causasse colapso das partículas.

7.3 Resultados e discussões

Estrutura inicial. As duas estruturas atômicas, de CCD e CBM, foram docadas rigidamente e independentemente no envelope SAXS através do programa SITUS [88]. De 20 resultados de cada um dos domínios duas combinações CCD-CBM com os menores valores de χ foram escolhidas, $\chi = 3,801$ e $\chi = 3,699$, com as orientações de CCD diferentes, Figura 7.6. A porcentagem de átomos pesados O, C, N e S dentro do envelope foi calculada definindo o envelope como a superfície do mapa volumétrico com isovalor de densidade eletrônica 0,8. Os modelos representam 81,19% e 79,72% de átomos pesados dentro do envelope, respectivamente. Este número não precisa ser necessariamente 100% porque o envelope representa uma média de conformações em solução e somente poucas conformações durante a simulação vão estar 100% dentro do envelope.

Na estrutura do CBM, existem três resíduos aromáticos que são expostos ao solvente e poderiam atuar no reconhecimento de substrato. É interessante observar que em todas as docagens o eixo dos três resíduos aromáticos do CBM (em vermelho) encontra-se perpendicular ao eixo da fenda catalítica do CCD. Neste modelo uma cadeia de celulose interagiria apenas com um dos três resíduos aromáticos, Figura 7.6.

As estruturas foram completadas com o linker modelado por ITASSER com pontas fixas às coordenadas dos terminais dos domínios já docados no envelope. Para escolher entre a estrutura A e B, uma modelagem do complexo enzima-cadeia de celulose foi feita, considerando a interação dos dois domínios com a mesma cadeia de celulose. A docagem da cadeia de celulose den-

tro da enzima intacta foi feita prolongando a celotetraose já docada e simulada (Capítulo 5) no sentido do CBM. Após uma longa minimização de 20 000 iterações, a celulose da estrutura B formou uma curva na saída do CCD. Esta saída é composta por um loop, chamado de loop5, que fecha a fenda catalítica impossibilitando assim passagem do substrato para o outro lado sem estar curvado, ver Figura 5.8 do Capítulo 5. Este loop é composto por uma parte altamente móvel e outra que nunca muda de posição e que fecha a fenda (azul). A estrutura onde o loop5 não impede o substrato de alcançar CBM, Figura 7.6A, foi escolhida. Entre as enzimas da mesma família (a família 5 das glicosidasas) cristalizadas com substrato, todas possuem esse tipo de loop e o substrato não passa além dele, Figura 7.6C. Entretanto, um substrato pode formar curvatura na região do loop, similar ao loop5, como no caso da endoglucanase E5 de *Thermobifida Fusca* (PDBid: 2CKR). A estrutura escolhida (A) após acréscimo do linker apresentou $\chi = 3,660$. Esta estrutura foi usada como ponto inicial para as simulações de MDFF.

Simulações de MDFF. A simulação MDFF no envelope foi feita definindo o fator de força $\xi = 0,3$. Este fator é arbitrário e não possui uma importância desde que seja diferente de zero, porque como a força que um átomo sente depende somente do gradiente do mapa volumétrico e dentro do envelope o valor da densidade é constante, esta força serve apenas para átomos fora do envelope entrarem no envelope. Dentro do volume do envelope não existe gradiente. Portanto, não existe a força MDFF e os átomos são submetidos a uma simulação convencional. Porém, esta força não deveria ser grande demais para não perturbar de modo significativo a energia potencial do sistema. Aqui o valor de 0,3 foi escolhido de modo que represente uma força da ordem das forças interatômicas.

Simulações direcionadas. O alvo é chegar a uma boa concordância com

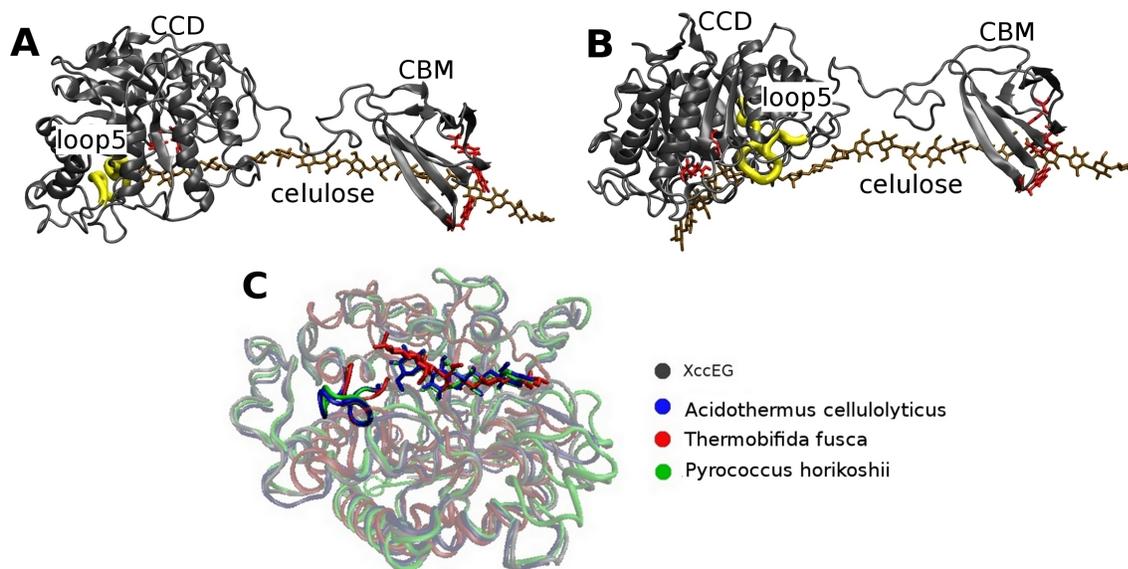


Figura 7.6. Modelos feitos pela docagem dos domínios separadamente. Resíduos catalíticos e resíduos aromáticos do CBM são representados em vermelho. **A:** O loop5 não impede a saída do substrato. **B:** O loop5 impede a saída do substrato. **C:** Enzimas da família 5 de glicosidases possuem um análogo ao loop5 e substratos que não ultrapassam esta barreira.

a curva SAXS experimental, ou seja, a um baixo valor de χ , simulando com vínculos de envelope. O valor ideal de χ seria $\chi=1$ o que significa que o modelo tem acurácia dentro do erro experimental de curva unidimensional SAXS. Na prática, é razoável chegar a um valor de χ similar ao da modelagem do próprio envelope tridimensional, $\chi= 2,299$.

Podemos ver na Figura 7.7, a primeira curva preta, que à temperatura ambiente e com a MD convencional, o χ não consegue abaixar em tempo razoável. Uma maneira de estimular amostragem é aumentar a temperatura do sistema. O aumento da temperatura aumenta módulo de velocidade média de sistema e assim os átomos ganham energia cinética para superar barreiras potenciais entre conformações diferentes, e conseguem amostrar mais o espaço dentro do envelope. Para amostragem ampliada não se pode usar temperatura alta

porque esta proteína possui estrutura secundária e pode acontecer desenovelamento (parcial ou total). No máximo podemos aplicar temperatura alta ao linker que não é bem estruturado. Uma vez atingido o subespaço de conformações de

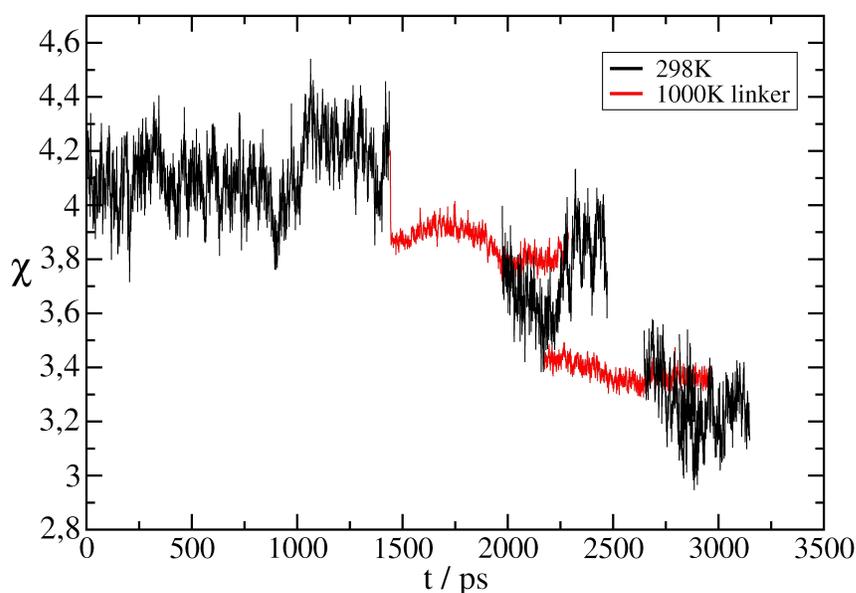


Figura 7.7. Concordância com a curva SAXS experimental durante o tempo de simulação MDFF. As curvas vermelhas representam simulação à 1000K de temperatura em que somente o linker foi livre para movimentos.

baixo χ , podemos resfriar o sistema à temperatura ambiente de novo e assim “congelar” a estrutura favorável, deixando o sistema inteiro executar movimentos dentro do espaço do envelope. Desta simulação à temperatura ambiente, a conformação com o χ mínimo é escolhida como ponto inicial para simular, esquentando somente o linker a 1000K de novo e assim por diante favorecendo a ida em direção ao χ baixo, sem usar nenhum potencial tendencioso. Este procedimento aqui usado assemelha à Amostragem de Fluxo de Ida [89]. A técnica de Amostragem de Fluxo de Ida escolhe uma grandeza discreta que represente a coordenada desejada de um evento simulado e para cada passo ao longo da coordenada, os trechos das simulações que não atingiram um valor da coord-

nada são descartados. A diferença é que aqui a coordenada χ foi tratada como uma grandeza contínua.

Durante o procedimento de modelagem por MDFF, as estruturas internas dos domínios são preservadas, que mostra valor baixo de RMSD e a porcentagem dos átomos pesados dentro do envelope cresce até 82%, Figura 7.8. O valor de RMSD foi calculado em relação à estrutura inicial da Figura 7.6A.

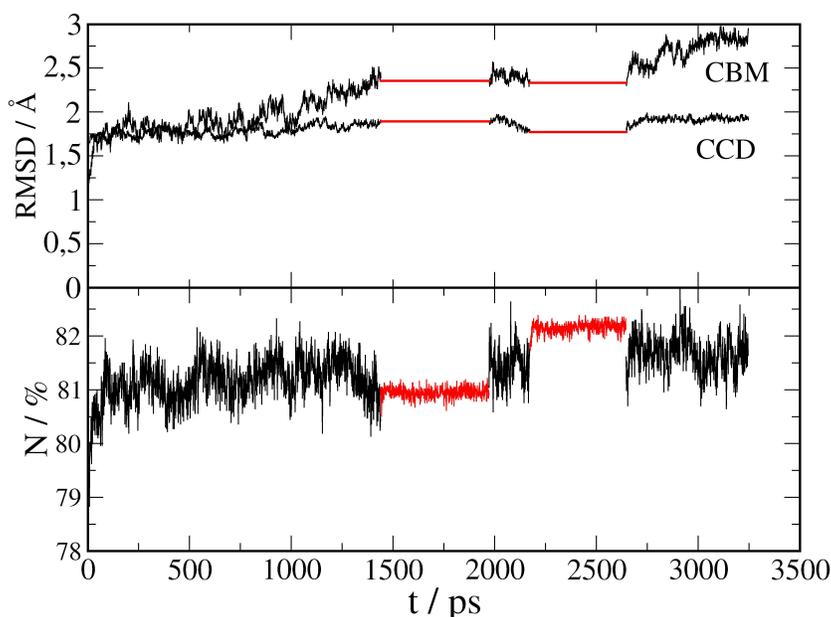


Figura 7.8. Durante o procedimento de modelagem por MDFF, estruturas internas dos domínios são preservadas, que mostra valor baixo de RMSD e a porcentagem dos átomos pesados dentro do envelope cresce até 82%.

Equilíbrio. A fim de testar o modelo obtido é necessário checar a sua estabilidade estrutural após interromper a aplicação da força vinculada ao envelope SAXS. Após o valor mínimo de χ obtido por esta técnica de escolher ponto de χ_{min} de cada trecho de trajetória, foi realizada uma simulação convencional sem força aplicada. A simulação foi realizada até chegar no equilíbrio representado

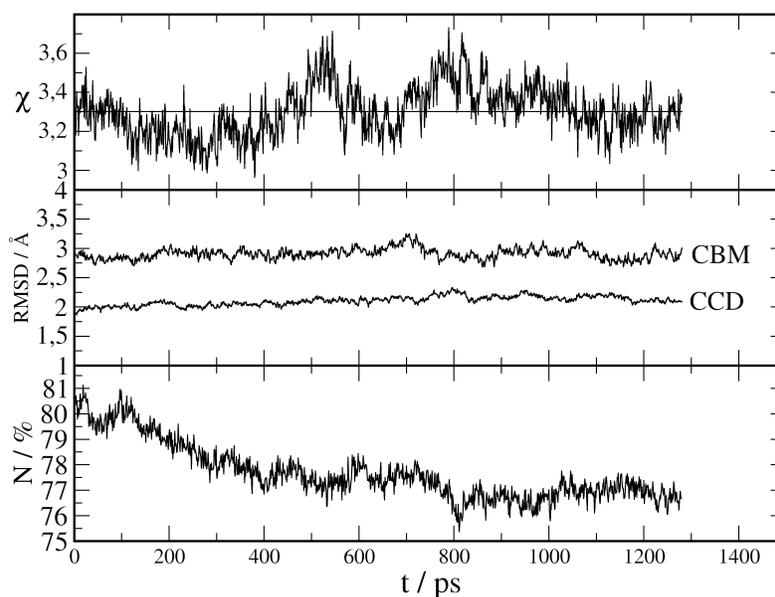


Figura 7.9. Sem vínculos de envelope, o sistema é capaz de manter χ em torno de 3,3 enquanto o RMSD e a porcentagem de átomos pesados dentro do envelope exibiram patamar de valores razoáveis.

pelo patamar de RMSD e a porcentagem de átomos pesados dentro do envelope. O sistema manteve χ em torno de 3,3 e um valor baixo de RMSD dos dois domínios. A porcentagem de átomos pesados dentro do envelope alcança 77% após atingir patamar em menos de 1ns, Figura 7.9. A Figura 7.10 mostra a estrutura inicial e final depois da modelagem com a técnica MDFF.

O modelo resultante. Depois das simulações, os resíduos aromáticos do CBM permaneceram igual à docagem inicial, ou seja, perpendiculares à fenda do CCD, Figura 7.11. Cada subsítio em CCD consiste em pelo menos um resíduo aromático. Estes resíduos foram identificados calculando a energia de interação com substrato, Figura 5.7 do Capítulo 5, e são: Phe50, Tyr337, Trp237 e Phe268, representados em vermelho.

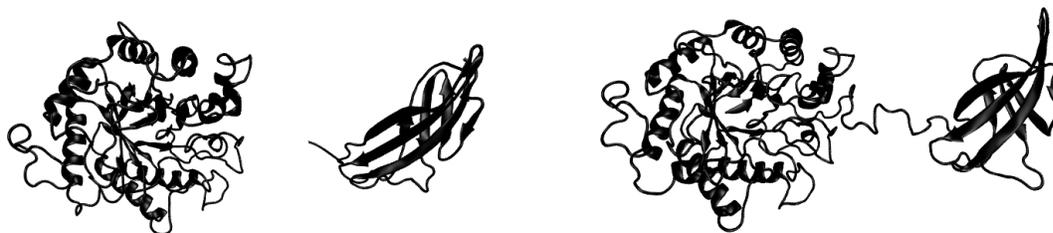


Figura 7.10. A estrutura inicial (à esquerda) e final (à direita) da modelagem da XccEG intacta em solução.

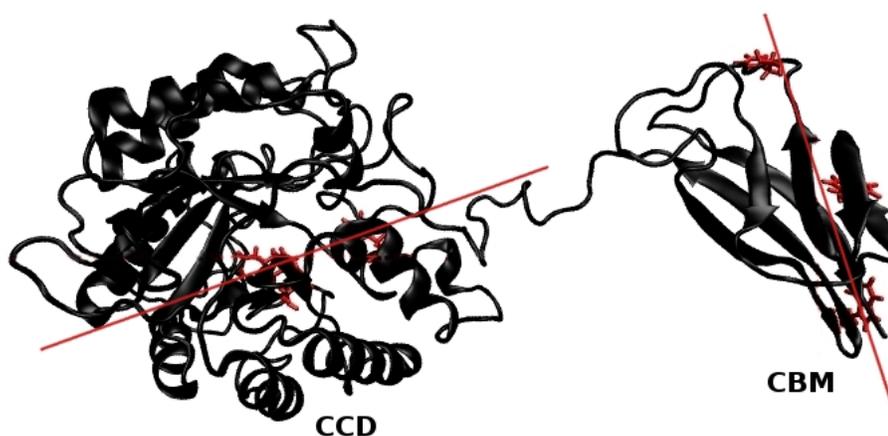


Figura 7.11. O modelo final. Os resíduos aromáticos estão em vermelho. As linhas vermelhas representam os eixos de CCD e CBM.

O domínio CBM da XccEG apresenta um arranjo de resíduos aromáticos característico para celulasas ativas em celulose cristalina [66]. Em contato com a celulose cristalina, a XccEG poderia interagir com mais de uma cadeia da celulose, Figura 7.12A. O modelo obtido da XccEG foi posicionado em cima da celulose cristalina de forma que o eixo do CCD, representado pelos quatro resíduos aromáticos da fenda catalítica, esteja alinhado com uma cadeia da celulose. Nesta situação, o eixo do CBM encontra-se perpendicular às cadeias da celulose abrangendo assim cinco cadeias da celulose. O modelo da celulose foi construído através do programa Cellulose-builder [90].

Esta orientação relativa dos domínios é diferente daquela do modelo pro-

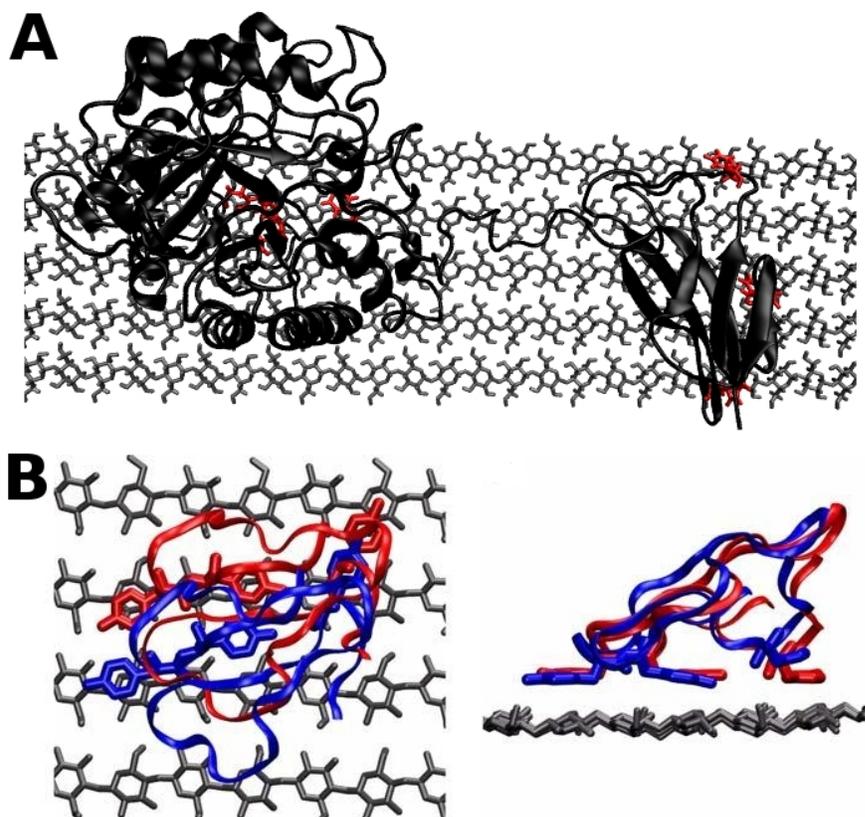


Figura 7.12. **A:** O modelo final da modelagem da XccEG posicionado em cima da celulose cristalina. **B:** O modelo de CBM da celobiohidrolase I de *Trichoderma reesei* [92] inicial (azul) e após simulação (vermelho). CBM desenhado como faixa e os resíduos aromáticos como barras.

posto para a celobiohidrolase I de *Trichoderma reesei*, [91, 92]. A celobiohidrolase I de *Trichoderma reesei* foi modelada conectando manualmente as pontas do CCD e CBM com um linker previamente modelado e relaxado pela MD. Este modelo apresentou os eixos de CCD e CBM paralelos. A Figura 7.12B foi adaptada de [92] e representa a estrutura apenas do CBM em cima da celulose cristalina. A estrutura azul é a estrutura inicial docada na celulose e a estrutura vermelha é a estrutura final depois de simulação de 1,5ns. Partindo da estrutura inicial que interage com duas cadeias da celulose, o CBM deslocou-

se para a posição em que interage com apenas uma cadeia da celulose através dos resíduos aromáticos característicos.

A fim de testar estabilidade das interações com a celulose na orientação proposta pelo nosso modelo, simulamos somente o domínio CBM em cima de uma camada de celulose. O CBM foi posicionado manualmente na superfície de uma camada de celulose cristalina de tamanho de sete cadeias contendo sete unidades glicosídicas. O posicionamento foi feito de forma que os três resíduos aromáticos característicos se encontrem na distância de no mínimo 2 Å do substrato, Figura 7.13(azul). O complexo CBM-camada de celulose foi hidratado por água em caixa de aproximadamente $(45 \times 62 \times 50) \text{Å}^3$. A posição da camada de celulose estava fixa durante a simulação.

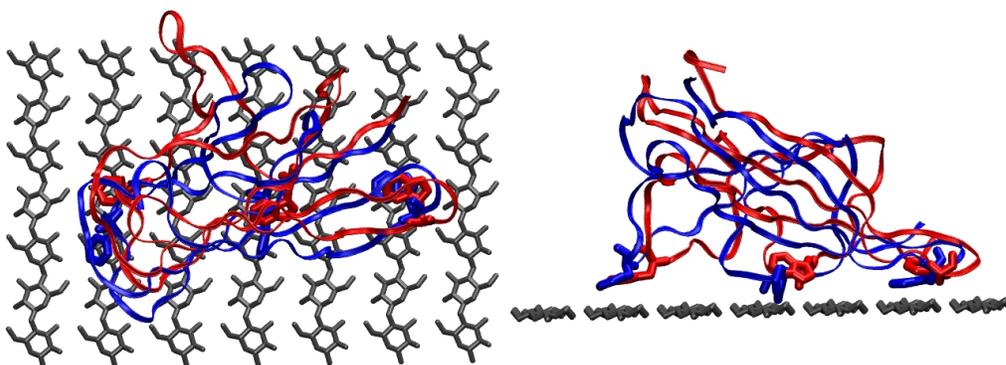


Figura 7.13. Posição do CBM da endoglucanase de *Xantomonas campestris* pv. *campestris* em cima da camada da celulose, inicial (azul) e após 40ns de simulação (vermelho). CBM desenhado como faixa e os resíduos aromáticos como barras.

Depois de 40ns de simulação o CBM permaneceu na posição perpendicular às cadeias da celulose, Figura 7.13(vermelho), ou seja, perpendicular à fenda catalítica do CCD. Os três resíduos aromáticos permaneceram na distância de no mínimo 3 Å do substrato. As simulações sugerem que tanto em solução quanto em complexo com substrato, a orientação do CBM da endoglucanase de *Xantomonas campestris* pv. *campestris* é perpendicular à fenda catalítica do

CCD.

Comparação de CCD em solução e cristal. Para investigar as diferenças entre a estrutura cristalográfica e o resultado da modelagem pelo ajuste ao envelope SAXS, algumas análises comparativas de estrutura e dinâmica foram realizadas. As trajetórias usadas são a parte equilibrada do modelo após ajuste com SAXS, Figura 7.9, e a simulação partindo da estrutura cristalográfica, Capítulo 5. Para a atribuição da diferença estrutural, a função RMSF foi empregada, e a estrutura de referência para as duas trajetórias foi a estrutura cristalográfica. As mudanças estruturais são presentes especialmente nas regiões de contato com outras moléculas da unidade simétrica em cristal, Figura 7.14 em cores. A análise de RMSF por blocos não mostra maiores diferenças na mobilidade da enzima, Figura 7.14, painel de baixo.

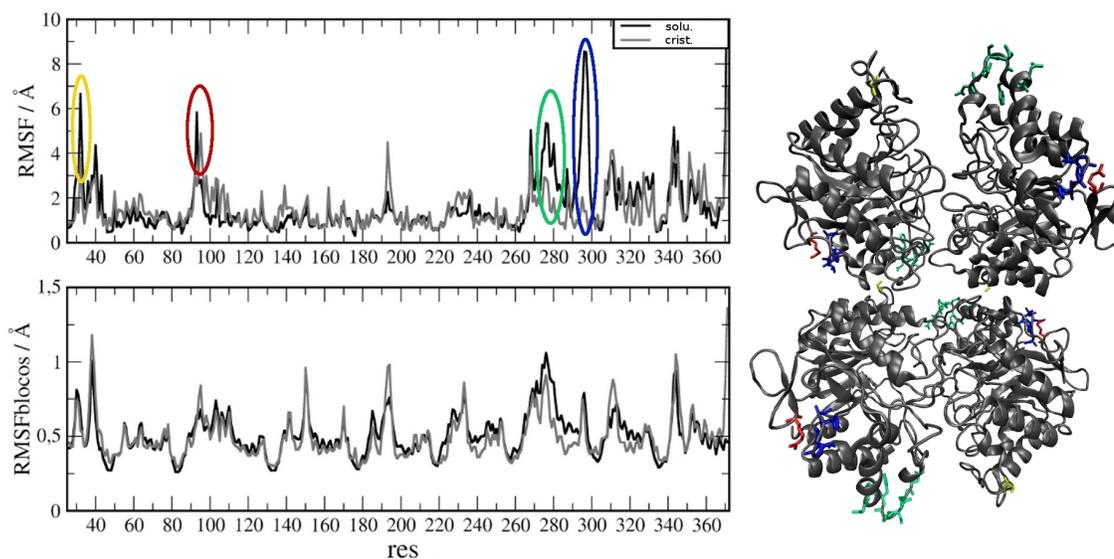


Figura 7.14. Diferenças estruturais (RMSF) e dinâmicas (RMSFblocos) entre o modelo após ajuste com SAXS e cristalográfico do CCD da XccEG. A unidade simétrica cristalográfica (direito).

Docagem alternativa. Uma modelagem em que o eixo dos três resíduos aromáticos do CBM é paralelo com o eixo da fenda do CCD foi realizada ao fim

de descartar a possibilidade desta posição, Figura 7.15.

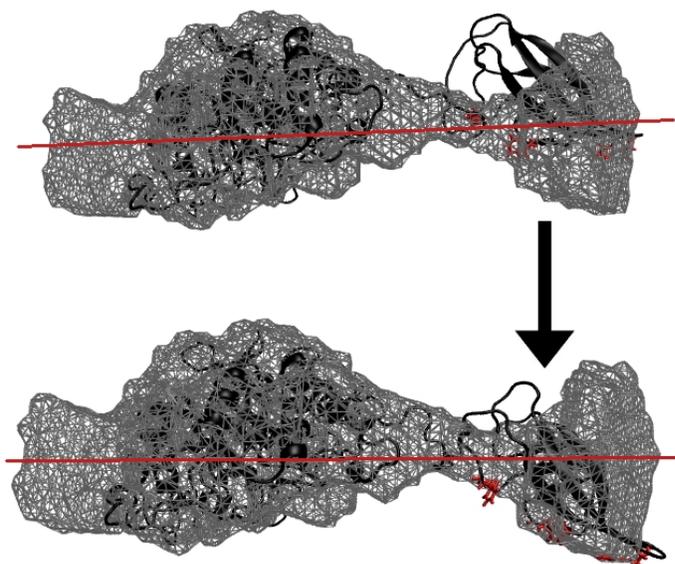


Figura 7.15. O instante inicial e final de uma simulação MDFF de 1ns, partindo do modelo do CBM paralelo à fenda do CCD. Eixo do CCD e resíduos aromáticos do CBM são representados em vermelho.

Nesta posição, todos os três resíduos aromáticos de CBM interagiriam com a mesma cadeia da celulose. A porcentagem dos átomos pesados dentro do envelope neste modelo inicial é 44,7. O mesmo procedimento de simulação por MDFF foi realizado por 1ns. Durante esta simulação, o CBM abandonou a posição inicial entrando no envelope até 73,98 % dos átomos pesados. A mudança da estrutura do CBM foi maior que no caso da docagem original, representando o valor de RMSD em relação ao modelo inicial de eixos paralelos igual a 3,60 Å. A mudança da estrutura do CBM na docagem original é no máximo 3,25 Å. Isto significa que o CBM saiu de uma posição não-favorável em pouco tempo e distorceu sua estrutura em relação à cristalográfica.

As simulações realizadas aqui sugerem que o CBM da endoglucanase de *Xanthomonas campestris pv. campestris* prefere a posição em que o seu eixo

é perpendicular ao eixo do CCD e assim possa interagir com cinco cadeias celulósicas diferentes. Isto se contrasta ao modelo da celobiohidrolase I de *Trichoderma reesei* e uma expectativa geral de que os dois domínios sejam ligados a uma cadeia de celulose.

7.4 Conclusão

O conhecimento da estrutura de uma enzima intacta, contendo todos os domínios, é fundamental para o conhecimento do seu mecanismo. Atualmente não existem estruturas intactas para celulasas com resolução atômica. Para este fim, a combinação de uma estrutura intacta de baixa resolução e uma estrutura incompleta de alta resolução pode ser útil.

A técnica de Ajuste Flexível por Dinâmica Molecular (MDFF) originalmente foi desenvolvida para as estruturas de crio-eletromicroscopia, mas aqui foi adaptada a ser usada com as estruturas de espalhamento de raios X de baixo ângulo (SAXS). Uma estrutura de SAXS apresenta resolução mais baixa do que uma estrutura de crio-eletromicroscopia, mais especificamente, ela apresenta apenas a superfície de molécula sem a sua estrutura interna. Por isso, no procedimento de criação do mapa volumétrico usado nas simulações de MDFF, toda a estrutura interna do mapa foi retirado: os valores do mapa que representam a densidade eletrônica SAXS maior que na superfície foram substituídos pelo valor da superfície.

Além disso, uma continuação do procedimento da modelagem com SAXS foi proposto. Ela consiste da verificação da concordância do modelo obtido com os dados experimentais, em cada passo da simulação de MDFF. Esta concordância é medida pela função χ , que mede a diferença entre a curva unidimensional experimental de SAXS e a curva de espalhamento teórico obtido com base na estrutura modelada.

A modelagem da endoglucanase de *Xanthomonas campestris pv. campestris* (XccEG) intacta foi realizada combinando a estrutura cristalográfica e a estrutura de SAXS da XccEG, utilizando a técnica de MDFF adaptada. O ponto inicial para simulações normalmente afeta muito os resultados porque a MD convencional não é capaz de amostrar o espaço de fase inteiro em tempo razoável. Por isso, primeiro foi docada a estrutura de resolução atômica no envelope SAXS e a orientação relativa dos domínios foi escolhida com cuidado. Depois disso, a simulação de MDFF foi realizada. A fim de acelerar a modelagem, um procedimento de direcionar simulações no sentido de melhorar a concordância com a curva 1D experimental foi feito. A meta foi atingida: foi obtida a estrutura da XccEG em solução em concordância com curva experimental razoável em tempo menor que 4ns. A estrutura final da XccEG apresenta uma orientação do CBM perpendicular ao eixo do CCD. Assim, a enzima poderia interagir com mais que uma cadeia da celulose no estado cristalino.

Capítulo 8

Conclusões gerais

Foram investigados quatro sistemas relacionados aos mecanismos enzimáticos para a hidrólise da celulose. Um deles, o estudo dos mecanismos de reconhecimento de substrato por endoglucanase 3 de *Trichoderma harzianum* foi publicado [57]. O trabalho sobre a seletividade da Endoglucanase de *Xanthomas campestris pv. campestris* foi concluído e será submetido em breve para publicação. O trabalho sobre a elasticidade de linker e o trabalho sobre a modelagem com SAXS estão em fase de preparação.

8.1 Endoglucanase sem CBM

A presença do domínio de ligação ao carboidrato (CBM) nas celulasas mostrou-se importante para o mecanismo de ligação não-covalente ao substrato. Simulações de Dinâmica Molecular foram realizadas com uma celulase que não possui CBM, a Endoglucanase 3 de *Trichoderma harzianum*. As análises de interações não covalentes enzima-substrato mostraram que a ausência de CBM é compensada pela presença de resíduos similares aos resíduos que reconhecem o substrato em um CBM.

Além da explicação da atividade na ausência de CBM, foi explicado porque esta enzima age em substratos com no mínimo cinco unidades glicosídicas. A tríade catalítica encontra-se espacialmente longe dos resíduos aromáticos que mostram-se importantes para a interação enzima-substrato, e por isso o

substrato precisa ter um comprimento mínimo.

8.2 Seletividade da Endoglucanase Xcc

Uma questão importante para desenhar enzimas de maior eficácia é entender as causas de seletividade e eventualmente mudá-las. As simulações de MD explicaram a causa da seletividade da Endoglucanase de *Xanthomonas campestris pv. campestris*. Esta enzima cataliza a hidrólise de cadeias de celulose de comprimento de quatro ou mais unidades glicosídicas. A fenda da enzima é composta por quatro subsítios de ligação à celulose que são constituídos por loops altamente móveis. Para o substrato permanecer ligado dentro da fenda catalítica, e assim na posição certa para a hidrólise acontecer, todos os subsítios precisam estar estabilizados pelas interações com substrato.

8.3 Propriedades mecânicas do *linker* da XccEG

A sequência específica de 12 blocos de Thr-Pro do linker da Endoglucanase de *Xanthomonas campestris pv. campestris* (XccEG) foi investigada calculando propriedades mecânicas extraídas da simulação de amostragem ampliada. O linker foi comparado com um linker de sequência habitual contendo os resíduos P,G,A,S,T,R e V, da celobiohidrolase 2 de *Trichoderma reesei* (CBH2). O linker (PT)₁₂ é mais rígido e essa rigidez cresce ligeiramente com aumento de temperatura. O resíduo rígido Pro e a Thr induzem uma rigidez do linker como todo. Existem evidências computacionais provas de que a glicosilação aumenta a extensão do linker CBH2 [77]. Nossos estudos apontam que o linker (PT)₁₂, que não é glicosilado, chega a ser estendido igualmente por causa da sua sequência específica.

8.4 Modelagem com SAXS

A técnica de Ajuste Flexível por Dinâmica Molecular originalmente desenvolvida para as estruturas da crio-eletromicroscopia foi aplicada para a estrutura SAXS. Como esta estrutura representa apenas a superfície da macromolécula preenchida pelos átomos fictícios, toda a estrutura interna de mapa volumétrico é retirada no procedimento de preparação de força adicional. O procedimento aqui desenvolvido segue no sentido de comparação com a curva SAXS experimental e maximização desta similaridade obtida por uma simulação direcionada. Este procedimento foi aplicado à modelagem da endoglucanase de *Xanthomonas campestris pv. campestris* XccEG intacta combinando a estrutura cristalográfica com a estrutura de SAXS. A meta foi atingida: foi obtida estrutura da XccEG em solução com uma concordância com curva experimental razoável. O modelo aqui proposto, ao contrário do outro disponível na literatura, tem o CBM interagindo com três cadeias de celulose paralelas.

Alanina	Ala	A	Leucina	Leu	L	Glutamina	Gln	Q	Serina	Ser	S
Arginina	Arg	R	Lisina	Lys	K	Glutamato	Glu	E	Treonina	Thr	T
Asparagina	Asn	N	Metionina	Met	M	Glicina	Gly	G	Triptofano	Trp	W
Aspartato	Asp	D	Fenilalanina	Phe	F	Histidina	His	H	Tirosina	Tyr	Y
Cisteína	Cys	C	Prolina	Pro	P	Isoleucina	Ile	I	Valina	Val	V

Apêndice A

Códigos de aminoácidos de uma e três letras

Referências Bibliográficas

- [1] Moon R. J. Martini A., Nairn J. Simonsen J. and Youngblood J. Cellulose nanomaterials review: structure, properties and nanocomposites. *Chemical Society Reviews*, 40:3941–399, 2011.
- [2] Ozioko P. C., Ikeyi Adachukwu I. P. and Ugwu O. P. C. Review article: Cellulases, their substrates, activity and assay methods. *The Experiment*, 12(2):778–785, 2013.
- [3] Voet D., Voet J. G. and Pratt C. W. Fundamentos de Bioquímica. *Trad. Porto Alegre: Artmed Editora.*, 12(2):778–785, 2000.
- [4] Stam M. R. Blanc E, Coutinho PM and Henrissat B. Evolutionary and mechanistic relationships between glycosidases acting on α - and β - bonds. *Carbohydrate Research*, 340:2728–2734, 2005.
- [5] Davies G and Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure*, 15.3:853–859, 1995.
- [6] Isom D. G., Castañeda C. A., Cannon B. R. and García-Moreno B. E. Large shifts in pKa values of lysine residues buried inside a protein. *PNAS*, 108:5260–5265, 2011.
- [7] Wang Q., Graham R. W., Trimbur D., Warren R. A. J. and Withers S. G. Changing enzymatic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. *Journal of the American Chemical Society*, 116:11594–11595, 1994.
- [8] Koshland M. E. The origin of fecal antibody and its relationship to immunization with adjuvant. *Journal of Immunology*, 70:359–365, 1953.
- [9] Vasella A., Davies G. J. and Bohm M. Glycosidase mechanisms. *Current Opinion in Chemical Biology*, 6:619–629, 2002.
- [10] Henrissat B. and Davies G. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology*, 7:637–644, 1997.

- [11] Teeri T. T. *Crystalline cellulose degradation: new insight into the function of cellobiohydrolases.*, volume 15. Elsevier Science Ltd, 1997.
- [12] McIntosh L. P. Hand G., Johnson P. E., Joshi M. D., Korner M., Plesniak L. A., Ziser L., Wakarchuk W. W. and Withers S. G. The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a ¹³C-NMR study of *Bacillus circulans* xylanase. *Biochemistry*, 35:9958–9966, 1996.
- [13] Nidetzky B. Steiner W., Hayn M. and Claeysens M. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: model for synergistic interaction. *Biochemical Journal*, 298:705–710, 1994.
- [14] Kostylev M. and Wilson D. Synergistic interactions in cellulose hydrolysis. *Biofuels*, 3:61–70, 2012.
- [15] Ganner T., Bubner P., Eibinger M., Mayrhofer C., Plank H., Nidetzky B. Dissecting and reconstructing synergism - in situ visualization of cooperativity among cellulases. *The Journal of Biological Chemistry*, 287: 43215–43222, 2012.
- [16] Arantes V. and Saddler J. N. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for Biofuels*, 3:4, 2010.
- [17] Irwin D. C. Spezio M., Walker L. P. and Wilson D. B. Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. *Biotechnology and Bioengineering*, 42:1002–1013, 1993.
- [18] <http://www.pdb.org/pdb/home/home.do>.
- [19] Merzel F. and Smith J. C. SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. *Biological Crystallography*, 58:242–249, 2002.
- [20] Lindner B. and Smith J. C. SASSIM - X-ray and neutron scattering calculated from molecular dynamics trajectories using massively parallel computers. *Computer Physics Communications*, 183:1491–1501, 2012.
- [21] Trabuco L. G., Villa E., Mitra K., Frank J. and Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, 16:673–683, 2008.

- [22] Oroguchi T., Ikeguchi M. and Sato M. Towards the structural characterization of intrinsically disordered proteins by SAXS and MD simulation. *Journal of Physics*, 272, 2011.
- [23] Lima L. H. F., Serpa V. I., Rosseto F. R., Sartori G. R., de Oliveira Neto M., Martínez L. and Polikarpov I. Small-angle X-ray scattering and structural modeling of full-length: cellobiohydrolase I from *Trichoderma harzianum*. *Cellulose*, pages doi:10.1007/s10570-013-9933-3, 2013.
- [24] McCammon J. A. Theory of biomolecular recognition. *Current Opinion in Structural Biology*, 8:245-249, 1998.
- [25] Alder B. J. and Wainwright T. E. Phase transition for a hard sphere system. *The Journal of Physical Chemistry*, 27:1208-1209, 1957.
- [26] Levitt M. and Warshel A. Computer simulation of protein folding. *Nature*, 253:694-698, 1975.
- [27] Levitt M. and Warshel A. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227-492, 1976.
- [28] The Nobel Prize in Chemistry 2013 (Press release). *Royal Swedish Academy of Sciences. October 9, 2013.*
- [29] Rossky P.J. and Rahman A. Molecular Dynamics of a Dipeptide in Water. *Report on the 1976 workshop, Models for Protein Dynamics, Centre European de Calcul Atomique et Moleculaire, Universite de Paris XI, France (1977).*
- [30] Durrant J. D. and McCammon J. A. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9:71, 2011.
- [31] Verlet L. Computer "experiments" on classical fluids. II. Equilibrium correlation functions. *Physical Review*, 165:201-214, 1967.
- [32] Allen M. P. and Tildesley D. *Computer Simulations of Liquids*. Clarendon Press, Oxford, 1987.
- [33] Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R. D., Kalé L. and Schulten K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781-1802, 2005.

- [34] Humphrey W., Dalke A. and Schulten K. VMD - visual molecular dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [35] <http://biophysics.cs.vt.edu/h++>.
- [36] Bashford D. and Karplus M. pKa of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, 29:10219–10225, 1990.
- [37] Gordon J. C., Myers J. B., Folta T., Shoja V., Heath L. S. and Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, 33:W368–71, 2005.
- [38] MacKerell A. D. Jr., Bashford D., Dunbrack M. R. L., Evanseck J. D. Jr., Field M. J., Fischer S., Gao J., Guo H., Ha S., Joseph-McCarthy D., Kuchnir L., Kuczera K., Lau F. T. K., Mattos C., Michnick S., Ngo T., Nguyen D. T., Prodhom B., Reiher W. E., Roux B., Schlenkrich M., Smith J. C., Stote R., Straub J., Watanabe M., Wiórkiewicz-Kuczera J., Yin D. and Karplus M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *Journal of Physical Chemistry B*, 102:3586–3616, 1998.
- [39] Guvench O. Greene S. N., Kamath G., Brady J. W., Venable R. M., Pastor R. W. and Mackerell A. D. Jr. Additive empirical force field for hexopyranose monosaccharides. *Journal of Computational Chemistr*, 29:2543, 2008.
- [40] Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., S. Swaminathan S. and Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.
- [41] Frenkel D. and Smit B. *Understanding Molecular Simulations: from Algorithms to Applications*, volume 2nd ed. Academic Press, London, 2002.
- [42] A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education, Harlow, 2001.
- [43] Tuckerman M. E. and Berne B. J. Molecullar dynamics algorithm for multiple time scales: Systems with disparate masses. *Journal of Chemical Physics*, 94(2):1465–1469, 1991.
- [44] Quigley D. and Probert M. I. J. Langevin Dynamics in Constant Pressure Extended Systems. *Chem Phys*, 120:11432, 2004.

- [45] Andersen H. C. Rattle: A 'velocity' version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52:24–34, 1983.
- [46] Miyamoto S. and Kollman P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.
- [47] Chipot C. and Pohorille A. *Free Energy Calculations - Theory and Applications in Chemistry and Biology*. Springer, 2007.
- [48] Schlick T. Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *Biology Reports*, 1:51, 2009.
- [49] Zuckerman D. M. Equilibrium Sampling in Biomolecular Simulations. *Annual Review of Biophysics*, 40:41–62, 2011.
- [50] Sugita Y. and Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141–151, 1999.
- [51] Hess B., van der Spoel D. and Lindahl E. *Gromacs User Manual*. Department of Biophysical Chemistry, University of Groningen. Nijenborgh 4, 9747 AG Groningen, The Netherlands., 2006.
- [52] Abraham M. J. and Gready J. E. Molecular dynamics simulations and drug discovery. *Journal of Chemical Theory and Computation*, 4:1119–1128, 2008.
- [53] Svergun D., Barberato C. and Koch M. H. J. Crysol - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, 28:768–773, 1995.
- [54] Cowley J.M., Peng L. M., Ren G., Dudarev S. L. and Whelan M. J. Parameterizations of electron atomic scattering factors. *International Tables for Crystallography*, 4.3:262, 2006.
- [55] Nimlos M. R., Matthews J. F., Crowley M. F., Walker R. C., Chukkapalli G., Brady J. W., Adney W. S., Cleary J. M., Zhong L. and Himmel M. E. Molecular modeling suggests induced fit of family I carbohydrate-binding modules with a broken-chain cellulose surface. *Protein Engineering, Design & Selection*, 20:179–187, 2007.

- [56] Várnai A., Siika-aho M. and Viikari L. Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnology for Biofuels*, 6:30, 2013.
- [57] Prates É. T., Stankovic I., Silveira R. L., Liberato M. V., Henrique-Silva F., Pereira N. Jr., Polikarpov I., and Skaf M. S. X-ray structure and molecular dynamics simulations of endoglucanase 3 from *Trichoderma harzianum*: Structural organization and substrate recognition by endoglucanases that lack cellulose binding module. *Plos One*, 8:e59069, 2013.
- [58] Grupo colaborador de cristalografia, Centro de biotecnologia molecular estrutural, Instituto de Física de São Carlos, Universidade de São Paulo.
- [59] Sandgren M. Shaw A., Ropp T. H., Wu S., Bott R., Cameron A. D., Stahlberg J., Mitchinson C. and Jones T. A. The X-ray crystal structure of the *Trichoderma reesei*, family 12 endoglucanase 3, Cel12A, at 1.9 Å resolution. *Journal of Molecular Biology*, 308:295–310, 2001.
- [60] Boraston A.B., Nurizzo D., Notenboom V., Ducros V., Rose D. R., Kilburn D. G. and Davies G. J. Differential oligosaccharide recognition by evolutionarily-related β -1,4 and β -1,3 glucan-binding modules. *Journal of Molecular Biology*, 319:1143–1156, 2002.
- [61] Karlsson J., Siika-ahob M., Tenkanenb M. and Tjerneldet F. Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. *Journal of Biotechnology*, 99:63–68, 2002.
- [62] Boraston A. B., Boraston A. B., Bolam D. N., Gilbert H. J. and Davies G. J. Carbohydrate binding modules: fine tuning polysaccharide recognition. *Biochemical Journal*, 382:769–781, 2004.
- [63] Payne C. M., Bomble Y. J., Taylor B. C., McCabe C., Himmel M. E., Crowley M. F. and Beckham G. T. Multiple functions of aromatic carbohydrate interactions in a processive cellulose examines with molecular simulation. *The Journal of Biological Chemistry*, 286:41028–41035, 2011.
- [64] Laughrey Z. R., Kiehna S. E., Riemen A. J. and Waters M. L. Carbohydrate - π interactions: what are they worth? *Journal of the American Chemical Society*, 130:14625–14633, 2008.
- [65] Cheng Y. S., Ko T. P., Wu T. H., Ma Y., Huang C. H., Lai H. L., Wang A. H., Liu J. R. and Guo R. T. Crystal structure and substrate binding mode

of cellulase 12A from *Thermotoga maritima*. *Proteins: Structure, Function, and Bioinformatics*, 79:1193–1204, 2011.

- [66] <http://www.cazy.org>
- [67] Yoshizawa T., Shimizu T., Hirano H., Sato M. and Hashimoto H. Structural basis for inhibition of xyloglucan-specific endo- β -1,4-glucanase (XEG) by XEG-protein inhibitor. *The Journal of Biological Chemistry*, 287:18710–18716, 2012.
- [68] Saburi W., Hondoh H., Kim Y., Mori H., Okuyama M. and Kimura A. Structure-function relationship of substrate length specificity of dextran glucosidase from *Streptococcus mutans*. *Biologia*, 63/6:1000–1005, 2008.
- [69] Bianchetti C. M., Brumm P., Smith R. W., Dyer K., Hura G. L., Rutkoski T. J. and Phillips G. N. Jr. Structure, dynamics, and specificity of endoglucanase D from *Clostridium cellulovorans*. *Journal of Molecular Biology*, 425:4267–4285, 2013.
- [70] Sakon J., Adney W. S., Himmel M. E., Thomas S. R. and Karplus P. A. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry*, 35:10648–10660, 1996.
- [71] Martinez L. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, 2009.
- [72] Kim H. W. and Ishikawa K. Functional analysis of hyperthermophilic endocellulase from *Pyrococcus horikoshii* by crystallographic snapshots. *Biochemical Journal*, 437:223–230, 2011.
- [73] Wriggers W., Chakravarty S. and Jennings P. A. Control of protein functional dynamics by peptide linkers. *Protein Functional Dynamics*, 80:736–746, 2005.
- [74] Romanowski M.J. and Burley S.K. Crystal structure of the *Escherichia coli* shikimate kinase I (AroK) that confers sensitivity to mecillinam. *Proteins*, 47: 558–562, 2002.
- [75] Hura G. L., Menon A. L., Hammel M., Rambo R. P., Poole F. L., Tsutakawa S. E., Jenney F. E. Jr., Classen S., Frankel K. A., Hopkins R. C., Yang S., Scott J. W., Dillard B. D., Adams M. W. W. and Tainer J. A. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature Methods*, 6(8):606–612, 2009.

- [76] Beckham G. T., Bomble Y. J., Matthews J. F., Taylor C. B., Resch M. G., Yarbrough J. M., Decker S. R., Bu L., Zhao X., McCabe C., Wohlert J., Bergenstrahle M., Brady J. W., Adney W. S., Himmel M. E. and Crowley M. F. The O-glycosylated linker from the *Trichoderma reesei* family 7 cellulase is a flexible, disordered protein. *Biophysical Journal*, 99:3773-3781, 2010.
- [77] Sammond D. W., Payne C. M., Brunecky R., Himmel M. E., Crowley M. F. and Beckham G. T. Cellulase linkers are optimized based on domain type and function: Insights from sequence analysis, biophysical measurements, and molecular simulation. *Plos One*, 7:e48615, 2012.
- [78] Gustavsson M., Lehtio J., Denman S., Teeri T. T., Hult K. and Martinelle M. Stable linker peptides for a cellulose-binding domain-lipase fusion protein expressed in *Pichia pastoris*. *Protein Engineering*, 14:711–715, 2001.
- [79] Da Silva A. C., Ferro J. A., Reinach F. C., Farah C. S., Furlan L. R., Quaggio R. B., Monteiro-Vitorello C. B., Van Sluys M. A., Almeida N. F., Alves L. M., do Amaral A. M., Bertolini M. C., Camargo L. E., Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L. P., Cicarelli R. M., Coutinho L. L., Cursino-Santos J. R., El-Dorry H., Faria J. B., Ferreira A. J., Ferreira R. C., Ferro M. I., Formighieri E. F., Franco M. C., Greggio C. C., Gruber A., Katsuyama A. M., Kishi L. T., Leite R. P., Lemos E. G., Lemos M. V., Locali E. C., Machado M. A., Madeira A. M., Martinez-Rossi N. M., Martins E. C., Meidanis J., Menck C. F., Miyaki C. Y., Moon D. H., Moreira L. M., Novo M. T., Okura V. K., Oliveira M. C., Oliveira V. R., Pereira H. A., Rossi A., Sena J. A., Silva C., de Souza R. F., Spinola L. A., Takita M. A., Tamura R. E., Teixeira E. C., Tezza R. I., Trindade dos Santos M., Truffi D., Tsai S. M., White F. F., Setubal J. C. and Kitajima J. P. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature*, 417:459–463, 2002.
- [80] Gilkes N. R., Kilburn D. G., Miller R. C. Jr. and Warren R. A. Structural and functional analysis of a bacterial cellulase by proteolysis. *The Journal of Biological Chemistry*, 30:17802–17808, 1989.
- [81] Warren R. A., Beck C. F., Gilkes N. R., Kilburn D. G., Langsford M. L., Miller R. C. Jr., O'Neill G. P., Scheufens M. and Wong W. K. Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. *Proteins*, 4:335–341, 1986.
- [82] Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9:40, 2008.

- [83] Zhang Y. Roy A. and Kucukural A. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5:725–738, 2010.
- [84] Hui J. P., White T. C. and Thibault P. Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from *Trichoderma reesei*. *Glycobiology*, 12:837–849, 2002.
- [85] <http://www.ks.uiuc.edu/research/gpu/files/icms-edinburgh/gb.pdf>.
- [86] Amadei A., Linssen A. B. M. and Berendsen Dr. H. J. C. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.
- [87] Zhao G., Perilla J. R., Yufenyuy E. L., Meng X., Chen B., Ning J., Ahn J., Gronenborn A. M., Schulten K., Aiken K. and Zhang P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643–646, 2013.
- [88] Wriggers W. Conventions and workflows for using situs. *Acta Crystallographica D*, 68:344–351, 2012.
- [89] Allen R. J., Valeriani C. and ten Wolde P. R. Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter*, 21:463–102, 2009.
- [90] Gomes T. C. and Skaf M. S. Cellulose-builder: a toolkit for building crystalline structures of cellulose. *Journal of Computational Chemistry*, 33(14):1338–1346, 2012.
- [91] Zhong L., Matthews J. F., Hansen P. I., Crowley M. F., Cleary J. M., Walker R. C., Nimlos M. R., Brooks C. L. 3rd, Adney W. S., Himmel M. E. and Brady J. W. Computational simulations of the *Trichoderma reesei* cellobiohydrolase I acting on microcrystalline cellulose I β : the enzyme-substrate complex. *Cellulose*, 15:261–273, 2008.
- [92] Zhong L., Matthews J. F., Crowley M. F., Rignall T., Talón C., Cleary J. M., Walker R. C., Chukkapalli G., McCabe C., Nimlos M. R., Brooks C. L., Himmel M. E. and Brady J. W. Interactions of the complete cellobiohydrolase I from *Trichoderma reesei* with microcrystalline cellulose I β . *Cellulose*, 15:261–273, 2008.