

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS

# Estudos Estruturais de Glicosídeos de Fungos

Adriana Lucely Rojas Cardona

Tese apresentada ao Instituto de Física  
de São Carlos, da Universidade de São  
Paulo, para obtenção do título de  
Doutor em Ciências na área de  
concentração: Física Aplicada.

USP/IFSC/SBI



8-2-001671

Orientador: Prof. Dr. Igor Polikarpov

São Carlos – 2005

IFSC-USP SERVIÇO DE BIBLIOTECA  
INFORMAÇÃO

Cardona, Adriana Lucely Rojas

“Estudos estruturais de glicosidases de fungo”

Adriana Lucely Rojas Cardona – São Carlos, 2005

Tese (Doutorado) – Área de Física Aplicada da Universidade de São Paulo,  
São Carlos  
2005 - Páginas: 150

Orientador: Prof. Dr. Igor Polikarpov

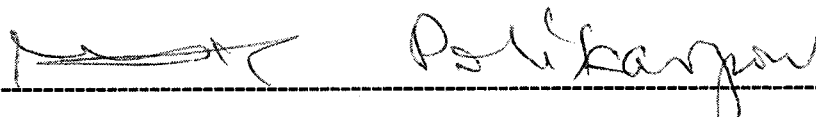
1. Glicosidases de carboidratos. 2.  $\beta$ -galactosidase 3. Exo-inulinase. 4.  $\beta$ -xilosidase. 5. Cristalografia. 6. SAXS.

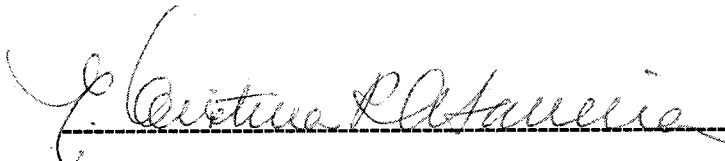
I. Título

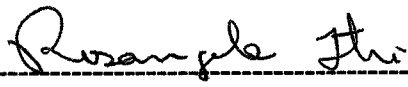


MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE ADRIANA LUCELY ROJAS CARDONA APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, UNIVERSIDADE DE SÃO PAULO, EM 08/06/2005.

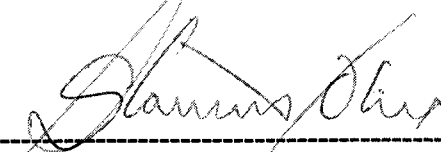
COMISSÃO JULGADORA:

  
-----  
Prof. Dr. Igor Polikarpov (Orientador e Presidente) – IFSC/USP

  
-----  
Prof. Dra. Maria Cristina Roque Antunes Barreira – FMRP/USP

  
-----  
Prof. Dra. Rosângela Itri – IF/USP

  
-----  
Prof. Dr. Ronaldo Alves Pinto Nagem – UFMG

  
-----  
Prof. Dr. Glaucius Oliva – IFSC/USP

---

# D EDICATÓRIA

*À minha filha Diana Carolina,  
por todas as noites que dormiu na minha mesa,  
esperando por mim, para sair do laboratório,  
e pelas que teve que dormir sozinha em casa,  
por que eu não consegui sair dele.*

---

# **A** GRADECIMENTOS

*A Deus pela oportunidade de cumprir um sonho, obrigada Pai pelo tempo da disciplina e pelo tempo da misericórdia.*

*Aos meus pais e meus irmãos por terem sempre acreditado em mim, e serem uma fonte inesgotável de amor. Se eu voltasse a nascer e pudesse escolher de novo a minha família, com certeza seriam vocês.*

*A Sandra Martha pela amizade e pelas frutíferas discussões acadêmicas ao longo do doutorado.*

*Ao Mario Sanches pela amizade, pelas discussões científicas e pelo prazer de tomar chá juntos.*

*A Monique pela amizade e por todas as vezes que acreditou que resolveríamos a *A.D.S.L.**

*Ao Ney Leite pela amizade e particularmente pela leitura cuidadosa deste manuscrito, obrigada pela paciência em corrigir meu português.*

*Ao Dr. Ronaldo Nagem pelas enriquecedoras discussões sobre cristalografia e pela sua disposição para ajudar e ensinar de uma maneira generosa.*

*Ao Dr. Alexander Golubev pela orientação na parte da cristalização.*

*Ao Dr. Igor Polikarpov por acolher-me no seu grupo e brindar-me com confiança e independência no trabalho.*

*Aos membros da banca pela leitura do manuscrito e suas valiosas contribuições para a apresentação final do trabalho.*

*A Beatriz Guimarães, uma amiga que a vida me presenteou em Paris, quem me aconselhou a fazer o doutorado no Brasil e fez o primeiro contato com o Dr. Polikarpov por mim.*

*Aos professores do Instituto, Richard, Otavio e Glaucius que sempre estiveram dispostos a colaborar. Foi um prazer trabalhar no grupo que vocês formaram e espero algum dia poder contribuir para que na Colômbia também se forme um grupo de Biologia Estrutural.*

*Ao André Ambrosio e João Renato pela amizade e a sua disposição para me ajudar, principalmente nas minhas mudanças.*

*Ao Alessandro meu filho amado, foi um prazer ter ensinado você no tempo da sua iniciação científica no grupo, eu tenho certeza que você será um grande cristalógrafo.*

*Ao Lucas Bleicher por os serviços de consultoria em gramática portuguesa e inglesa, obrigada.*

*A Carol Valadão, a Nadia Helena e o Lucas Bleicher por cuidarem de meu gato no tempo da escrita desta tese.*

*Ao Feko pelo prazer da dança e a sua companhia nos meus últimos meses no Brasil.*

*Ao Geraldo por tudo que me ensinou, com respeito à manutenção do anôdo rotatório e ao Augusto por todas as vezes que saiu de sua casa para vir me socorrer quando o anôdo deu pau.*

*Ào pessoal da secretaria, Fernando, Luciana, Fernanda e Jaciara por sempre estarem prontos a ajudar ao longo destes anos, principalmente na época dos relatórios da Fapesp.*

*Em geral a todo o grupo de Cristalografia do UFGC por ter sido uma família para minha filha e para mim aqui no Brasil*

*Àos meus amigos de toda a vida, William Herrera e Alejandro Mora que foram um apoio fundamental nos tempos difíceis. Após o terremoto sem vocês teria sido muito difícil retomar minha vida.*

*Ào Beynor Paez pelo apoio espiritual e material na fase mais difícil da minha vida, sem você eu não poderia sequer ter viajado para o Brasil, espero que algum dia a vida me dê a oportunidade de te retribuir de alguma forma.*

*À FAPESP pelo apoio financeiro ao longo do doutorado. Eu não sendo brasileira tenho que fazer minha gratidão ao governo brasileiro pela sua política generosa com os seus vizinhos latino-americanos, este tipo de apoio com certeza contribuirá no futuro desenvolvimento de nossos países.*

# A PRESENTAÇÃO

Esta tese foi desenvolvida ao longo dos últimos quatro anos (2001-2005), durante o curso do doutorado do Instituto de Física (IFSC) da Universidade de São Paulo campus São Carlos. O trabalho experimental foi realizado no grupo de cristalografia de proteínas. Embora seja muito difícil lembrar todos os caminhos que me trouxeram até aqui, gostaria só de remarcar alguns momentos que foram importantes. Meu primeiro contato com a cristalografia foi na minha iniciação científica enquanto cursei a graduação em física na Colômbia. Entretanto, naquela época, havia muito poucas pessoas que conhecessem o assunto no meu país, e, até hoje, a biologia estrutural é uma linha de pesquisa incipiente na Colômbia. Durante o final da minha graduação eu trabalhei nos processos de purificação, cristalização e resolução da estrutura da lectina P4 de *Dioclea lehmanni*. Nesta etapa de minha vida eu tive a oportunidade de trabalhar com o Dr. Gerardo Perez, quem me orientou nos procedimentos de purificação da proteína e me ensinou a trabalhar em um laboratório de Bioquímica; aproveito o ensejo para agradecer a ele pelos ensinamentos. Em 1997 tive a oportunidade de conhecer o Dr. Pedro Alzari em um dos primeiros cursos de cristalografia de proteínas organizado na Colômbia. Naquela época eu já tinha cristais da lectina, de maneira que fui convidada pelo Dr. Alzari para fazer um estágio de 3 meses no Instituto Pasteur em Paris, com o intuito de fazer a coleta de dados dos cristais e aprender a usar os programas necessários para a resolução da estrutura. Depois de finalizar minha graduação foi convidada pelo Dr. Alzari para realizar meu doutorado no seu grupo, de maneira que em setembro de 1998 comecei meu Doutorado em Paris. Considero um privilégio o tempo que trabalhe junto ao Dr. Alzari e seu grupo. O Dr. Alzari não só é um excelente cristalógrafo como um dos melhores mestres que tive na minha vida. Infelizmente por problemas pessoais tive que voltar para Colômbia em agosto de 1999. Após este evento, por recomendação de Beatriz Guimarães, que na época fazia pós-doutorado no grupo do Dr. Alzari, decidi vir fazer o doutorado no Brasil. Em março de 2001 me juntei ao grupo do Dr. Polikarpov no LNLS, estando, naquela época, meu doutorado ligado ao IFGW da UNICAMP. Alguns meses depois o Dr. Polikarpov



tornou-se professor do Instituto de Física da USP de São Carlos, tendo se mudado pra esta instituição com seu grupo de alunos. Assim passei a desenvolver meu trabalho de doutoramento no IFSC, onde tive a oportunidade de trabalhar em um grupo interdisciplinar que não tem nada que invejar em termos de infra-estrutura aos laboratórios do primeiro mundo. Sou grata a Deus por ter me dado a oportunidade de fazer meu doutorado em um país maravilhoso como o Brasil.

Gostaria de aproveitar esta seção para responder um comentário da Dra Maria Cristina Barreira no dia da defesa com respeito à quantidade de trabalho que foi feito, mas eu não teria conseguido fazer tudo isto sozinha. Quando eu me juntei ao grupo do Dr. Igor Polikarpov, o Dr. Ronaldo Nagem já tinha feito um grande avanço na resolução da estrutura da  $\beta$ -galactosidase e trabalhou junto comigo na resolução da estrutura da exo-inulinase, tendo ele sido um excelente parceiro de trabalho, uma pessoa com a qual aprendi muito, graças a sua generosidade e sua excelente capacidade didática. O trabalho de cristalização foi uma parceria com o Dr. Golubev e considero um privilégio os meses que trabalhei e aprendi ao lado de uma pessoa com tanta experiência na área da cristalização. O fato de conseguir um bom número de publicações em jornais de alto impacto na comunidade só reflete a orientação do Dr. Polikarpov que é um exemplo de produtividade, ele conseguiu concentrar nossos esforços em fechar os diferentes trabalhos e apresenta-los à comunidade científica em forma de artigos.

**Adriana Rojas**

# ÍNDICE

<b>CAPÍTULO 1: Introdução Geral</b>	<b>6</b>
<b>1.1. Revisão bibliográfica</b>	<b>7</b>
1.1.1. Classificação das enzimas	8
1.1.2. As hidrolases de carboidratos ou glicosidases	8
1.1.3. Módulos de ligação de carboidratos	17
<b>1.2. Objetivos</b>	<b>21</b>
1.2.1. Objetivo Geral	21
1.2.2. Objetivos Específicos	21
<b>1.3. Métodos</b>	<b>22</b>
1.3.1. Cristalização	22
1.3.2. Caracterização estrutural	24
1.3.2.1. Difração de raios-X	24
1.3.2.2. Espalhamento de raios-X a baixo ângulo (SAXS)	31
<b>1.4. Referências</b>	<b>35</b>
<b>CAPÍTULO 2: <math>\beta</math>-Galactosidase de <i>Penicillium sp.</i></b>	<b>40</b>
<b>2.1. Introdução</b>	<b>41</b>
2.1.1. $\beta$ -galactosidases	41
2.1.2. $\beta$ -galactosidase de <i>Penicillium sp.</i>	42
<b>2.2. Metodologia</b>	<b>43</b>
2.2.1. Cristalização	43
2.2.2. Coleta e processamento de dados	43
2.2.3. Faseamento	44
<b>2.3 Resultados e Discussão</b>	<b>44</b>
2.3.1. Estrutura tridimensional da $\beta$ -galactosidase de <i>Penicillium sp.</i>	47
2.3.2. Sítio ativo	47
2.3.3. Sítios de glicosilação	50
<b>2.4 Conclusões</b>	<b>51</b>
<b>2.5. Referências</b>	<b>52</b>

<b>CAPÍTULO 3: Exo-Inulinase de <i>Aspergillus awamori</i></b>	<b>55</b>
<b>3.1. Introdução</b>	<b>56</b>
3.1.1. Inulinases	57
3.1.2. Exo-inulinase de <i>Aspergillus awamori</i>	58
<b>3.2 Metodologia</b>	<b>59</b>
3.2.1. Cristalização	59
3.2.2. Coleta e processamento de dados	59
3.3.3. Faseamento	60
<b>3.3 Resultados e Discussão</b>	<b>60</b>
3.3.4 Estrutura tridimensional da exo-inulinase de <i>Aspergillus awamori</i>	62
3.3.3 Sítio ativo.	63
3.3.4 Glicosilação da enzima e seus efeitos na cristalização.	64
<b>3.4. Conclusões</b>	<b>65</b>
<b>3.5. Referências</b>	<b>66</b>
<b>CAPÍTULO 4: <math>\beta</math>-Xilosidase De <i>Trichoderma reesei</i></b>	<b>70</b>
<b>4.1. Introdução</b>	<b>71</b>
4.1.1. Xilosidases	71
4.1.2. $\beta$ -xilosidase de <i>Trichoderma reesei</i>	72
<b>4.2. Metodologia</b>	<b>73</b>
4.2.1. Purificação	73
4.2.2 Cristalização	73
4.2.3 Coleta dos dados de SAXS	74
4.2.3 Dicroísmo circular	74
<b>4.3. Resultados e Discussão</b>	<b>75</b>
4.3.1. Determinação do envelope estrutural.	79
4.3.2 Estrutura secundária da $\beta$ -xilosidase de <i>T. reesei</i>	81
4.3.3. Enovelamento da $\beta$ -xilosidase de <i>T. reesei</i> .	82
<b>4.5 Conclusões</b>	<b>86</b>
<b>4.6 Referências</b>	

<b>CAPÍTULO 5: laminarinases</b>	<b>91</b>
<b>5.1. Endo-Glucanase De <i>Rhodotermus marinus</i></b>	<b>92</b>
<b>5.1.1. Metodologia</b>	<b>93</b>
5.1.1.1. Cristalização	93
5.1.1.2 Coleta e processamento de dados	94
<b>5.1.2. Resultados e discussão</b>	<b>95</b>
5.1.2.1 Cristalização	95
5.1.2.2. Coleta e processamento de dados	98
<b>5.2. Exo-Glucanase De <i>Trichoderma viride</i></b>	<b>100</b>
<b>5.2.2. Metodologia</b>	<b>100</b>
5.2.2.1. Purificação	100
5.2.2.2. Dicroísmo Circular	100
<b>5.2.3. Resultados e Discussão</b>	<b>101</b>
5.2.3.1. Purificação	101
5.2.3.2. Dicroísmo circular	103
<b>5.3. Conclusões</b>	<b>104</b>
<b>5.4. Referências</b>	<b>105</b>

## LISTA DE FIGURAS

PG

- Figura 1.1.** Representação esquemática do mecanismo de catálise das enzimas inversoras. 9.
- Figura 1.2.** Representação esquemática do mecanismo de catálise das enzimas retentoras. 9.
- Figura 1.3.** Representação esquemática da topologia do enovelamento tipo *TIM Barrel*. Figura adaptada de Branden e Tooze. 10.
- Figura 1.4.** a. Representação do enovelamento tipo rocambole (*Jelly roll*) b. Diagrama topológico do rocambole (*Jelly roll*). Adaptado de Branden e Tooze. 11.
- Figura 1.5.** a. Representação do enovelamento tipo  $\beta$ -*propeller*, aonde cada pá é representada em uma cor diferente. b. Diagrama topológico do  $\beta$ -*propeller*, cada pá é assinalada com a letra W. 11.
- Figura 1.6.** Morfologia geral das glicosil hidrolases, os resíduos catalíticos se apresentam em vermelho. a. Sítio ativo tipo bolsão (glucoamilase de *A. awamori*) b. Sítio ativo tipo sulco (endoglucanase E2 de *T. fusca*). c. Sítio ativo tipo túnel (Celobiohidrolase II de *T. reesei*). Figura adaptada de Davies e Henrissat. 16.
- Figura 1.7.** Microscopia eletrônica de uma fibra de algodão antes e depois de ser tratada com um domínio da celobiohidrolase I de *Penicillium janthinellum* (*PjCBM1*). a. Fibra de algodão (6000 $\times$ ). b. Fibra de algodão (2000 $\times$ ). c. Fibra de algodão depois de ter sido incubada 6 horas com  $CBD_{CBHI}$  (6000 $\times$ ). d. Fibra de algodão depois ter sido incubada 10 horas com *PjCBM1* (2000 $\times$ ). Figura obtida de Gao e colaboradores. 18.
- Figura 1.8.** Classificação dos CBMs por superfamílias e por tipos. Os quadros em linhas descontínuas apresentam os domínios agrupados por tipo. Os colchetes com números mostram exemplos das diferentes superfamílias. a. *CcCBM17* de *Clostridium Cellulovorans* em complexo com celulose (PDB:1J84). b. *TmCBM4-2* de *T. maritima* em complexo com laminohexose (PDB:1GUI). c. *CjCBM15* de *Cellvibrio japonicus* em complexo com xylopentose (PDB:1GNY). d. *CtCBM3* de *Clostridium thermocellum* (PDB:1NBC). e. *CjCBM2* de *Cellulomonas fimi* (PDB:1EXG). f. *TmCBM9-2* de *T. maritima* em complexo com celobiose (PDB:1I82). g. *MvCBM32* de *Micromonospora viridifaciens* em complexo com galactose (PDB: 1EUU). h. *EcCBM5* de *Erwinia chrysanthemi* (PDB:1AIW). i. *SiCBM13* de *S. lividans* em complexo com xylopentose (PDB:1MC9). j. *TrCBM1* de *Trichoderma reesei* (PDB: 1CBH). k. *CjCBM10* de *Cellvibrio japonicus* (PDB:1E8R). l. *UdCBM18* de *Urtica dioica* em complexo com quititriose (PDB:1EN2). m. *TtCBM14* de *Tachypleus tridentatus* (PDB: 1DQC). Figura adaptada de Boraston e colaboradores. 19.

- Figura 1.9.** Representação esquemática da localização do sítio de ligação nos CMBs tipo B que apresentam enovelamento  $\beta$ -sandwich. a. Nesta figura vários domínios tipo B com seus ligantes foram superpostos, mas só a estrutura terciária do domínio CfCBM4 da Cel9B de *Cellulomonas fimi* é apresentada. b. Celupentose ligada no sulco do domínio CfCBM4; a superfície dos resíduos aromáticos que participam da interação é apresentada em roxo. 20.
- Figura 1.10.** À esquerda placa utilizada nos experimentos de cristalização e à direita, um dos poços está amplificado mostrando como a lamínula, contendo a gota composta por solução do poço e solução da proteína, é colocada sobre o reservatório. 22.
- Figura 1.11.** Diagrama de fase que representa a solubilidade da proteína em função do agente precipitante a. Representação esquemática do que acontece em um experimento de difusão de vapor onde não se formam cristais b. Representação esquemática do processo de formação do cristal. 24.
- Figura 1.12.** Diagrama de Argand mostrando as contribuições de  $F_P$  e  $F_H$  para o valor de  $F_{PH}$ . Na figura  $\alpha_{PH}$ ,  $\alpha_P$ ,  $\alpha_H$ , representam a fase do fator de estrutura do cristal derivado, do cristal nativo e da subestrutura dos átomos pesados respectivamente. 28.
- Figura 1.13.** Representação dos fatores de estrutura, mostrando as contribuições de  $F_P(+)$ ,  $F'_H(+)$  e  $F''_H(+)$  para o valor de  $F_{PH}(+)$ . Idem para  $F_{PH}(-)$ . Na figura da direita, apenas espelhamos (-h-k-l) em relação ao eixo positivo real. Adaptado de Nagem. 30.
- Figura 1.14.** Intensidades de espalhamento e função distribuição de distâncias  $p(r)$  de diferentes corpos geométricos. Figura adaptada de Svergun e Koch. 33.
- Figura 2.1** Diagrama de Ramachandran da estrutura nativa da  $\beta$ -galactosidase de *Penicillium sp.* 46.
- Figura 2.2.** Representação estereográfica da estrutura da Psp- $\beta$ -gal. De acordo com PROCHECK (Laskowski et al., 1993) os domínios 1, 2, 3, 4 e 5 são apresentados em ciano, vermelho, amarelo, verde e magenta, respectivamente. Em azul se apresenta a cadeia polipeptídica que conecta os domínios 3 e 4. As figuras foram feitas usando os programas Molscript (Kraulis, 1991), Bobscrip (Esnouf, 1997) e Raster3D (Merritt and Bacon, 1997). 47.

- Figura 2.3.** Alinhamento da estrutura primária das  $\beta$ -galactosidases de *Penicillium canescens* (171 aa), *Talaromyces emersonii* (218 aa), *Aspergillus candidus* (1005 aa), *Aspergillus niger* (1006 aa) e *Penicillium* sp. (determinada a partir do mapa de densidade). A similaridade entre as seqüências foi calculada usando o programa ALSCRIPT (Barton, 1993) e se mostra em diferentes grus de azul. A cor mais escura corresponde a uma similaridade alta nas seqüências. O Bfactor médio, o coeficiente de correlação (CC) entre o mapa de densidade predito e observado e a superfície exposta ao solvente por resíduo são mostrados em uma escala de cor de amarelo a vermelho. Quatro cores diferentes foram usadas para os intervalos diferentes de Bfactor, CC e a área da superfície exposta ao solvente. Vermelho (Bfactor maior que 30 Å<sup>2</sup>; CC menor que 95%; área exposta maior que 90 Å<sup>2</sup>). Vermelho claro (Bfactor maior que 25 e menor que 30 Å<sup>2</sup>; CC menor que 96 e maior que 95%; área exposta maior que 60 e menor que 90 Å<sup>2</sup>). Laranja (Bfactor maior que 20 e menor que 25 Å<sup>2</sup>; CC menor que 97 e maior que 96%; área exposta maior que 30 e menor que 60 Å<sup>2</sup>). Amarelo (Bfactor menor que 20 Å<sup>2</sup>; CC maior que 97%; área exposta menor que 30 Å<sup>2</sup>). Os elementos da estrutura secundária da Psp- $\beta$ -gal são apresentados em ciano, vermelho, amarelo, verde e magenta para os domínios 1, 2, 3, 4 e 5, respectivamente. Em azul se apresenta a cadeia polipeptídica que conecta os domínios 3 e 4. 48.
- Figura 2.4.** Mapa de densidade eletrônica no sítio catalítico indicando a presença da galactose. Este mapa foi gerado utilizando um modelo sem galactose. 49.
- Figure 2.5.** Estrutura dos 7 sítios de N-glicosilação (I-VII) na Psp- $\beta$ -gal. Foram identificados cinco tipos de ligações glicosídicas (Man $\beta$ 1-4GlcNAc, GlcNAc $\beta$ 1-4GlcNAc, Man $\alpha$ 1-3Man, Man $\alpha$ 1-2Man e Man $\alpha$ 1-6Man). 50.
- Figura 2.6.** Representação da posição dos açúcares com respeito aos domínios da Psp- $\beta$ -gal. Os oligossacarídeos estão principalmente localizados ao redor dos domínios 2 (verde), 3(laranja) e 5(vermelho). 50.
- Figura 3.1** Representação da estrutura da exo-inulinase de *Aspergillus awamori*. As folhas beta do domínio tipo “ $\beta$ -propeller” foram nomeadas com a letra W. Cada folha é representada por uma cor diferente (W<sub>1</sub> amarelo, W<sub>2</sub> azul marinho, W<sub>3</sub> vermelho, W<sub>4</sub> verde, W<sub>5</sub> laranja). O segundo domínio que apresenta 12 fitas beta é apresentado em roxo. 62.
- Figura 3.2.** (a) Representação da superfície da Exo-inulinase *Aspergillus awamori* com a frutose no bolsão catalítico e os açúcares da glicosilação, (b) Mapa de densidade eletrônica no sítio catalítico indicando a presença da frutose, Este mapa foi gerado utilizando um modelo sem frutose. 63.
- Figura 4.1.** Representação esquemática da hidrolise de xilano. 71.
- Figura 4.2.** Representação esquemática da presença das hemiceluloses na parede celular. 71.

- Figura 4.3.** Perfil da Cromatografia de troca iônica (MonoQ 5/5) para a  $\beta$ -xilosidase. 75.
- Figura 4.4.** Eletroforese em gel de poliacrilamida (SDS-PAGE). No poço 1 observa-se a proteína antes da MonoQ; no poço 6 o último pico da cromatografia que corresponde à  $\beta$ -xilosidase de *Trichoderma reesei*. 75.
- Figura 4.5.** Cristais de  $\beta$ -xilosidase crescidos em 18% de PEG 3350, acetato de sódio 50mM pH 6.0 e 8mg/ml de proteína. 76.
- Figura 4.6.** Cristais de  $\beta$ -xilosidase crescidos em 18% de PEG 3350, acetato de sódio 50mM pH 5.0, e 8mg/ml de proteína. 76.
- Figura 4.7.** Cristais de  $\beta$ -xilosidase crescidos em 15% de PEG 3350, acetato de sódio 50mM pH 4.0 e 8mg/ml de proteína. 77.
- Figura 4.8.** Curva de espalhamento a baixo ângulo da  $\beta$ -xilosidase de *Trichoderma reesei* em um tampão 20 mM TrisHCl, pH 6. 79.
- Figura 4.9.** Distribuição de distâncias  $P(r)$  da  $\beta$ -xilosidase de *Trichoderma reesei* calculada usando o programa GNOM. 80.
- Figure 4.10.** Representação estereográfica de três diferentes vistas do envelope da  $\beta$ -xilosidase de *T. reesei*. Em b. e c. as figuras foram rodadas 90° ao redor de Y e Z. Com respeito a figura a, respectivamente. 80.
- Figure 4.11.** a. Espectro de dicroísmo circular da  $\beta$ -xilosidase *T. reesei*. b. Ajuste dos dados experimentais e a curva estimada pelo programa CONTINLL c. Ajuste dos dados experimentais e a curva estimada pelo programa CDSSTR. 82.
- Figure 4.12.** a. Estrutura cristalográfica da  $\beta$ -D-glucanase de cevada membro da família GHF3. b. Comparação da curva de espalhamento de raios-X a baixos ângulos da  $\beta$ -xilosidase *T. reesei* (pontos) com a curva teórica da  $\beta$ -D-glucanase de cevada (linha sólida) c. Superposição do envelope da  $\beta$ -xilosidase *T. reesei* e a estrutura da  $\beta$ -D-glucanase. 83.
- Figure 4.13.** Alinhamento da seqüência da  $\beta$ -Xilosidase de *T. reesei* (797aa) e  $\beta$ -D-glucanase de cevada (605aa). A similaridade entre as seqüências foi calculada com ajuda do programa ALSCRIPT e é apresentada em diferentes tons de azul. Azul escuro corresponde a uma alta similaridade e o mais claro a uma baixa. Os domínios 1-3 da  $\beta$ -xilosidase de *T. reesei* são apresentados em vermelho, verde claro e ciano, respectivamente. Os domínios 1-2 da  $\beta$ -D-glucanase são apresentados em laranja e verde, respectivamente. 85.
- Figura 5.1.** Agulhas de laminarinase de *Rhodotermus marinus*. Gota: 6% PEG 3350 em 10mM fosfato de potássio a pH 5.0. 95.



- Figura 5.2.** Cristais da laminarinase de *Rhodotermus marinus*. Gota: 6% PEG 3350 em 90mM fosfato de potássio a pH 6.5. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0. 96.
- Figura 5.3.** Cristais da laminarinase do *Rhodotermus marinus* (*macroseeding*). Gota: 6% PEG 3350 em 90mM fosfato de potássio a pH 6.5, glucose 5mM. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0. 96.
- Figura 5.4.** Cristais da laminarinase de *Rhodotermus marinus*. Gota: 6% PEG 3350 em 65mM fosfato de potássio a pH 6.5. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0 e etanol 2%. 97.
- Figura 5.5.** Cristal da laminarinase de *Rhodotermus marinus*, usados para a coleta de dados de raios-x. Gota: 6% PEG 3350 em 120mM fosfato de potássio a pH 6.0. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0 e etanol 2%. 97.
- Figura 5.6.** Imagem de difração do cristal nativo da laminarinase de *Rhodotermus marinus*. Devido ao parâmetro da rede as reflexões aparecem muito próximas. 98.
- Figura 5.7.** Imagens de difração do cristal derivado da laminarinase de *Rhodotermus marinus*. a. imagem a  $\text{PHI}=0^\circ$  b. Zoom da imagem a  $\text{PHI}=90^\circ$ , apresenta reflexões duplas. 99.
- Figura 5.8.** Perfil da cromatografia de troca iônica (MonoQ 5/5) para a exo-1,3- $\beta$ -glucanase do *Trichoderma viride*. 101.
- Figura 5.9.** Eletroforese em gel de poliacrilamida (SDS-PAGE) das amostras coletadas depois da MonoQ, no poço 4 verifica-se a amostra eluída no pico 1 e nos poços 5 e 6 a proteína eluída no segundo pico. A amostra que foi injetada na coluna se observa no poço 9. 101.
- Figura 5.10.** Perfil da cromatografia de gel filtração (Superdex 200 HR10/30) para a exo-1,3- $\beta$ -glucanase de *Trichoderma viride*. 102.
- Figura 5.11.** Eletroforese em gel de poliacrilamida (SDS-PAGE), corado com prata. Poços 2 e 3: frações correspondentes ao segundo pico da cromatografia, que corresponde a um peso molecular de 67 kDa. No poço 9 a amostra que foi injetada na coluna. 102.
- Figure 5.12.** a. Espectro de dicroísmo circular da Exo-1,3- $\beta$ -glucanase do *Trichoderma viride*. b. Ajuste dos dados experimentais e a curva estimada pelo programa CONTINLL c. Ajuste dos dados experimentais e a curva estimada pelo programa CDSSTR. 103.

<b>LISTA DE TABELAS</b>	<b>PG.</b>
<b>Tabela 1.1.</b> Na classificação internacional sistemática, as enzimas são divididas em seis grupos de acordo com o tipo de reação catalisada.	8.
<b>Tabela 1.2.</b> Classificação das GHs por Clãs.	12-16.
<b>Tabela 1.3.</b> Classificação dos CBMs por tipos.	19.
<b>Tabela 2.1.</b> Estatísticas da coleta de dados de difração de raios-X dos cristais da $\beta$ -galactosidase de <i>Penicillium</i> sp. Os valores para alta resolução encontram-se entre parênteses.	45.
<b>Tabela 2.2.</b> Distâncias entre os átomos da proteína e os grupos OH da galactose.	49.
<b>Tabela 3.1.</b> Estatísticas da coleta dos dados de difração de raios-X dos cristais da exo-inulinase de <i>Aspergillus awamori</i> . Os valores para alta resolução encontram-se entre parênteses.	61.
<b>Tabela 3.2.</b> Primeiras 8 posições achadas pelo programa SHELX para o derivado de iodo.	61.
<b>Tabela 3.3.</b> Distâncias entre os átomos da proteína e os grupos OH da frutose.	63.
<b>Tabela 4.1.</b> Ensaio de cristalização para a $\beta$ -xilosidase.	76.
<b>Tabela 4.2.</b> Ensaio de cristalização para a $\beta$ -xilosidase variando o pH e a concentração do acetato de sódio.	77.
<b>Tabela 4.3.</b> Ensaio de cristalização para a $\beta$ -xilosidase trocando o tampão acetato de sódio por fosfato de sódio.	77.
<b>Tabela 4.4.</b> Ensaio de cristalização para a $\beta$ -xilosidase, usando a proteína parcialmente pura.	78.
<b>Tabela 4.5.</b> Ensaio de cristalização para a $\beta$ -xilosidase com xilose.	78.
<b>Tabela 4.6.</b> Ensaio de cristalização para a $\beta$ -xilosidase com sulfato de amônio.	78.

<b>Tabela 5.1.</b> Concentrações de PEG e de proteína usados no teste de precipitação para a laminarinase de <i>R. marinus</i> .	<b>94.</b>
<b>Tabela 5.2.</b> Ensaios de cristalização da laminarinase de <i>Rhodotermus marinus</i> .	<b>95.</b>
<b>Tabela 5.3.</b> Ensaios de cristalização da laminarinase de <i>Rhodotermus marinus</i> , adicionando aditivos.	<b>96.</b>
<b>Tabela 5.4.</b> Estatística do processamento dos cristais nativo e derivado da laminarinase de <i>Rhodotermus marinus</i> .	<b>99.</b>
<b>Tabela 5.5.</b> Conteúdo de estrutura secundária da Exo-1,3- $\beta$ -glucanase do <i>Trichoderma viride</i> , determinado a partir do espectro de dicroísmo circular.	<b>103.</b>

## Resumo

As glicosidases são enzimas que apresentam uma grande variedade de enovelamentos, assim como uma alta especificidade frente a diferentes substratos. Estas enzimas têm em comum a presença de dois resíduos catalíticos, responsáveis pela clivagem das ligações glicosídicas. O uso de glicosidases nas indústrias têxtil e alimentícia, no processamento de polpa de papel e na síntese de oligossacarídeos tem incentivado a engenharia destas proteínas no sentido de melhorar suas propriedades catalíticas e estabilidade. Estudos estruturais das glicosidases têm aumentado nosso entendimento de seus mecanismos de ação catalítica, assim como dos processos de interação proteína-carboidrato. Neste trabalho apresentamos os estudos cristalográficos de duas glicosidases de fungos, sendo elas a  $\beta$ -galactosidase de *Penicillium sp.* e a Exo-inulinase de *Aspergillus awamori*, assim como estudos por espalhamento de raios-X a baixos ângulos (SAXS) da  $\beta$ -xylosidase de *Trichoderma reesei*.

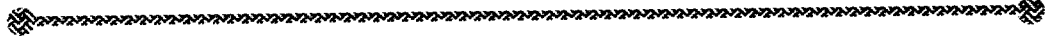
As estruturas cristalográficas da  $\beta$ -galactosidase e de seu complexo com galactose foram determinadas pela técnica de substituição isomórfica simples com espalhamento anômalo (SIRAS) até 1.9 Å de resolução para a estrutura sem substrato e 2.0 Å de resolução para o complexo. A estrutura do complexo com galactose foi usada para identificar os resíduos catalíticos, sendo o resíduo Glu 200 identificado como doador de próton e o resíduo Glu 299 como o nucleófilo.

As estruturas cristalográficas da Exo-inulinase de *Aspergillus awamori* e de seu complexo com frutose foram também determinadas pela técnica de substituição isomórfica simples com espalhamento anômalo (SIRAS) até 1.55 Å e 1.8 Å de resolução, respectivamente. A partir da estrutura do complexo foi possível identificar os resíduos Asp41 e Glu241 como o nucleófilo e o doador de próton, respectivamente. Além disto, foi possível verificar que o Asp189, o qual faz parte do motivo conservado Arg-Asp-Pro (RDP), é importante no reconhecimento do substrato através de duas pontes de hidrogênio.

Com o intuito de obter informações estruturais sobre a  $\beta$ -xylosidase seu envelope foi determinado a partir dos dados do espalhamento de raios-X a baixos ângulos. O envelope da  $\beta$ -xylosidase em solução foi calculado a 20 Å de resolução, sendo o raio de giro e a dimensão máxima 36.9 Å e 90 Å, respectivamente. Usando algoritmos de reconhecimento de possíveis domínios foi determinado que esta proteína apresenta, além dos dois domínios característicos da família GHF3, um barril TIM e um domínio  $\alpha/\beta$ , um terceiro domínio. A predição da estrutura secundária e os dados de dicroísmo circular indicam que este terceiro domínio apresentaria um enovelamento tipo beta.

## Abstract

Glycosidases belong to a group of enzymes displaying a great variety of protein folds and substrate specificities. Two critically located acidic residues make up the catalytic machinery of these enzymes, responsible for the cleavage of glycosidic bonds. The applications of glycosidases in textile, food, and pulp processing, as well as in catalysts and oligosaccharide synthesis have encouraged the engineering of these proteins in order to obtain improved catalytic properties and stability. Furthermore, structural studies extend our understanding of the catalytic mechanism and the role of glycosidases in the recognition processes of their different substrates. In this work, we describe crystallographic studies of two fungi glycosidases,  $\beta$ -galactosidase from *Penicillium sp* and Exo-inulinase from *Aspergillus awamori*, and the small-angle x-ray scattering (SAXS) studies of another glycosidase,  $\beta$ -xylosidase (from *Trichoderma reesei*). The crystallographic structures of  $\beta$ -galactosidase its complex with galactose were solved by single isomorphous replacement with anomalous scattering (SIRAS) using the quick cryo-soaking technique, at 1.90 Å and 2.10 Å resolution, respectively. The X-ray structure of the enzyme-galactose complex was useful in identifying the residue Glu200 as the proton donor and residue Glu299 as the nucleophile involved in catalysis. The x-ray structure of exo-inulinase and its complex with fructose were also solved by SIRAS using the quick cryo-soaking technique at 1.55 Å and 1.8 Å resolutions, respectively. The solved structure of the enzyme-fructose complex revealed two catalytically important residues, Asp41 and Glu241, as nucleophile and proton donor, respectively. It was also possible to see that residue Asp189, which belongs to the Arg-Asp-Pro motif, provides hydrogen bonds important for substrate recognition. In order to gain structural insights about the  $\beta$ -Xylosidase from *Trichoderma reesei*, we calculated their SAXS envelope. The low resolution shape of this enzyme in solution was obtained from synchrotron x-ray scattering data at 20 Å resolution. The radii of gyration and the maximum dimension of the  $\beta$ -Xylosidase were calculated to be 36.9 Å and 90 Å, respectively. In contrast to the fold of the only structurally characterized member of GHF-3, the  $\beta$ -D-glucan exohydrolase, which has

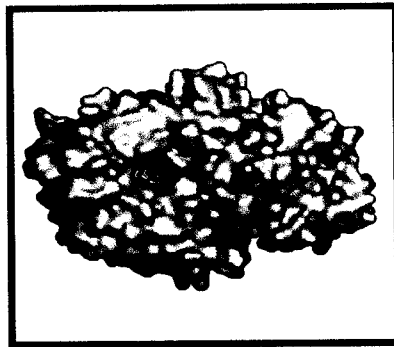
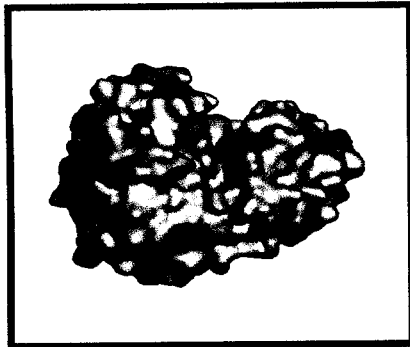
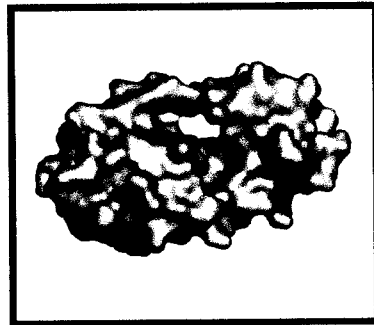


two distinct domains, the shape of the  $\beta$ -xylosidase indicates the presence of three domains located in the same plane. Domain recognition algorithms were used to show that the C-terminal part of the amino acid sequence of the protein forms the third domain. Circular dichroism spectroscopy and secondary structure prediction programs show that this additional domain adopts predominantly the  $\beta$ -conformation.



# CAPÍTULO 1.

**Introdução Geral: hidrolases de carboidratos.**



Morfologia geral das hidrolases de carboidratos.



## 1.1. Revisão bibliográfica

Depois dos antibióticos, as enzimas são os produtos microbianos mais explorados na indústria biotecnológica. As enzimas microbianas apresentam uma série de vantagens sobre seus equivalentes de origem animal ou vegetal, tais como: custos de produção relativamente baixos, possibilidade de produção em larga escala em fermentadores industriais, amplo espectro de condições físico-químicas em que elas são estáveis, geralmente relacionadas ao habitat e fisiologia do microorganismo produtor (por exemplo, organismos termofílicos produtores de enzimas termo-resistentes); susceptibilidade de manipulação genética além de se apresentarem como recursos renováveis.

Atualmente, as enzimas microbianas são amplamente utilizadas no processamento de alimentos, na produção de detergentes (biológicos), nas indústrias têxtil e farmacêutica, na química bioorgânica, na biologia molecular e em aplicações médicas. A ampla utilização destas enzimas é um reflexo da elevada especificidade em suas ações como biocatalizadores. Contudo, enzimas com perfis equivalentes de atuação sobre um mesmo substrato podem apresentar funcionamento ótimo em pH, temperatura e concentração iônica diferentes. Este fato limita, naturalmente, o número de enzimas disponíveis para aplicações biotecnológicas. Os estudos estruturais destas enzimas podem, entretanto, oferecer uma compreensão ao nível atômico de seus mecanismos de ação, possibilitando com que mudanças sítio dirigidas venham a modificar a ação das mesmas em condições específicas para a indústria.

No Instituto de Física de São Carlos da Universidade de São Paulo, o grupo do Dr. Igor Polikarpov tem trabalhado na caracterização estrutural de hidrolases de carboidratos como  $\beta$ -galactosidases,  $\alpha$ -galactosidases, Exo-inulinases, Laminarinases e  $\beta$ -xilosidases. Algumas destas enzimas têm sido caracterizadas por difração de raios-X<sup>1,2,3</sup> e outras pelo uso de técnicas de espalhamento dinâmico de luz (DLS), dicróismo circular (CD) e espalhamento de raios-X a baixo ângulo (SAXS). Estas últimas técnicas têm sido aplicadas geralmente nos casos onde a determinação das estruturas cristalográficas ainda não foi possível<sup>4,5</sup>.

No presente trabalho apresenta-se a caracterização estrutural por difração de raios-X da  $\beta$ -galactosidase de *Penicillium sp* e da Exo-inulinase de *Aspergillus*

*awamori*, assim como a caracterização da  $\beta$ -xilosidase de *Trichoderma reesei* pelas técnicas de CD e SAXS.

### 1.1.1. Classificação das enzimas

Tradicionalmente, a *International Union of Biochemistry and Molecular Biology (IUBMB)* adotou um sistema de nomenclatura identificando as enzimas em seis classes, de acordo com a natureza da reação química que catalisam (Tabela 1.1). Para cada enzima, são atribuídos dois nomes e um número de classificação de quatro dígitos que identificam as classes, as subclasses e as sub-subclasses. Por exemplo, a enzima  $\beta$ -galactosidase tem como código de classificação EC 3.2.1.23. Os números representam a classe (3, hidrolase), subclasse (2, glicosilase), sub-subclasse (1, glicosidase, enzimas que hidrolisam ligações O- e S-glicosil) e seu número de série dentro da sub-subclasse (23). A sigla “EC” é a abreviatura de *Enzyme Commission*.

**Tabela 1.1.** Na classificação internacional sistemática, as enzimas são divididas em seis grupos de acordo com o tipo de reação catalisada.

Classe da enzima	Tipo de reação catalisada
Oxidoredutases	Reações de oxidação-redução
Hidrolases	Reações de hidrólise
Transferases	Transferência de grupos funcionais
Liases	Remoção de grupos para formar ligações duplas
Isomerases	Isomerização
Ligases	Formação de ligações entre duas moléculas

### 1.1.2. As hidrolases de carboidratos ou glicosidases

Cerca de 3000 glicosil hidrolases já foram identificadas; entretanto, existem por volta de 5500 ORFs (*Open Reading Frames*) registradas no *Carbohydrate Active Enzyme database (CAZY)*<sup>6</sup>. Este grande volume de informação gerou diferentes classificações. Tradicionalmente, as glicosidases foram agrupadas de acordo com sua capacidade de hidrolisar substratos similares; por exemplo, enzimas que clivam a galactose presente na lactose, no PNPG (para-nitrofenilgalactose) ou em XGAL (5-bromo-4-cloro-3-indolil- $\beta$ -D-galactoside) foram classificadas como  $\beta$ -galactosidases.

Contudo, este tipo de classificação não é capaz de classificar adequadamente uma enzima que atua sobre diferentes substratos.

Em vista disto, um segundo tipo de divisão foi proposto com base nos dois mecanismos de ação catalítica observados nestas enzimas. Na primeira classe estão compreendidas as enzimas que hidrolisam a ligação glicosídica com inversão da configuração anomérica e na segunda estão aquelas que agem sem modificar esta configuração. Contudo, ambos os mecanismos utilizam um par de resíduos catalíticos: um doador de próton e um nucleófilo.

Nas duas classes de mecanismos o doador de prótons encontra-se a uma distância que lhe permite fazer uma ponte de hidrogênio com o oxigênio da ligação glicosídica. Nas enzimas inversoras da configuração anomérica, numa reação de etapa única, uma molécula de água é usada como nucleófilo após sua ativação por um carboxilato que age como base catalítica (Figura 1.1). Nas enzimas retentoras, o segundo resíduo catalítico é um grupo nucleofílico carboxilato que age numa reação de dupla dissociação, atuando tanto como um par iônico, estabilizando um estado de transição oxíônio transiente, ou como um grupo de saída nucleofílico para um intermediário covalente transiente (Figura 1.2).

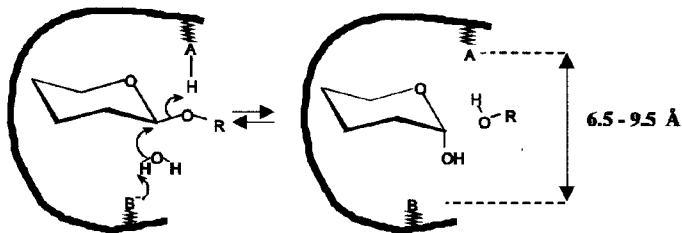


Figura 1.1. Representação esquemática do mecanismo de catálise das enzimas inversoras.

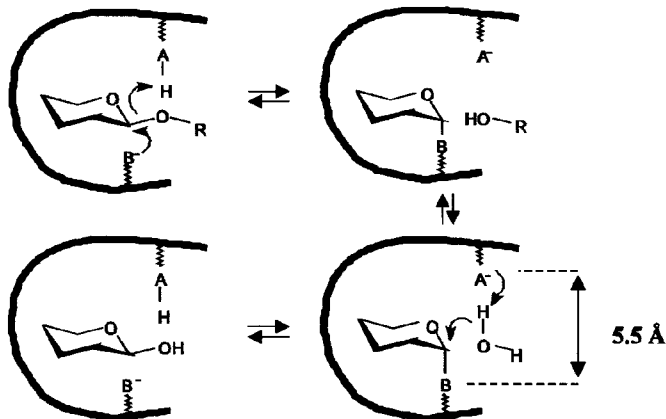


Figura 1.2. Representação esquemática do mecanismo de catálise das enzimas retentoras.

Dado que existe uma relação direta entre a sequência de aminoácidos e o enovelamento das proteínas, foi sugerida, em 1991, uma outra classificação para as glicosidases. Desta vez, a classificação foi baseada na similaridade entre as sequências de aminoácidos das enzimas<sup>7</sup>. Atualmente, as glicosil hidrolases (GHs) estão classificadas em 99 famílias<sup>8,9,10,11</sup>.

As GHs também podem ser classificadas em várias superfamílias chamadas clãs. Esta classificação é baseada na análise de *clusters* hidrofóbicos (HCA) das estruturas das proteínas<sup>9</sup>. Um clã é um grupo de famílias que tem um ancestral comum e as enzimas pertencentes a este clã são reconhecidas por apresentarem similaridades na estrutura terciária junto à conservação dos resíduos catalíticos e do mecanismo de ação. O maior membro destas superfamílias é o clã GH-A, o qual compreende proteínas de 17 famílias (Tabela 1.2). Por enquanto apenas 40 famílias foram classificadas em 14 clãs.

As enzimas que pertencem aos clãs GH-A, GH-D, GH-H e GH-K se caracterizam por terem enovelamento tipo *TIM Barrel* (Triose fosfato isomerase)<sup>12</sup> no domínio catalítico, onde oito  $\alpha$ -hélices e oito fitas  $\beta$  paralelas conformam uma estrutura em forma de barril e o sítio ativo é formado pelos *loops* que fazem parte do C-terminal das fitas do barril (Figura 1.3). Estas enzimas clivam as ligações glicosídicas através do mecanismo de retenção da configuração anomérica<sup>13</sup>. O clã GH-A é muitas vezes descrito como a superfamília 4/7 porque o doador de próton e o nucleófilo encontram-se nas fitas 4 e 7 do  $(\beta/\alpha)_8$  barril, respectivamente<sup>14</sup>. Uma lista atualizada das GHs, classificadas tanto por família quanto por clãs, pode ser encontrada em “<http://afmb.cnrs-mrs.fr/CAZY/>”. Até a presente data (Abril, 2005), já foi reportada pelo menos uma estrutura representativa para 59 famílias das 99 GHFs.



Figura 1.3. Representação esquemática da topologia do enovelamento tipo *TIM Barrel*. Figura adaptada de Branden e Tooze<sup>15</sup>.

Os clãs GH-B e GH-C apresentam um enovelamento tipo  $\beta$ -sandwich, especificamente denominado rocambole (*Jelly roll*), o qual é formado basicamente por motivos tipo chave grega. Este motivo consiste em quatro fitas  $\beta$  formando uma folha antiparalela onde a quarta fita encontra-se no começo da folha em posição antiparalela à fita 1 e afastada da fita 3 (figura 1.4). No rocambole duas chaves gregas são dispostas uma junto da outra<sup>15</sup>.



Figura 1.4. a. Representação do enovelamento tipo rocambole (*Jelly roll*) b. Diagrama topológico do rocambole (*Jelly roll*). Adaptado de Branden e Tooze<sup>15</sup>.

Os clãs GH-E e GH-F apresentam um enovelamento tipo  $\beta$ -propeller, no qual as folhas  $\beta$  independentes se arranjam em torno de um eixo formando uma estrutura parecida a uma hélice de barco. Esta hélice pode ter de cinco a oito pás e o sítio ativo é formado pelos *loops* que se encontram de um lado da hélice ao redor do eixo central, semelhante ao que acontece nos barris  $\alpha/\beta$ . Finalmente, os clãs GH-L e GH-M apresentam um enovelamento chamado barril ( $\alpha/\alpha$ )<sub>6</sub>, onde sete grampos de cabelo tipo  $\alpha$  se arranjam em forma de barril, sendo a diferença em relação aos outros barris dado pelo núcleo central formado por  $\alpha$ -hélices.

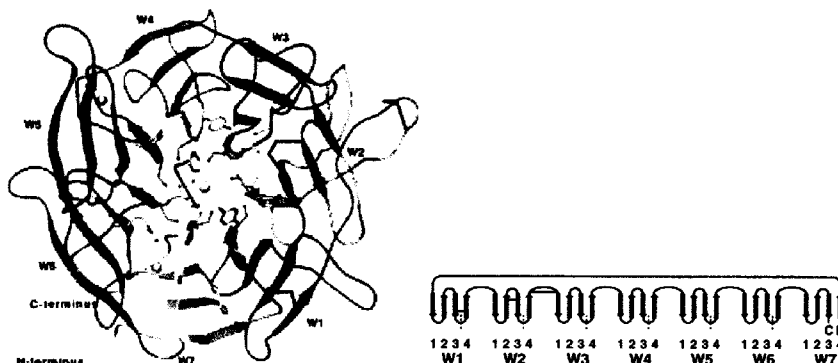


Figura 1.5. a. Representação do enovelamento tipo  $\beta$ -propeller, aonde cada pá é representada em uma cor diferente. b. Diagrama topológico do  $\beta$ -propeller, cada pá e assinalada com a letra W.

Tabela 1.2. Classificação das GHs por Clãs.

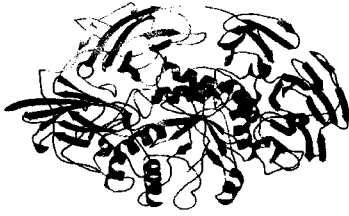


Clã	Famílias	Membro representativo	Mecanismo
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86.	<p><math>\beta</math>-galactosidase de <i>Penicillium sp</i></p>  <p>Família: 35, EC 3.2.1.23, PDB:1TG7 Enovelamento: <math>(\beta/\alpha)_8</math></p>	Retenção
		<p>Endo-<math>\beta</math>-Galactosidase de <i>Clostridium Perfringens</i></p>  <p>Família: 16, EC 3.2.1.-, PDB:1UPS Enovelamento: <math>\beta</math>/Jelly roll</p>	
GH-B	7, 16.	<p>Xilanase de <i>Aspergillus niger</i></p>  <p>Família: 11, EC 3.2.1.8, PDB:1T6G Enovelamento: <math>\beta</math>/Jelly roll</p>	Retenção

Tabela 1.2. Continuação




Clã	Famílias	Membro representativo	Mecanismo
GH-D	27, 36.	<p><math>\alpha</math>-D-Galactosidase de <i>Homo sapiens</i></p> 	Retenção
		<p>Família: 27, EC 3.2.1.22, PDB:1R46 Enovelamento: <math>(\beta/\alpha)_8</math></p>	
GH-E	33, 34, 83.	<p>Sialidase de <i>Trypanosoma rangeli</i></p> 	Retenção
		<p>Família: 33, EC 3.2.1.18, PDB:1MZ5 Enovelamento: 6-fold <math>\beta</math>-propeller</p>	
GH-F	43, 62.	<p><math>\beta</math>-1,4-Xilosidase de <i>Bacillus subtilis</i></p> 	Inversão
		<p>Família: 43, EC 3.2.1.37, PDB:1YIF Enovelamento: 5-fold <math>\beta</math>-propeller</p>	

Tabela 1.2. Continuação

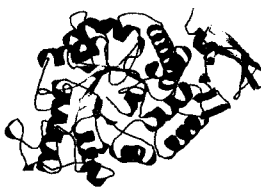


Clã	Famílias	Membro representativo	Mecanismo
GH-G	37, 63.	-	Inversão
GH-H	13, 70, 77.	<p><math>\alpha</math>-amilase de <i>Aspergillus niger</i></p>  <p>Família: 13, EC 3.2.1.1, PDB:2AAA Enovelamento: <math>(\beta/\alpha)_8</math></p>	Retenção
GH-I	24, 46, 80.	<p>Lisozima de <i>Bacteriófago T4</i></p>  <p>Família: 24, EC 3.2.1.17, PDB:1KW5 Enovelamento: <math>\beta+\alpha</math></p>	Inversão
GH-J	32, 68.	<p>Exo-inulinase de <i>Aspergillus awamori</i></p>  <p>Família: 32, EC 3.2.1.80, PDB:1Y4W Enovelamento: 5-fold <math>\beta</math>-propeller</p>	Retenção



Tabela 1.2. Continuação

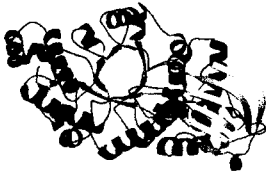

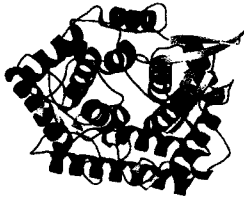
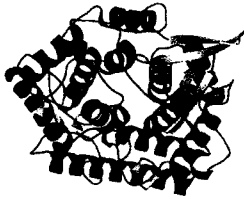

Clã	Famílias	Membro representativo	Mecanismo
GH-K	18, 20.	<p><math>\beta</math>-N-Acetylhexosaminidase de <i>Homo sapiens</i></p>  <p>Família: 18, EC 3.2.1.52, PDB: 1NOU Enovelamento: <math>(\beta/\alpha)_8</math></p>	Retenção
		<p>Glucoamilase de <i>Aspergillus awamori</i></p>  <p>Família: 15, EC 3.2.1.3, PDB: 1AGM Enovelamento: <math>(\alpha/\alpha)_6</math></p>	
GH-L	15, 65.	<p>Endo-1,4- <math>\beta</math>-Xilanase de <i>Pseudoalteromonas haloplanktis</i></p>  <p>Família: 8, EC 3.2.1.8, PDB: 1H12 Enovelamento: <math>(\alpha/\alpha)_6</math></p>	Inversão
		<p>Endo-1,4- <math>\beta</math>-Xilanase de <i>Pseudoalteromonas haloplanktis</i></p>  <p>Família: 8, EC 3.2.1.8, PDB: 1H12 Enovelamento: <math>(\alpha/\alpha)_6</math></p>	

Tabela 1.2. Continuação

Clã	Famílias	Membro representativo	Mecanismo
GH-N	28, 49.	Dextranase de <i>Penicillium minioluteum</i>	Inversão
		 <p>Família: 49, EC 3.2.1.11, PDB: 1OGM Enovelamento: <math>\beta</math>-helix</p>	

Apesar da variedade de enovelamentos encontrados nas GHs, todas as estruturas podem ser agrupadas em três tipos morfológicos segundo seu mecanismo de ação (figura 1.6). O primeiro tipo é conhecido como bolsão (*pocket*); os membros desta classe são geralmente exo-hidrolases cujo bolsão tem a capacidade de reconhecer o monossacarídeo terminal de um oligossacarídeo. O segundo tipo é uma conformação aberta em forma de sulco na superfície da proteína, o que permite a ligação de um polissacarídeo; os membros desta classe são tipicamente endo-hidrolases. A terceira classe se forma pela presença de um *loop* fechando o sulco da classe anterior, o que cria uma estrutura em forma de túnel. Esta morfologia é presente em hidrolases que têm como substratos longas cadeias polipeptídicas.

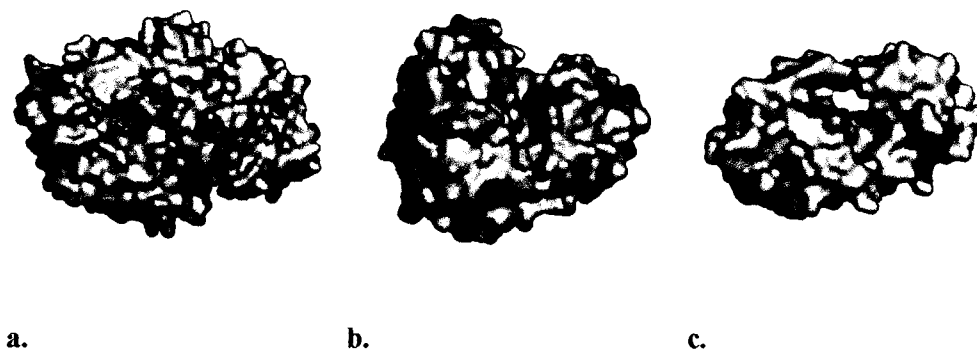


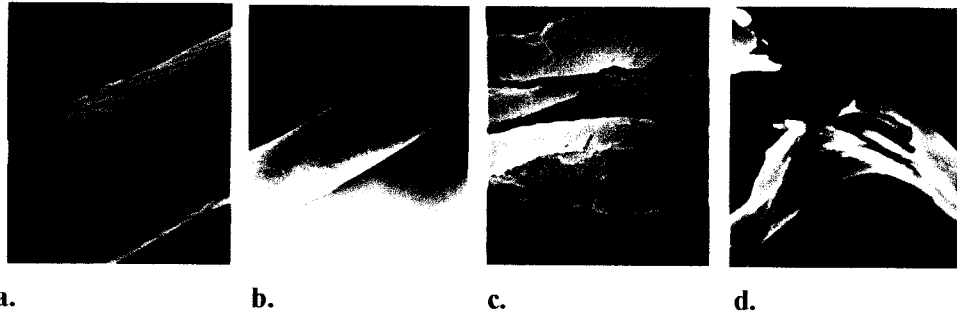
Figura 1.6. Morfologia geral das glicosil hidrolases, os resíduos catalíticos se apresentam em vermelho. a. Sítio ativo tipo bolsão (glucoamilase de *A. awamori*) b. Sítio ativo tipo sulco (endoglucanase E2 de *T. fusca*).c. Sítio ativo tipo túnel (Celobiohidrolase II de *T. reesei*). Figura adaptada de Davies e Henrissat<sup>11</sup>.

### 1.1.3. Módulos de ligação de carboidratos

Muitas glicosidases possuem, além do domínio catalítico, um ou mais domínios que participam da ligação dos carboidratos, ajudando na eficiência de ligação da enzima com seus substratos. Estes domínios são conhecidos como CBMs (*Carbohydrate-binding modules*) e originalmente foram definidos como CBDs (*Cellulose-binding domains*), pelo fato dos primeiros domínios deste tipo que foram estudados ligarem celulose como seu primeiro ligante<sup>16,17,18</sup>. Posteriormente foi encontrada uma grande diversidade de ligantes para estes módulos<sup>19</sup>, muitos dos quais têm sido estudados tanto por ressonância magnética nuclear<sup>20</sup> quanto por difração de raios-X<sup>21,22</sup>. Assim como os domínios catalíticos das glicosil hidrolases, os CBMs foram divididos em famílias com base na sua similaridade seqüencial. Atualmente, 43 famílias têm sido classificadas no CAZy (*Carbohydrate Active enZYmes*; <http://afmb.cnrs-mrs.fr/CAZY/>)<sup>6</sup>. Estes tipos de domínios reconhecem polissacarídeos tais como celulose cristalina, celulose não cristalina, quitina,  $\beta$ -1,3-glucanas e misturas de  $\beta$ -1,3-1,4-glucanas, xilano, manana, galactana e amido. Enquanto que alguns CBMs apresentam especificidades similares às das lectinas e se ligam a uma variedade de glicanos na superfície celular, a maioria encontra-se em glicosil hidrolases que degradam polissacarídeos insolúveis. Algumas famílias de CBMs reconhecem um único polissacarídeo enquanto que outras famílias se ligam a diferentes carboidratos. Estes domínios servem como um excelente modelo para o estudo dos mecanismos de reconhecimento proteína-carboidrato, assim como uma ferramenta em aplicações biotecnológicas<sup>23</sup>.

Seguindo a nomenclatura adotada para as glicosil hidrolases, os domínios são nomeados de acordo com a família a que pertencem, isto é, um domínio da família 17 presente em Cel5A de *Clostridium cellulovorans* pode ser chamado como CBM17, mais especificamente também pode ser incluído o organismo de onde provém a enzima da qual o domínio faz parte. Assim, no exemplo anterior o domínio pode ser nomeado como CcCBM17. Se a enzima possui mais de um domínio da mesma família, o número da posição do CBM na enzima relativo ao N-terminal deve ser incluído. Por exemplo, *Clostridium stercoarium* contém uma enzima com três módulos da família 6, o primeiro é nomeado como CsCBM6-1, o segundo como CsCBM6-2 e o terceiro como CsCBM6-3.

Os domínios que ligam carboidratos participam na interação da enzima com seu substrato de três formas diferentes: primeiro estes domínios geram o chamado efeito de proximidade, pelo qual a concentração da enzima aumenta na superfície do polissacarídeo<sup>24</sup> levando assim a uma mais rápida degradação do substrato. Depois, podem influenciar na capacidade da enzima para hidrolisar um determinado polissacarídeo<sup>25</sup>. Carrard e colaboradores<sup>26</sup> mostram que quando módulos da família 3 e da família 1 foram ligados ao mesmo domínio catalítico, a enzima apresentou diferenças na sua habilidade para degradar celulose cristalina. Por último, alguns CBMs têm a capacidade de quebrar a estrutura do polissacarídeo (Figura 1.7); este efeito não catalítico ainda não é entendido completamente e tem sido observado só na família 2<sup>27</sup> e na família 1<sup>28</sup>, não podendo, portanto, ser generalizado.



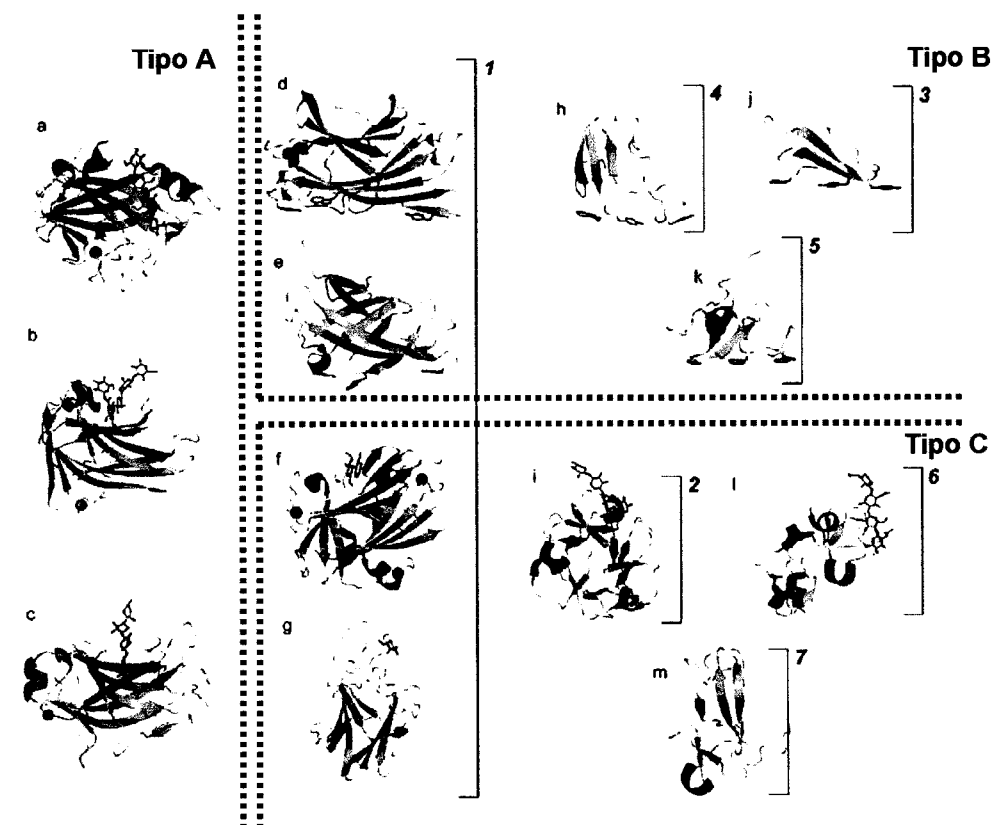
**Figura 1.7.** Microscopia eletrônica de uma fibra de algodão antes e depois de ser tratada com um domínio da celobiohidrolase I de *Penicillium janthinellum* (*PjCBM1*). **a.**Fibra de algodão (6000×). **b.** Fibra de algodão (2000×). **c.**Fibra de algodão depois de ter sido incubada 6 horas com  $CBD_{CBHI}$  (6000×). **d.** Fibra de algodão depois ter sido incubada 10 horas com *PjCBM1* (2000×). Figura obtida de Gao e colaboradores<sup>28</sup>.

Usando o critério da conservação da estrutura terciária, Boraston<sup>29</sup> e colaboradores propuseram uma divisão por tipo de enovelamentos para os CBMs análoga à classificação das GHFs. Infelizmente esta classificação é insuficiente para predizer a função dos diferentes domínios; portanto foi feita uma outra divisão tendo como critério a topologia e a similaridade na função. Assim, os CBMs foram agrupados em três tipos (Tabela 1.4 e Figura 1.6):

- Tipo A: domínios que ligam superfícies
- Tipo B: domínios que ligam cadeias de glicanos
- Tipo C: domínios que ligam oligossacarídeos curtos.

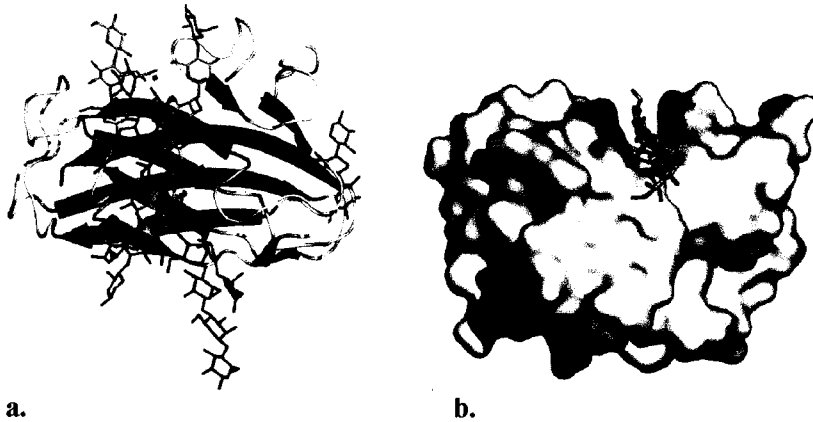
Tabela 1.3. Classificação dos CBMs por tipos.

Tipo	Superfamília	CBM-famílias
A	1, 3, 4, 5	1, 2a, 3, 5, 10
B	1	2b, 4, 6, 15, 17, 20, 22, 27, 28, 29, 34, 36
C	1, 2, 6, 7	9,13,14,18,32



**Figura 1.8.** Classificação dos CBMs por superfamílias e por tipos. Os quadros em linhas descontinuas apresentam os domínios agrupados por tipo. Os colchetes com números mostram exemplos das diferentes superfamílias. **a.** CcCBM17 de *Clostridium Cellulovorans* em complexo com celulose (PDB:1J84<sup>30</sup>). **b.** TmCBM4-2 de *T. maritima* em complexo com laminohexose (PDB:1GUI<sup>31</sup>). **c.** CjCBM15 de *Cellvibrio japonicus* em complexo com xylopentose (PDB:1GNY<sup>32</sup>). **d.** CtCBM3 de *Clostridium thermocellum* (PDB:1NBC<sup>33</sup>). **e.** CjCBM2 de *Cellulomonas fimi* (PDB:1EXG<sup>34</sup>). **f.** TmCBM9-2 de *T. maritima* em complexo com celobiose (PDB:1I82<sup>35</sup>). **g.** MvCBM32 de *Micromonospora viridifaciens* em complexo com galactose (PDB: 1EUU<sup>36</sup>). **h.** EcCBM5 de *Erwinia chrysanthemi* (PDB:1AIW<sup>37</sup>). **i.** SiCBM13 de *S. lividans* em complexo com xylopentose (PDB:1MC9<sup>38</sup>). **j.** TrCBM1 de *Trichoderma reesei* (PDB: 1CBH<sup>39</sup>). **k.** CjCBM10 de *Cellvibrio japonicus* (PDB:1E8R<sup>40</sup>). **l.** UdCBM18 de *Urtica dioica* em complexo com quititriose (PDB:1EN2<sup>41</sup>). **m.** TrCBM14 de *Tachypleus tridentatus* (PDB: 1DQC<sup>42</sup>). Figura adaptada de Boraston e colaboradores<sup>29</sup>.

Os CBMs tipo A mostram pouca ou nenhuma afinidade por carboidratos solúveis<sup>43</sup> apresentando uma preferência por celulose ou quitina cristalina. Na sua maioria os resíduos que participam da interação com o ligante são resíduos aromáticos que se dispõem em forma de um plano (Figura 1.8). Nos domínios tipo B a região de ligação com os carboidratos é um sulco de mais de 15 Å composto por vários sub sítios onde vários monossacarídeos podem se encaixar (Figura 1.9). Os domínios tipo C ligam mais eficientemente mono-, di ou tri-sacarídeos e apresentam sulcos menos extensos que os domínios tipo B. Em geral, fazem parte de xilanases e têm sido menos caracterizados que os domínios tipo A e B.



a.

b.

**Figura 1.9.** Representação esquemática da localização do sítio de ligação nos CBMs tipo B que apresentam enovelamento  $\beta$ -sandwich. a. Nesta figura vários domínios tipo B com seus ligantes foram superpostos, mas só a estrutura terciária do domínio CfCBM4 da Cel9B de *Cellulomonas fimi* é apresentada. b. Celupentose ligada no sulco do domínio CfCBM4<sup>31</sup>; a superfície dos resíduos aromáticos que participam da interação é apresentada em roxo.

## 1.2. Objetivos

### 1.2.1. Objetivo Geral

Cristalizar e determinar a estrutura tridimensional de algumas hidrolases de carboidratos de fungos.

### 1.2.2. Objetivos Específicos

- Reproduzir os cristais nativos da  $\beta$ -xilosidase de *Trichoderma reesei*, da  $\beta$ -galactosidase de *Penicillium sp* e da Exo-inulinase de *Aspergillus awamori*, e cristalizar a Endo-glucanase do *Rhodotermus marinus*.
- Obter derivados de átomos pesados ou espalhadores anômalos com os quais fosse possível determinar as fases dos cristais nativos das proteínas em estudo.
- Cristalizar a  $\beta$ -galactosidase de *Penicillium sp* em complexo com galactose e a Exo-inulinase de *Aspergillus awamori* em complexo com frutose.
- Calcular os mapas de densidade eletrônica das enzimas em estudo, construir e refinar suas estruturas tridimensionais.
- Além dos estudos cristalográficos, usar as técnicas de dicroísmo circular (CD) e espalhamento de raios-X a baixo ângulo (SAXS), como complemento para obter informações destas proteínas em solução.

### 1.3. Métodos

#### 1.3.1. Cristalização

Devido às propriedades físico-químicas das moléculas biológicas, ao número de parâmetros envolvidos no crescimento dos cristais e à sensibilidade das proteínas às condições externas, a cristalização se torna uma das etapas mais críticas na resolução da estrutura por raios-X de uma macromolécula biológica. A cristalização acontece através da diminuição lenta da solubilidade da proteína e está fortemente ligada à precipitação da mesma. O balanço das interações que controlam a solubilidade e/ou a conformação de uma macromolécula pode ser modificado, principalmente, por temperatura, concentração e natureza dos sais em solução, aditivos hidrofóbicos e solventes orgânicos<sup>44</sup>.

Existem muitos métodos usados na cristalização de proteínas. Dentre eles, os métodos mais frequentemente utilizados são as técnicas de difusão de vapor, em particular a técnica da gota pendurada. Nesta técnica uma pequena gota de 1-5  $\mu\text{l}$  de proteína é misturada com um volume igual ou similar de solução de cristalização (formada usualmente por tampão, sal e precipitante) e colocada sobre uma lamínula siliconizada. A lamínula é invertida e selada sobre um reservatório contendo 500-1000  $\mu\text{l}$  de solução de cristalização (Figura 1.10). A diferença de concentração entre a gota e o reservatório direciona o equilíbrio do sistema, que se dá por difusão através da fase de vapor da solução menos concentrada para a de maior concentração. Em um experimento perfeitamente delineado a proteína se torna supersaturada e o cristal começa a se formar quando a gota e o reservatório estão no equilíbrio ou perto dele.

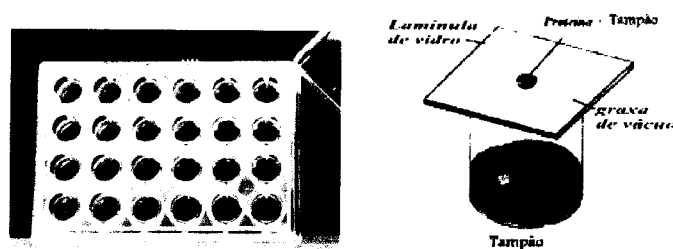


Figura 1.10. À esquerda placa utilizada nos experimentos de cristalização e à direita, um dos poços está amplificado mostrando como a lamínula, contendo a gota composta por solução do poço e solução da proteína, é colocada sobre o reservatório.

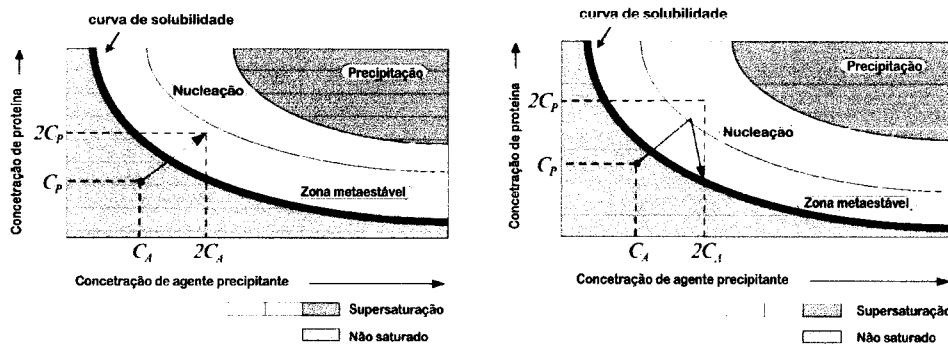


Na supersaturação as moléculas de proteína interagem entre si perdendo liberdade rotacional e translacional pela formação de pontes de hidrogênio estáveis, o que reduz a energia livre do sistema<sup>45,46</sup>. Os produtos mais favorecidos cineticamente por esta interação são os precipitados amorfos, que em algumas condições podem formar cristais<sup>47</sup>. Infelizmente, essas condições não podem ser preditas para uma dada proteína, e assim, esta metodologia está sujeita ao *screening*<sup>48</sup> de várias condições, o que pode ser uma tarefa árdua devido ao grande número de variáveis envolvidas.

A cristalização envolve três etapas: nucleação, crescimento e interrupção do crescimento. Uma explicação clássica do processo de formação dos núcleos e o crescimento dos cristais é dada pelo diagrama de solubilidade<sup>49</sup> apresentado na figura 1.11. A curva de solubilidade representa o limite entre a região de supersaturação da solução e a zona não saturada. Cada ponto da curva corresponde a uma concentração na qual proteína em solução está em equilíbrio com proteína que forma precipitados amorfos ou cristalinos. Na região não saturada a cristalização não pode acontecer. A região de supersaturação pode ser dividida em três zonas:

1. *zona de precipitação*: onde o excesso de moléculas de proteína se separa da solução formando agregados amorfos.
2. *zona de nucleação*: onde o excesso de moléculas de proteína agrega em forma de precipitado cristalino; perto da zona de precipitação microcristais podem ser confundidos com precipitado amorfo.
3. *zona metaestável*: nesta região uma solução supersaturada pode não nuclear por um longo período, a menos que a solução seja mecanicamente perturbada ou que a semente de um cristal seja introduzida.

Depois que os núcleos se formam na zona de nucleação, a solução pode retornar à região metaestável onde não aparecem mais núcleos e o cristal cresce ao mesmo tempo em que a concentração de proteína em solução diminui até atingir o equilíbrio (figura 1.11b).



**Figura 1.11.** Diagrama de fase que representa a solubilidade da proteína em função do agente precipitante a. Representação esquemática do que acontece em um experimento de difusão de vapor onde não se formam cristais b. Representação esquemática do processo de formação do cristal.

Em geral os parâmetros mais críticos na cristalização de uma macromolécula são: a solubilidade da proteína, a pureza e concentração da mesma; a concentração de sais e outros agentes precipitantes na solução de cristalização, a temperatura e o pH. Achar as condições nas quais uma proteína cristaliza implica fazer numerosos ensaios variando de maneira sistemática estes parâmetros.

### 1.3.2. Caracterização estrutural

Os processos de interação da radiação com a matéria constituem a base para a abordagem das técnicas espectroscópicas usadas na caracterização estrutural das proteínas. Uma breve revisão destes processos aqui se faz pertinente, a fim de contextualizar cada um dos métodos espectroscópicos que foram usados neste trabalho. Ressaltando que uma abordagem profunda tem sido amplamente discutida na literatura<sup>50</sup>.

#### 1.3.2.1. Difração de raios-X

O objetivo da cristalografia de monocristal é determinar a densidade eletrônica de um cristal a partir da intensidade dos raios-X difratados por ele. O problema teórico da interação radiação-matéria não tem uma solução exata, mas usando algumas aproximações é possível relacionar a intensidade do feixe espalhado  $I$  com a densidade eletrônica do cristal  $\rho$ . Quando um feixe de raios-X interage com um cristal ocorrem vários processos:

1. Absorção: principalmente por efeito fotoelétrico.

2. Espalhamento elástico: Alguns fótons são refletidos sem perda de energia, o que se constitui em radiação dispersa com o mesmo comprimento de onda da radiação incidente.
3. Espalhamento inelástico: Outros fótons são espalhados com perda de energia, estes constituem a radiação Compton, que tem um comprimento de onda um pouco maior que o comprimento da onda incidente.

Na difração de raios-X por um cristal o efeito cooperativo do espalhamento elástico é mais significativo que a soma das contribuições incoerentes. Portanto, no estudo cristalográfico somente o espalhamento elástico é levado em conta. No quadro do eletromagnetismo clássico, J.J. Thompson estudou o fato de que uma partícula carregada sob efeito de uma onda monocromática oscila com a mesma frequência da onda incidente. A oscilação implica aceleração e, portanto re-emissão da radiação. Thompson determinou que a intensidade do feixe espalhado por uma partícula de carga  $e$  e massa  $m$  a uma distância  $r$ , pode ser expressa pela seguinte equação:

$$I = I_0 \frac{e^4}{r^2 m^2 c^4} \left( \frac{1 + \cos^2 2\theta}{2} \right) \quad (1)$$

onde  $I_0$  é a intensidade do feixe incidente e  $2\theta$  o ângulo de espalhamento da radiação. Como a intensidade varia inversamente com o quadrado da massa, no processo de espalhamento de uma onda eletromagnética, apenas os elétrons contribuem efetivamente para a intensidade da onda espalhada. Embora os núcleos atômicos também sejam carregados e, portanto, também sofram a ação do campo elétrico da onda incidente, sua contribuição para a intensidade total é desprezível. A determinação das posições atômicas mediante cristalografia de raios-X está baseada na hipótese da coincidência entre o centróide da densidade de carga atômica e a correspondente posição nuclear.

Considerando um átomo esférico com o seu centro coincidente com a origem do sistema de coordenadas, temos que, a onda total espalhada por ele é dada por:

$$f(\mathbf{s}) = \int_{V. do. atomo} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r} \cdot \mathbf{s}) dv . \quad (2)$$

A equação anterior define o que se convencionou chamar de fator de espalhamento atômico e que tem sido determinado para diversos átomos<sup>51</sup>. Para um conjunto de  $N$  átomos a onda total espalhada será a soma das contribuições de cada átomo:

$$G(\mathbf{s}) = \sum_{j=1}^N f_j \exp(2\pi i \mathbf{r}_j \cdot \mathbf{s}) \quad (3)$$

Para um cristal que tem  $n_1$  células unitárias na direção  $\mathbf{a}$ ,  $n_2$  na direção  $\mathbf{b}$ , e  $n_3$  na direção  $\mathbf{c}$ , e que contem  $N$  átomos em cada uma delas, a onda espalhada será a soma das contribuições de cada cela unitária.

$$F(\mathbf{s}) = G(\mathbf{s}) \sum_{t=0}^{n_1} \exp(2\pi i t \mathbf{a} \cdot \mathbf{s}) \sum_{u=0}^{n_2} \exp(2\pi i u \mathbf{b} \cdot \mathbf{s}) \sum_{v=0}^{n_3} \exp(2\pi i v \mathbf{c} \cdot \mathbf{s}) \quad (4)$$

onde  $t\mathbf{a} + u\mathbf{b} + v\mathbf{c}$  é o vetor que determina a posição de uma célula com relação à origem. As somatórias da equação (4) são diferentes de zero só quando  $\mathbf{a} \cdot \mathbf{s} = h$ ,  $\mathbf{b} \cdot \mathbf{s} = k$ ,  $\mathbf{c} \cdot \mathbf{s} = l$ , sendo  $h, k, l$  números inteiros. A equação (4) se conhece como fator de estrutura já que depende do arranjo (estrutura) dos átomos no cristal, e pode ser escrita como:

$$F(hkl) = \sum_{j=1}^N f_j \exp 2\pi i (hx_j + ky_j + lz_j) = |F(hkl)| \exp[i\alpha(hkl)] \quad (5)$$

onde  $|F(hkl)|$  e  $\alpha(hkl)$  são chamados, respectivamente, de amplitude e fase do fator de estrutura  $F(hkl)$ ,  $x_j, y_j, z_j$  são as coordenadas fracionárias dos átomos e  $f_j$  é o fator de espalhamento atômico do  $j$ -ésimo átomo. Outra maneira de escrever a equação (5) é:

$$F(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp[2\pi i (hx + ky + lz)] dx dy dz \quad (6)$$

onde é fácil ver que a o fator de estrutura é a transformada de Fourier da densidade eletrônica, por tanto, é possível calcular  $\rho$  como a transformada inversa do fator de estrutura:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) \exp[-2\pi i (hx + ky + lz)] \quad (7)$$

Assim se o fator de estrutura é conhecido em magnitude e fase, é possível calcular a densidade eletrônica para cada ponto xyz que pertence à célula unitária do cristal.

$$[F(hkl)]^2 = F(hkl) \times F(hkl)^* \propto I(hkl) \quad (8)$$

A magnitude do fator de estrutura pode ser determinada a partir das intensidades  $I(hkl)$ , (equação (8)), mas a fase não pode ser determinada diretamente dos dados experimentais. Este problema se conhece como o *problema das fases*. Existem três métodos para a solução deste problema:

1. Métodos diretos; com os quais é possível determinar de forma indireta as fases a partir das intensidades. Estes métodos são usados normalmente para a solução de estruturas cristalográficas pequenas (aproximadamente 200 átomos) quando os dados de difração têm resolução atômica (superior ou próxima a 1Å).
2. Método de substituição molecular; com o qual as fases são determinadas a partir das fases de um modelo estrutural de uma proteína homóloga à proteína de interesse.
3. Método de substituição isomorfa; neste método para determinar as fases são incorporados átomos pesados ou espalhadores anômalos no cristal nativo. Como estes átomos apresentam um espalhamento diferenciado dos raios-X isso pode ser usado para determinar de maneira indireta as fases do cristal nativo.

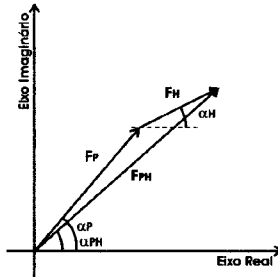
#### ***1.3.2.1.1. Substituição isomorfa múltipla***

O primeiro passo no uso da técnica de substituição isomorfa é a preparação de cristais derivados. Chama-se de derivado o cristal nativo no qual foi introduzido um ou vários átomos pesados ligados a cada molécula protéica. Normalmente utilizam-se neste procedimento de derivatização átomos de elevado peso atômico. Uma vez que os átomos pesados são introduzidos na estrutura cristalina de um composto, coletam-se dois conjuntos de dados de difração: um do cristal nativo e outro do cristal derivado. Se a introdução dos átomos não deformatar a estrutura do cristal, toda a

diferença entre os fatores de estrutura dos cristais nativo e derivado será devida aos átomos pesados. Neste caso é possível escrever para cada uma das reflexões hkl:

$$F_{PH} = F_P + F_H \quad (9)$$

onde  $F_{PH}$  é o fator de estrutura para o cristal derivado,  $F_P$  é o fator de estrutura para o cristal nativo e  $F_H$  é o fator de estrutura para o cristal formado apenas pela subestrutura dos átomos pesados. A representação da equação acima em um diagrama de Argand pode ser vista na Figura 1.11.



**Figura 1.12.** Diagrama de Argand mostrando as contribuições de  $F_P$  e  $F_H$  para o valor de  $F_{PH}$ . Na figura  $\alpha_{PH}$ ,  $\alpha_P$ ,  $\alpha_H$ , representam a fase do fator de estrutura do cristal derivado, do cristal nativo e da subestrutura dos átomos pesados respectivamente.

Aplicando a regra dos cossenos na figura 1.12 podemos encontrar a seguinte relação matemática:

$$\alpha_P = \alpha_H + \cos^{-1} \left( \frac{(F_{PH})^2 - (F_P)^2 - (F_H)^2}{2(F_P)(F_H)} \right) = \alpha_H \pm \alpha' \quad (10)$$

Como  $|F_P|$  e  $|F_{PH}|$  são calculados diretamente dos dados experimentais e  $\alpha_H$  e  $F_H$  podem ser calculados a partir das posições dos átomos pesados, na célula unitária, o valor da fase  $\alpha_P$  pode ser determinado para cada reflexão. Dado que  $\alpha_P$  na equação (10) pode assumir 2 valores distintos, se faz necessário o uso de um ou mais derivados para resolver esta ambigüidade. Esta técnica se conhece como substituição isomorfa múltipla MIR (*Multiple Isomorphous Replacement*).

1.3.2.1.2. *Substituição isomorfa simples com espalhamento anômalo*

Durante todo o desenvolvimento teórico realizado até o momento foi assumido que os elétrons dos átomos espalhadores eram considerados elétrons livres; ou seja, eles estavam fracamente ligados aos núcleos atômicos. Acontece, entretanto que quanto mais pesado for o núcleo atômico, mais fortemente ligados estão os elétrons; principalmente aqueles presentes nas camadas K e L. Uma vez que o processo de derivatização de cristais protéicos introduz na estrutura cristalina átomos pesados (Hg, Gd, I, Cs, Fe, Zn e Au, etc), deve-se levar em conta o comportamento individual destes átomos no processo de espalhamento de raios-X.

O chamado espalhamento anômalo em difração de raios-X consiste em um efeito de ressonância. A radiação incidente sobre o cristal apresenta uma frequência próxima de uma frequência natural do sistema elétron-núcleo de um determinado tipo de átomo (do átomo pesado no nosso caso). Uma das conseqüências imediatas do efeito anômalo é o surgimento de um fator de espalhamento atômico complexo para o tipo de átomo em questão. Desta forma o fator de espalhamento atômico para o espalhador anômalo deve ser escrito como:

$$f = f^o + \Delta f'(\lambda) + if''(\lambda) = f' + if'' \quad (11)$$

onde  $f^o$  é o fator de espalhamento atômico da equação (2), e  $\Delta f'(\lambda)$  e  $f''(\lambda)$  são chamadas as correções de dispersão, real e imaginária respectivamente. Assim as fases dos fatores de estrutura do cristal derivado podem ser determinadas se quando se coleta um conjunto de dados de um cristal derivado se sintoniza o comprimento de onda da radiação incidente para produzir o maior efeito anômalo possível. Mesmo não tendo coletado um conjunto de dados de um cristal nativo, é possível escrever a relação:

$$F_{PH} = F_P + F_H \quad (12)$$

Agora, como  $F_P$  só leva em conta os átomos leves da estrutura nativa:

$$F_P(+) = \sum_{j=1}^P f_j^0 \exp[2\pi i(hx_j + ky_j + lz_j)] \quad \text{para } hkl \quad (13)$$

e

$$F_P(-) = \sum_{j=1}^P f_j^0 \exp[-2\pi i(hx_j + ky_j + lz_j)] \quad \text{para } \overline{hkl} \quad (14)$$

Já a contribuição dos átomos pesados deve ser escrita como:

$$F'_H(+) = \sum_{j=1}^H f'_j \exp[2\pi i(hx_j + ky_j + lz_j)], \quad F''_H(+) = \sum_{j=1}^H f''_j \exp[2\pi i(hx_j + ky_j + lz_j)] \quad (15)$$

e

$$F'_H(-) = \sum_{j=1}^H f'_j \exp[-2\pi i(hx_j + ky_j + lz_j)], \quad F''_H(-) = \sum_{j=1}^H f''_j \exp[-2\pi i(hx_j + ky_j + lz_j)] \quad (16)$$

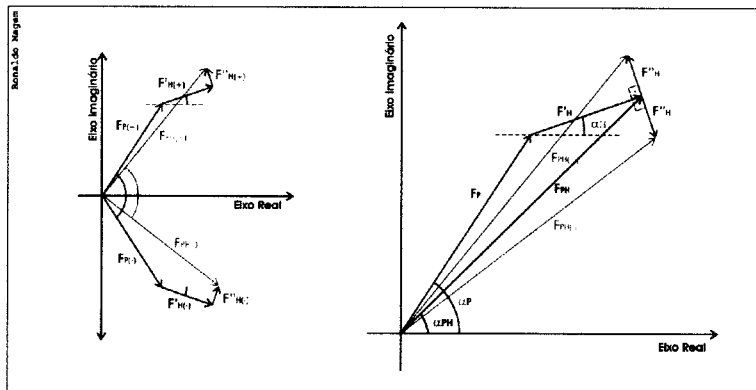
Combinando as equações (13) até (16) com (12), tem-se:

$$F_{PH}(+) = F_P(+) + F'_H(+) + iF''_H(+) \quad (17)$$

$$F_{PH}(-) = F_P(-) + F'_H(-) + iF''_H(-) \quad (18)$$

A representação das equações (17) e (18) no diagrama de Argand está mostrada na Figura 1.13. Uma das primeiras conseqüências da existência do efeito anômalo é a violação da lei de Friedel; ou seja:

$$F_{hkl} \neq F_{-h-k-l} \quad (19)$$



**Figura 1.13.** Representação dos fatores de estrutura, mostrando as contribuições de  $F_P(+)$ ,  $F'_H(+)$  e  $F''_H(+)$  para o valor de  $F_{PH}(+)$ . Idem para  $F_{PH}(-)$ . Na figura da direita, apenas espelhamos  $(-h-k-l)$  em relação ao eixo positivo real. Adaptado de Nagem<sup>52</sup>.



É possível obter o valor das fases  $\alpha_{PH}$  mediante os valores de  $F_{PH}(+)$  e  $F_{PH}(-)$  sem necessidade de usar os fatores de estrutura de um cristal nativo, porém é preciso encontrar as posições dos átomos pesados:

$$\alpha_{PH} = \alpha_H + \frac{\pi}{2} + \cos^{-1} \left( \frac{k[(F_{PH}(+)) - (F_{PH}(-))]}{2(F_H)} \right). \quad (20)$$

A técnica que utiliza apenas um cristal derivado com sinal anômalo para obtenção das fases dos fatores de estrutura recebe o nome de dispersão anômala simples (SAD). A utilização de um cristal nativo e um derivado com sinal anômalo recebe o nome de substituição isomorfa simples com espalhamento anômalo (SIRAS). Se além de um cristal nativo, dois ou mais derivados forem usados, a técnica recebe o nome de substituição isomorfa múltipla com espalhamento anômalo (MIRAS).

### 1.3.2.2. Espalhamento de raios-X a baixo ângulo (SAXS)

A técnica de espalhamento de raios-X a baixo ângulo (*Small Angle X-Ray Scattering*: SAXS) é uma técnica que fornece informações estruturais de flutuações ou heterogeneidades da densidade eletrônica com dimensões características da ordem de 10 a 1000Å. A forma das curvas de SAXS contém informações sobre o tamanho e a forma de partículas e também sobre a correlação espacial das mesmas. Uma partícula espalhadora pode compreender um único objeto espalhador ou um aglomerado de objetos formando domínios dispersos numa matriz homogênea. Para serem detectadas por esta técnica as partículas espalhadoras têm que possuir uma densidade eletrônica  $\rho$  que se diferencie da matriz  $\rho_o$ .

As técnicas de SAXS e cristalografia de monocristal compartilham o mesmo fundamento teórico, a grande diferença se encontra na estrutura dos elementos espalhadores. Em cristalografia estes se encontram arrançados numa rede tridimensional periódica e nos estudos de SAXS estão distribuídos de maneira aleatória. Assim a equação (6) que descreve a onda total espalhada deve ser escrita como

$$F(q) = \int_V \rho(r) \exp[-iq \cdot r] dV \quad (21)$$

onde  $q=2\pi S$  e  $V$  é o volume da amostra. Assim a intensidade espalhada pode ser calculada como

$$I(q) = F(q) \cdot F^*(q) = \int \int_{V_1, V_2} \rho(r_1) \rho(r_2) \exp[-i(r_1 - r_2) \cdot q] dV_1 dV_2 \quad (22)$$

Se considerarmos uma partícula isolada, esta integral depende apenas das distâncias relativas entre os elementos de volume da mesma e do produto das respectivas densidades eletrônicas. Assim é possível definir uma função de autocorrelação conhecida como função de Patterson ou função de distribuição de pares  $P(r)$ :

$$P(r_1 - r_2) = \int_V \rho(r_1) \rho(r_2) dV \quad (24)$$

Para sistemas suficientemente diluídos tal que os objetos espalhadores não interagem é possível promediar todas as possíveis orientações com relação ao feixe incidente e  $I(q)$  pode ser escrito através de uma transformada de Fourier da função distribuição de distância,  $p(r)$ <sup>53</sup>, tal que:

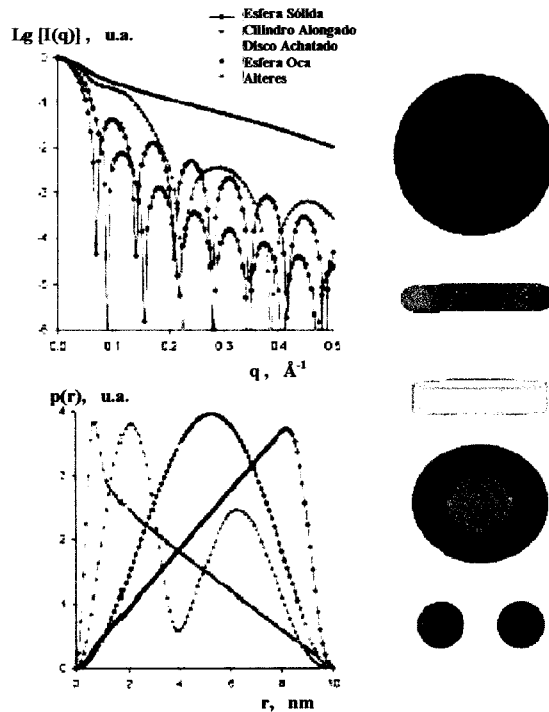
$$I(q) = 4\pi \int_0^{D_{max}} p(r) \left[ \frac{\text{sen}(q \cdot r)}{q \cdot r} \right] dr \quad (25)$$

e a transformada inversa da equação (25), permite calcular a função  $p(r)$  a partir da intensidade espalhada:

$$p(r) = \frac{1}{2\pi^2} \int_0^{\infty} I(q) [(q \cdot r) \cdot \text{sen}(q \cdot r)] dq \quad (26)$$

Na equação (25),  $D_{max}$  corresponde à dimensão máxima do objeto espalhador de tal forma que  $p(r)$  vai a zero para  $r \geq D_{max}$ . A princípio, a função  $p(r)$  contém as mesmas informações que a intensidade de espalhamento  $I(q)$ , mas no espaço real a morfologia da partícula espalhadora (forma e tamanho) pode ser inferida da função  $p(r)$ . Como

exemplo, apresentamos na Figura 1.14 curvas de espalhamento típicas e as respectivas funções  $p(r)$  de diferentes corpos geométricos com a mesma dimensão máxima.



**Figura 1.14.** Intensidades de espalhamento e função distribuição de distâncias  $p(r)$  de diferentes corpos geométricos. Figura adaptada de Svergun e Koch<sup>54</sup>.

No caso de partículas globulares (curva vermelha) a  $p(r)$  tem a forma de sino com o máximo em torno de  $D_{max}/2$ . Para partículas alongadas o máximo aparece a pequenas distâncias correspondendo ao raio da seção transversal (curva verde). Para partículas em forma de disco o máximo da  $p(r)$  é muito alargado (curva amarela), e também deslocado para distâncias menores que  $D_{max}/2$ . Na  $p(r)$  um máximo deslocado para distâncias maiores que  $D_{max}/2$  pode ser indicativo de partículas ocas (curva azul). Múltiplos máximos podem indicar partículas com subunidades; na figura o primeiro máximo corresponde às distâncias intra-subunidades e o segundo à distância entre as subunidades (curva roxa).

A partir da função  $p(r)$ , na ausência de interferência entre partículas, também é possível calcular o raio de giro  $R_g$ , da partícula espalhadora por:

$$R_g^2 = \frac{\int_0^{D_{\max}} p(r)r^2 dr}{2 \int_0^{D_{\max}} p(r) dr} \quad (27)$$

O conhecimento do raio de giro pode ser muito útil na investigação estrutural de uma partícula em solução, pois um aumento ou diminuição do mesmo provocado por qualquer variação nas condições da solução indicaria uma mudança nas posições relativas entre os átomos que compõem a partícula. O valor de  $R_g$  pode alternativamente ser obtido a partir da porção da curva de espalhamento a  $q$ 's muito baixos para a qual é válida a aproximação de Guinier:

$$I(q) = N(\Delta m)^2 e^{-\frac{1}{3}q^2 R_g^2} \quad (28)$$

onde  $N$  é o número de espalhadores e  $\Delta m$  o contraste da partícula com o solvente. A função  $p(r)$  contém informação mais completa que o raio de giro. Porém, em soluções de proteínas muito diluídas (menos de  $\sim 1\text{mg/mL}$ ), os erros estatísticos na medida da intensidade de espalhamento são muito grandes e não se consegue assim coletar informação de qualidade suficiente numa ampla faixa de  $q$  para poder determinar  $p(r)$ . Neste caso a curva experimental fica estatisticamente confiável somente a pequenos ângulos ( $q \rightarrow 0$ ) e somente o  $R_g$  obtido a partir da lei de Guinier pode ser determinado. No caso de proteínas dissolvidas numa solução próxima às condições fisiológicas (soluções predominantemente compostas por água, contendo diferentes sais ou açúcares em pequenas quantidades), a densidade eletrônica média do solvente é  $\rho_o = 0,33 \text{ e}^-/\text{\AA}^3$  (água). A densidade média de todas as proteínas é aproximadamente igual a  $\rho_l = 0,40 \text{ e}^-/\text{\AA}^3$ . As macromoléculas biológicas são essencialmente compostas por átomos leves (H, C, N, O e P no caso de ácidos nucleicos) e uma vez estando imersas em água ocorre pouca diferença na densidade eletrônica (ou no contraste) entre a molécula e o solvente. Portanto este tipo de solução produz um espalhamento muito fraco.

## 1.4. Referências

- <sup>1</sup> Golubev, A. M., Nagem, R. A. P., Brandao Neto, J. R., Neustroev, K. N., Eneyskaya, E. V., Kulminskaya, A. A., Shabalin, K. A., Savel'ev, A. N. and Polikarpov I. (2004) Crystal Structure of alpha-Galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism. *J Mol Biol.* **339**(2):413-22.
- <sup>2</sup> Rojas, A. L., Nagem, R. A. P., Neustroev, K. N., Golubev, A. M., Arand, M., Adamska, M., Eneyskaya, E. V., Kulminskaya, A. A. and Polikarpov, I. (2004) Crystal structures of  $\beta$ -galactosidase from *Penicillium sp.* and its complex with galactose. *J. Mol. Biol.* **343**: 1281-1292.
- <sup>3</sup> Nagem R. A. P., Rojas A. L., Golubev A. M., Korneeva O.S. Eneyskaya E. V., Kulminskaya A. A., Neustroev K. N. and Polikarpov I. (2004) Crystal Structure of exo-Inulinase from *Aspergillus awamori* : The Enzyme Fold and Structural Determinants of Substrate Recognition. *J. Mol. Biol.* **344**: 471-480.
- <sup>4</sup> Aparicio, R., Fischer, H., Scott, D.J., Verschueren, K.H.G., Kulminskaya, A., Eneyskaya, E., Neustroev, K.N., Craievich, A. F., Golubev, A., and Polikarpov I. (2002). Structural insights into the  $\alpha$ -Mannosidase from *T. reesei* obtained by synchrotron small-Angle X-ray solution scattering enhanced by X-ray Crystallography. *Biochemistry.* **41**, 9370-9375
- <sup>5</sup> Rojas, A. L., Fischer, H., Eneyskaya, E. V., Kulminskaya, A. A., Shabalin, K. A., Neustroev, K. N., Golubev A. M., Polikarpov I.(2005). Structural insights into the  $\beta$ -Xylosidase from *Trichoderma reesei*, obtained by Synchrotron Small-Angle X-ray and Circular Dichroism Spectroscopy. *Biochemistry*. In press.
- <sup>6</sup> Coutinho, P.M. and Henrissat, B. (1999) Carbohydrate-Active Enzymes: an integrated database approach. In Svensson, B. (ed.), *Recent Advances in Carbohydrate Bioengineering*. The Royal Society of Chemistry, Cambridge, pp. 3-12.
- <sup>7</sup> Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
- <sup>8</sup> Henrissat, B. and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788.
- <sup>9</sup> Henrissat, B. and Bairoch, A. (1996). Updating the sequence based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696.
- <sup>10</sup> Henrissat, B. and Davies, G (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**,637-644.

- <sup>11</sup> Davies, G. and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Struct.* **3**, 853–859.
- <sup>12</sup> Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature*, **255**, 609-614.
- <sup>13</sup> Durand, P., Lehn, P., Callebaut I, Fabrega, S., Henrissat, B., Mornon, J.P., (1997). Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiol.* **7**, 277-284.
- <sup>14</sup> Jenkins J, Leggio LL, Harris G, Pickersgill R. (1995).  $\beta$ -glucosidase,  $\beta$ -galactosidase, family A cellulases, family F xylanases and two barley glucanases form a superfamily of enzymes with 8-fold  $\beta/\alpha$  architecture and with two conserved glutamates near the carboxy-terminal ends of  $\beta$ -strands four and seven. *FEBS Lett.* **362**, 281-285.
- <sup>15</sup> Branden, C. and Tooze, J. (1991) Introduction to Protein Structure. Garland Publishing, New York.
- <sup>16</sup> Van Tilbeurgh, H., Tomme, P., Claeysens, M., Bhikhabhai, R. and Pettersson, G. (1986). Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. *FEBS Lett.* **204**, 223–227.
- <sup>17</sup> Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeysens, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei*QM 9414: analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* **170**, 575–581
- <sup>18</sup> Gilkes, N. R., Warren, R. A. J., Miller, Jr, R. C. and Kilburn, D. G. (1988) Precise excision of the cellulose binding domains from two *Cellulomonas fimicellulases* by a homologous protease and the effect on catalysis. *J. Biol. Chem.* **263**, 10401–10407
- <sup>19</sup> Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G. and Warren, R. A. J. (1999) Carbohydrate-binding modules: diversity of structure and function. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B. and Svensson, B., eds.), pp. 202–211, Royal Society of Chemistry, Cambridge.
- <sup>20</sup> Sorimachi, K., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1997). Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to beta-cyclodextrin. *Structure* **5**, 647–661
- <sup>21</sup> Rutenber, E. and Robertus, J. D. (1991) Structure of ricin B-chain at 2.5 Å resolution. *Proteins* **10**, 260–269.

- <sup>22</sup> Bains, G., Lee, R. T., Lee, Y. C. and Freire, E. (1992) Microcalorimetric study of wheat germ agglutinin binding to N-acetylglucosamine and its oligomers. *Biochemistry* **31**, 12624–12628.
- <sup>23</sup> Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes, C. A., Warren, R. A. J. and Kilburn, D. G. (1998). Characterization and affinity applications of cellulose-binding domains. *J. Chromatogr. B Biomed. Sci. Appl.* **715**, 283–296.
- <sup>24</sup> Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. and Gilbert, H. J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* **331**, 775–781.
- <sup>25</sup> Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A. J. and Kilburn, D. G. (2003). Recognition and hydrolysis of noncrystalline cellulose. *J. Biol. Chem.* **278**, 6120–6127.
- <sup>26</sup> Carrard, G., Koivula, A., Soderlund, H. and Beguin, P. (2000). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10342–10347.
- <sup>27</sup> McCartney, L., Gilbert, H. J., Bolam, D. N., Boraston, A. B. and Knox, J. P. (2004). Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Anal. Biochem.* **326**, 49–54.
- <sup>28</sup> Gao, P. J., Chen, G. J., Wang, T. H., Zhang, Y. S. and Liu, J. (2001). Non-hydrolytic disruption of crystalline structure of cellulose by cellulose binding domain and linker sequence of cellobiohydrolase I from *Penicillium janthinellum*. *Shengwu Huaxue Yu Shengwu Wuli Xuebao* **33**, 13–18.
- <sup>29</sup> Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* (2004) **382**, 769–781.
- <sup>30</sup> Notenboom, V., Boraston, A. B., Chiu, P., Frelove, A. C., Kilburn, D. G. and Rose, D. R. (2001). Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J. Mol. Biol.* **314**, 797–806.
- <sup>31</sup> Boraston, A. B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D. R., Kilburn, D. G. and Davies, G. J. (2002) Differential oligosaccharide recognition by evolutionarily-related  $\beta$ -1,4 and  $\beta$ -1,3 glucan-binding modules. *J. Mol. Biol.* **319**, 1143–1156.
- <sup>32</sup> Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose: evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* **276**, 49061–49065.

- <sup>33</sup> Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. and Steitz, T. A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751
- <sup>34</sup> Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E. and Harvey, T. S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993–7009.
- <sup>35</sup> Notenboom, V., Boraston, A. B., Kilburn, D. G. and Rose, D. R. (2001) Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry* **40**, 6248–6256.
- <sup>36</sup> Gaskell, A., Crennell, S. and Taylor, G. (1995) The three domains of a bacterial sialidase: a  $\beta$ -propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **3**, 1197–1205
- <sup>37</sup> Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F. and Marion, D. (1997). Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*. *Biochemistry* **36**, 16074–16086.
- <sup>38</sup> Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G. and Rose, D. R. (2002) High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry* **41**, 4246–4254.
- <sup>39</sup> Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. and Gronenborn, A. M. (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28**, 7241–7257.
- <sup>40</sup> Raghothama, S., Simpson, P. J., Szabo, L., Nagy, T., Gilbert, H. J. and Williamson, M. P. (2000) Solution structure of the CBM10 cellulose binding module from *Pseudomonas* xylanase A. *Biochemistry* **39**, 978–984.
- <sup>41</sup> Saul, F. A., Rovira, P., Boulot, G., Damme, E. J., Peumans, W. J., Truffa-Bachi, P. and Bentley, G. A. (2000). Crystal structure of *Urtica dioica* agglutinin, a superantigen presented by MHC molecules of class I and class II. *Structure Fold. Des.* **8**, 593–603.
- <sup>42</sup> Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K. and Kawano, K. (2000). Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *J. Biol. Chem.* **275**, 17929–17932.
- <sup>43</sup> Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I. and Mizuno, H. (2002). Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. *J. Mol. Biol.* **316**, 65–78.



<sup>44</sup> von Hippel, P.H. and T.W. Schleich. (1969). The effects of neutral salts on the structure and conformational stability of macromolecules in solution. In "Biological Macromolecules," Vol. II, G.D. Fasman and S.N. Timasheff, eds., Marcel Dekker, Inc., New York, pp. 417-584.

<sup>45</sup> McPherson, A. (1982). Preparation and Analysis of Protein Crystals. John Wiley & Sons, New York.

<sup>46</sup> McPherson, A. (1999) Crystallisation of biological macromolecules, Cold Spring Harbor Press, New York.

<sup>47</sup> McPherson, A. and Weickmann, J. (1990). X-ray analysis of new crystal forms of the sweet protein thaumatin. *J Biomol Struct Dyn*, 7, 1053-1060.

<sup>48</sup> Jancarik, J. and Kim, S. H. (1991) Sparse matrix sampling: a screening method for crystallisation of proteins. *J Appl Crystallogr*, 24, 409 - 411.

<sup>49</sup> Ducruix, A. and Giege, R. (1992) Crystallization of nucleic acids and proteins. A practical approach, Oxford University Press, New York.

<sup>50</sup> Giacovazzo, C. et al. Fundamentals of Crystallography. Oxford: Oxford University Press, 1992.

<sup>51</sup> International Tables For X-Ray Crystallography (1974). Vol. III. The Kynoch Press, Birmingham, England.

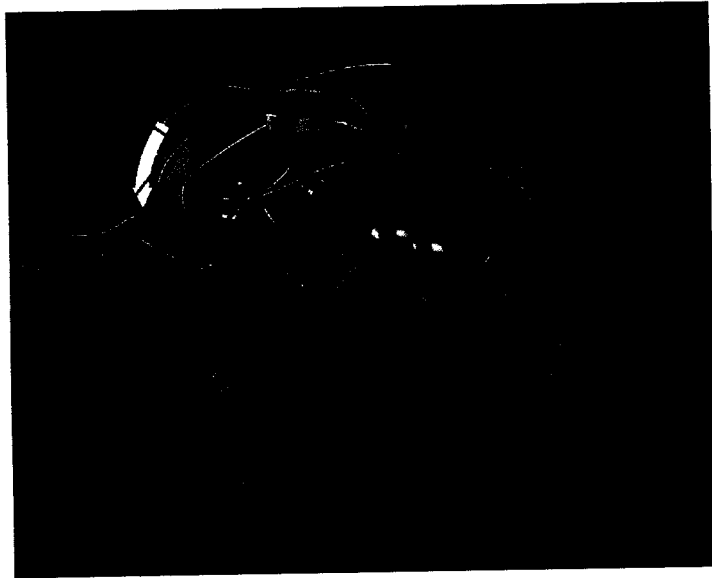
<sup>52</sup> Nagem, R. A. P. Otimização das condições para a aquisição de dados de derivados e para a determinação das fases dos fatores de estrutura de cristais de proteínas por meio da difração da luz síncroton. Tese de Doutorado. UNICAMP-IFGW, Campinas, SP (2003).

<sup>53</sup> O. Glatter and O. Kratky, *Small Angle X-Ray Scattering*. Academic Press: New York, 1982.

<sup>54</sup> D. Svergun, M.H.J. Koch, *Reports on Progress in Physics*, 66, 1735-1782 (2003).

# CAPÍTULO 2.

**$\beta$ -Galactosidase de *Penicillium sp***



## 2.1. Introdução

### 2.1.1. $\beta$ -galactosidases

A  $\beta$ -galactosidase (E.C 3.2.1.23) é uma enzima que hidrolisa as ligações  $\beta(1-3)$ - $\beta(1-4)$  de galactose em poli e oligossacarídeos. É sabido que muitas carbohidrases catalisam não só a reação de hidrólise, mas também a reação reversa de condensação ou transglicosilação. Esta propriedade, observada em  $\beta$ -galactosidases de vários organismos, tem sido usada para a síntese enzimática de oligossacarídeos que contêm galactose. Seu alto grau de especificidade permite que sejam usadas na síntese quimio-enzimática de oligossacarídeos contendo galactose com as configurações  $\beta(1-3)$  e/ou  $\beta(1-4)$  para as ligações químicas<sup>1,2</sup>. Assim, as enzimas de *Bacillus circulans* e de bovinos têm sido usadas com sucesso na síntese de Gal- $\beta$ -1,3-Glc-Nac, um oligossacarídeo fundamental do *sialyl-Lewis antigen*<sup>3,4</sup>. A mesma atividade de transglicosilação tem sido usada na síntese de Gal- $\beta$ -1,3-Gal-Nac, um importante constituinte das glicoproteínas (mucintype glicoproteins) e também na obtenção de N-acetil-lactosamina (D-Gal- $\beta$ -1,6-Glc-NAc)<sup>5</sup>. Já para a síntese de sialil-N-acetil-lactosamina, a  $\beta$ -galactosidase de *Diplococcus pneumoniae* tem sido usada em conjunto com sialidasas<sup>6</sup>. A habilidade de transglicosilação das  $\beta$ -galactosidases tem sido utilizada para a obtenção de derivados de  $\beta$ -galactosil-serina<sup>7</sup>.

$\beta$ -Galactosidases são usadas na hidrólise de lactose em produtos lácteos, tais como leite e soro de queijo, e  $\beta$ -galactosidases termoestáveis em particular têm um grande potencial em processos industriais destes produtos. Além disso,  $\beta$ -galactosidases têm sido usadas como uma ferramenta na biologia molecular devido à alta especificidade frente a ligações galactosil amplamente independente da variação dos aglicons. Tal comportamento permite determinar a presença de  $\beta$ -galactosidase com um ensaio colorimétrico simples baseado em substratos tais como para-nitrophenil- $\beta$ -D-galactopiranoside (PNPG) e  $\beta$ -D-galactopiranoside com aglicons cromogênicos (XGal). A  $\beta$ -galactosidase de *Escherichia coli* (Ec- $\beta$ -gal), da família GH2, é uma das  $\beta$ -galactosidases mais estudadas, tendo sua estrutura tridimensional determinada assim como as bases de seu mecanismo de reação<sup>8,9,10</sup>. Embora outras  $\beta$ -galactosidases importantes tenham sido encontradas em diversos organismos, além de

*Thermus thermophilus* A4  $\beta$ -galactosidase (A4- $\beta$ -gal) da família GH42, cuja estrutura tridimensional foi reportada<sup>11</sup>, nenhuma outra estrutura de  $\beta$ -galactosidase tinha sido reportada antes deste trabalho.

### 2.1.2. $\beta$ -galactosidase de *Penicillium sp.*

No presente trabalho apresentamos a análise estrutural da  $\beta$ -galactosidase de *Penicillium sp* e seu complexo com galactose. Os primeiros cristais desta proteína foram obtidos em 1999, fato que resultou em uma publicação internacional relatando os primeiros estudos cristalográficos desta proteína<sup>12</sup>. Desde então, nosso grupo tem trabalhado na resolução da estrutura desta proteína de aproximadamente 120 kDa. Infelizmente, as primeiras tentativas de Substituição Molecular falharam e a estrutura desta proteína só pode ser resolvida mediante a obtenção de cristais derivados. Este processo consumiu um tempo considerável até o momento em que se utilizou a técnica de derivatização rápida (*quick cryo soaking*)<sup>13,14</sup> para obtenção de derivados. A partir deste momento, obtivemos derivados altamente isomorfos e capazes de produzirem excelentes mapas de densidade eletrônica que nos permitiram resolver a estrutura cristalina desta proteína.

A  $\beta$ -galactosidase extracelular de *Penicillium sp* (Psp- $\beta$ -gal) possui uma alta atividade de transglicosilação para p-nitrofenil,  $\beta$ -D-galactopiranosose, lactose e metil  $\beta$ -D-galactopiranosose<sup>12</sup>, mostrando-se uma ferramenta poderosa para diversas reações de síntese enzimática.

Quando o presente estudo foi iniciado, a seqüência de aminoácidos da  $\beta$ -galactosidase de *Penicillium sp* era desconhecida, entretanto, a qualidade dos mapas de densidade eletrônica obtidos permitiu determinar a estrutura primária desta proteína. A seqüência assim determinada foi usada para o desenho de vários pares de oligonucleotídeos, os quais foram empregados para a amplificação do gene através da técnica de PCR.

Comparando a seqüência de aminoácidos com  $\beta$ -galactosidasas homólogas<sup>15</sup>, foi concluído que Psp- $\beta$ -gal pertence à família GH 35. Antes do presente estudo nenhuma estrutura tridimensional desta família era conhecida. Os resíduos Glu200 e Glu299 foram identificados como os resíduos catalíticos desta enzima sendo Glu200 o doador de prótons e o Glu299 o nucléofilo. Apesar da baixa similaridade da estrutura primária com outros membros da superfamília 4/7, Psp- $\beta$ -gal apresenta características

estruturais desta superfamília, incluindo a localização dos resíduos catalíticos e a distância entre os grupos carboxílicos.

Os resultados deste estudo publicados em um artigo na revista *Journal Molecular Biology*<sup>16</sup>, que se encontra no **Anexo 1**. Uma vez que toda a metodologia empregada, assim como os resultados e conclusões, estão descritos no artigo, optou-se por não reproduzi-los. Será feita neste capítulo somente uma breve descrição das metodologias e os resultados obtidos. Para uma discussão mais detalhada vide **Anexo 1**.

## **2.2. Metodologia**

### **2.2.1. Cristalização**

Ensaio de cristalização foram conduzidos para a  $\beta$ -galactosidase de *Penicillium sp* utilizando a técnica da gota pendurada, como descrito previamente por Neustroev e colaboradores<sup>12</sup>. A proteína foi extraída do fungo e liofilizada, pelo grupo do Dr Alexander M. Golubev, no Instituto de física nuclear, divisão de biofísica molecular (*Gatchina st Petersburg, Russia*). O liofilizado da proteína foi dissolvido em água para atingir concentrações entre 5 e 10 mg/ml e depois centrifugado a 3000 g por 5 minutos. Gotas de 1-5  $\mu$ l de volume desta solução foram misturadas com volumes iguais de uma solução de fosfato de sódio 50mM pH 4.0 e 15% de PEG 8000. As gotas foram equilibradas contra 1 ml da mesma solução a 18 °C.

### **2.2.2. Coleta e processamento de dados**

Foram coletados 4 conjuntos de dados, um nativo e três derivados (iodo, céσιο e urânio), na linha de cristalografia de proteínas do LNLS (Laboratório Nacional de Luz Síncrotron)<sup>17</sup>, usando um detector *Image Plate Mar345*. Os cristais foram congelados com nitrogênio depois de serem mergulhados em soluções que continham átomos pesados. Os derivados de iodo e céσιο foram preparados com o método do quick cryo-soaking<sup>13,14</sup> enquanto que o derivado de urânio foi obtido por meio da técnica tradicional de derivatização para proteínas. As imagens de difração foram processadas com os programas DENZO e SCALEPACK<sup>18</sup>.

### 2.2.3. Faseamento

O problema das fases foi resolvido pelo método de SIRAS usando o conjunto de dados nativo e o derivado de iodo. A posição dos átomos de iodo foi determinada por métodos diretos com ajuda dos programas DREAR<sup>19</sup> e SnB 2.1<sup>20</sup>. Essas posições foram usadas pelo programa SHARP<sup>21</sup> para o cálculo das fases. Usando o programa ARP/wARP<sup>22</sup> e um refinamento iterativo com o programa REFMAC<sup>23</sup>, construiu-se um modelo inicial da estrutura formado por glicinas. Como não foi possível determinar as cadeias laterais no mapa de densidade obtido pelo método de SIRAS, foram recalculadas as fases pelo método de MIRAS usando todos os conjuntos de dados. O modelo de poliglicinas foi usado para fazer *Solvent flattening* com ajuda do programa SOLOMON<sup>24</sup>.

Os dados dos derivados de céscio e urânio não foram usados na construção do modelo inicial, embora tenham sido utilizados com os conjuntos nativo e derivado de iodo para o cálculo das fases pelo método de MIRAS. O cálculo completo das fases usando 74 átomos pesados com diferentes valores de ocupância (0.8 – 0.1) foi feito com SHARP. Mais de 155.000 reflexões foram faseadas com uma figura de mérito de 0.57 em uma faixa de 111.0 – 2.40 Å.

Usando o programa ARP/wARP<sup>22</sup> e um refinamento iterativo com REFMAC<sup>23</sup>, construiu-se um modelo inicial da estrutura da  $\beta$ -galactosidase. Usando o programa O<sup>25</sup> construiu-se as cadeias laterais (970 resíduos), dado que não se conhecia a estrutura primária da  $\beta$ -galactosidase de *Penicillium* sp. A seqüência de aminoácidos determinada a partir do mapa da densidade eletrônica foi alinhada com seqüências conhecidas de outras  $\beta$ -galactosidases com ajuda do programa BLAST<sup>15</sup>. Nas cadeias laterais que apresentavam ambigüidade, a informação obtida pelo alinhamento com as outras  $\beta$ -galactosidases foi usada junto com mapas de diferença anômala, potências pontes de hidrogênio e interações proteína-oligossacarídeo para determinar o melhor tipo de resíduo no modelo.

## 2.3 Resultados e Discussão

Estudos de difração de raios-X mostram que a  $\beta$ -Galactosidase do *Penicillium* sp cristaliza-se no grupo espacial tetragonal P4<sub>3</sub> ou P4<sub>1</sub> com parâmetros de cela a = b = 110.96Å, c = 161.05 Å, e difrata até 1.90 Å de resolução. O cálculo do coeficiente de Matthews<sup>26</sup> sugere a presença de um dímero na unidade assimétrica ( $V_M = 2.3 \text{ \AA}^3\text{Da}^{-1}$ )

e um conteúdo de solvente de aproximadamente 45%.

**Tabela 2.1.** Estatísticas da coleta de dados de difração de raios-X dos cristais da  $\beta$ -galactosidase de *Penicillium sp.* Os valores para alta resolução encontram-se entre parênteses.

	Nat- $\beta$ -gal	I- $\beta$ -gal	Cs- $\beta$ -gal	U- $\beta$ -gal
Comprimento de onda (Å)	1.54	1.54	1.54	1.54
Grupo espacial	P4 <sub>3</sub>	P4 <sub>3</sub>	P4 <sub>3</sub>	P4 <sub>3</sub>
Dimensões da célula (Å)	a=b=110,96 c=161,05	a=b=110,62 c=159,89	a=b=110,83 c=160,92	a=b=111,01 c=161,33
Resolução (Å)	78.5 - 1.90 (1.93 - 1.90)	111.0 - 2.10 (2.15-2.10)	111.0 - 2.10 (2.15 - 2.10)	91.4 - 2.40 (2.46 - 2.40)
No. de observações	639065	944162	614561	608673
No. de únicas <sup>1</sup>	155025	224036	223388	148232
$\langle I/\sigma(I) \rangle$	13.9 (3.2)	12.5 (3.1)	8.4 (2.4)	10.0 (2.8)
Multiplicidade	4.1 (3.9)	4.2 (4.2)	2.8 (2.7)	4.1 (4.0)
Completeza	99.9 (100.0)	99.9 (100.0)	99.6 (99.6)	98.6 (95.7)
R <sub>merge</sub> <sup>2</sup>	10.5 (40.0)	10.7 (48.1)	12.3 (47.0)	13.0 (53.1)
Graus coletados	105	240	134	215
Solução crioprotetora	solução mãe 15% etileno glicol	solução mãe 0.5 M NaI 15% etileno glicol	solução mãe 0.33 M CsCl 15% etileno glicol	solução mãe 50 mM UO <sub>2</sub> Ac <sub>2</sub> 15% etileno glicol
tempo de banho	30 segundos	330 segundos	60 segundos	12 horas

<sup>1</sup> Multiplicidade do conjunto de dados derivado (nativo) calculada separando as reflexões que correspondem a pares de Friedel.

<sup>2</sup>  $R_{merge} = \sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \sum_{hkl} I_{hkl}$

Os resultados mostram que a  $\beta$ -galactosidase do *Penicillium sp* apresenta uma similaridade importante na estrutura primária com  $\beta$ -galactosidasas de *Aspergillus candidus*, *Aspergillus niger*, *Penicillium canescens* e *Talaromyces emersonii*. Nenhuma destas quatro  $\beta$ -galactosidasas têm estrutura tridimensional determinada.

O modelo final é um monômero com 971 resíduos, 7 cadeias de oligossacarídeos formados por 16 manoses e 12 NAG (N-acetil- $\alpha$ -D-glucosamina), 3 íons de sódio, 4 íons de fosfato 9 moléculas de etileno glicol e 1252 moléculas de água. O modelo foi refinado com o conjunto de dados nativo e os coeficientes de Hendrickson-Lattman derivados do MIRAS, obtendo-se estatísticas finais de R<sub>factor</sub> e R<sub>free</sub> de 16.5% e 18.5%, respectivamente. A qualidade da estereoquímica do modelo pode verificar-se no diagrama de Ramachandran (Figura 2.1).

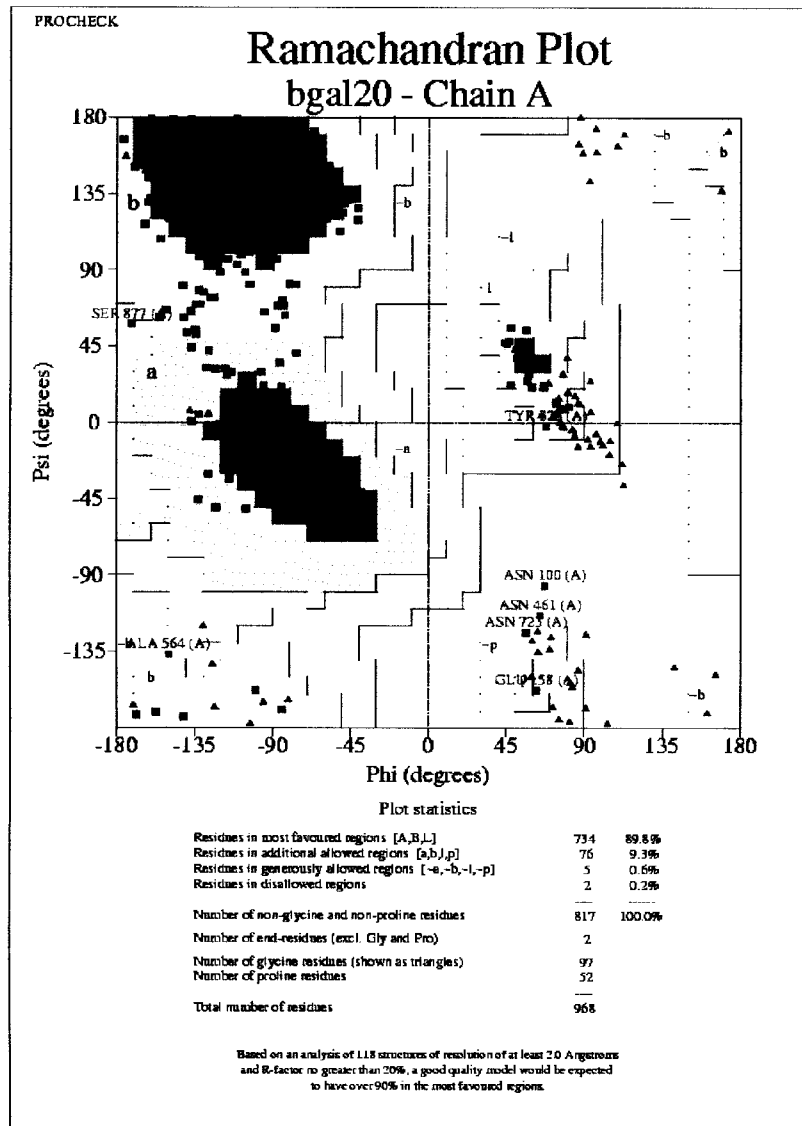


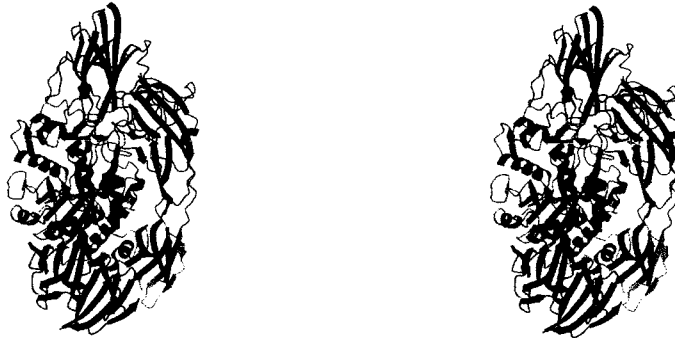
Figura 2.1 Diagrama de Ramachandran da estrutura nativa da  $\beta$ -galactosidase de *Penicillium sp*



### 2.3.1. Estrutura tridimensional da $\beta$ -galactosidase de *Penicillium*

*sp*

$\beta$ -Galactosidase de *Penicillium sp* é um monômero de 120kDa composto de cinco domínios. A estrutura tridimensional dobra-se ao redor do primeiro domínio o qual é chamado  $(\alpha/\beta)_8$ ,  $(\beta/\alpha)_8$ , ou *TIM barrel*. Este domínio contém a maioria dos resíduos do sítio ativo. Os outros domínios são constituídos na sua maioria de folhas  $\beta$  sendo que o domínio 2 apresenta um subdomínio com enovelamento tipo imunoglobulina e o domínio 3 é um  $\beta$ -sandwich. Os domínios 4 e 5 possuem um enovelamento tipo *jelly-roll*. (figura 2.2)



**Figura 2.2.** Representação estereográfica da estrutura da Psp- $\beta$ -gal. De acordo com PROCHECK (Laskowski et al., 1993)<sup>27</sup> os domínios 1, 2, 3, 4 e 5 são apresentados em ciano, vermelho, amarelo, verde e magenta, respectivamente. Em azul se apresenta a cadeia polipeptídica que conecta os domínios 3 e 4. As figuras foram feitas usando os programas Molscrip (Kraulis, 1991)<sup>28</sup>, Bobscrip (Esnouf, 1997)<sup>29</sup> e Raster3D (Merritt and Bacon, 1997)<sup>30</sup>.

### 2.3.2. Sítio ativo

Uma única molécula de galactose foi encontrada ligada ao domínio “TIM barrel” em uma conformação de cadeira com oxigênio do carbono anomérico em  $\beta$ . A Figura 2.4 mostra a galactose e os resíduos que interagem com ela. As distâncias de ligação são mostradas na tabela 2.2.

O alinhamento da seqüência de Psp- $\beta$ -gal com  $\beta$ -galactosidasas de outros organismos (Figure 2.3) indica que os 9 resíduos que participam da ligação da galactose são conservados em todas as espécies.



**Figura 2.3.** Alinhamento da estrutura primária das  $\beta$ -galactosidases de *Penicillium canescens* (171 aa), *Talaromyces emersonii* (218 aa), *Aspergillus candidus* (1005 aa), *Aspergillus niger* (1006 aa) e *Penicillium* sp. (determinada a partir do mapa de densidade). A similaridade entre as seqüências foi calculada usando o programa ALSCRIPT (Barton, 1993)<sup>31</sup> e se mostra em diferentes grus de azul. A cor mais escura corresponde a uma similaridade alta nas seqüências. O Bfactor médio, o coeficiente de correlação (CC) entre o mapa de densidade predito e observado e a superfície exposta ao solvente por resíduo são mostrados em uma escala de cor de amarelo a vermelho. Quatro cores diferentes foram usadas para os intervalos diferentes de Bfactor, CC e a área da superfície exposta ao solvente. Vermelho (Bfactor maior que 30 Å<sup>2</sup>; CC menor que 95%; área exposta maior que 90 Å<sup>2</sup>). Vermelho claro (Bfactor maior que 25 e menor que 30 Å<sup>2</sup>; CC menor que 96 e maior que 95%; área exposta maior que 60 e menor que 90 Å<sup>2</sup>). Laranja (Bfactor maior que 20 e menor que 25 Å<sup>2</sup>; CC menor que 97 e maior que 96%; área exposta maior que 30 e menor que 60 Å<sup>2</sup>). Amarelo (Bfactor menor que 20 Å<sup>2</sup>; CC maior que 97%; área exposta menor que 30 Å<sup>2</sup>). Os elementos da estrutura secundária da Psp- $\beta$ -gal são apresentados em ciano, vermelho, amarelo, verde e magenta para os domínios 1, 2, 3, 4 e 5, respectivamente. Em azul se apresenta a cadeia polipeptídica que conecta os domínios 3 e 4.

Tabela 2.2. Distâncias entre os átomos da proteína e os grupos OH da galactose.

Pontes de hidrogênio		
Átomos da galactose	Átomos da proteína	Distancia (Å)
O1	Glu299 O <sup>e1</sup>	2.75
	Glu200 O <sup>δ1</sup>	2.19
O2	Glu299 O <sup>δ1</sup>	1.93
	Asn199 N <sup>δ2</sup>	3.23
O3	Tyr96 O <sup>η</sup>	2.92
	Tyr96 O <sup>η</sup>	2.92
	Ala141N	3.06
	Ile139 O	3.03
O4	Asn 140 N <sup>δ2</sup>	3.02
	Glu 142 O <sup>e2</sup>	2.59
O6	Glu 142 O <sup>δ1</sup>	2.32
	Tyr 365 O <sup>η</sup>	2.51
Contact	Tyr 261	
	Tyr 343	

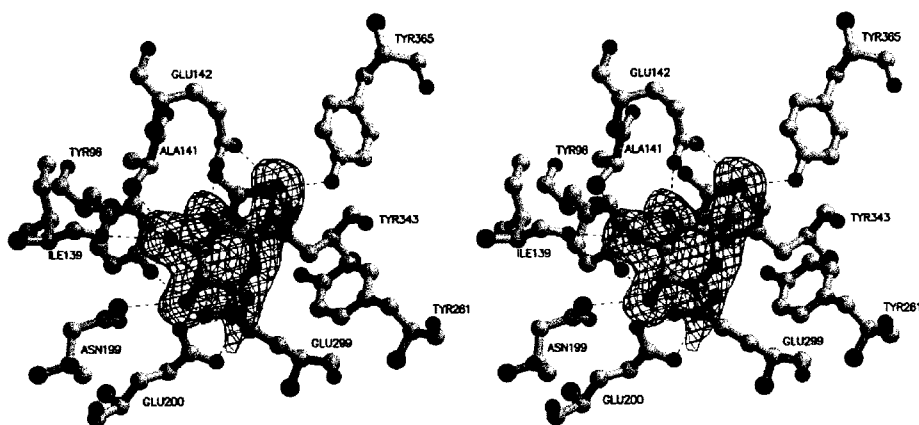


Figura 2.4. Mapa de densidade eletrônica no sítio catalítico indicando a presença da galactose. Este mapa foi gerado utilizando um modelo sem galactose.

Com base na superposição das estruturas tridimensional de Psp-β-gal e Ec-β-gal, foi inferido que os resíduos Glu200 e Glu299 localizados no final do C-terminal da fita β-4 e da fita β-7 do domínio “TIM barrel”, respectivamente são os resíduos catalíticos, aonde o Glu200 é o doador de próton e o Glu299 o nucleófilo. Esta é uma característica importante da superfamília 4/7. A distância média entre os oxigênios dos grupos carboxilatos do Glu200 e Glu299 são da ordem de 4.8 Å, o que é uma característica das enzimas retentoras da configuração anomérica do substrato.

2.3.3. Sítios de glicosilação

Foram encontrados sete sítios de N-glicosilação no mapa de densidade eletrônica, três destes com 5, 7 e 9 monossacarídeos, respectivamente (figura 2.5). Os oligossacarídeos estão localizados principalmente ao redor dos domínios 3, 4 e 5. No total é possível observar 28 monossacarídeos no mapa de densidade eletrônica (Figura 2.6).

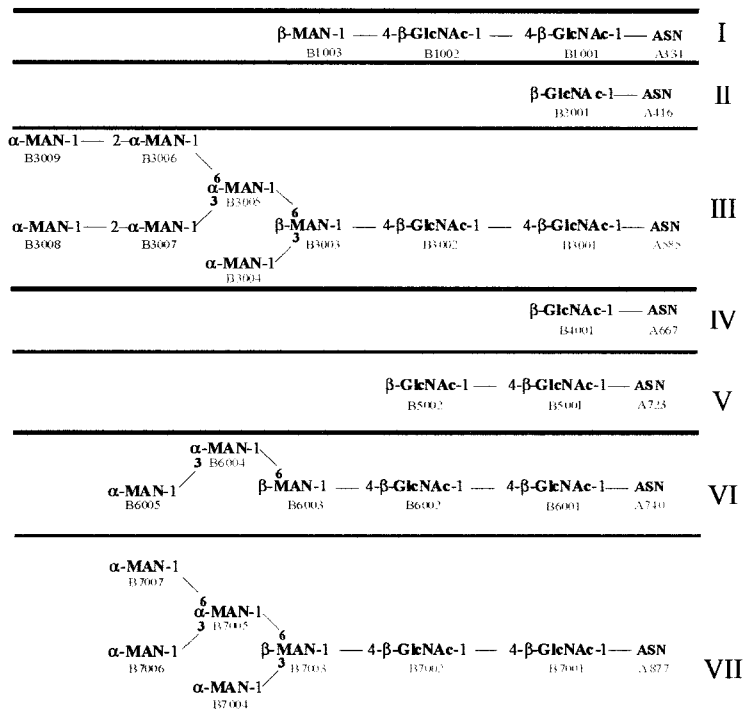


Figure 2.5. Estrutura dos 7 sítios de N-glicosilação (I-VII) na Psp-β-gal. Foram identificados cinco tipos de ligações glicosídicas (Manβ1-4GlcNAc, GlcNAcβ1-4GlcNAc, Manα1-3Man, Manα1-2Man e Manα1-6Man).



Figura 2.6. Representação da posição dos açúcares com respeito aos domínios da Psp-β-gal. Os oligossacarídeos estão principalmente localizados ao redor dos domínios 2 (verde), 3(laranja) e 5(vermelho).

## 2.4 Conclusões

- Determinou-se que a  $\beta$ -Galactosidase do *Penicillium sp* é um monômero de 120kDa composto de cinco domínios. A estrutura tridimensional enovela-se ao redor de um domínio catalítico conhecido como *TIM barrel*. Os outros domínios são constituídos na sua maioria de folhas  $\beta$  sendo que o segundo domínio apresenta um subdomínio tipo imunoglobulina, o terceiro é um  $\beta$ -*sandwich* e os domínios 4 e 5 tem enovelamento tipo *jelly rolls*.
- A  $\beta$ -galactosidase de *Penicillium sp* é o primeiro membro da família 35 das glicosil hidrolases cuja estrutura tridimensional foi determinada.
- Foram encontrados sete sítios de N-glicosilação no mapa de densidade eletrônica da  $\beta$ -galactosidase de *Penicillium sp*, três destes têm 5, 7 e 9 monossacarídeos respectivamente. Os oligossacarídeos estão principalmente localizados ao redor dos domínios 3, 4 e 5. No total é possível observar 28 monossacarídeos no mapa de densidade eletrônica.
- Foi encontrada uma única molécula de galactose ligada no domínio *TIM barrel* em uma conformação de cadeira com oxigênio do carbono anomérico em configuração  $\beta$ .
- O alinhamento da seqüência de aminoácidos de Psp- $\beta$ -gal com  $\beta$ -galactosidasas de outros organismos indica que dos 9 resíduos que participam da ligação da galactose são conservados em todas as espécies.

## 2.5. Referências

- <sup>1</sup> Usui, T., Kuboto, S. and Ohi, H. (1983). A convenient synthesis of  $\beta$ -D-galactosyl disaccharide derivatives using the  $\beta$ -D-galactosidase from *Bacillus circulans*. *Carbohydr. Res.* **244**, 315-323.
- <sup>2</sup> Mori, T., Fujita, S. and Okahata, Y. (1997). Transglycosylation in a two-phase aqueous-organic system with catalysis by a lipid-coated  $\beta$ -D-galactosidase. *Carbohydr. Res.* **298**, 65-73.
- <sup>3</sup> Fujimoto, H., Miyasato, M., Ito, Y., Sasaki, T. and Ajisaka, K. (1998). Purification and properties of recombinant  $\beta$ -galactosidase from *Bacillus circulans*. *Glycoconj. J.* **15**, 155-160.
- <sup>4</sup> Hedbys, L., Johansson, E., Mosbach, K., Larsson, P. O., Gunnarsson, A., Svensson, S. and Lonn, H. (1989). Synthesis of Gal beta 1-3GlcNAc and Gal beta 1-3GlcNAc beta-SEt by an enzymatic method comprising the sequential use of beta-galactosidases from bovine testes and *Escherichia coli*. *Glycoconj. J.* **6**, 161-168.
- <sup>5</sup> Vetere, A. and Paoletti, S. (1996). High-Yield Synthesis of *N*-Acetyllactosamine by Regioselective Transglycosylation. *Biochem. Biophys. Res. Commun.* **219**, 6-13.
- <sup>6</sup> Ajisaka, K., Fujimoto, H. and Isomura, M. (1994). Regioselective transglycosylation in the synthesis of oligosaccharides: comparison of  $\beta$ -galactosidases and sialidases of various origins. *Carbohydr. Res.* **259**, 103-115.
- <sup>7</sup> Cantacuzene, D. and Attal, S. (1991). Enzymic synthesis of galactopyranosyl-L-serine derivatives using galactosidases. *Carbohydr. Res.* **211**, 327-331.
- <sup>8</sup> Jacobson, R. H., Zhang, X-J., DuBose, R. F. and Matthews, B. W. (1994). Three-dimensional structure of  $\beta$ -galactosidase from *E. coli*. *Nature* **369**, 761-766.
- <sup>9</sup> Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X. J., Huber, R. E., Tronrud, D. E. and Matthews, B. W. (2000). High resolution refinement of  $\beta$ -galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for  $\alpha$ -complementation. *Protein Sci.* **9**, 1685-1699.
- <sup>10</sup> Juers, D. H., Huber, R. E. and Matthews, B. W. (1999). Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between  $\beta$ -galactosidase and other glycohydrolases. *Protein Sci.* **8**, 122-136.
- <sup>11</sup> Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, Shoun, H. and Wakagi T. (2002). Trimeric crystal structure of the glycoside hydrolase family 42 beta-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J. Mol. Biol.* **322**, 79-91.
- <sup>12</sup> Neustroev, K. N., de Sousa, E. A., Golubev, A. M., Brandão Neto, J. R., Eneyskaya, E. V., Kulminkaya, A. A. and Polikarpov, I. (2000). Purification,

crystallization and preliminary diffraction study of beta-galactosidase from *Penicillium sp.* *Acta Cryst.* D56, 1508-1509.

<sup>13</sup> Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta Cryst.* D57, 996-1002.

<sup>14</sup> Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst.* D56, 232-237.

<sup>15</sup> Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.

<sup>16</sup> Rojas, A. L., Nagem, R. A. P., Neustroev, K. N., Golubev, A. M., Arand, M., Adamska, M., Eneyskaya, E. V., Kulminskaya, A. A. and Polikarpov, I. (2004) Crystal structures of  $\beta$ -galactosidase from *Penicillium sp.* and its complex with galactose. *J. Mol. Biol.* 343: 1281-1292

<sup>17</sup> Polikarpov, I., Perles, L. A., *et al.* & Craievich, A. (1997). Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *J. Synchrotron Rad.* 5, 72-76

<sup>18</sup> Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326.

<sup>19</sup> Blessing, R. H. & Smith, G. D. (1999). Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. *J. Appl. Cryst.* 32, 664-670

<sup>20</sup> Weeks, C. M. & Miller, R. (1999). The design and implementation of SnB version 2.0. *J. Appl. Cryst.* 32, 120-124.

<sup>21</sup> La Fortelle, E. de & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* 276, 472-494.

<sup>22</sup> Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* 6, 458-463

<sup>23</sup> Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* D53, 240-255.

<sup>24</sup> Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Cryst.* D52, 30-42.

<sup>25</sup> Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst.* A47, 110-119.

- 
- <sup>26</sup> Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
- <sup>27</sup> Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
- <sup>28</sup> Kraulis, P. J. (1991). Molscript – A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
- <sup>29</sup> Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133-138.
- <sup>30</sup> Merritt, E. A. & Bacon, D. J. (1997). Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.
- <sup>31</sup> Barton G. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.



# Annexo 1

Rojas, A. L., Nagem, R. A. P., Neustroev, K. N., Golubev, A. M., Arand, M., Adamska, M., Eneyskaya, E. V., Kulminskaya, A. A. and Polikarpov, I. (2004) Crystal structures of  $\beta$ -galactosidase from *Penicillium sp.* and its complex with galactose. *J. Mol. Biol.* 343: 1281-1292

## Crystal Structures of $\beta$ -Galactosidase from *Penicillium* *sp.* and its Complex with Galactose

A. L. Rojas<sup>1</sup>†, R. A. P. Nagem<sup>1</sup>†, K. N. Neustroev<sup>2</sup>, M. Arand<sup>3</sup>  
M. Adamska<sup>3</sup>, E. V. Eneyskaya<sup>2</sup>, A. A. Kulminskaya<sup>2</sup>, R. C. Garratt<sup>1</sup>  
A. M. Golubev<sup>2</sup> and I. Polikarpov<sup>1\*</sup>

<sup>1</sup>Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, CEP 13560-970 São Carlos, SP Brazil

<sup>2</sup>Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg 188300, Russia

<sup>3</sup>Institute of Pharmacology and Toxicology, University of Wuerzburg, Versbacher Str. 9 D-97078 Wuerzburg, Germany

$\beta$ -Galactosidases catalyze the hydrolysis of  $\beta$ (1-3) and  $\beta$ (1-4) galactosyl bonds in oligosaccharides as well as the inverse reaction of enzymatic condensation and transglycosylation. Here we report the crystallographic structures of *Penicillium* *sp.*  $\beta$ -galactosidase and its complex with galactose solved by the SIRAS quick cryo-soaking technique at 1.90 Å and 2.10 Å resolution, respectively. The amino acid sequence of this 120 kDa protein was first assigned putatively on the basis of inspection of the experimental electron density maps and then determined by nucleotide sequence analysis. Primary structure alignments reveal that *Penicillium* *sp.*  $\beta$ -galactosidase belongs to family 35 of glycosyl hydrolases (GHF-35). This model is the first 3D structure for a member of GHF-35. Five distinct domains which comprise the structure are assembled in a way previously unobserved for  $\beta$ -galactosidases. Superposition of this complex with other  $\beta$ -galactosidase complexes from several hydrolase families allowed the identification of residue Glu200 as the proton donor and residue Glu299 as the nucleophile involved in catalysis. *Penicillium* *sp.*  $\beta$ -galactosidase is a glycoprotein containing seven *N*-linked oligosaccharide chains and is the only structure of a glycosylated  $\beta$ -galactosidase described to date.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** glycosyl hydrolases;  $\beta$ -galactosidases; crystal structure; *Penicillium* *sp.*; quick cryo-soaking

\*Corresponding author

### Introduction

The enzyme  $\beta$ -galactosidase (EC 3.2.1.23) hydrolyses  $\beta$ (1-3) and  $\beta$ (1-4) galactosyl bonds in poly- and oligosaccharides and also catalyzes enzymatic condensation and transglycosylation.<sup>1–7</sup>  $\beta$ -Galactosidases are used for the hydrolysis of lactose in dairy products, such as milk and cheese whey, and thermostable  $\beta$ -galactosidases recently attracted special interest due to their potential usefulness in the industrial processing of lactose-containing products.<sup>8</sup> Furthermore,  $\beta$ -galactosidases

have also been used as a tool in molecular biology due to the fairly strict specificity towards galactosyl bonds but not to a wide variety of aglycons. This feature allows the use of simple, colorimetric assays based on substrates such as *p*-nitrophenyl- $\beta$ -D-galactopyranoside (PNPG) and  $\beta$ -D-galactopyranoside with chromogenic aglycons (XGal). The extracellular  $\beta$ -galactosidase from *Penicillium* *sp.* (Psp- $\beta$ -gal) possesses a high trans-glycosylation activity towards PNPG, lactose and methyl  $\beta$ -D-galactopyranoside.

Glycosidases may be classified on the basis of two independent criteria, either *via* their EC numbers (derived from the chemistry of the catalyzed reaction) or on the basis of their evolutionary relationships. The second criterion, based on sequence homology, divides glycosyl hydrolases into 90 families (GHFs)<sup>9–13</sup> of which the known  $\beta$ -galactosidases belong to GHF-1, GHF-2, GHF-35 and GHF-42. These are all members of a superfamily (or clan) termed GH-A, which comprises families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51 and 53. GH-A enzymes all

† A.L.R. and R.A.P.N. contributed equally to this work.

Abbreviations used: PNPG, *p*-nitrophenyl- $\beta$ -D-galactopyranoside; SIRAS, single isomorphous replacement with anomalous scattering; MIRAS, multiple isomorphous replacement with anomalous scattering; rms, root mean square; PDB, Protein Data Bank; MR, molecular replacement.

E-mail address of the corresponding author: ipolikarpov@if.sc.usp.br

cleave glycosidic bonds *via* a retaining mechanism and possess a catalytic domain which is based on a TIM barrel fold. Two glutamic acid residues act as proton donor and nucleophile<sup>14</sup> and emanate from strands 4 and 7 of the barrel and for this reason this clan is sometimes referred to as the 4/7 superfamily. At present, 3D models are only available for enzymes from families 1, 2, 5, 10, 17 and 42.<sup>15</sup>

The  $\beta$ -galactosidase from *Escherichia coli* (Ec- $\beta$ -gal), an enzyme which belongs to GHF-2, is one of the most widely studied  $\beta$ -galactosidases. The 3D structure of Ec- $\beta$ -Gal is known and the structural basis for its reaction mechanism has been reported.<sup>16-18</sup> The only other  $\beta$ -galactosidase of known crystal structure is that from *Thermus thermophilus* (A4- $\beta$ -gal), a member of GHF-42.<sup>8</sup> Ec- $\beta$ -gal forms a 464 kDa homotetramer in solution, whereas A4- $\beta$ -gal is a trimer.

In this work, we describe the crystal structure of *Penicillium* sp.  $\beta$ -galactosidase and its complex with galactose determined by quick cryo-soaking and SIRAS approaches. Prior to our studies, neither the primary nor the tertiary structures of Psp- $\beta$ -gal had been reported. Therefore, a putative amino acid sequence of the enzyme was obtained initially by inspection of the electron density maps. This information was then used to correctly identify the  $\beta$ -galactosidase gene from *Penicillium* sp. and the amino acid sequence of the enzyme was determined by translation of the DNA sequence. Remarkably, comparison of these two sequences<sup>19</sup> reveals that 91% of the amino acid sequence was correctly identified by the visual inspection of the electron density maps and classifies Psp- $\beta$ -gal as a member of the GH family 35. The Psp- $\beta$ -gal catalytic residues Glu200 and Glu299 were identified as the proton donor and the nucleophile of the reaction, respectively.

## Results and Discussion

### Primary and tertiary structures determination

All attempts to solve the crystal structure of Psp- $\beta$ -gal using the molecular replacement method with the *E. coli* and *T. thermophilus*  $\beta$ -galactosidase structures as search models were unsuccessful. The phase problem was solved by the single isomorphous replacement with anomalous scattering (SIRAS) method. An initial hybrid model, consisting of a poly-glycine peptide chain and water molecules, was built automatically in the SIRAS-derived electron density map. Since the amino acid sequence of this enzyme was unknown, the identity of the amino acid residues was derived from the experimental electron density map. To improve the phase estimates and the quality of the electron density map, a complete multiple isomorphous replacement with anomalous scattering (MIRAS) phase calculation using four data sets up to 1.9 Å resolution (one native and three derivative sets) was performed with SHARP.<sup>20</sup> The MIRAS derived

phases along with the molecular envelope obtained from the poly-glycine model were used in a density modification procedure with SOLOMON.<sup>21</sup> The final phases were of excellent quality and allowed the calculation of the electron density maps that were then used for the primary structure assignment.

Based on this putative primary structure, different sets of degenerate primers were constructed and used to generate gene specific fragments from genomic *Penicillium* sp. DNA *via* conventional PCR. The Psp- $\beta$ -gal gene sequence, containing a total of seven protein coding exons and six introns, was obtained (see Materials and Methods). The amino acid sequence translated from the Psp- $\beta$ -gal gene has 1011 residues. However, careful inspection of the Psp- $\beta$ -gal primary structure revealed that the 40 N-terminal amino acid residues are involved in signaling and are cleaved from the mature protein. Remarkably, for the remaining 971 amino acid residues, 91% of the primary structure identified from the electron density inspection was correctly determined. In Figure 2, the primary structure of Psp- $\beta$ -gal is aligned with a number of homologous members of the GH family 35. Finally, the Psp- $\beta$ -gal amino acid sequence was used to construct the definitive crystallographic model. The significant sequence similarity with  $\beta$ -galactosidases from *Aspergillus candidus*, *Aspergillus niger* and *Talaromyces emersonii* enables us to affirm that these proteins should share significant structural homology with Psp- $\beta$ -gal (Figure 2).

### Description of the 3D model

The crystallographic structure of Psp- $\beta$ -gal was solved at 1.90 Å resolution with a final  $R_{\text{factor}}$  and  $R_{\text{free}}$  of 16.5% and 18.5%, respectively. The model has good stereochemical parameters, with a root mean square (rms) deviation of 0.005 Å for bond distances, of 1.5° for bond angles, and an average  $B$ -factor of 23.0 Å<sup>2</sup> for all atoms. Only two amino acid residues (Asn140 and Asn501) are found in the disallowed regions of the Ramachandran plot. The hydrogen bond interactions between Asn140 O<sup>δ1</sup> and the Tyr138 hydroxyl (2.59 Å) and between this hydroxyl and the Asp808 main-chain amide (2.69 Å) may explain the observed energetically unfavorable conformation of residue Asn140. In contrast, Asn501 is located at the protein surface and the electron density for this amino acid residue is poor. Four cysteine residues form two disulfide bridges (Cys205-Cys206 and Cys267-Cys316). Two alternative conformations are observed for the side-chains of the residues Ile62, Val653, and Tyr930.

The refined Psp- $\beta$ -gal model comprises 971 amino acid residues, 16 mannoses (MAN) and 12 *N*-acetylglucosamines (NAG) in seven oligosaccharide chains, 1252 water molecules, three sodium ions, four phosphate ions and nine ethylene glycol molecules. Its structure can be divided into five domains. The first domain, containing the catalytic site, is a distorted TIM barrel comprising 355 amino

acid residues (Leu41-Gly395). Unlike representative TIM barrel, which consists of eight  $\beta/\alpha$  repeats, the  $\beta/\alpha$  barrel in Psp- $\beta$ -gal lacks the fifth helix. Furthermore, the hydrogen-bonding pattern around the barrel is irregular on one side. This is the consequence of several factors including the shortness of some of the strands and the presence of distortions introduced by proline residues and a  $\beta$ -bulge in the seventh strand. A similar feature was observed in the crystal structure of Ec- $\beta$ -gal<sup>16</sup> in which this strand was also described as distorted. The structure of this domain is completed by an antiparallel two-stranded  $\beta$ -sheet, formed by the N-terminal region (Leu41-Leu60), which seals off the base of the barrel.

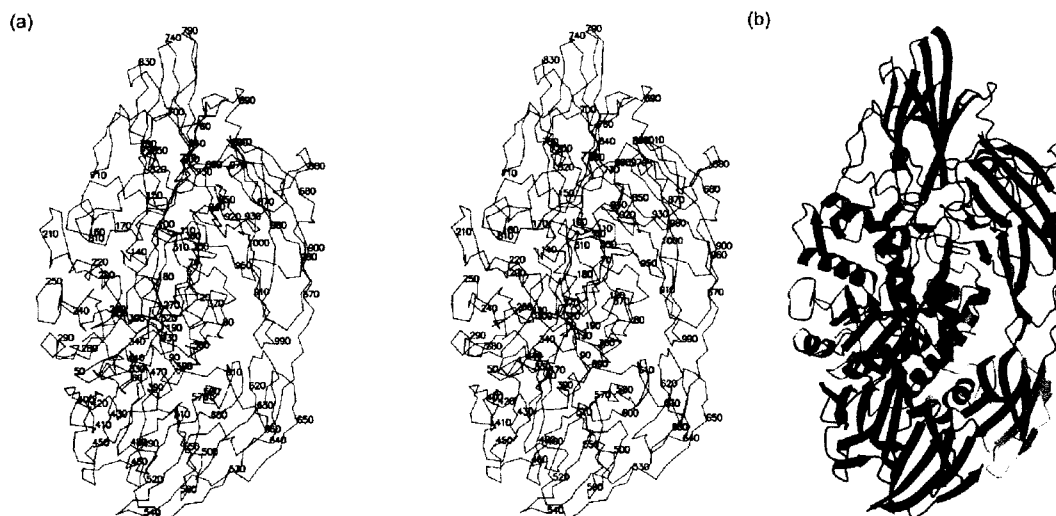
The second domain, comprising amino acid residues Tyr396-Tyr576, consists of 16 antiparallel  $\beta$ -strands and an  $\alpha$ -helix at its C terminus. The fold of this domain appears to be unique as evidenced by a DALI<sup>22</sup> structural similarity search. Notwithstanding this general observation, the last seven strands of the domain form a subdomain with an immunoglobulin-like fold. More specifically, this can best be described as being an I-type Ig fold in which the first strand (conventionally labeled the *a* strand) is divided between the two  $\beta$ -sheets. In the Psp- $\beta$ -gal structure this strand is interrupted by a 12-residue insertion which forms an additional edge-strand to the second  $\beta$ -sheet of the subdomain. The remainder of the second domain forms a series of  $\beta$ -hairpins at its N terminus, four strands of which are contiguous with part of the Ig-like sub-domain forming in total a seven-stranded antiparallel  $\beta$ -sheet.

The third domain (Trp577-Tyr665) is much

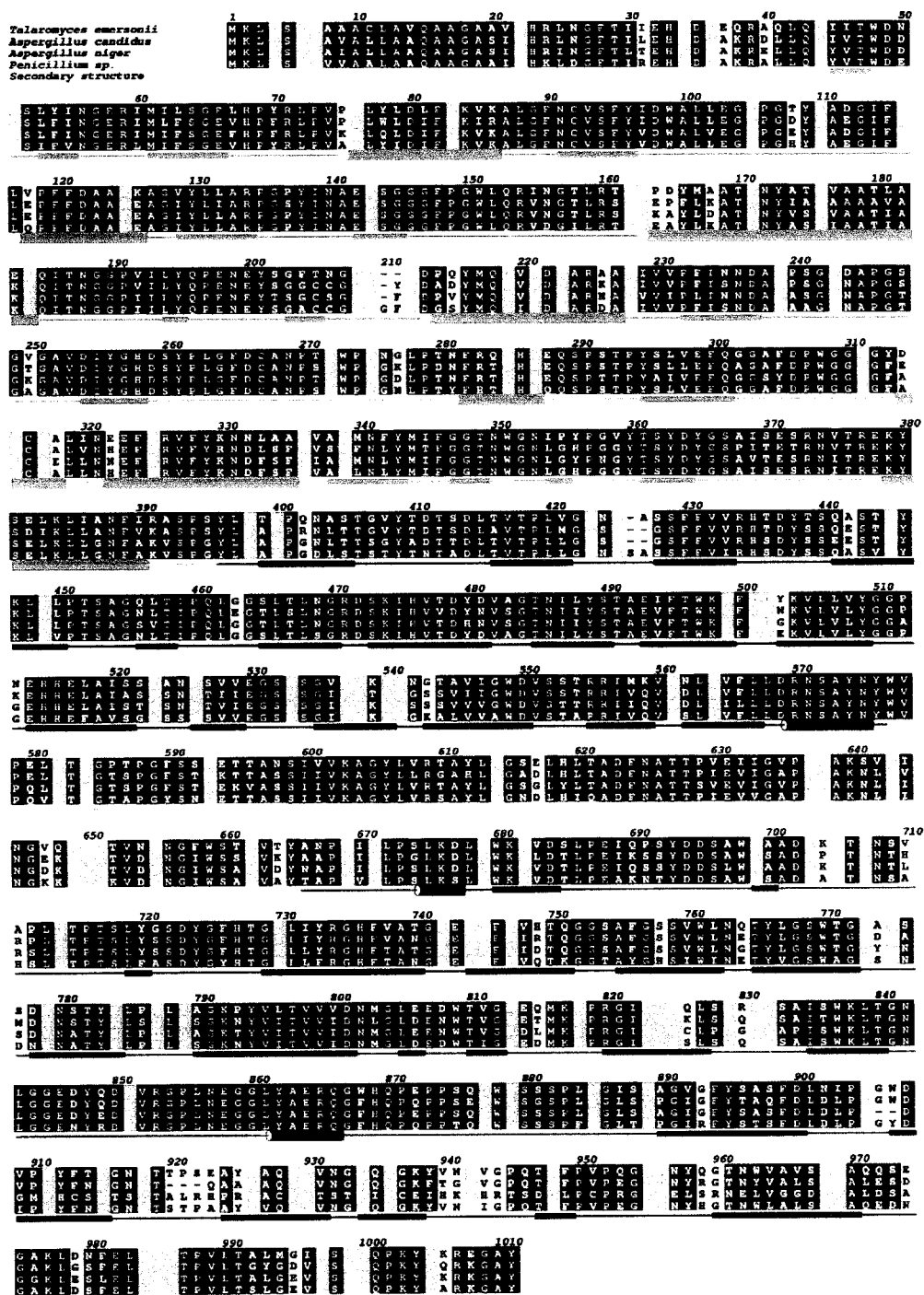
smaller than the second and consists of an  $\alpha$ -helix at its N terminus followed by eight antiparallel  $\beta$ -strands arranged as a Greek key  $\beta$ -sandwich. As with the second domain, a DALI search against known structures revealed that despite possessing a commonly observed architecture (the  $\beta$ -sandwich) this domain has a previously unobserved topology. Here, the terms architecture and topology are used in the sense defined by the CATH classification.<sup>23</sup>

On exiting the third domain, a short stretch of the polypeptide chain (residues Thr666-Pro687) passes through domain 5, forming a short  $\beta$ -strand prior to entering domain 4 (Figure 1). The latter comprises amino acid residues Glu688-Leu861 and is composed of eight  $\beta$ -strands in a  $\beta$ -sandwich which is best described as a class II right-handed jelly roll.<sup>24</sup> The jelly roll sandwich is somewhat unusual in that it is composed of one five-stranded sheet and one three-stranded sheet rather than the more regular two four-stranded sheets. This is a consequence of one of the  $\beta$ -strands of the jelly roll hairpin switching from one sheet to the other. Edge strand switching is a common phenomenon observed, for example, in variants of the classical immunoglobulin fold but we are unaware of a similar example in class II jelly rolls described to date. A large loop between  $\beta$ -strands 6 and 7 of the jelly roll protrudes over the top of the TIM barrel to participate in the formation of the active site, as described below.

The fifth domain is based on a class I jelly roll and consists of a total of eight strands divided into five and three-stranded  $\beta$ -sheets. The first strand of a conventional jelly roll is missing explaining why one of the sheets possesses only three strands. The other sheet includes an additional strand formed by



**Figure 1.** Psp- $\beta$ -gal is a 120 kDa monomer composed of five distinct structural domains. The overall structure is built around the first, TIM barrel, domain. Domain 2 is an all  $\beta$ -sheet domain containing an immunoglobulin-like subdomain, domain 3 is based on a Greek-key  $\beta$ -sandwich and domains 4 and 5 are jelly rolls. (a) Stereo view of the Psp- $\beta$ -gal  $C_{\alpha}$  trace. (b) Ribbon representation of the secondary structure elements of Psp- $\beta$ -gal, defined by the program PROCHECK.<sup>47</sup> Domains 1–5 are colored in cyan, red, yellow, green and magenta, respectively. The long linker peptide connecting domains 3 and 4 is depicted in blue. The Figures were drawn using the programs Molscript,<sup>48</sup> Bobscript<sup>49</sup> and Raster3D.<sup>50</sup>



**Figure 2.** Primary structure alignment of the  $\beta$ -galactosidases from *T. emersonii* (1008 aa), *A. candidus* (1005 aa), *A. niger* (1006 aa) and *Penicillium* sp. (1011 aa). The amino acid similarity between primary structures as calculated by the program ALSCRIPT<sup>51</sup> is shown in three different shades of blue. The darker color represents higher sequence similarity. Secondary structure elements of the Psp- $\beta$ -gal domains 3–6 are depicted in cyan, red, yellow, green and magenta, respectively. The linker peptide connecting domains 3 and 4 is in blue.

part of the connecting peptide which runs between domains 3 and 4, as described above (Figure 2).

### Comparison with two known $\beta$ -galactosidase structures

Like Psp- $\beta$ -gal, the structure of Ec- $\beta$ -gal can be divided into five domains. However, with the exception of the catalytic domain, there is little or no structural similarity between Ec- $\beta$ -gal and Psp- $\beta$ -gal in terms of the individual domain folds. Only the catalytic domain, based on a TIM barrel, is genuinely comparable between the two enzymes and even so, its relative position both in terms of sequence and domain orientation is very different in the two molecules. For example, in Ec- $\beta$ -gal the catalytic domain is the third domain in the sequence whilst in Psp- $\beta$ -gal it is the first.

The overall structure of both enzymes is built around the TIM barrel with the remaining four domains being based on different variations of  $\beta$ -sandwiches. One of the most notable differences between the two enzymes is the relative spatial disposition of these four domains with respect to the TIM barrel, which occupies a central position in both structures. Despite this basic difference, a common feature in both molecules is the fact that one side of the TIM barrel makes no contact with the remaining domains of the same subunit. This is the side of the barrel which shows distortions to the  $\beta$ -strands and which lies between the active-site residues (Glu200 and Glu299 in Psp- $\beta$ ) which emanate from the C termini of barrel strands 4 and 7, respectively. In Ec- $\beta$ -gal this side of the barrel is part of one of the inter-subunit interfaces but in Psp- $\beta$ -gal which is monomeric, lies exposed to solvent.

Unlike both Psp- $\beta$ -gal and Ec- $\beta$ -gal, the recombinant A4- $\beta$ -gal structure is trimeric. Each A4- $\beta$ -gal monomer is composed of only three domains, a TIM barrel, an  $\alpha/\beta$  fold and a  $\beta$  fold domain.<sup>8</sup> The TIM barrel is more regular than that seen in either Psp- $\beta$ -gal or Ec- $\beta$ -gal. Neither of the latter two domains bare any resemblance to the non-catalytic domains observed in Psp- $\beta$ -gal.

As a result of the structural differences between the crystallographic structures of *Penicillium* sp., *E. coli* and *T. thermophilus*  $\beta$ -galactosidases, a reasonable superposition of the monomers was not possible. Nevertheless, a superposition of TIM barrel domains of the three  $\beta$ -galactosidases could be performed (Figure 3). An rms deviation of 2.8 Å and 3.3 Å was calculated for 236 and 291 C $\alpha$  atoms pairs for, respectively, Psp- $\beta$ -gal and Ec- $\beta$ -gal, and for Psp- $\beta$ -gal and A4- $\beta$ -gal.

### Galactose-binding site

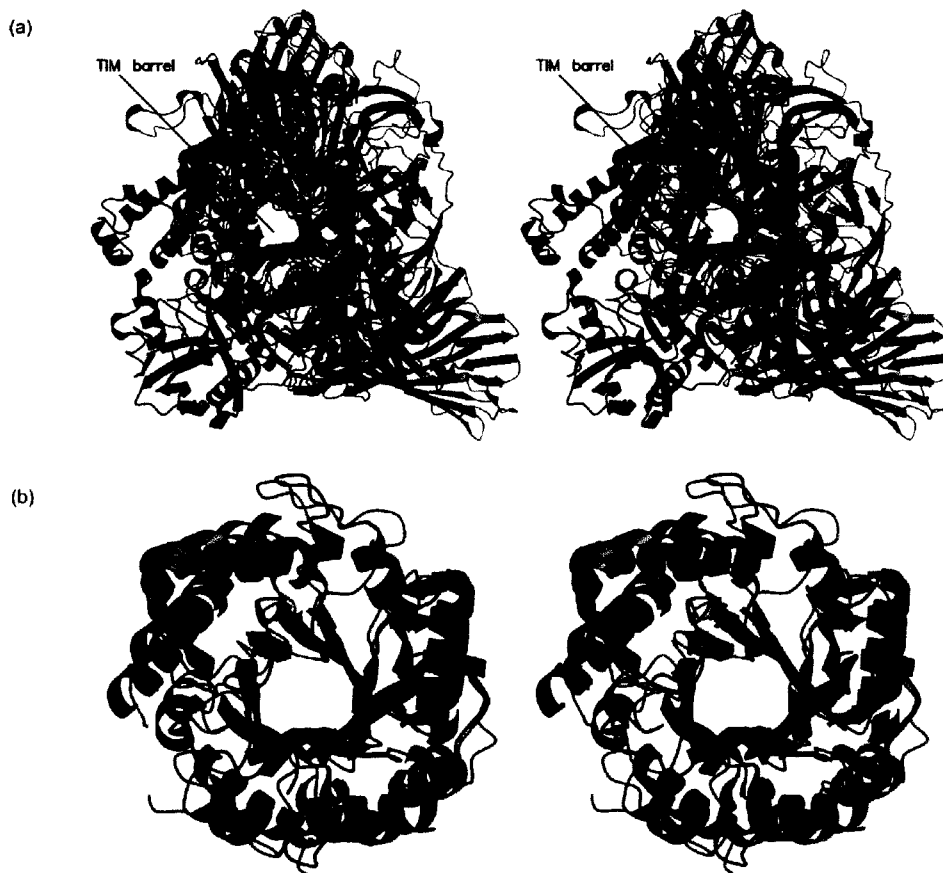
The structural analysis of the galactose-binding site was based on the comparison of the crystallographic models of the native enzyme and its complex with galactose. A single galactose molecule is bound to the TIM barrel domain of Psp- $\beta$ -gal in the chair conformation with its O1 in

the  $\beta$ -anomer configuration (Figure 4). The electron density around the sugar molecule, depicted in Figure 4, clearly indicates its presence in the catalytic site. Amino acid sequence comparison between  $\beta$ -galactosidases from fungi (Figure 2) shows that all the nine residues involved in galactose binding are well conserved among the different species. Even A4- $\beta$ -gal, whose TIM barrel catalytic domain is markedly different from that seen in Psp- $\beta$ -gal, presents a canonical heptapeptide, Asp-Ser-Tyr-Pro-Leu-Gly-Phe forming part of the sixth strand of the barrel. This is conserved in both sequence and conformation in Psp- $\beta$ -gal (residues 259–265). Its importance appears to lie in the adequate orientation of the Asp and Tyr residues for galactose binding (Figure 5). Ec- $\beta$ -gal does not have this heptapeptide sequence but has a tyrosine (Tyr503) in place of the serine whose orientation is inverted and points towards the galactose ligand substituting the tyrosine of the heptapeptide. It is interesting to note that Psp- $\beta$ -gal and A4- $\beta$ -gal belong to distinct families (35 and 42, respectively) within the GH-A clan, but nevertheless conserve this heptapeptide sequence despite showing an overall sequence identity of only 17% within the TIM barrel domains.

The enzymatic hydrolysis of sugar molecules catalyzed by glycoside hydrolases takes place *via* general acid catalysis that requires two critical chemical groups, a proton donor and a nucleophile/base,<sup>25</sup> respectively, Glu200 and Glu299 in the case of Psp- $\beta$ -gal. These residues reside on the fourth and seventh  $\beta$ -strand of the TIM barrel and the average distance between the four pairs of side-chain oxygen atoms is 4.9 Å, which is within the appropriate range for retaining enzymes<sup>26</sup> as anticipated for a GHF-35 member. Furthermore, this is in agreement with the recent experimental determination of the stereochemistry of the reaction catalyzed by Psp- $\beta$ -gal using <sup>1</sup>H NMR spectroscopy.<sup>27</sup>

In Psp- $\beta$ -gal the proton donor, Glu200 is oriented correctly with respect to the substrate by forming a hydrogen bond with the side-chain nitrogen of Trp809, which comes from the tip of the large finger-like loop between strands 6 and 7 of the fourth domain jelly roll. This large loop protrudes from the jelly roll and enters the top of the TIM barrel domain from above its third  $\beta$ -strand. A loop in this position would be impossible in the structure of Ec- $\beta$ -gal due to the conformation of the connection between  $\beta$ -strand 3 and  $\alpha$ -helix 3 which occludes this region of space. However, in Ec- $\beta$ -gal this connection itself provides His418 which, by interacting with the proton donor Glu461, appears to play a role analogous to Trp809. Curiously, in Psp- $\beta$ -gal, the  $\beta_3\alpha_3$  connection is actually larger but assumes a completely different conformation, forming a platform on which the finger-like protrusion from the jelly roll sits.

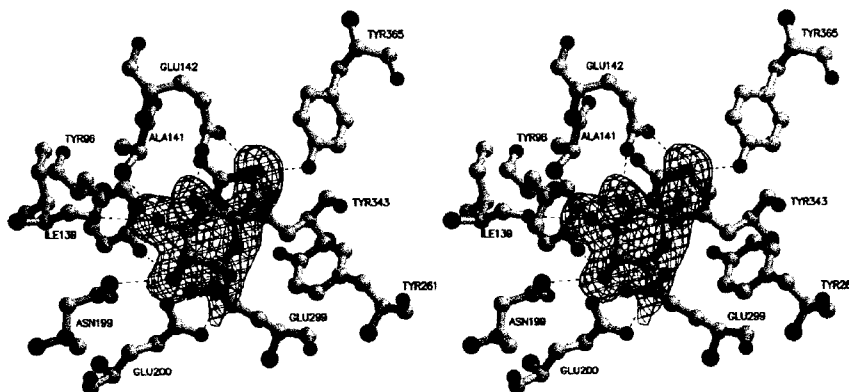
This region of the TIM barrel is important for subunit-subunit contacts in the Ec- $\beta$ -gal tetramer. Specifically, a loop from domain 2 of a 2-fold related



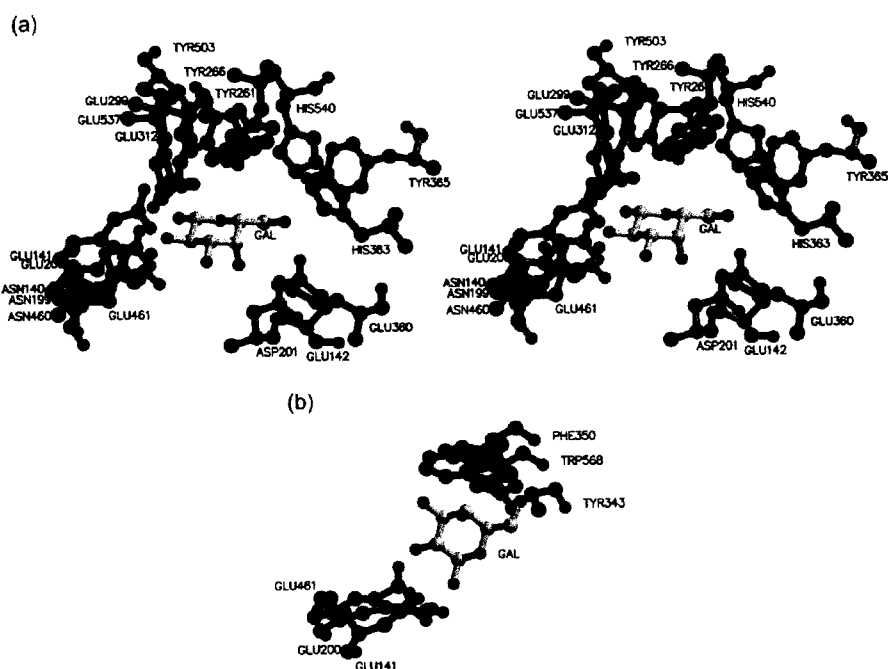
**Figure 3.** (a) Stereo view of the structural superposition of Psp- $\beta$ -gal (in cyan), Ec- $\beta$ -gal (in orange) and A4- $\beta$ -gal (in brown). (b) Stereo view of the superposition of their respective TIM barrel domains.

subunit in Ec- $\beta$ -gal enters the top of the catalytic domain and it has been suggested that this is essential for composing part of the active site.<sup>16</sup> This would justify the need for quaternary structural integrity in Ec- $\beta$ -gal for catalytic activity. In Psp- $\beta$ -gal these interactions are prevented by a large insertion

after the fourth strand of the barrel with respect to Ec- $\beta$ -gal which would contain such subunit-subunit contacts. In the case of Psp- $\beta$ -gal therefore, the entirety of the catalytic machinery would appear to be provided by a single subunit. Of the sequences aligned in Figure 2, all contain the features



**Figure 4.** Stereo representation of the galactose-protein interactions in the catalytic site of Psp- $\beta$ -gal. The galactose molecule is superimposed with an omit electron density map ( $mF_{obs} - DF_{calc}$ ,  $\varphi_{calc}$ ) contoured at 3 sigma.



**Figure 5.** (a) Stereo view of the superposition of the catalytic sites in Psp- $\beta$ -gal (in cyan), Ec- $\beta$ -gal (in orange) and A4- $\beta$ -gal (in brown). (b) Detailed view of the active site showing the orientation of the aromatic rings of Tyr343 (Psp- $\beta$ -gal), Phe350 (A4- $\beta$ -gal) and Trp568 (Ec- $\beta$ -gal) with respect to the bound galactose. In both pictures only the galactose from the Psp- $\beta$ -gal complex is shown.

described above and are therefore compatible with being catalytically active as monomers.

#### Carbohydrate moieties

Even though an inspection of the Protein Data Bank (PDB) indicates that as many as 70% of the deposited proteins have potential *N*-glycosylation sites (Asn-X-Ser/Thr, where X is not proline), crystallographic structures of the glycosylated proteins are not very common. Crystallographic studies of oligosaccharides and glycan components of glycoproteins are known to be difficult.<sup>28</sup> Crystals of oligosaccharides are hard to obtain, whereas crystallographic disorder frequently leads to a lack of identifiable electron density for the glycan components of glycoproteins hampering their structural characterization. As a result, relatively few oligosaccharide structures have been determined to date, hindering the process of the structural characterization of the general conformational rules of glycosidic linkages. Because of the high degree of glycan mobility or disorder, about 50% of glycoprotein structures contain only up to three resolved sugar residues. There are very few glycoproteins in which seven or more residues of a single glycan have been structurally resolved.<sup>29</sup>

Seven *N*-glycosylation sites have been localized in the electron density map of Psp- $\beta$ -gal, and three of them have 5, 7 and 9 monosaccharide residues each. Several oligosaccharides are wrapped around

domains 3, 4 and 5 and a number of hydrogen bonds between amino acid residues at the protein surface and the carbohydrate moieties are observed. The electron density map for a representative *N*-linked carbohydrate chain is shown in Figure 6.

Search for potential *N*-glycosylation sites in  $\beta$ -galactosidases used in the primary structure comparison (Figure 2), indicates that all of them possess several NXS/T motifs. However, the only two  $\beta$ -galactosidase structures available in the PDB are from bacterial sources and, hence, are non-glycosylated. Therefore, this is the first structural report of a glycosylated  $\beta$ -galactosidase with several long oligosaccharides attached to it.

The conformation of the linkage between an *N*-glycan and corresponding asparagine side-chain has been well characterized by NMR studies of glycopeptides in solution.<sup>30,31</sup> These studies demonstrate that the asparagine residues linkage is relatively rigid and planar, with a tendency to extend the first glycan residue away from the peptide backbone and into the solvent. In Psp- $\beta$ -gal five different glycosidic linkages (Man $\beta$ 1-4GlcNAc, GlcNAc $\beta$ 1-4GlcNAc, Man $\alpha$ 1-3Man, Man $\alpha$ 1-2Man, Man $\alpha$ 1-6Man) were identified. Their topology is shown in detail in Figure 7.

Petrescu and co-workers composed a database containing crystallographic information about glycans on the basis of the 3D structures of oligosaccharides covalently attached to proteins deposited in the PDB.<sup>29,9</sup> Using simple statistical



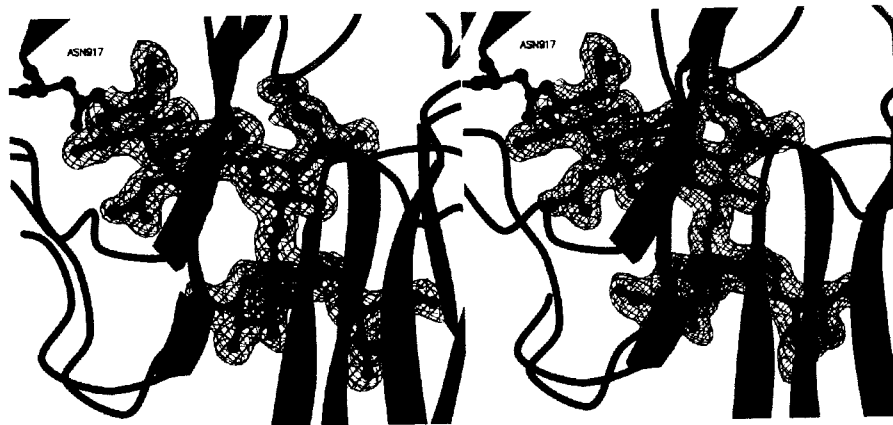


Figure 6. Stereo view of the electron density map around a representative oligosaccharide chain.

analysis, they were able to identify distinct conformers for each linkage type, providing average structures for glycosidic linkages. For example, the glycosidic linkages Man $\beta$ 1-4GlcNAc, GlcNAc $\beta$ 1-4GlcNAc and Man $\alpha$ 1-3Man are found to have a single conformer in Petrescu's classification, while Man $\alpha$ 1-2Man and Man $\alpha$ 1-6Man have two and three distinct conformers, respectively (see Table 1 for details).

A comparison between Psp- $\beta$ -gal oligosaccharides and Petrescu's database indicates that most of the glycosidic linkages in  $\beta$ -galactosidase are commonly found in other proteins. However, probably due to the large number of oligosaccharide chains visible in the Psp- $\beta$ -gal electron density map and the limited size of the original database described by Petrescu, a number of novel conformers were observed and are highlighted in Table 1.

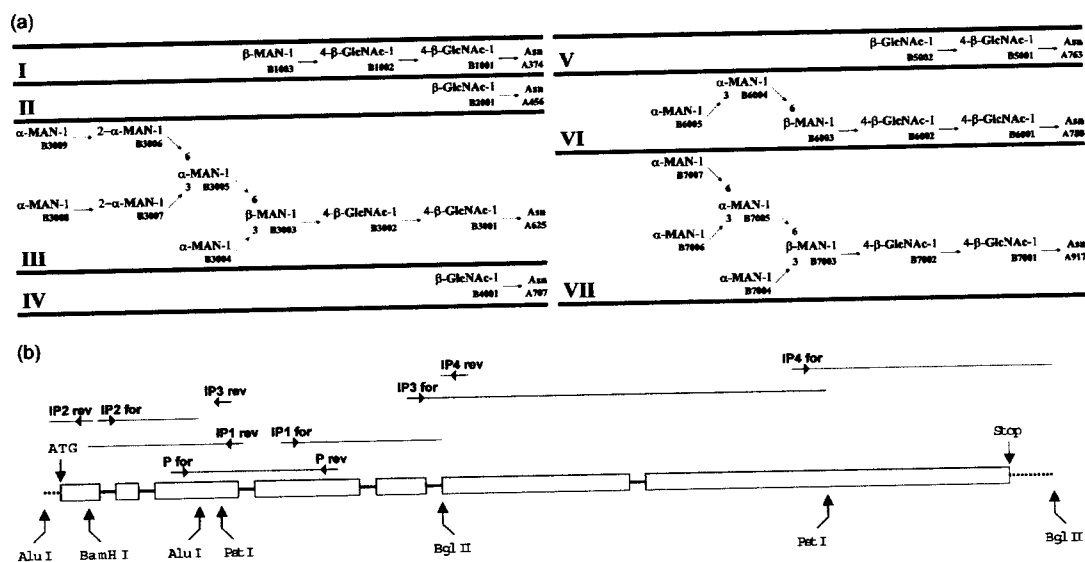


Figure 7. (a) Schematic representations of the seven N-linked oligosaccharides found in the Psp- $\beta$ -gal crystallographic structure. Five different glycosidic linkages (Man $\beta$ 1-4GlcNAc, GlcNAc $\beta$ 1-4GlcNAc, Man $\alpha$ 1-3Man, Man $\alpha$ 1-2Man and Man $\alpha$ 1-6Man) were identified on the basis of electron density maps and previous biochemical studies. (b) The  $\beta$ -galactosidase gene. Exons are represented by open boxes connected by thick lines (introns). The hatched lines at both ends of the gene represent non-coding sequences. The beginning (ATG) and end (Stop) of the protein coding sequence is marked by vertical arrows, as are the restriction sites that were used for circularization of the gene fragments in the course of the inverse PCR. The primers employed for the PCR are shown as horizontal arrows. The amplified sequences obtained with the respective primers are represented by thin lines. P for and P rev is the primer pair used for the initial, conventional PCR. IP1 for and IP1 rev is the primer pair used for the amplification of the BamHI/BglII fragment. IP2 for and IP2 rev is the primer pair used for the amplification of the PstI fragment. IP3 for and IP3 rev is the primer pair used for the amplification of the AluI fragment. IP4 for and IP4 rev is the primer pair used for the amplification of the BglII fragment. Further explanations are given in the text.

**Table 1.** Psp- $\beta$ -gal glycosidic linkages and Petrescu's conformer classification

Glycosidic linkage	Average linkage torsion angles for distinct conformers Petrescu's classification <sup>a</sup>			Residues	Linkage torsion angles for distinct conformers Psp- $\beta$ -gal		
	$\phi$	$\psi$	$\Omega$		$\phi$	$\psi$	$\omega$
GlcNAc $\beta$ 1-4GlcNAc	$-73.7 \pm 8.4$	$116.8 \pm 15.6$	-	1001/1002	$-82.5$	$94.2$	-
				3001/3002	$-76.5$	$118.4$	-
				5001/5002	$-61.1$	$135.4$	-
				6001/6002	$-70.2$	$134.8$	-
				7001/7002	$-80.5$	$110.0$	-
Man $\beta$ 1-4GlcNAc	$-88.0 \pm 10.8$	$107.9 \pm 20.3$	-	1002/1003	$-83.4$	$136.3$	-
				3002/3003	$-68.6$	$134.2$	-
				6002/6003	$-95.2$	$97.7$	-
				7002/7003	$-61.2$	$132.2^b$	-
				3006/3009	$84.2$	$-100.2$	-
Man $\alpha$ 1-2Man	$71.9 \pm 13.1$	$-104.4 \pm 15.4$	-	3007/3008	$65.9$	$-125.9$	-
				3003/3004	$-109.8$	$-104.9^b$	-
Man $\alpha$ 1-3Man	$72.5 \pm 11.0$	$-112.3 \pm 22.5$	-	3005/3007	$67.3$	$-131.6$	-
				6004/6005	$90.6$	$-74.4$	-
				7003/7004	$72.9$	$-114.5$	-
				7005/7006	$80.7$	$-93.2$	-
				3003/3005	$70.0$	$100.3$	$39.7^b$
Man $\alpha$ 1-6Man	$65.4 \pm 9.0$	$182.6 \pm 5.1$	$66.4 \pm 10.2$	7005/7007	$51.9$	$171.2$	$53.1$
				7005/3006	$80.6$	$224.5$	$178.9^b$
				6003/6004	$66.7$	$190.2$	$188.3$
				7003/7005	$52.8$	$188.3$	$46.9$

<sup>a</sup>  $\phi = \text{O5-C1-O-C}(x)'$ ,  $\psi = \text{C1-O-C}(x)'\text{-C}(x-1)$  for 1-2, 1-3 and 1-4 linkages  $x=2, 3$  or 4 and  $\phi = \text{O5-C1-O-C}(6)'$ ,  $\psi = \text{C1-O-C}(6)'\text{-C}(5)$ ,  $\omega = \text{O-C6'-C5'-C4}'$  for 1-6 linkages.

<sup>b</sup> Values deviating significantly from the standard conformers.

## Materials and Methods

### Protein purification, crystallization and data collection

Extracellular  $\beta$ -galactosidase from *Penicillium* sp. was purified to homogeneity and crystallized as described.<sup>32</sup> The molecular mass estimate by SDS-PAGE yielded a value of  $120(\pm 5)$  kDa. This value coincides with the mass calculated from the elution profile on a Sephacryl S-300 column assuming that the enzyme exists as a monomer in solution.

Preliminary X-ray diffraction studies revealed that  $\beta$ -galactosidase from *Penicillium* sp. crystallizes in the tetragonal space group  $P4_1$  or  $P4_3$ , with the unit cell parameters  $a=b=110.96$  Å,  $c=161.05$  Å. The crystals diffracted beyond 2.0 Å resolution. Matthews coefficient<sup>33</sup> calculations suggested the presence in the asymmetric unit of either a dimer ( $V_M=2.1$  Å<sup>3</sup> Da<sup>-1</sup>, a solvent content of approximately 40%) or a monomer (a solvent content of 70%). Further crystallographic analysis confirmed that the asymmetric unit contained one enzyme molecule.

Four complete diffraction data sets (one native and three derivatives) were collected at the Protein Crystallography beamline<sup>34,35</sup> of the Brazilian National Synchrotron Light Laboratory using a MAR345 image plate detector. The crystals were immersed in different soaking solutions for different periods of time (Table 2), and mounted in rayon loops. Finally, samples were flash-cooled to 100 K in a cold nitrogen stream. The iodine and cesium derivatives (Table 2) were prepared according to the quick cryo-soaking protocol of derivatization,<sup>36-38</sup> whereas the uranium derivative was obtained by soaking a Psp- $\beta$ -gal crystal in a solution of  $\text{UO}_2\text{Ac}_2$  overnight. Diffraction images were processed and scaled with the

programs DENZO and SCALEPACK.<sup>39</sup> Data collection statistics for each data set are shown in Table 2.

### Phases calculation and model building

All attempts to solve the  $\beta$ -galactosidase structure with the molecular replacement (MR) method using two available crystal structures of bacterial  $\beta$ -galactosidases did not provide reasonable solutions. Therefore, the structure was solved by the SIRAS method, using the native and the iodine derivative data sets. The major heavy-atom positions of the iodine derivative were determined by direct methods with the programs DREAR<sup>40</sup> and SnB.<sup>41</sup> This heavy-atom substructure was then set as the input to SHARP for phase calculation. Density modification with solvent flattening was performed with the program SOLOMON. At this point, the correct space group was identified as  $P4_3$ .

A solvent flattened I-SIRAS derived electron density map and structure factor amplitudes from the native data set were used for automatic construction of the  $\beta$ -galactosidase crystal structure with the ARP/wARP program.<sup>42</sup> A hybrid model of the  $\beta$ -galactosidase was finally obtained after a few thousand cycles of iterative refinement with REFMAC<sup>43</sup> and hundreds of cycles of the automatic main-chain tracing. Preliminary inspection of this model revealed a few main-chain gaps where proline residues could be placed. The Fourier maps with the coefficients  $2mF_{\text{obs}} - DF_{\text{calc}}$  and model derived phases enabled construction of missing residues using the program O.<sup>44</sup> The resulting hybrid model was formed of a single poly-glycine chain of 971 amino acid residues and a few thousand water molecules.

In order to establish the primary structure of the enzyme from the electron density maps a complete

**Table 2.** Details of the data-collection and the refinement statistics of the Psp- $\beta$ -galactosidase crystal structure solution

	Nat- $\beta$ -gal	l- $\beta$ -gal <sup>a</sup>	Cs- $\beta$ -gal <sup>a</sup>	U- $\beta$ -gal
<i>Crystal preparation</i>				
Cryoprotectant solution	Mother liquor, 15% ethylene glycol	Mother liquor, 0.5 M NaI, 15% ethylene glycol	Mother liquor, 0.33 M CsCl, 15% ethylene glycol	Mother liquor, 50 mM UO <sub>2</sub> Ac <sub>2</sub> , 15% ethylene glycol
Soaking time	30 seconds	330 seconds	60 seconds	12 hours
<i>Data collection</i>				
Wavelength (Å)	1.54	1.54	1.54	1.54
Space group	<i>P</i> 4 <sub>3</sub>	<i>P</i> 4 <sub>3</sub>	<i>P</i> 4 <sub>3</sub>	<i>P</i> 4 <sub>3</sub>
Unit cell parameters (Å)	<i>a</i> = <i>b</i> = 110.96, <i>c</i> = 161.05	<i>a</i> = <i>b</i> = 110.62, <i>c</i> = 159.89	<i>a</i> = <i>b</i> = 110.83, <i>c</i> = 160.92	<i>a</i> = <i>b</i> = 111.01, <i>c</i> = 161.33
Resolution (Å)	22.3–1.90 (1.93–1.90)	23.4–2.10 (2.15–2.10)	27.0–2.10 (2.15–2.10)	22.2–2.40 (2.46–2.40)
No. of reflections	639,065	944,162	614,561	608,673
No. of unique reflections <sup>b</sup>	155,025	224,036	223,388	148,232
<i>I</i> / <i>σ</i> ( <i>I</i> )	13.9 (3.2)	12.5 (3.1)	8.4 (2.4)	10.0 (2.8)
Multiplicity	4.1 (3.9)	4.2 (4.2)	2.8 (2.7)	4.1 (4.0)
Completeness	99.9 (100.0)	99.9 (100.0)	99.6 (99.6)	98.6 (95.7)
<i>R</i> <sub>merge</sub> <sup>c</sup>	10.5 (40.0)	10.7 (48.1)	12.3 (47.0)	13.0 (53.1)
Data collected (deg.)	105	240	134	215
<i>Structure refinement</i>		Native structure	Psp- $\beta$ -gal:galactose complex	
<i>R</i> -factor (%)		16.5	17.1	
<i>R</i> -free (%)		18.5	18.5	
No. of protein atoms		7484	7484	
No. of water molecules		1252	952	
No. of sugar atoms		344	356	
No. of phosphate ions		4	1	
No. of sodium ions		3	1	
No. of iodine ions <sup>d</sup>		0	52	
No. of ethyl glycol molecules		9	11	
rmsd bond distances (Å)		0.005	0.005	
rmsd bond angles (°)		1.5	1.3	
Average <i>B</i> -factors (Å <sup>2</sup> )		23.0	22.6	

Statistical values for the highest resolution shells are shown in parentheses.

<sup>a</sup> Data set collected from a crystal of the protein complexed with galactose.

<sup>b</sup> Multiplicities of the derivative data sets were calculated with the Friedel-related reflections treated separately. Multiplicity of the native data set was calculated with the Friedel-pairs treated as equivalent.

<sup>c</sup>  $R_{\text{merge}} = \sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \sum_{hkl} I_{hkl}$ .

<sup>d</sup> Iodine sites were not fully occupied.

MIRAS phases calculation with the four data sets and a total of 74 heavy atoms with varying occupancies (0.8–0.1) was carried out using the SHARP program. This process took approximately seven days on a Pentium IV 1.9 GHz. More than 155,000 reflections were phased with an overall figure of merit of 0.61 in the range of 27.0 Å–2.40 Å. The MIRAS phases with the initial envelope derived from the poly-glycine model were used to perform density modification with SOLOMON. The resulting map was of high quality and enabled the identification of a putative primary structure composed of 971 amino acid residues.

#### Cloning and sequence analysis of the $\beta$ -galactosidase gene

The complete protein coding sequence of the  $\beta$ -galactosidase gene from *Penicillium* sp. was obtained by inverse PCR using thermostable DNA Taq polymerase. Amplified fragments were directly sequenced without prior subcloning in order to guarantee the absence of artificial Taq-derived alterations in the obtained sequence. Initially, different sets of degenerate primers deduced from peptide sequences obtained from the experimental *ab initio* electron density map were employed to generate gene specific fragments *via* conventional PCR

from genomic *Penicillium* sp. DNA. Using the primer pair 5'-GTCGATTGGGCTTTGTGGAGG-3' and 5'-GGTGCRTTRTGCTCCTGCTGCCA-3' deduced from the peptide sequences VDWALLEG and WAAGHNAP, a 513 bp PCR fragment was obtained. Its sequence served as the basis for subsequent iterative extension of the sequence information by inverse PCR, which was carried out essentially as described.<sup>45</sup>

The complete sequence was obtained in four separate steps. First, a 994 bp fragment was amplified after digestion of genomic DNA with BamHI and BglII and circularization with T4 DNA ligase, using the respective primers. Next, a 531 bp fragment was obtained in a similar manner after restriction with AluI and circularization, which covered the start codon of the  $\beta$ -galactosidase gene as deduced from the encoded peptide sequence. Finally, a fragment of 1474 bp, extending the 3' end of the sequence obtained so far, was likewise generated after PstI digestion. The 3' end of the  $\beta$ -galactosidase gene was finally obtained after BglII digestion, circularization and amplification. The positions of the relevant restriction and priming sites in the  $\beta$ -galactosidase gene are shown in Figure 7. The  $\beta$ -galactosidase gene derived primary structure (NCBI accession number CAF32457) has a total of 1011 amino acid residues, with a 40 amino acid long signal peptide. Comparison of the

amino acid sequence with available protein sequences using the BLAST program revealed that the sequence of Psp- $\beta$ -gal is similar to those of  $\beta$ -galactosidases from *A. candidus* (1005 aa), *A. niger* (1006 aa) and *T. emersonii* (1008 aa) (Figure 2). None of these  $\beta$ -galactosidases have had their 3D structure elucidated so far.

#### Structure refinement

After correct identification of the amino acid side-chains using the gene translated primary structure, the 3D model underwent many cycles of refinement with and without simulated annealing, iterations of energy minimization, water molecule insertion and oligosaccharide addition using the CNS package.<sup>46</sup> The final native model includes one  $\beta$ -galactosidase polypeptide chain of 971 amino acid residues, seven oligosaccharides comprising 16 MAN and 12 NAG, three sodium and four phosphate ions, nine ethylene glycol molecules and 1252 water molecules.

The analysis of the electron density maps for the iodine and cesium derivative data sets revealed clear electron density for the galactose bound to the active site of the enzyme, whereas in the native crystal and uranium derivative, galactose was not observed. The native  $\beta$ -galactosidase present in the crystals used in the iodine and cesium derivatizations was obtained from a separate purification procedure, leading to the conclusion that in this case, galactose was fortuitously bound to the enzyme in the fungal extract and was not washed out in the process of purification.

The structure of the galactose-containing complex was refined against the iodine derivative data set in a manner similar to that described for the native structure. The final refinement statistics for both structures are given in Table 2.

#### Protein Data Bank accession codes

Atomic coordinates and structure factors have been deposited in the Protein Data Bank for release upon publication. The PDB entries corresponding to the structures of native Psp- $\beta$ -gal and Psp- $\beta$ -gal:galactose complex are 1TG7 and 1XC6, respectively.

#### Acknowledgements

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via grants #99/03387-4, 02/14208-8 and 01/07014-0, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via grant 300220/96-0 (to I.P.) grants of Russian Academy of Sciences through the Basic Research program "Physical-chemistry and biology", and Russian Foundation Basic Research #03-04-48756 (to K.N.N. and A.M.G.).

#### References

1. Usui, T., Kuboto, S. & Ohi, H. (1983). A convenient synthesis of  $\beta$ -D-galactosyl disaccharide derivatives using the  $\beta$ -D-galactosidase from *Bacillus circulans*. *Carbohydr. Res.* **244**, 315–323.
2. Mori, T., Fujita, S. & Okahata, Y. (1997). Transglycosylation in a two-phase aqueous-organic system with catalysis by a lipid-coated  $\beta$ -D-galactosidase. *Carbohydr. Res.* **298**, 65–73.
3. Fujimoto, H., Miyasato, M., Ito, Y., Sasaki, T. & Ajisaka, K. (1998). Purification and properties of recombinant  $\beta$ -galactosidase from *Bacillus circulans*. *Glycoconj. J.* **15**, 155–160.
4. Hedbys, L., Johansson, E., Mosbach, K., Larsson, P. O., Gunnarsson, A., Svensson, S. & Lonn, H. (1989). Synthesis of Gal beta 1-3GlcNAc and Gal beta 1-3GlcNAc beta-SEt by an enzymatic method comprising the sequential use of beta-galactosidases from bovine testes and *Escherichia coli*. *Glycoconj. J.* **6**, 161–168.
5. Vetere, A. & Paoletti, S. (1996). High-yield synthesis of N-acetylglucosamine by regioselective transglycosylation. *Biochem. Biophys. Res. Commun.* **219**, 6–13.
6. Ajisaka, K., Fujimoto, H. & Isomura, M. (1994). Regioselective transglycosylation in the synthesis of oligosaccharides: comparison of  $\beta$ -galactosidase and sialidases of various origins. *Carbohydr. Res.* **259**, 103–115.
7. Cantacuzene, D. & Attal, S. (1991). Enzymic synthesis of galactopyranosyl-L-serine derivatives using galactosidases. *Carbohydr. Res.* **211**, 327–331.
8. Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H. & Wakagi, T. (2002). Trimeric crystal structure of the glycoside hydrolase family 42 beta-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J. Mol. Biol.* **322**, 79–91.
9. Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
10. Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788.
11. Henrissat, B. & Bairoch, A. (1996). Updating the sequence based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696.
12. Henrissat, B. & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644.
13. Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859.
14. Jenkins, J., Leggio, L. L., Harris, G. & Pickersgill, R. (1995).  $\beta$ -Glucosidase,  $\beta$ -galactosidase, family A cellulases, family F xylanases and two barley glucanases form a superfamily of enzymes with 8-fold  $\beta/\alpha$  architecture and with two conserved glutamates near the carboxy-terminal ends of  $\beta$ -strands four and seven. *FEBS Letters*, **362**, 281–285.
15. Appel, R. D., Bairoch, A. & Hochstrasser, D. F. (1994). A new generation of information retrieval tools for biologists: the example of the ExpASY WWW server. *Trends Biochem. Sci.* **19**, 258–260.
16. Jacobson, R. H., Zhang, X.-J., DuBose, R. F. & Matthews, B. W. (1994). Three-dimensional structure of  $\beta$ -galactosidase from *E. coli*. *Nature*, **369**, 761–766.
17. Juers, D. H., Huber, R. E. & Matthews, B. W. (1999). Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between  $\beta$ -galactosidase and other glycohydrolases. *Protein Sci.* **8**, 122–136.
18. Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X. J., Huber, R. E., Tronrud, D. E. & Matthews, B. W. (2000). High resolution refinement of  $\beta$ -galactosidase in a

- new crystal form reveals multiple metal-binding sites and provides a structural basis for  $\alpha$ -complementation. *Protein Sci.* **9**, 1685–1699.
19. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
  20. La Fortelle, E. & de Briconne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
  21. Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallog. sect. D*, **52**, 30–42.
  22. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595–602.
  23. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure*, **5**(8), 1093–1108.
  24. Stirk, H. J., Woolfson, D. N., Hutchinson, E. G. & Thornton, J. M. (1992). Depicting topology and handedness in jellyroll structures. *FEBS Letters*, **308**(1), 1–3.
  25. Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* **90**, 1171–1202.
  26. McCarter, J. D. & Withers, S. G. (1994). Mechanisms of enzymatic glycosidase hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885–892.
  27. Zinin, A. I., Eneyskaya, E. V., Shabalin, K. A., Kulminskaya, A. A., Shishlyannikov, S. M. & Neustroev, K. N. (2002). 1-O-Acetyl- $\beta$ -D-galactopyranose: a novel substrate for the transglycosylation reaction catalyzed by the  $\beta$ -galactosidase from *Penicillium* sp. *Carbohydr. Res.* **337**, 635–642.
  28. Wormald, M. R. & Raymod, A. D. (1999). Glycoproteins: glycan presentation and protein-fold stability. *Structure*, **7**, 155–160.
  29. Petrescu, A. J., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. (1999). A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology*, **9**, 343–352.
  30. Davis, J. T., Hirani, S., Barlett, C. & Reid, B. R. (1994).  $^1\text{H}$  NMR studies on an Ans-linked glycopeptide. GlcNAc-1 C2–N2 bond is rigid in  $\text{H}_2\text{O}$ . *J. Biol. Chem.* **269**, 3331–3338.
  31. Imberty, A. & Perez, S. (1995). Stereochemistry of the N-glycosylation sites in glycoproteins. *Protein Eng.* **8**, 699–709.
  32. Neustroev, K. N., de Sousa, E. A., Golubev, A. M., Brandão Neto, J. R., Eneyskaya, E. V., Kulminskaya, A. A. & Polikarpov, I. (2000). Purification, crystallization and preliminary diffraction study of beta-galactosidase from *Penicillium* sp. *Acta Crystallog. sect. D*, **56**, 1508–1509.
  33. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
  34. Polikarpov, I., Oliva, G., Castellano, E. E., Garratt, R., Arruda, P., Leite, A. & Craievich, A. (1997). The protein crystallography beamline at LNLS, the Brazilian National Synchrotron Light Source. *Nucl. Instrum. Methods A*, **405**, 159–164.
  35. Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. & Craievich, A. (1997). Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *J. Synchrotron Rad.* **5**, 72–76.
  36. Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallog. sect. D*, **56**, 232–237.
  37. Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta Crystallog. sect. D*, **57**, 996–1002.
  38. Nagem, R. A. P., Polikarpov, I. & Dauter, Z. (2003). Phasing on rapidly soaked ions. *Methods Enzymol.* **374**, 120–137.
  39. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
  40. Blessing, R. H. & Smith, G. D. (1999). Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. *J. Appl. Crystallog.* **32**, 664–670.
  41. Weeks, C. M. & Miller, R. (1999). The design and implementation of SnB version 2.0. *J. Appl. Crystallog.* **32**, 120–124.
  42. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
  43. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
  44. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
  45. Arand, M., Hemmer, H., Dürk, H., Baratti, J., Archelas, A., Furstoss, R. & Oesch, F. (1999). Cloning and molecular characterization of a soluble epoxide hydrolase from *Aspergillus niger* that is related to mammalian microsomal epoxide hydrolase. *Biochem. J.* **344**, 273–280.
  46. Brünger, A. T., Adams, P. D. G., Clore, M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
  47. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
  48. Kraulis, P. J. (1991). Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
  49. Esnouf, R. M. (1997). An extensively modified version of molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133–138.
  50. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
  51. Barton, G. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40.

Edited by R. Huber

(Received 5 May 2004; received in revised form 7 September 2004; accepted 9 September 2004)

# CAPÍTULO 3.

## **Exo-inulinase de *Aspergillus awamori***



### 3.1. Introdução

As inulinas são carboidratos polidispersos que consistem em uma mistura de polímeros de frutose geralmente terminados por uma molécula de glicose (GFn) com ligações principalmente de  $\beta$ -2,1 frutossil-frutose. Apresentam cadeias de tamanho variado, contendo de 2 a 60 unidades, com grau de polimerização médio de 10. Além de atuarem como carboidratos de reserva para as plantas<sup>1</sup>, acredita-se que os frutanos estejam relacionados a mecanismos adaptativos em resposta ao estresse ambiental, oferecendo resistência à seca e ao congelamento através de sua contribuição na mudança do potencial osmótico da célula<sup>2</sup>.

Os frutooligossacarídeos (FOS) podem ser produzidos industrialmente por meio de extração ou pela ação de fructosiltransferases. As propriedades funcionais dos FOS são decorrentes de sua não digestibilidade pelo intestino humano e por serem consumidos preferencialmente pelas bifidobactérias, cujo crescimento é então estimulado<sup>3</sup>. Os ácidos graxos de cadeias curtas, produtos deste processo fermentativo, contribuem para a redução da pressão sanguínea, de taxas de triglicéridos, colesterol e glicemia<sup>4</sup>.

Os benefícios dos FOS incluem a redução de bactérias nocivas e de metabólicos tóxicos, prevenção da diarreia, efeito anticarcinogênico, redução na lipídemia e aumento na absorção de cálcio, magnésio e ferro no cólon<sup>5</sup>. Os efeitos adversos relatados têm sido a flatulência e a diarreia quando consumidos em quantidades superiores a 30g/dia. A inulina não hidrolisada é utilizada na indústria alimentícia, devido ao sabor suave e à alta solubilidade sob aquecimento<sup>6</sup>, e na indústria farmacêutica, em testes de função renal, sendo considerada a substância ideal para a determinação do ritmo de filtração glomerular<sup>7</sup>.

Os frutanos são, também, de grande interesse econômico para as indústrias alimentícia e farmacêutica, na forma de frutose livre produzida a partir da sua hidrólise. A frutose é um monossacarídeo obtido originalmente de plantas que contêm inulina, como a alcachofra de Jerusalém ou tupinambor (*Helianthus tuberosus*). A frutose é um substituto da sacarose em alimentos para diabéticos, pois entra no metabolismo humano sem a liberação inicial de insulina e sua absorção pelo sangue é mais demorada do que a da glicose. A insulina não é necessária no transporte da

frutose do fígado para os tecidos periféricos e para sua fosforilação e transformação em glicogênio. O efeito sinérgico da frutose com a sacarina é maior do que o da sacarose com esta última, ampliando ainda mais o seu uso em alimentos dietéticos. Em mistura com a sacarina, pode se tornar sete vezes mais doce que a sacarose, com apenas 1/7 de seu teor calórico, além de ser um excelente mascarador do sabor amargo deixado pela sacarina. Reduz em até 25% a incidência de cáries quando se substitui a sacarose completamente da alimentação humana. Adicionalmente, este monossacarídeo acelera o metabolismo do etanol e aumenta a absorção de ferro através da formação de um complexo quelante ferro-frutose.

Quando a frutose é adicionada a fontes protéicas ela reage com os grupamentos amina formando complexos aromáticos agradáveis. Por ser agente redutor, reage com aminoácidos (reação de Maillard) gerando compostos de cor amarelo dourado a marrom escuro, de odor e sabor típicos. Assim, quando usada em panificação intensifica a coloração típica destes produtos.

### 3.1.1. Inulinases

As inulinases são enzimas que degradam ligações  $\beta$ -2,1 de cadeias de fructanos *via* dois mecanismos. Exo-inulinases cortam sequencialmente ligações terminais  $\beta$ -2,1-fructopiranosídeo, gerando moléculas de frutose a partir do terminal não redutor da cadeia da inulina, enquanto que endo-inulinases hidrolisam as ligações internas na inulina gerando inulo-oligossacarídeos como produto principal da reação. Antes de nossos estudos nenhuma informação estrutural destas enzimas tinha sido reportada.

As endo-inulinases são classificadas pela formação de oligossacarídeos produto da síntese tendo como substrato a sacarose<sup>8</sup>. Diversas inulinases microbianas têm sido parcialmente purificadas e suas propriedades revisadas por Vandamme e Derycke<sup>8</sup>. Em contraste com as enzimas de plantas, as inulinases microbianas apresentam atividade sobre a sacarose, bem como sobre a inulina. Usualmente, estas inulinases são classificadas como 2,1- $\beta$ -D frutano frutohidrolase (EC3.2.1.7), mas a diferença com as invertases (EC 3.2.1.26) não é clara para enzimas microbianas.

Inulinases de fungo são frequentemente misturas de frutano-hidrolases com alta atividade e estabilidade<sup>9</sup>. As inulinases de fungo mais estudadas são as produzidas pelas espécies de *Penicillium*<sup>10,11</sup>, *Aspergillus*<sup>12, 13</sup> e *Kluyveromyces*<sup>14</sup>. Têm sido



reportadas as seqüências de nucleotídeos e aminoácidos para exo-inulinases de chicória<sup>15</sup>, levedura<sup>16</sup>, *Aspergillus awamori*<sup>13</sup> e *Geobacillus stearothermophilus*<sup>17</sup>.

### 3.1.2. Exo-inulinase de *Aspergillus awamori*

A exo-inulinase de *Aspergillus awamori* é uma enzima monomérica com um peso molecular de  $69 \pm 1$  kDa e um pI de 4.4. Esta enzima hidrolisa  $\beta$ -1,2-fructana via exoclivagem, gerando frutose. A atividade enzimática é inibida entre 40 e 50% por íons de  $Mn^{2+}$  e  $Cu^{2+}$ , mas os íons  $Hg^{2+}$  não afetam sua atividade, o que sugere que a proteína não tem grupos SH no sítio ativo.

A seqüência de aminoácidos indica que a proteína pertence à família 32 das glicosil hidrolases (GHs), e mostra similaridades com a seqüência da frutossil transferase de *Aspergillus foeticus*. A família 32 das GHs inclui também invertases, levanases e dois tipos de frutossil-transferases. Comparada com outras exo-inulinases de origem bacteriana, a exo-inulinase de *Aspergillus awamori* é mais termoestável, apresentando um grande potencial para aplicações biotecnológicas. Os parâmetros cinéticos para esta enzima foram determinados para diferentes substratos (tabela 3.1).

**Tabela 3.1.** Parâmetros cinéticos da exo-inulinase de *Aspergillus awamori* com diferentes substratos. Adaptada de Arand e colaboradores<sup>13</sup>.

Substrato	$K_m$	$V_{max}$ [ $\mu\text{mol}/(\text{min}.\text{mg})$ ]
Inulina	$0.003 \pm 0.0001$ mM	$175 \pm 5$
Levano	$2.08 \pm 0.04$ mg/ml	$80 \pm 0.8$
Staquiase	$10.1 \pm 0.1$ mM	$1.2 \pm 0.02$
Rafinose	$8 \pm 0.2$ mM	$120 \pm 4$

Os valores de  $K_m$  e  $V_{max}$  da exo-inulinase para levano mostram que a enzima é capaz de degradar especificamente ligações  $\beta(2-1)$  assim como  $\beta(2-6)$  de fruto-oligosacarídeos, o que é surpreendente dada a sua alta homologia com a inulinase de *Aspergillus niger*, para a qual não foi reportada atividade detectável de hidrolisar levano. Arand e colaboradores observaram que a exo-inulinase de *Aspergillus awamori* apresentava tanto atividade hidrolítica quanto atividade de transfrutossilacção, sendo assim uma enzima produtora de frutooligosacarídeos<sup>13</sup>.

A estrutura tridimensional da exo-inulinase de *Aspergillus awamori*, resolvida e descrita no presente trabalho, é a primeira estrutura de uma inulinase reportada. Este trabalho foi publicado na revista *Jornal Molecular Biology*<sup>18</sup> e o artigo referente esta incluído no Anexo 2. Uma vez que toda a metodologia empregada, assim como os resultados e conclusões, estão descritos no artigo, optou-se por não reproduzi-los. Será feita neste capítulo somente uma breve descrição das metodologias e os resultados obtidos. Para uma discussão mais detalhada vide Anexo 2.

## 3.2 Metodologia

### 3.2.1. Cristalização

A exo-inulinase de *Aspergillus awamori* foi cristalizada<sup>13</sup> pelo método da gota pendurada. A proteína liofilizada foi ressuspendida em água em uma concentração de 10 mg/ml. Gotas contendo 10 µl de solução da proteína e uma proporção 1:1 de 15% PEG 3350 em 20mM acetato de sódio a pH 4.2 e 0.7mM de cacodilato foram equilibradas contra 800µl da mesma solução de PEG, a temperatura ambiente. Cristais pequenos em forma de barra apareceram depois de três semanas. Estes cristais foram usados como sementes nas mesmas condições e depois de uma semana cristais maiores apareceram (figura 3.1).



**Figura 3.1.** Cristais de exo-inulinase de *Aspergillus awamori* crescidos em 15% de PEG 3350, acetato de sódio 100mM pH 4.4 depois da semeadura (*seeding*).

### 3.2.2. Coleta e processamento de dados

Foram coletados 5 conjuntos de dados, dois nativos e três derivados (dois com iodo e um com césio), no Laboratório de Cristalografia de Proteínas do Instituto de Física de São Carlos, usando um gerador de raios-X de anodo rotatório, que utiliza um cristal de grafite como monocromador para selecionar somente as linhas  $K\alpha_1$  e  $K\alpha_2$  do cobre ( $\lambda=1.5408\text{\AA}$ ). Os dados de difração foram coletados utilizando um detector *Image Plate Mar345*. Os cristais foram congelados com nitrogênio líquido, após

serem mergulhados em soluções que continham átomos pesados. Os derivados de iodo e cério foram preparados com o método do *quick cryo-soaking*<sup>19,20</sup>. As imagens de difração foram processadas com os programas DENZO/SCALEPACK<sup>21</sup> e MOSFLM/SCALA<sup>22</sup>.

### 3.3.3. Faseamento

O problema das fases foi resolvido pelo método de SIRAS, usando o conjunto de dados nativo e o derivado de iodo. A posição dos átomos de iodo foi determinada por métodos diretos com ajuda do programa SHELX<sup>23</sup>. O sinal anômalo do derivado de cério era muito baixo, não sendo possível determinar as posições dos átomos pesados neste cristal. As posições encontradas foram usadas pelo programa SHARP<sup>24</sup> para o cálculo das fases, *solvent flattening* foi feito usando o programa SOLOMON<sup>25</sup>. Usando o programa ARP/wARP<sup>26</sup> e um refinamento iterativo com REFMAC<sup>27</sup>, construiu-se um modelo inicial da estrutura da exo-inulinase de *Aspergillus awamori*. As cadeias laterais (517 resíduos) foram construídas usando o programa O<sup>28</sup>.

## 3.3 Resultados e Discussão

Estudos de difração de raios-X mostram que a exo-inulinase de *Aspergillus awamori* cristaliza-se no grupo espacial monoclinico P2<sub>1</sub> com parâmetros de célula  $a = 49.96 \text{ \AA}$ ,  $b = 93.35 \text{ \AA}$ ,  $c = 67.83 \text{ \AA}$ ,  $\alpha = 90^\circ$ ,  $\beta = 106.92^\circ$ ,  $\gamma = 90^\circ$  e difrata até  $1.87 \text{ \AA}$  de resolução. O cálculo do coeficiente de Matthews (Matthews, 1968)<sup>29</sup> sugere a presença de um monômero na unidade assimétrica e um conteúdo de solvente de aproximadamente 49.2%. Alguns dados estatísticos do processamento encontram-se na tabela 3.1

**Tabela 3.1.** Estatísticas da coleta dos dados de difração de raios-X dos cristais da exo-inulinase de *Aspergillus awamori*. Os valores para alta resolução encontram-se entre parênteses.

	<b>Cristais Nativos</b>	<b>Derivado Iodo</b>	<b>Complexo fructose</b>
<b>Solução crioprotetora</b>	solução mãe 20% glycerol 15% PEG 3350	solução mãe 0.5 M NaI 20% glycerol 15% PEG 3350	solução mãe 20% fructose 15% PEG 3350
<b>Tempo de banho</b>	-----	15 min	2 min
<b>Comprimento de onda (Å)</b>	1.42	1.54	1.54
<b>Grupo espacial</b>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
<b>Dimensões da célula (Å)</b>	64.82 Å, 82.13 Å, 136.22 Å	49.95 Å, 94.03 Å, 67.63 Å, β = 107.04°	49.97 Å, 93.67 Å, 83.36 Å, β = 106.92°
<b>Resolução (Å)</b>	19-1.89 (1.99-1.89)	17.8-1.55 (1.63-1.55)	26.6 - 1.87 (1.97-1.87)
<b>No. de observações</b>	310283	487330	834739
<b>&lt;I/(σ)&gt;</b>	8.0(3.8)	10.3(4.5)	12.8 (5.9)
<b>Multiplicidade</b>	5.5(5.0)	2.9 (2.9)	7.9 (7.4)
<b>Completeza</b>	99.8(99.6)	91.3 (85.7)	92.1(84.1)
<b>R<sub>merge</sub><sup>1</sup></b>	5.4(29.7)	4.5 (15.0)	4.2 (12.8)

$$^1R_{\text{merge}} = \sum_{\text{hkl}} |I_{\text{hkl}} - \langle I_{\text{hkl}} \rangle| / \sum_{\text{hkl}} I_{\text{hkl}}$$

As posições achadas usando o programa SHELX são apresentadas na tabela 3.2.

**Tabela 3.2.** Primeiras 8 posições achadas pelo programa SHELX para o derivado de iodo.

<b>Átomo</b>	<b>X</b>	<b>Y</b>	<b>Z</b>	<b>Ocupancia</b>
I001	0.652107	0.200470	0.173499	1.0000
I002	0.976807	0.185127	0.289618	0.9561
I003	1.150383	-0.019791	0.431075	0.5005
I004	0.507492	0.227219	-0.018396	0.4649
I005	0.280762	0.009659	0.073676	0.4375
I006	0.397507	0.222458	0.132246	0.4153
I007	1.333382	0.058899	0.500105	0.3836
I008	0.592522	0.134514	-0.167032	0.3784

### 3.3.4 Estrutura tridimensional da exo-inulinase de *Aspergillus awamori*

A exo-inulinase de *Aspergillus awamori* é um monômero de 69kDa composto de dois domínios. O primeiro domínio apresenta um enovelamento tipo “ $\beta$ -propeller”. O segundo domínio apresenta 12 fitas beta arranjadas em um enovelamento tipo “sandwich” (figura 3.1). O modelo final é um monômero com 517 resíduos, 5 cadeias de oligossacarídeos formados por diferentes números de NAG (N-acetil- $\alpha$ -D-glucosamina), e 768 moléculas de água. O modelo foi refinado com o conjunto de dados nativo obtendo-se estatísticas finais de  $R_{\text{factor}}$  e  $R_{\text{free}}$  de 16,5% e 18,2%, respectivamente.



**Figura 3.1** Representação da estrutura da exo-inulinase de *Aspergillus awamori*. As folhas beta do domínio tipo “ $\beta$ -propeller” foram nomeadas com a letra W. Cada folha é representada por uma cor diferente ( $W_1$  amarelo,  $W_2$  azul marinho,  $W_3$  vermelho,  $W_4$  verde,  $W_5$  laranja). O segundo domínio que apresenta 12 fitas beta é apresentado em roxo.

O domínio N-terminal contém o sítio catalítico. Este domínio é composto por 350 aminoácidos e pertence ao enovelamento tipo  $\beta$ -propeller, onde cinco folhas  $\beta$  se arranjam radial e pseudosimetricamente ao redor de um eixo central; cada folha contém 4 fitas  $\beta$ . As fitas 1,2,3, e 4 estão conectadas por grampos de cabelo e a fita 4 de uma folha esta ligada com a fita 1 da folha seguinte. As fitas 1 estão perto do eixo central e se dispõem paralelas a este. O enovelamento geral do domínio é cilíndrico, com as fitas 4 na parte de fora do cilindro. Utilizando a convenção padrão para o enovelamento tipo  $\beta$ -propeller, cada folha foi nomeada com a letra W.

O segundo domínio estrutural da exo-inulinase de *Aspergillus awamori*

compreende 156 resíduos e consiste em duas folhas beta com seis fitas em cada folha, apresentando um enovelamento tipo “sandwich”. Os dois domínios encontram-se separados por uma pequena cadeia polipeptídica.

### 3.3.3 Sítio ativo.

Foi encontrada uma única molécula de frutose ligada ao domínio  $\beta$ -propeller. A Figura 3.2b mostra a frutose e os resíduos que interagem com ela. As distâncias de ligação são mostradas na tabela 3.3.

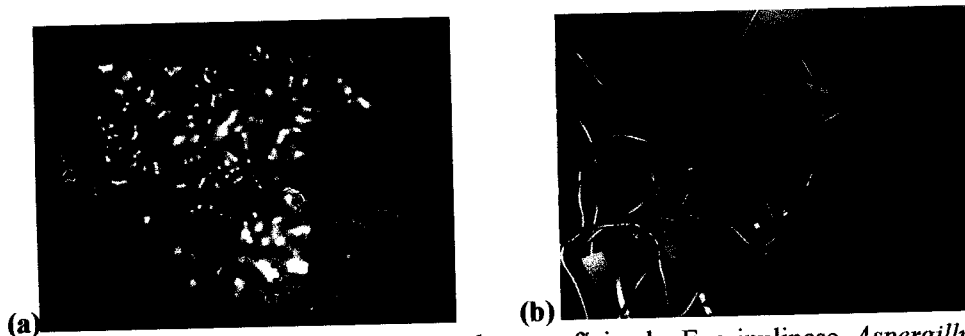


Figura 3.2. (a) Representação da superfície da Exo-inulinase *Aspergillus awamori* com a frutose no bolsão catalítico e os açúcares da glicosilação, (b) Mapa de densidade eletrônica no sítio catalítico indicando a presença da frutose, Este mapa foi gerado utilizando um modelo sem frutose.

Tabela 3.3. Distâncias entre os átomos da proteína e os grupos OH da frutose.

Pontes de hidrogênio		
Átomo do açúcar	Átomo da proteína	Distância (Å)
O1	Asp41OD1	2.49
	Trp335NE1	2.95
O2	Glu241 OE2	2.7
	Water429	2.72
O3	Asp189OD2	2.58
	Arg188NE	2.99
	Glu241OE2	3.05
O4	Ser103N	3.16
	Asp189 OD1	2.61
O5	Gln57NE2	2.59
	Trp65ND2	2.96
	Asn40ND2	2.97

Com base na distância do carbono anomérico foi inferido que os resíduos Glu241 e o Asp41 são os resíduos catalíticos, sendo o primeiro o doador de próton e o

último o nucleófilo. Os resíduos Glu241 e o Asp41 se encontram localizados no final do N-terminal da fita  $\beta$ -1 na folha W1 e da fita  $\beta$ -1 da folha W4 do domínio “ $\beta$ -propeller”, respectivamente.

A distância média entre os oxigênios dos grupos carboxilatos dos resíduos catalíticos (O $\delta$ 2 do Asp41 e O $\epsilon$ 2 do Glu241) e o C<sub>1</sub> da frutose são levemente diferentes (3.19Å e 3.26 Å respectivamente) do esperado para enzimas retentoras (3.0 Å). Provavelmente devido à posição da frutose, produto da reação, que pode encontrar-se em uma orientação diferente em comparação com o substrato natural que é um polissacarídeo.

### **3.3.4 Glicosilação da enzima e seus efeitos na cristalização.**

A exo-inulinase é uma glicoproteína com cinco sítios de N-glicosilação. No processo de cristalização foram encontrados cristais em dois sistemas cristalinos; monoclinico P2<sub>1</sub> e ortorrômico P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. Dados de difração de raios-x foram coletados para as duas formas cristalinas e, após a resolução da estrutura, cinco sítios de glicosilação foram encontrados nas asparaginas Asn48, Asn92, Asn 281, Asn379 e Asn411 nas duas estruturas cristalinas.

No sistema monoclinico só foi possível colocar no mapa de densidade eletrônica resíduos de N-acetilglucosamina, enquanto que no sistema ortorrômico foi encontrada uma cadeia de quatro manoses ligada a uma N-acetilglucosamina. Esta cadeia de oligossacarídeos impede a cristalização desta população de proteína no sistema monoclinico. A superposição das duas estruturas no grupo espacial P2<sub>1</sub> mostra que as manoses ligadas na Asn48 causam impedimento estérico com as moléculas relacionadas por simetria.

Naturalmente, todos os cinco sítios de glicosilação podem apresentar estas cadeias de manoses. Sua ausência pode ser explicada por clivagem durante o cultivo do *Aspergillus awamori* por  $\alpha$ -manosidases como foi descrito previamente para a glucoamilase do mesmo fungo. A seleção de moléculas no processo de cristalização pode acontecer devido à diferença de solubilidade entre moléculas com diferentes graus de glicosilação.

### 3.4. Conclusões

- Determinou-se que a exo-inulinase de *Aspergillus awamori* é um monômero de 69kDa composto de dois domínios, o primeiro domínio apresenta um enovelamento tipo “ $\beta$ -propeller” e o segundo domínio apresenta 12 fitas beta arranjadas em um enovelamento tipo “sandwich”.
- No processo de cristalização da exo-inulinase de *Aspergillus awamori* foram encontrados cristais em dois sistemas cristalinos; monoclinico  $P2_1$  e ortorrômbico  $P2_12_12_1$ , que correspondem a proteínas com diferentes graus de glicosilação.
- Foram identificados cinco sítios de N-glicosilação no mapa de densidade eletrônica da exo-inulinase de *Aspergillus awamori* e uma única molécula de frutose ligada no domínio  $\beta$ -propeller.
- Com base na distância do carbono anomérico foi inferido que o Glu241 e o Asp41 são os resíduos catalíticos, sendo o primeiro o doador de próton e o último o nucleófilo. Glu241 e Asp41 se encontram localizados no final do N-terminal da fita  $\beta$ -1 na folha W1 e da fita  $\beta$ -1 da folha W4 do domínio “ $\beta$ -propeller”, respectivamente.



### 3.5. Referências

- <sup>1</sup> Eldelman J. and Jefford T.G., (1964), The metabolism of fructose polymers in plants. *Biochem. J.* **93**:148-161.
- <sup>2</sup> Pontis H. G. and Del Campillo, E. (1985) Biochemistry of storage carbohydrates in green plants, New Cork, Academic Press, p.205-226.
- <sup>3</sup> Incoll L.D. and Bonnett G. D. (1996) Fructans in the compositae- a short review. *Biology and Utilization*, Kew: Proceedings of the international compositae conference, **V.2**, p.401-413.
- <sup>4</sup> Roberfroid M.B., (1993) Dietary fiber, inulin and oligofructose: a review comparing their physiological effects, *Critical reviews in Food Science and Nutrition*, **v.33**, p. 103-148.
- <sup>5</sup> Roberfroid M.B., Bornet F., Bouley C. and Cummings J. H. (1995) Colonic microflora: nutritional and health. *Nutrition Reviews*, **v.53**, p.127-130.
- <sup>6</sup> Dysseler P. and Hoffem D. (1995) Inulin, an alternative dietary fibre: properties and quantitative analysis. *European Journal of Clinical Nutrition*, **V.49**, p.145-152.
- <sup>7</sup> Aires M.M., (1991) Fisiologia renal. *Fisiologia*, Rio de Janeiro Editota Guanabara Koogan, p. 495-583.
- <sup>8</sup> Vandamme, E., J., & Derycke, D. G. (1983) Fermentation Process, Properties and Applications *Advances in Applied Microbiology*, **29**: 139-176.
- <sup>9</sup> Balayan, L.M., Pivazian, I.A., Khachaturian, R.N., Afrikian, I.G. and Abelian, V.A. (1996) Inulinases from *Penicillium palitans* and *Penicillium cyclopium*. *Biochemistry* **61**, 645-650.
- <sup>10</sup> Onodera, S. and Shiomi, N. (1992) Purification and subsite affinities of exo-inulinases from *Penicillium trzebinskii*. *Bioscience biotechnology and Biochemistry* **56**, 1443-1447.
- <sup>11</sup> Pessoni, R. A. B., Figueredo-Ribeiro, R. C. L. and Braga, M. R. (1999) Extracellular inulinases from *Penicillium janczemsii*, fungus isolated from the rhizosphere of *Vernonia herbacea* (Asteracea). *J. Appl. Microbiol.* **87**, 141-147.
- <sup>12</sup> Ettalibi, M. and Baratti, J.C. (1990) Molecular and kinetic properties of *Aspergillus ficuum* inulinases. *Agricultural and Biological Chemistry* **54**, 611-668.
- <sup>13</sup> Arand M, Golubev A, Neto J.R.B, Polikarpov I, Wattiez R, Korneeva O, Eneyskaya E. V., Kulminkaya A.A, Shabalin K. A., Shishliannikov S.M., Chepurnaya O. V. And Neustroev K.N. (2002) Purification, characterization, gene cloning and preliminary X-ray data of the exo-inulinase from *Aspergillus awamori* *Biochem. J.* **362**, 131-135.

- <sup>14</sup> Gupta, A.K., Singh, D.P., Kaur, N. and Singh, R. (1994) Production, purification and immobilization of inulinases from *Kluyveromyces fragilis*. *Journal of chemical Technology and Biotechnology* **59**, 377-385.
- <sup>15</sup> Van Den Ende, W., Michiels, A., De Roover, J., Verhaert, P. and Van Laere, A. (2000) Cloning and functional analysis of chicory root fructan1-exohydrolase I (1-FEH I) : a vacuolar enzyme derived from a cell-wall invertase ancestor mass fingerprint of the 1-FEH I enzyme. *Plant J.* **24**, 447-456.
- <sup>16</sup> Burne, R. A. and Penders, J. E. (1992) Characterization of the *Streptococcus mutans* GS-5 fruA gene encoding exo-beta-D-fructosidase. *Infect. Immun.* **60**, 4621-4632.
- <sup>17</sup> Tsujimoto, Y., Watanabe, A., Nakano, K., Watanabe K., Matsui, H., Tsuji, K., Tsukihara, T., Suzuki, Y. (2003). Gene cloning, expression, and crystallization of a thermostable exo-inulinase from *Geobacillus stearithermophilus* KP1289. *Appl. Microbiol. Biotechnol.* **62**, 180-185.
- <sup>18</sup> Nagem R. A. P., Rojas A. L., Golubev A. M., Korneeva O.S. Eneyskaya E. V., Kulminskaya A. A., Neustroev K. N. and Polikarpov I. (2004) Crystal Structure of exo-Inulinase from *Aspergillus awamori* : The Enzyme Fold and Structural Determinants of Substrate Recognition. *J. Mol. Biol.* **344**: 471-480.
- <sup>19</sup> Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst. D***56**, 232-237.
- <sup>20</sup> Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta Cryst. D***57**, 996-1002.
- <sup>21</sup> Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
- <sup>22</sup> Leslie, A.G.W., (1992), Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
- <sup>23</sup> Schneider, T. R.; Sheldrick, G. M. Substructure solution with shelxd (2002) *Acta Cryst. D*, v. **58**, p. 1772-1779.
- <sup>24</sup> La Fortelle, E. de & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.
- <sup>25</sup> Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Cryst. D***52**, 30-42.
- <sup>26</sup> Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458-463.

- 
- <sup>27</sup> Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* D53, 240-255.
- <sup>28</sup> Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst.* A47, 110-119.
- <sup>29</sup> Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* 33, 491-497.

# Annexo 2

Nagem R. A. P., Rojas A. L., Golubev A. M., Korneeva O.S. Eneyskaya E. V., Kulminskaya A. A., Neustroev K. N. and Polikarpov I. (2004) Crystal Structure of exo-Inulinase from *Aspergillus awamori*: The Enzyme Fold and Structural Determinants of Substrate Recognition. *J. Mol. Biol.* 344: 471-480.

## Crystal Structure of Exo-inulinase from *Aspergillus awamori*: The Enzyme Fold and Structural Determinants of Substrate Recognition

R. A. P. Nagem<sup>1</sup>†, A. L. Rojas<sup>1</sup>†, A. M. Golubev<sup>2</sup>, O. S. Korneeva<sup>3</sup>  
E. V. Eneyskaya<sup>2</sup>, A. A. Kulminskaya<sup>2</sup>, K. N. Neustroev<sup>2</sup> and  
I. Polikarpov<sup>1</sup>\*

<sup>1</sup>Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, CEP 13560-970, São Carlos, SP Brazil

<sup>2</sup>Petersburg Nuclear Physics Institute, Gatchina, St Petersburg, 188300, Russia

<sup>3</sup>Voronezh State Technological Academy, pr. Revolutsii 19 Voronezh, 394017, Russia

Exo-inulinases hydrolyze terminal, non-reducing 2,1-linked and 2,6-linked  $\beta$ -D-fructofuranose residues in inulin, levan and sucrose releasing  $\beta$ -D-fructose. We present the X-ray structure at 1.55 Å resolution of exo-inulinase from *Aspergillus awamori*, a member of glycoside hydrolase family 32, solved by single isomorphous replacement with the anomalous scattering method using the heavy-atom sites derived from a quick cryo-soaking technique. The tertiary structure of this enzyme folds into two domains: the N-terminal catalytic domain of an unusual five-bladed  $\beta$ -propeller fold and the C-terminal domain folded into a  $\beta$ -sandwich-like structure. Its structural architecture is very similar to that of another member of glycoside hydrolase family 32, invertase ( $\beta$ -fructosidase) from *Thermotoga maritima*, determined recently by X-ray crystallography. The exo-inulinase is a glycoprotein containing five N-linked oligosaccharides. Two crystal forms obtained under similar crystallization conditions differ by the degree of protein glycosylation. The X-ray structure of the enzyme:fructose complex, at a resolution of 1.87 Å, reveals two catalytically important residues: Asp41 and Glu241, a nucleophile and a catalytic acid/base, respectively. The distance between the side-chains of these residues is consistent with a double displacement mechanism of reaction. Asp189, which is part of the Arg-Asp-Pro motif, provides hydrogen bonds important for substrate recognition.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** exo-inulinase; *Aspergillus awamori*; glycoside hydrolase; crystallographic structure; X-ray structure

\*Corresponding author

### Introduction

Most plants store starch or sucrose as reserve carbohydrates; however, about 15% of flowering plant species store fructans,<sup>1</sup> which are branched polymers of fructose. Inulin is a widespread plant polyfructan that has linear chains of  $\beta$ -(2,1)-linked fructose residues attached to a terminal sucrose residue.<sup>2</sup> Inulin serves as a storage polysaccharide in the Compositae and Gramineae, and is accumulated in the underground parts of several plants of the Asteraceae, including *Vernonia herbacea*,

Jerusalem artichoke (*Helianthus tuberosus*) and chicory (*Cichorium endivia*).<sup>3</sup> Levan, a similar polyfructan, which consists of linear  $\beta$ -(2,6)-linked fructofuranosyl units, is produced in dicotyledonous species, and is found in several types of grass, e.g. *Dactylis glomerata*.<sup>4</sup> Mixed levan, consisting of both 2→1 and 2→6 linked  $\beta$ -D-fructosyl residues, is found in most plant species belonging to the Poales family, including wheat and barley.<sup>4,5</sup>

Inulin, due to its role as a relatively inexpensive and abundant substrate for production of rich fructose syrups, is of considerable industrial interest.<sup>3,6</sup> This type of syrup may be obtained by acid hydrolysis of inulin at 80–100 °C; however, under these conditions fructose easily degrades, producing colored products, such as difructose anhydrides.<sup>7</sup> The enzymatic hydrolysis with inulinases and exo-inulinases (EI) offers an alternative approach to

† R.A.P.N. and A.L.R. contributed equally to this work.

Abbreviations used: EI, exo-inulinase; GHF, glycoside hydrolase family; Bsl, *Bacillus subtilis* levansucrase.

E-mail address of the corresponding author: ipolikarpov@if.sc.usp.br

produce fructose free of undesirable by-products. Successful biotechnological applications of inulinases have brought significant attention to these enzymes.<sup>8</sup>

Inulinases expressed by *Penicillium*,<sup>9–11</sup> *Kluyveromyces*,<sup>12</sup> and *Aspergillus*<sup>13,14</sup> are the most intensively studied fungal enzymes of this class. Exo-inulinases (EC 3.2.1.80, also known as fructan  $\beta$ -fructosidase or exo- $\beta$ -D-fructosidase) hydrolyze terminal, non-reducing 2,1-linked and 2,6-linked  $\beta$ -D-fructofuranose residues in fructans, with concomitant release of  $\beta$ -D-fructose. The natural substrates of exo-inulinases include inulin, levan and sucrose. Nucleotide and amino acid sequences were reported for exo-inulinases from yeast,<sup>15</sup> chicory,<sup>16</sup> *Aspergillus awamori*<sup>14</sup> and *Geobacillus stearothermophilus*.<sup>17</sup>

Based on amino acid sequence comparisons, it has been demonstrated that the exo-inulinase (EI) from *A. awamori* belongs to the glycoside hydrolase family (GHF) 32.<sup>18</sup> This family includes invertase, inulinase, levanase and two types of 1-fructosyl transferase. The catalytic activity of the enzymes from GHF 32 is based on a similar mechanism that involves overall retention of the configuration at the anomeric carbon atom of the substrate. The similarities of the amino acid sequences and activity toward the fructose containing poly- and oligosaccharides possibly reflect an evolutionary relationship among GHF 32 members.

Here, we describe the first crystal structure of exo-inulinase in a ligand-free and fructose-bound form, and discuss a structural basis of the catalytic mechanism of the enzyme.

## Results and Discussion

### Quality of the model and a three-dimensional structure of exo-inulinase

Two crystal forms of the native exo-inulinase were obtained under the same crystallization conditions. Crystallization of the enzyme in the primitive orthorhombic  $P2_12_12_1$  space group has been reported.<sup>14</sup> The new crystal form belongs to the monoclinic  $P2_1$  space group. The crystallographic structure of the enzyme was solved by the SIRAS method, using a quick cryo-soaking technique,<sup>19–21</sup> at 1.55 Å resolution in the monoclinic

form and was refined to the final  $R_{\text{work}}$  and  $R_{\text{free}}$  of 17.0% and 19.8%, respectively. The model has good stereochemical parameters, with root-mean-square deviations (rmsd) of 0.007 Å for bond distances, and 1.17° for bond angles. The crystallographic model was also refined in the orthorhombic crystal form at 1.89 Å resolution, with a final  $R_{\text{work}}$  and  $R_{\text{free}}$  of 16.5% and 19.9%, respectively (rmsd of 0.011 Å for bond distances and 1.31° for bond angles). These structures were found to be essentially the same, with the exception of the degree of glycosylation. The crystallographic structure of EI:fructose complex was refined at 1.87 Å resolution with final  $R_{\text{work}}$  and  $R_{\text{free}}$  of 14.5% and 16.9%, respectively. All residues but one fall in the allowed regions of the Ramachandran plot, as determined by PROCHECK.<sup>22</sup> Ser348 lies outside of the allowed regions, probably because its O $\gamma$  forms hydrogen bonds with both Asn333 N $\delta^2$  (2.80 Å) and Trp331 N $\epsilon^1$  (3.15 Å), which results in energetically unfavorable conformation of this residue.

The enzyme folds into two domains (Figure 1). The N-terminal domain, containing the catalytic site, comprises 353 amino acid residues (Phe20 to Gln372). It belongs to the  $\beta$ -propeller fold with five  $\beta$ -sheets arranged like blades of a propeller, radially and pseudosymmetrically around a central axis. Each  $\beta$ -sheet contains four antiparallel  $\beta$ -strands. Strands 1, 2, 3, and 4 are connected by hairpin turns, and strand 4 of one sheet is connected to strand 1 of the next. Strand 1, in each "blade" of the  $\beta$ -propeller, is the closest to the central axis and runs parallel with it. The overall shape of the domain can be described as cylindrical, with strand 4 of each blade being on the outside of the cylinder.

The five-bladed  $\beta$ -propeller fold is rare, and is exemplified by tachylectin-2,<sup>23</sup> *Cellvibrio japonicus*  $\alpha$ -L-arabinanase A43<sup>24</sup> and *Bacillus subtilis* levansucrase.<sup>25</sup> Remarkably, the latter two enzymes are members of GHF 43 and 68, respectively. Common sequence motifs have been identified between GHF 32 and 68,<sup>26</sup> as well as between GHF 68 and 43<sup>27</sup> despite low (~15%) sequence identity between these families.

The EIC-terminal domain, comprising 156 amino acid residues (Arg382 to Asn537), consists of two  $\beta$ -sheets, each composed of six  $\beta$ -strands, assembled into a sandwich-like structure. The first and second domains are connected by a short

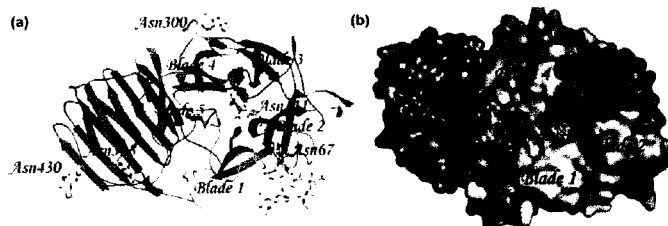
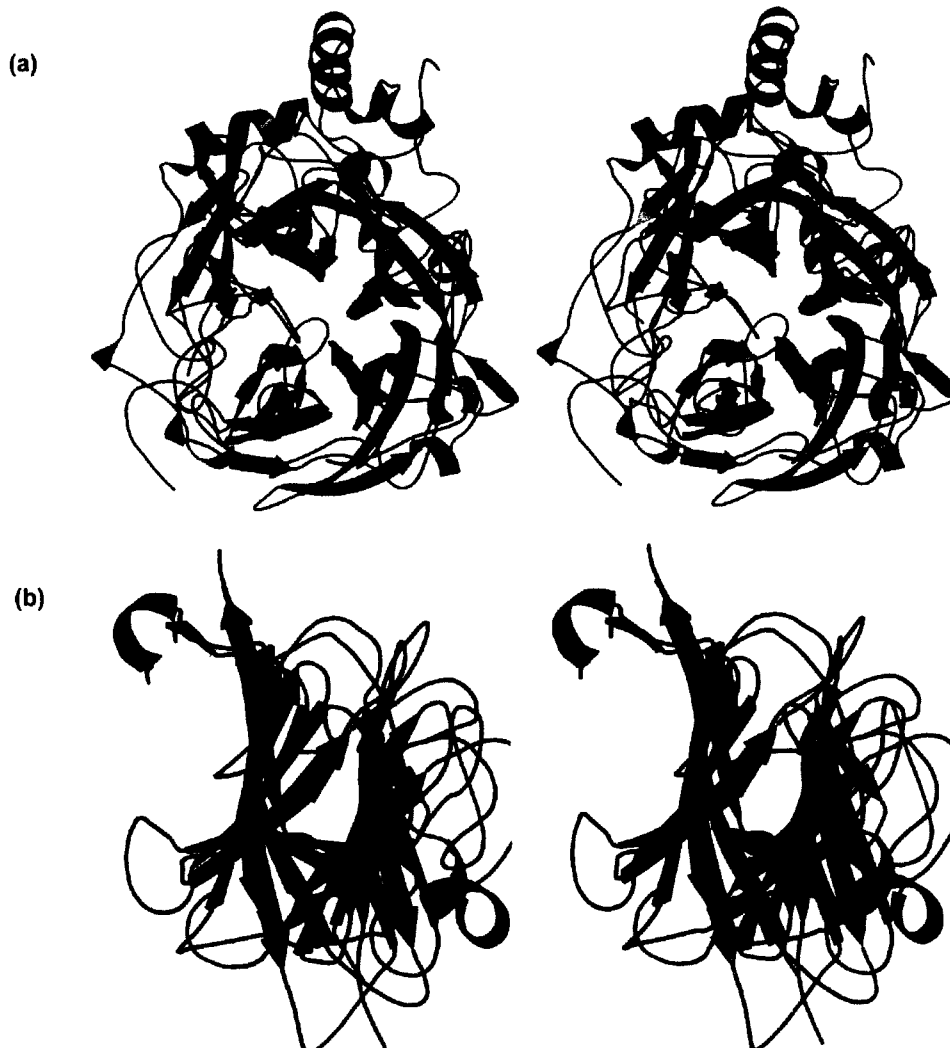


Figure 1. (a) Ribbon representation of the secondary structure elements of exo-inulinase. The N-terminal domain belongs to the five-bladed  $\beta$ -propeller fold. Each blade is shown in a different color (blade 1 is yellow, blade 2 is marine, blade 3 is red, blade 4 is lime and blade 5 is orange). The second domain, colored in slate, consists of 12  $\beta$ -strands arranged in a  $\beta$ -sandwich fold. The short polypeptide chain connecting the two domains is depicted in pink. The five N-linked oligosaccharides present in the orthorhombic crystal are represented as stick. (b) Representation of exo-inulinase surface with the fructose molecule in the activity site.

polypeptide linker, with relative positions and orientations of the domains stabilized by multiple hydrogen bonds and hydrophobic interactions.

Comparison of the present EI model with the structures deposited with the Protein Data Bank, carried out using the DALI server,<sup>28</sup> reveals a significant structural similarity between the EI N-terminal  $\beta$ -propeller domain and the N-terminal domain of levansucrase from *B. subtilis* (PDB code 1OYG),<sup>25</sup> a member of GHF 68. The rmsd values for 189 C $^{\alpha}$  atom pairs is 1.98 Å, while the sequence identity between these two enzymes is only 13% (Figure 2(a)). An additional search shows structural similarity of the inulinase C-terminal domain with carbohydrate recognition domain of rat calcium-dependent lectin p58 (PDB code 1GV9).<sup>29</sup> Strikingly,

although the amino acid sequences of these enzymes hardly bear any similarity (amino acid sequence identity is only 12%), the fold of both protein domains is very similar with rmsd between 105 C $^{\alpha}$  atom pairs of 1.97 Å (Figure 2(b)). Overall, the fold of EI is essentially the same as that of *Thermotoga maritima* invertase, another member of GHF 32, which has been determined very recently by protein crystallography (PDB code 1UYP).<sup>30</sup> Their amino acid sequence is 29% identical; the structural similarity extends over the entire core of two enzymes and the rmsd values between 392 correspondent C $^{\alpha}$  positions is only 1.6 Å. The close structural homology of two distinct members of the GHF 32 enables us to assert that the observed fold represents a template for other members of



**Figure 2.** (a) Stereoview of a superposition of the EI N-terminal domain (blue) with levansucrase from *Bacillus subtilis* (red). (b) Stereoview of the EI C-terminal domain, depicted in blue, superimposed with the carbohydrate recognition domain of rat lectin p58, painted in red.

this family. Figure 3 shows primary structure alignments of representative members of GHF 32 together with the secondary structure elements of exo-inulinase from *A. awamori*. Alignment of amino acid sequences of *A. awamori* exo-inulinase (GenBank™ CAC44220.1), *Aspergillus foetidus*

sucrose:sucrose 1-fructosyl transferase (GenBank™ CAA04131.1), *Gluconacetobacter diazotrophicus* levanase (GenBank™ AAF16405.1), *Aspergillus ficuum* inulinase (GenBank™ CAA07345.1) and *T. maritima* invertase (GenBank™ AAD36485.1) was performed with program CLUSTAL W.<sup>31</sup>

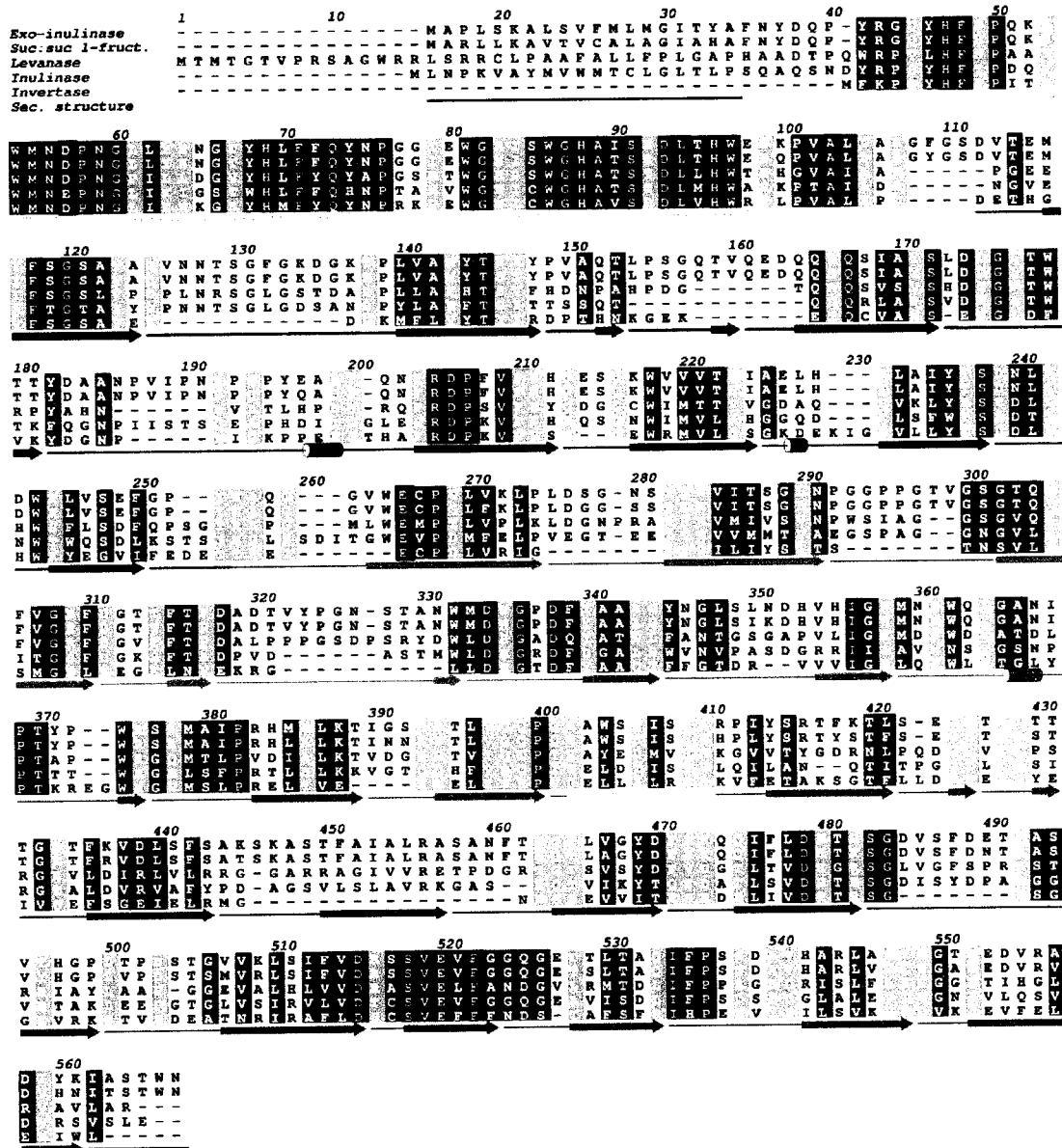


Figure 3. Sequence alignment of representative members of glycoside hydrolase family 32. The following sequences were used in the alignment: Exo-inulinase, *Aspergillus awamori* exo-inulinase; Suc:suc 1-fruct., *Aspergillus foetidus* sucrose:sucrose 1-fructosyl transferase; Levanase, *Gluconacetobacter diazotrophicus* levanase; Inulinase, *Aspergillus ficuum* inulinase; Invertase, *Thermotoga maritima* invertase. The amino acid similarity between primary structures as calculated by the program ALSCRIPT<sup>55</sup> is shown in three different shades of blue. The darker color corresponds to higher levels of sequence similarity. Secondary structure elements for EI as suggested by the program PROCHECK are shown below the amino acid sequences (the same color codes as those used in Figure 1 were applied). Catalytic residues and the conserved RDP motif are highlighted by green and red boxes, respectively.



### Fructose-binding site and catalytic residues

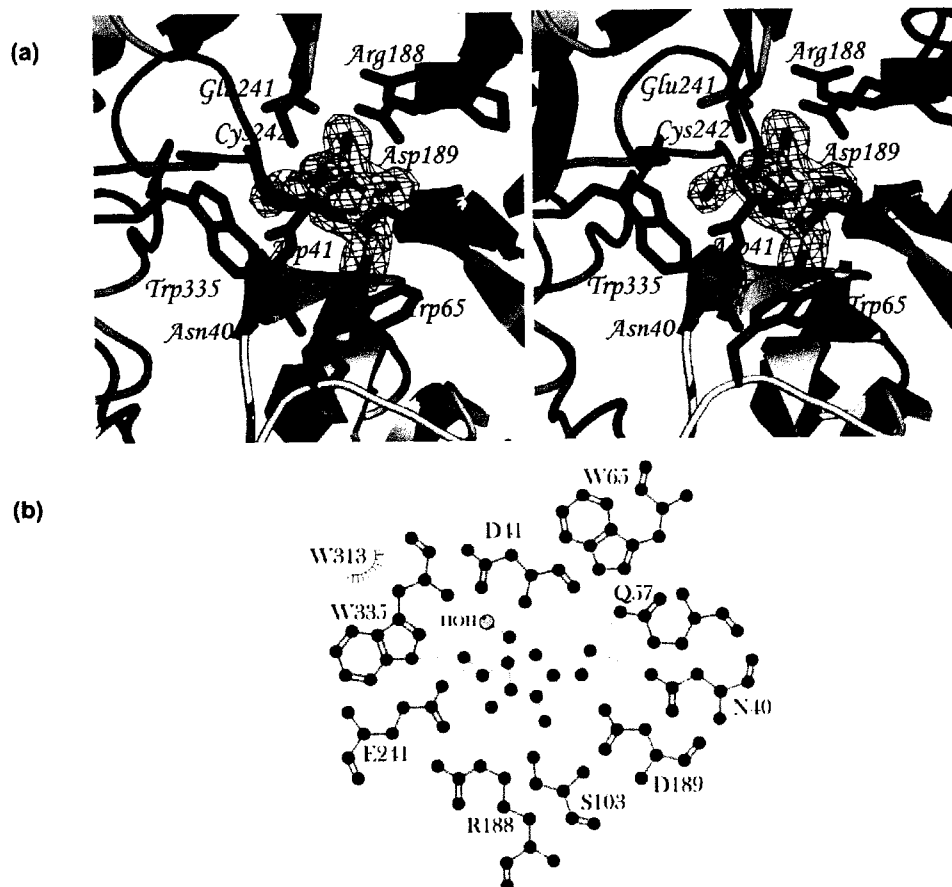
Superposition of the native EI structure with the crystallographic model of the fructose-containing complex of enzyme enables 3D analysis of the fructose-binding site. A single fructose molecule, clearly defined by the difference Fourier electron density map, is bound between the blades of the  $\beta$ -propeller almost at the central axis of the N-terminal domain (Figure 1(b)). The angle between the plane of the sugar ring and the central axis is approximately  $70^\circ$ . Figure 4(a) shows the fructose molecule bound to EI, together with the amino acid residues of the active site. Direct hydrogen bond distances between the protein, water molecules and OH groups of fructose, depicted in Figure 4(b), are given in Table 1.

The enzymatic hydrolysis catalyzed by glycoside hydrolases takes place *via* a general acid catalysis that requires two catalytic residues: a proton donor and a nucleophile.<sup>32</sup> The hydrolysis occurs *via* two major mechanisms giving rise to either an overall

**Table 1.** Close contacts between protein and fructose non-carbon atoms

Fructose hydroxyl groups	Direct hydrogen bond	
	Protein atom	Distance (Å)
OH-1	Asp41 OD1	2.49
	Trp335 NE1	2.95
OH-2	Glu241 OE2	2.70
	Water	2.72
OH-3	Asp189 OD2	2.58
	Arg188 NE	2.99
	Glu241 OE2	3.05
OH-4	Ser103 N	3.16
	Asp189 OD1	2.61
OH-5	Gln57 NE2	2.59
	Trp65 ND2	2.96
	Asn40 ND2	2.97

A distance cut-off of 3.3 Å was used.



**Figure 4.** (a) Stereoview of fructose:enzyme interactions in the catalytic site of exo-inulinase. The fructose and the neighboring residues are shown. An omit electron density map ( $mF_{obs} - DF_{calc}$ ,  $\phi_{calc}$ ), contoured at  $3\sigma$ , is displayed around the catalytic site of EI. (b) Schematic design of the interactions between the active-site residues and the fructose molecule (in pink). Hydrogen bonds are shown as cyan dotted lines. The Figure was drawn using the program LIGPLOT.<sup>56</sup>

retention or inversion of the anomeric configuration of the substrate.<sup>33</sup> In both, retaining and inverting glycosidases, the proton donor is found within a hydrogen bond distance from the glycosidic oxygen atom. In retaining enzymes, the nucleophile/base is in close vicinity of the sugar anomeric carbon atom. However, the nucleophile is more distant in inverting enzymes that must accommodate a water molecule between the catalytic base and sugar. Consequently, the average distances of approximately 5.5 Å and 10 Å are commonly observed between the two catalytic residues in the retaining and inverting enzymes, respectively.<sup>34</sup> On the basis of the 3D superposition of the EI structure and its complex with fructose (Figure 4(a)), we identified Asp41 and Glu241 as two putative catalytic residues. The former is presumably a nucleophile, whereas the latter is a proton donor. This structure-based assignment of the catalytic residues is in agreement with the site-directed mutagenesis studies of another member of GHF 32, yeast invertase.<sup>35,36</sup> Amino acid sequence alignment reveals 36% identity and 52% similarity of these two enzymes. EI Asp41, the residue of the conserved sequence motif 38-WMNDPNG-44, and Glu241 of the conserved motif 241-ECPGL-245 (EI amino acid numbering), are equivalent, respectively, to Asp23 and Glu204 of yeast invertase. Mutation of Asp23 to Asn rendered the enzyme basically inactive,<sup>35</sup> whereas mutagenizing Glu204 to Ala resulted in a 3000-fold reduction in the  $k_{cat}$  of invertase, indicating that these two residues play a major role in catalysis.<sup>36</sup> This conclusion is strengthened further by the recent results of site-directed mutagenesis and structural analysis of *B. subtilis* levansucrase (*Bsl*).<sup>25</sup> Mutation of Asp86 (nucleophile) or Glu342 (general acid), that are structurally analogous to Asp41 and Glu241 in EI, to alanine completely abrogates *Bsl* enzymatic activity. Moreover,

positions and conformations of the active-site residues are conserved between EI and *T. maritima* invertase. The putative reaction mechanism involving Asp41 and Glu241 in hydrolysis of fructose containing oligosaccharides is depicted in Figure 5.

Asp41 and Glu241 are located, respectively, at the  $\beta$ -1 strand of the first blade of the  $\beta$ -propeller and the  $\beta$ -1 strand of the fourth blade. Although EI belongs to the group of enzymes that retain the overall anomeric configuration of the substrate, the distances from Asp41 O<sup>δ2</sup> and Glu241 O<sup>δ2</sup> to C1 of fructose are slightly larger (3.19 Å and 3.26 Å, respectively) than those found normally in the retaining carbohydrases (approximately 3.0 Å). The possible reason for this observation is related to the fact that the position and orientation of the fructose, the inhibitor and a product of the enzymatic reaction, may be slightly different from those of the natural substrate, polyfructan.

#### The conserved RDP motif

The multiple sequence alignment of fructosyl transferases, invertases, levanases, inulinases and sucrose-6-phosphate hydrolases revealed the presence of a conserved Arg-Asp-Pro (RDP) motif, which is presumed to have a common functional role.<sup>37-39</sup> The amino acid sequence of EI contains the conserved RDP motif (Arg188-Asp189-Pro190). Residues Arg188 and Asp189 of this motif participate in substrate binding, and are involved in hydrogen bond interactions with the hydroxyl groups 3 and 4 of fructose. It seems plausible, therefore, that these interactions are important for recognition of the sugar ring and might be responsible for specificity of the enzyme toward fructopyranosyl residues (Figure 4).

Recently, the role of Asp of the RDP motif of

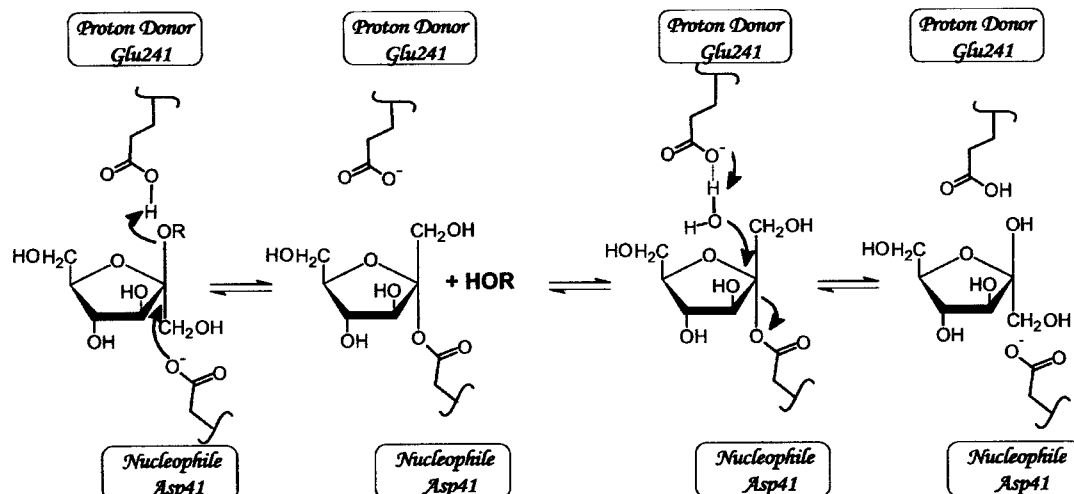


Figure 5. A representation of a double displacement mechanism of reaction with an overall retention of the configuration at the anomeric carbon atom of the substrate, catalysed by exo-inulinase from *Aspergillus awamori*. The R group may represent fructose units of inulin.

levansucrase from *Acetobacter diazotrophicus* SRT4 (LsdA) was studied by site-directed mutagenesis.<sup>40</sup> The D309N LsdA mutant showed considerable reduction in catalytic activity. The  $k_{\text{cat}}$  of the D309N LsdA was decreased 75-fold, while  $K_m$  value was similar to those of the wild-type enzyme, indicating that RDP motif Asp plays an essential role in the catalytic mechanism.<sup>40</sup> Along the same line, mutation of Asp274 to alanine resulted in inactivation of Bsl.<sup>25</sup> Asp274 makes part of the RDP conserved sequence motif found also in GHF 68.

In order to elucidate the result of a similar mutation within the RDP motif on catalytic activity of EI, we replaced Asp189 by Asn in the crystallographic model of the enzyme (D189N mutation). This mutation does not significantly affect overall positions or orientations of the active-site residues, but it opens a possibility for a new hydrogen bond between catalytic Asp41 and Asn189 that does not exist in the native EI. Such hydrogen bond interaction would change the nucleophilicity of Asp41, causing the electron density of the hydroxyl groups of the Asp41 side-chain to shift towards the Asn189 amide group. This would render Asp41 less nucleophilic in character, affecting the efficiency of sucrose hydrolysis, but not the enzyme specificity. This hypothesis is in agreement with the site-directed mutagenic studies of levansucrase from *A. diazotrophicus*,<sup>40</sup> and provides a structural basis for the importance of the conserved RDP motif for glycosyl transferase specificity toward fructosylsaccharides.

#### Carbohydrate moiety and its effect on crystallization

The inulinase is a glycoprotein containing five sites of *N*-glycosylation. During the inulinase crystallization, two crystal forms, primitive monoclinic,  $P2_1$ , and primitive orthorhombic,  $P2_12_12_1$ , were obtained in the same crystallization drops. X-ray diffraction data for both crystal forms were collected and the difference in glycosylation between EI structures determined in the two forms was analyzed. Five glycosylation sites in both crystal forms are located at asparagine residues; Asn67, Asn111, Asn300, Asn398, and Asn430. In the monoclinic form, only *N*-acetylgalactosamine residues were located in the electron density map. In the orthorhombic form, a branched tetramannosaccharide attached to two *N*-acetylgalactopyranoside residues was found at Asn67. It is tempting to conclude that this antenna-like sugar structure prevented crystallization in the monoclinic space group. Indeed, a superposition of the X-ray structure refined in space group  $P2_12_12_1$  onto the 3D model determined in the monoclinic crystal form proves that the tetramannosaccharide would cause a clash with the symmetry-related molecules. In nature, all five *N*-glycosylation sites could contain antenna-like mannosaccharides. Their absence from the crystal may be explained by presuming that they were cleaved with *A. awamori*

$\alpha$ -mannosidases during cultivation of the fungi. A similar process has been observed for *Trichoderma reesei*.<sup>41</sup> Further selection of the less glycosylated molecules could occur during crystallization as a consequence of variation of protein solubility with different content of glycosyl moieties. It is known that mannose oligosaccharides attached to other glycoside hydrolases in the process of post-translational modification aid in self-association into tetramers, hexamers and other oligomeric assemblies.<sup>36,42,43</sup> It is, therefore, not surprising that protein glycosylation interferes strongly in its assembly during nucleation and crystal growth.

#### Conclusions

The high-resolution X-ray model of the exo-inulinase from *A. awamori* presented here reveals that the protein is composed of two separated domains, the larger N-terminal catalytic domain of the rare five-bladed  $\beta$ -propeller fold and the smaller C-terminal domain, folded as a  $\beta$ -sandwich.

Structural comparison of the native enzyme with the enzyme:fructose complex enables the identification of the active site, and residues Glu241 as the proton donor and Asp41 as the nucleophile. Asp41 and Glu241 are located, respectively, at the blades 1 and 4 of the N-terminal  $\beta$ -propeller domain. The average distance of 6.1 Å between the four pairs of side-chain oxygen atoms of the catalytic residues is similar to that (5.5 Å) found in enzymes with the overall retention of the anomeric configuration of the substrate. The EI structures shed light on the putative role of the Asp189 residue that is part of the conserved RDP motif. This residue is found in the active site of the enzyme and is important for substrate recognition. We speculate that the D189N mutation might result in a new hydrogen bond between Asn189 and catalytic nucleophile Asp41, affecting its nucleophilic properties, but not changing its conformation. This could decrease the catalytic efficiency significantly, but not the specificity, of the enzyme toward sucrose hydrolysis, which is in agreement with biochemical studies of structurally similar enzymes.

#### Materials and Methods

##### Protein purification, crystallization and data collection

Extracellular *A. awamori* EI was purified and preliminary X-ray data were collected from an orthorhombic crystal form as described.<sup>14</sup> A new monoclinic form of crystals was found under similar crystallization conditions. X-ray diffraction studies revealed that EI crystallizes in the monoclinic space group  $P2_1$ , with the unit cell parameters  $a = 49.95$  Å,  $b = 94.03$  Å,  $c = 67.63$  Å,  $\beta = 107.04^\circ$  and the crystals diffracted beyond 1.55 Å resolution. Calculation of the Matthews coefficient suggested the presence of a monomer ( $V_M = 2.6$  Å<sup>3</sup> Da<sup>-1</sup>) in the asymmetric unit with a solvent content of approximately 52%

(v/v). Native EI crystals (monoclinic) were used to prepare an iodine derivative according to the quick cryo-soaking approach for derivatization<sup>9-21</sup> by immersing a crystal in the cryogenic solution containing 0.5 M sodium iodide for a few minutes and then flash-cooled to 100 K in a stream of cold nitrogen. Similarly, the fructose:enzyme complex was obtained by soaking native EI crystals in 20% (w/v) fructose solution. X-ray diffraction data were collected on a MAR345 image plate using osmic mirror-focused Cu K $\alpha$  X-rays, generated by a RIGAKU ultra X 18 rotating anode operating at 80 mA and 50 kV.

In total, three complete data sets (monoclinic form) were collected: native, iodine derivative and fructose complex. Diffraction images were processed and scaled with the programs MOSFLM<sup>44</sup> and SCALA,<sup>45</sup> respectively. Data collection statistics of each data set, including the previously collected orthorhombic form, are shown in Table 2.

#### Phasing procedure

Due to the strong incorporation of iodine anions into the  $P2_1$  crystallographic structure, the phase problem in the monoclinic form could be solved either by single anomalous dispersion (SAD) or single isomorphous replacement with anomalous scattering (SIRAS) methods. Nevertheless, only more robust SIRAS phases were used for structure determination. The majority of the iodine positions were determined by direct methods with the programs DREAR<sup>46</sup> and Shake-and-Bake.<sup>47</sup> This heavy-atom substructure was then set as an input into the SHARP program<sup>48</sup> for phase calculations. Density

modification with solvent flattening was performed with the program SOLOMON.<sup>49</sup> In the orthorhombic case ( $P2_12_12_1$ ), initial phases for native structure were obtained by the molecular replacement (MR) method with program AMoRe<sup>50</sup> using the refined  $P2_1$  structure as search model.

#### Initial model building in monoclinic form and structure refinement

Solvent-flattened, iodine-derived electron density map and structure factor amplitudes from the native monoclinic data set were used by the program ARP/wARP<sup>51</sup> combined with cycles of REFMAC<sup>52</sup> for an automatic build of the EI 3D structure. Electron density maps with coefficients  $2mF_{obs} - DF_{calc}$  and model-derived phases enabled the identification of the side-chains of the protein. The manual model construction and the initial side-chain assignment were performed using the programs O<sup>53</sup> and XtalView.<sup>54</sup>

After careful identification of the amino acid sequence, the 3D model went through several cycles of refinement with REFMAC and water molecules insertion with ARP-WATERS from the CCP4 program package. The N-linked oligosaccharides were added using the Xfit program from the XtalView suite.

Analysis of the electron density maps of the EI:fructose complex revealed clear electron density for the fructose bound to the active site of the enzyme. The structures of the fructose:enzyme complex and the iodine derivative were refined using the same protocols as those used for the native structure. The structure in the orthorhombic form was refined in a very similar way, after identification

**Table 2.** Data collection and refinement statistics of exo-inulinase (EI) crystals

Data set	Native protein		Iodine derivative	EI:fructose
<b>A. Crystal preparation</b>				
Cryoprotectant solution	Mother liquor <sup>a</sup> 20% glycerol		Mother liquor 20% glycerol 0.5 M NaI	Mother liquor 20% fructose
Soaking time	A few seconds		15 minutes	2 minutes
<b>B. Data collection</b>				
Wavelength (Å)	1.54	1.42	1.54	1.54
Space group	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_1$
Unit cell parameters				
<i>a</i> (Å)	49.95	64.82	49.97	49.83
<i>b</i> (Å)	94.03	82.13	93.67	93.56
<i>c</i> (Å)	67.63	136.22	83.36	67.58
$\beta$ (deg.)	107.04		106.92	107.29
Resolution (Å)	17.8–1.55 (1.63–1.55)	19.9–1.89 (1.99–1.89)	26.6–1.87 (1.97–1.87)	26.6–1.87 (1.97–1.87)
No. reflections	487,330	310,283	834,739	365,375
$I/\sigma$	10.3 (4.5)	8.0 (3.8)	12.8 (5.9)	11.7 (4.0)
Multiplicity	2.9 (2.9)	5.5 (5.0)	7.9 (7.4)	3.5 (3.3)
Completeness (%)	91.3 (85.7)	99.8 (99.6)	92.1 (84.2)	99.6 (97.6)
$R_{merge}$ <sup>b</sup> (%)	4.5 (15.0)	5.4 (29.7)	4.2 (12.8)	5.7 (19.0)
<b>C. Structure refinement</b>				
<i>R</i> -factor (%)	17.0	16.5	17.8	14.5
<i>R</i> -free (%)	19.8	19.9	21.6	16.9
No. water molecules	921	639	549	755
No. carbohydrate residues	5	11	3	6
No. iodine ions <sup>c</sup>	–	–	25	–
No. glycerol molecules	2	4	1	3
Rmsd values				
Bond lengths (Å)	0.007	0.011	0.012	0.013
Bond angles (deg.)	1.17	1.31	1.34	1.39

Statistical values for the highest-resolution shells are shown in parentheses.

<sup>a</sup> Mother liquor consists of 15% (w/v) PEG 3350, 1 mM sodium cacodylate, 20–100 mM sodium acetate (pH 4.0–5.0).

<sup>b</sup>  $R_{merge} = \sum_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \sum_{hkl} I_{hkl}$ .

<sup>c</sup> Iodine sites were not fully occupied.

of the correct molecular replacement solution. The final refinement statistics for each data set are given in Table 2.

## References

- Hendry, G. (1993). Evolutionary origins and natural functions of fructans. A climatological, biogeographic and mechanistic appraisal. *New Phytol.* **123**, 3–14.
- Edelman, J. & Jefford, T. G. (1964). The metabolism of fructose polymers in plants. 4. Beta fructofuranosidases of tubers of *Helianthus tuberosus* L. *Biochem. J.* **93**, 148–161.
- Vandamme, E. J. & Derycke, D. G. (1983). Microbial inulinases: fermentation process, properties, and applications. *Advan. Appl. Microbiol.* **29**, 139–176.
- Bonnett, G. D., Sims, I. M., Simpson, R. J. & Cairns, A. J. (1997). Structural diversity of fructan in relation to the taxonomy of the *Poaceae*. *New Phytol.* **136**, 11–17.
- Carpita, N. C., Kanabus, J. & Housley, T. L. (1989). Linkage structure of fructans and fructan oligomers from *Triticum aestivum* and *Festuca arundinaceae* leaves. *J. Plant Physiol.* **134**, 162–168.
- Pandey, A., Soccol, C. R., Selvakumar, P., Soccol, V. T., Krieger, N. & Fontana, J. D. (1999). Recent developments in microbial inulinases: its production, properties, and industrial applications. *Appl. Biochem. Biotechnol.* **81**, 35–52.
- Barthomeuf, C., Regeat, F. & Pourrat, J. (1991). Production of inulinases by a new mold of *Penicillium rugulosum*. *J. Ferment. Bioengng.* **72**, 491–494.
- Ohta, K., Hamada, S. & Nakamura, T. (1993). Production of high concentrations of ethanol from inulin by simultaneous saccharification and fermentation using *Aspergillus niger* and *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **59**, 729–733.
- Onodera, S. & Shiomi, N. (1992). Purification and subsite affinities of exo-inulinases from *Penicillium trzebinskii*. *Biosci. Biotechnol. Biochem.* **56**, 1443–1447.
- Pessoni, R. A. B., Figueredo-Ribeiro, R. C. L. & Braga, M. R. (1999). Extracellular inulinases from *Penicillium janczewskii*, fungus isolated from the rhizosphere of *Vernonia herbacea* (Asteraceae). *J. Appl. Microbiol.* **87**, 141–147.
- Balayan, L. M., Pivazian, I. A., Khachaturian, R. N., Afrikan, I. G. & Abelian, V. A. (1996). Inulinases from *Penicillium palitans* and *Penicillium cyclopium*. *Biochemistry*, **61**, 645–650.
- Gupta, A. K., Singh, D. P., Kaur, N. & Singh, R. (1994). Production, purification and immobilization of inulinases from *Kluyveromyces fragilis*. *J. Chem. Technol. Biotechnol.* **59**, 377–385.
- Ettalibi, M. & Baratti, J. C. (1990). Molecular and kinetic properties of *Aspergillus ficuum* inulinases. *Agric. Biol. Chem.* **54**, 611–668.
- Arand, M., Golubev, A. M., Neto, J. R., Polikarpov, I., Wattiez, R., Korneeva, O. S. *et al.* (2002). Purification, characterization, gene cloning and preliminary X-ray data of the exo-inulinase from *Aspergillus awamori*. *Biochem. J.* **362**, 131–135.
- Burne, R. A. & Penders, J. E. (1992). Characterization of the streptococcus mutans GS-5 fruA gene encoding exo-beta-D-fructosidase. *Infect. Immun.* **60**, 4621–4632.
- Van Den Ende, W., Michiels, A., De Roover, J., Verhaert, P. & Van Laere, A. (2000). Cloning and functional analysis of chicory root fructan1-exohydrolase I (1-FEH I): a vacuolar enzyme derived from a cell-wall invertase ancestor mass fingerprint of the 1-FEH I enzyme. *Plant J.* **24**, 447–456.
- Tsujimoto, Y., Watanabe, A., Nakano, K., Watanabe, K., Matsui, H., Tsuji, K. *et al.* (2003). Gene cloning, expression, and crystallization of a thermostable exo-inulinase from *Geobacillus stearothermophilus* KP1289. *Appl. Microbiol. Biotechnol.* **62**, 180–185.
- Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G., Henrissat, B. & Svensson, B., eds), pp. 3–12. The Royal Society of Chemistry, Cambridge, UK.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta. Crystallog. sect. D*, **56**, 232–237.
- Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta. Crystallog. sect. D*, **57**, 996–1002.
- Nagem, R. A. P., Polikarpov, I. & Dauter, Z. (2003). Phasing on rapidly soaked ions. *Methods Enzymol.* **374**, 120–137.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
- Beisel, H. G., Kawabata, S., Iwanga, S., Huber, R. & Bode, W. (1999). Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO J.* **18**, 2313–2322.
- Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A. *et al.* (2002). *Cellvibrio japonicus*  $\alpha$ -L-arabinanase A43 has a novel five-blade  $\beta$ -propeller fold. *Nature Struct. Biol.* **9**, 665–668.
- Meng, G. & Futterer, K. (2003). Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. *Nature Struct. Biol.* **10**, 935–941.
- Pons, T., Hernandez, L., Batista, F. R. & Chinea, G. (2000). Prediction of a common  $\beta$ -propeller catalytic domain for fructosyltransferases of different origin and substrate specificity. *Protein Sci.* **9**, 2285–2291.
- Naumoff, D. G. (1999). Conserved sequence motifs in levansucrases and bifunctional  $\beta$ -xylosidases and  $\alpha$ -L-arabinases. *FEBS Letters*, **448**, 177–179.
- Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
- Velloso, L. M., Svensson, K., Schneider, G., Pettersson, R. F. & Lindqvist, Y. (2002). Crystal structure of the carbohydrate recognition domain of p58/ERGIC-53, a protein involved in glycoprotein export from the endoplasmic reticulum. *J. Biol. Chem.* **277**, 15979–15984.
- Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B. & Czjzek, M. (2004). The three-dimensional structure of invertase  $\beta$ -fructosidase from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J. Biol. Chem.* **279**, 18903–18910.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

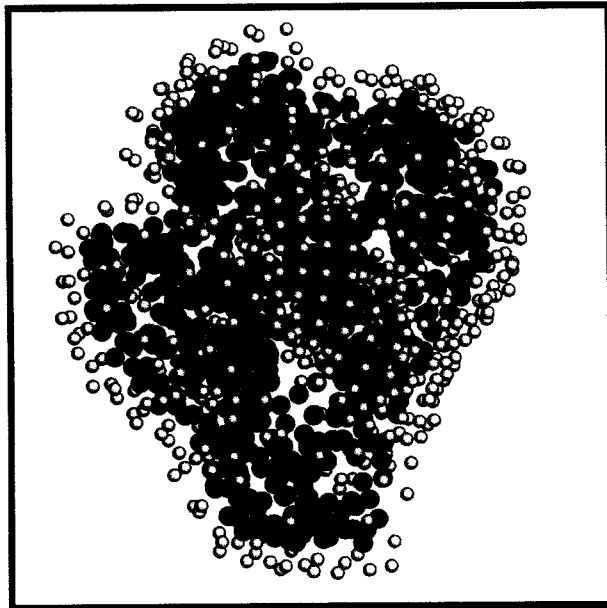
32. Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* **90**, 1171–1202.
33. Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Cambridge Phil. Soc.* **28**, 416–436.
34. McCarter, J. D. & Withers, S. G. (1994). Mechanisms of enzymatic glycosidase hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885–892.
35. Reddy, A. & Maley, F. (1990). Identification of an active-site residue in yeast invertase by affinity labeling and site-directed mutagenesis. *J. Biol. Chem.* **265**, 10817–10820.
36. Reddy, A. & Maley, F. (1996). Studies on identifying the catalytic role of Glu-204 in the active site of yeast invertase. *J. Biol. Chem.* **271**, 13953–13958.
37. Henikoff, S. & Henikoff, J. G. (1991). Automated assembly of protein blocks for database searching. *Nucl. Acids Res.* **19**, 6565–6572.
38. Sprenger, N., Bortlik, K., Brandt, A., Boller, T. & Wiemken, A. (1995). Purification, cloning, and functional expression of sucrose:fructan 6-fructosyltransferase, a key enzyme of fructan synthesis in barley. *Proc. Natl Acad. Sci. USA*, **92**, 11652–11656.
39. Chávez, F. P., Pons, T., Delgado, J. M. & Rodriguez, L. (1998). Isolation and sequence analysis of the orotidine-5'-phosphate decarboxylase gene (URA3) of *Candida utilis*. Comparison with the OMP decarboxylase gene family. *Yeast*, **14**, 1223–1232.
40. Batista, F. R., Hernández, L., Fernández, J. R., Arrieta, J., Menéndez, C., Gómez, R. *et al.* (1999). Substitution of Asp-309 by Asn in the Arg-Asp-Pro (RDP) motif of *Acetobacter diazotrophicus* levansucrase affects sucrose hydrolysis, but not enzyme specificity. *Biochem. J.* **337**, 503–506.
41. Eneyskaya, E. V., Kulminskaya, A. A., Savel'ev, A. N., Shabalin, K. A., Golubev, A. M. & Neustroev, K. N. (1998).  $\alpha$ -Mannosidase from *Trichoderma reesei* participates in the postsecretory deglycosylation of glycoproteins. *Biochem. Biophys. Res. Commun.* **245**, 43–49.
42. Williams, R. S., Trumbly, R. J., MacColl, R., Trimble, R. B. & Maley, F. (1985). Comparative properties of amplified external and internal invertase from the yeast SUC2 gene. *J. Biol. Chem.* **260**, 13334–13341.
43. Reddy, A., MacColl, R. & Maley, F. (1990). Effect of oligosaccharides and chloride on the oligomeric structures of external, internal, and deglycosylated invertase. *Biochemistry*, **29**, 2482–2487.
44. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. In *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography*. SERC Laboratory, Daresbury, Warrington, UK.
45. Evans, P. R. (1997). Scaling of MAD data. In *Proceedings of CCP4 Study Weekend on Recent Advances in Phasing* (Wilson, K. S., Davies, G., Ashton, A. W. & Bailey, S., eds), pp. 97–102, CCLRC Daresbury Laboratory, Warrington, UK.
46. Blessing, R. H. & Smith, G. D. (1999). Difference structure-factor normalization for heavy-atom or anomalous-scattering substructure determinations. *J. Appl. Crystallog.* **32**, 664–670.
47. Weeks, C. M. & Miller, R. (1999). The design and implementation of SnB version 2.0. *J. Appl. Crystallog.* **32**, 120–124.
48. La Fortelle, E. & de Briconne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
49. Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallog. sect. D*, **52**, 30–42.
50. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
51. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
52. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
53. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
54. McRee, D. E. (1999). XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
55. Barton, G. (1993). ALSRIPT: a tool to format multiple sequence alignments. *Protein Engng.* **6**, 37–40.
56. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engng.* **8**, 127–134.

Edited by I. Wilson

(Received 8 June 2004; received in revised form 26 August 2004; accepted 14 September 2004)

# CAPÍTULO 4.

## $\beta$ -Xilosidase de *Trichoderma reesei*



## 4.1. Introdução

### 4.1.1. Xilosidases

A hidrólise de xilano (figura 4.1), o principal componente das hemiceluloses, é importante no processo de degradação de material vegetal. Hemiceluloses são cadeias curtas de heteropolissacarídeos ramificados, compostas de hexoses e pentoses que em conjunto com a celulose e lignina, são os principais componentes da madeira (Figura 4.2). As hemiceluloses são o segundo grupo mais abundante de polissacarídeos na natureza e as hemicelulases mais importantes são as xilanases e as mananases<sup>1</sup>.

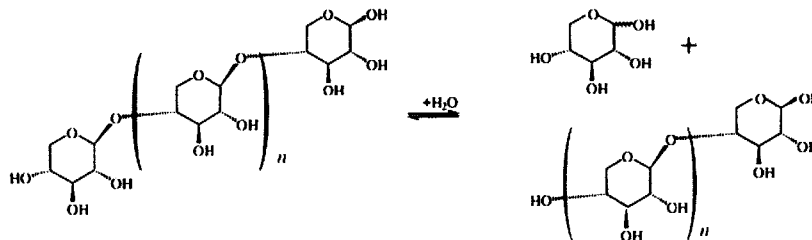


Figura 4.1. Representação esquemática da hidrólise de xilano.

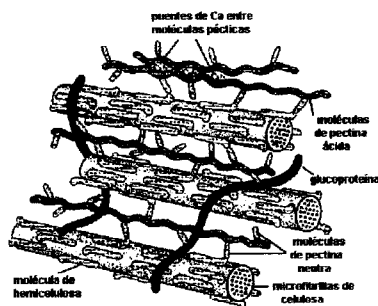


Figura 4.2. Representação esquemática da presença das hemiceluloses na parede celular.

As hemiceluloses estão presentes em resíduos agrícolas, representando, em média, 30% de sua composição<sup>2</sup>. A biotecnologia baseada nas hemicelulases tem aplicações na indústria de polpa de papel, de alimentos e têxtil, na bioconversão de açúcares poliméricos para a produção de etanol e outros produtos químicos, dentre outras.

As enzimas envolvidas na biodegradação das hemiceluloses são hidrolases específicas que clivam determinados tipos de ligações existentes no polímero. Assim, as xilanases rompem ligações glicosídicas entre unidades monoméricas de xilose, as



mananases atuam sobre ligações glicosídicas entre moléculas de manose e as glucuronidases sobre ligações de ácidos urônicos com moléculas de açúcares. Assim como as celulases, as hemicelulases são divididas em dois grupos: endo-hemicelulases, que hidrolisam o polímero ao acaso e geram fragmentos de menor massa molecular; e exo-hemicelulases que hidrolisam os fragmentos gerados pelas endo-hemicelulases.

As  $\beta$ -xilosidases (1,4- $\beta$ -D-xilano; xylanohidrolases; EC 3.2.1.37) clivam a xilose terminal num xylo-oligossacarídeo<sup>3</sup> e são produzidas por fungos, bactérias, leveduras, algas marinhas, crustáceos, insetos e sementes. A principal fonte comercial são os fungos filamentosos.  $\beta$ -xilosidases de vários organismos têm sido classificadas nas famílias das glicosil hidrolases 3, 39, 43, 52 e 54. A estrutura tridimensional desta classe de enzimas é conhecida para  $\beta$ -xilosidases da família 39 ( $\beta$ -xilosidase de *Thermoanaerobacterium saccharo*, PDB:1UHV), e da família 43 ( $\beta$ -xilosidase de *Bacillus subtilis*, PDB:1YIF). Contudo, já foram reportados cristais para a família 52 ( $\beta$ -xilosidase de *Geobacillus steonothermo*) e a família 3 ( $\beta$ -xilosidase de *Trichoderma reesei*).

#### 4.1.2. $\beta$ -xilosidase de *Trichoderma reesei*

A  $\beta$ -xilosidase de *Trichoderma reesei* é uma glicoproteína da família GHF3 que apresenta os dois tipos de glicosilação (N e O) e seu peso molecular estimado é de  $95 \pm 5$  kDa, podendo diferir dependendo das condições de cultivo do fungo<sup>4</sup>. Foi reportado que esta proteína possui atividade  $\alpha$ -L-arabinofuranosídeo e foi sugerido que os dois sítios de atividade enzimática são diferentes<sup>5</sup>. Por seu mecanismo de ação catalítica é classificada como uma enzima retentora<sup>6</sup>.

Existem mais de 180 membros conhecidos da família GHF3; a maioria dos quais são  $\beta$ -D-glucosidases,  $\beta$ -D-xilosidases ou N-acetyl  $\beta$ -D-glucosaminidases. A  $\beta$ -D-glucana de cevada é o membro melhor caracterizado desta família, tanto funcional como estruturalmente<sup>7</sup>. Esta enzima apresenta dois domínios, o primeiro tem um enovelamento tipo barril TIM, enquanto o segundo é um  $\beta$ -sandwich. Embora seja conhecido que os membros pertencentes a uma mesma família de GHs apresentam um enovelamento similar, os dados de SAXS da  $\beta$ -xilosidase de *T. reesei* e a curva

teórica da  $\beta$ -D-glucana de cevada mostram uma grande diferença, sugerindo diferenças nas estruturas tridimensionais destas proteínas. A pesquisa dos possíveis domínios da  $\beta$ -xilosidase de *T. reesei* a partir da seqüência de aminoácidos sugere que os primeiros 643 resíduos podem ser divididos em dois domínios característicos da família GHF3. Contudo, os últimos 146 resíduos do C-terminal devem fazer parte de um terceiro domínio. A predição da estrutura secundária junto com os espectros de dicroísmo circular indicam que este terceiro domínio tem uma conformação predominantemente de folhas beta.

## 4.2. Metodologia

### 4.2.1. Purificação

O protocolo para a purificação da  $\beta$ -xilosidase descrito por Golubev e colaboradores<sup>4</sup>, envolve três colunas de cromatografia, sendo a primeira uma coluna de Sefadex, seguida por uma Sephacryl S-300, e finalmente uma MonoQ HR5/5. A proteína foi fornecida ao nosso laboratório parcialmente pura (faltando o último passo cromatográfico), em forma liofilizada, pelo grupo do Dr. Alexander M. Golubev. A proteína liofilizada foi ressuspensa em tampão acetato de sódio 0.5mM pH 4.2 para atingir concentrações entre 10 e 15 mg/ml. Em seguida foi centrifugada a 3000 g por 5 minutos à 4° C e o sobrenadante foi aplicado em uma coluna MonoQ HR5/5 previamente equilibrada com tampão Tris 20 mM pH 7,2 e eluído com um gradiente linear de NaCl de 0 a 0.5 M com um fluxo de 1 ml/min. O conjunto das frações correspondentes ao último pico foi dialisado por 12 horas à 4° C contra água destilada e concentrado com microcom (50kDa). A pureza da enzima foi verificada através de eletroforese em gel de poliacrilamida (SDS-PAGE).

### 4.2.2. Cristalização

A  $\beta$ -xilosidase de *Trichoderma reesei* foi cristalizada a partir de uma solução de proteína de 10mg/ml em tampão 18% de PEG 3350, acetato de sódio (NaAc) 50mM em pH 5,0 a 18°C. Os cristais começaram a crescer em quatro ou cinco dias e alcançaram seu tamanho máximo depois de sete dias. Estudos preliminares de difração de raios-X revelaram que são ortorrômbicos e difratam até uma resolução de

7Å. Apesar de nosso grupo já ter conseguido um cristal que difratava até 2.7 Å, usando as primeiras remessas de proteína enviadas pelo grupo do Dr. Golubev, tentativas de reproduzir estes cristais não foram bem sucedidas; mesmo utilizando condições similares e diferentes das condições reportadas.

#### 4.2.3 Coleta dos dados de SAXS

Todas as medidas foram realizadas no Laboratório Nacional de Luz Síncrotron (LNLS) em Campinas, onde foi usada a linha de SAXS. O comprimento de onda do feixe incidente foi de  $\lambda=1.608\text{\AA}$ . Para a aquisição de dados foi utilizado um detector ID sensível à posição (DSP). Com ele foi obtida a curva de intensidade de espalhamento em função do módulo do vetor de espalhamento  $q$ . O espalhamento parasita do ar, janelas e fendas foi subtraído da intensidade total espalhada. Houve a necessidade de fazer um *desmearing*, isto é, corrigir os efeitos de borramento (*smearing*) dos dados experimentais introduzidos pela janela de entrada dos fótons do DSP. Utilizou-se a fenda de 8 mm e a distância entre amostra-detector (719mm) foi escolhida de modo a adquirir o espalhamento de 0.01 a 0.25  $\text{\AA}^{-1}$ . As amostras foram isoladas dentro de duas janelas de mica para que as medidas fossem realizadas com o mínimo de espalhamento parasita possível.

#### 4.2.3 Dicroísmo circular

O espectro de dicroísmo circular foi obtido em um espectropolarímetro J715 (Jasco corp., Japão). A proteína, em um tampão de Tris 10 mM, pH 7 na concentração de 0.1mg/ml, foi colocada em uma cuveta de 10 mm de caminho ótico. O espectro do tampão sem proteína foi medido e subtraído daquele obtido para a  $\beta$ -xilosidase de *T. reesei*. O conteúdo de estrutura secundária foi analisado usando os programas CDSSTR<sup>8,9</sup>, CONTINLL<sup>10</sup>, K2D<sup>11</sup> e SELCON3<sup>12</sup> no *website* DichroWeb (<http://www.cryst.bbk.ac.uk/cdweb/html/home.html>)<sup>13</sup>. Foram utilizados dados entre 190-240 nm. Os menores valores de NRMSD<sup>14</sup> foram obtidos com os programas CDSSTR (0.013) e CONTIN (0.038). Nos dois casos o conjunto 7 da base de dados do DichroWeb foi usado como referência.

### 4.3. Resultados e Discussão

Um cromatograma típico da purificação da  $\beta$ -xilosidase com a coluna MonoQ HR5/5 é mostrados na Figura 4.3. A proteína é eluída em 40% de concentração de NaCl 1M, que corresponde ao último pico do espectro. As frações correspondentes a cada um dos picos observados do cromatograma foram reunidas e alíquotas de 7 $\mu$ l aplicados no gel de SDS-PAGE (Figura 4.4)

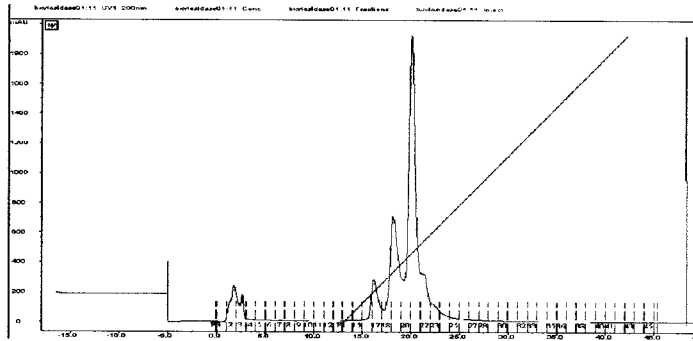


Figura 4.3. Perfil da Cromatografia de troca iônica (MonoQ 5/5) para a  $\beta$ -xilosidase.

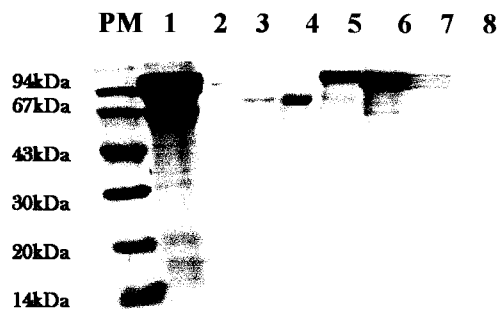


Figura 4.4. Eletroforese em gel de poliacrilamida (SDS-PAGE). No poço 1 observa-se a proteína antes da MonoQ; no poço 6 o último pico da cromatografia que corresponde à  $\beta$ -xilosidase de *Trichoderma reesei*.

Inicialmente foram obtidos cristais que difratavam a baixa resolução, observando-se variações para cristais que provinham de uma mesma gota. Observou-se também se observaram variações para cristais crescidos na mesma condição de cristalização, mas com proteína de diferentes remessas. A falta de reprodutibilidade no processo de cristalização pode ser explicada pelo grau de glicosilação da enzima, que depende das condições de cultivo do fungo que não podem ser completamente controladas<sup>4</sup>.

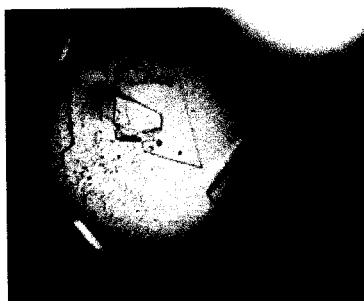
Foi observada uma diferença na solubilidade das diferentes remessas de proteína, e não foi possível atingir concentrações de 10mg/ml para nenhuma delas. Esta concentração foi reportada como requerida para a cristalização da proteína. Os melhores valores medidos pelo método de Bradford foram de 6 a 8mg/ml. Os cristais crescidos usando estas concentrações foram muito pequenos ou placas muito finas. Foi feito um *screening* ao redor das condições reportadas, visando melhorar a qualidade dos cristais (Tabela 4.1). Os melhores resultados se observam nas figuras 4.5 e 4.6.

**Tabela 4.1.** Ensaio de cristalização para a  $\beta$ -xilosidase.

Tampão	Agente precipitante	pH	Concentração da proteína
NaAc 50mM	PEG 3350 10 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 3350 12%	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 3350 14 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 3350 16 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 3350 18 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 3350 20 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 6000 10 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 6000 15 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 4000 10 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 4000 15 %	5.0, 5.2, 5.4, 5.6, 5.8, 6.0	6mg/ml e 8mg/ml



**Figura 4.5.** Cristais de  $\beta$ -xilosidase crescidos em 18% de PEG 3350, acetato de sódio 50mM pH 6.0 e 8mg/ml de proteína.



**Figura 4.6.** Cristais de  $\beta$ -xilosidase crescidos em 18% de PEG 3350, acetato de sódio 50mM pH 5.0, e 8mg/ml de proteína

Novos ensaios foram realizados variando o pH e a concentração do acetato de sódio (tabela 4.2). Os melhores resultados são apresentados na figura 4.7.

**Tabela 4.2.** Ensaio de cristalização para a  $\beta$ -xilosidase variando o pH e a concentração do acetato de sódio.

Tampão	Agente precipitante	pH	Concentração da proteína
NaAC 10mM	PEG 3350 15 %	5.0, 5.2,5.4,5.6,5.8,6.0	6mg/ml e 8mg/ml
NaAC 20mM	PEG 3350 15%	5.0, 5.2,5.4,5.6,5.8,6.0	6mg/ml e 8mg/ml
NaAC 30mM	PEG 3350 15 %	5.0, 5.2,5.4,5.6,5.8,6.0	6mg/ml e 8mg/ml
NaAC 40mM	PEG 3350 15 %	5.0, 5.2,5.4,5.6,5.8,6.0	6mg/ml e 8mg/ml
NaAC 50mM	PEG 3350 15 %	4.0, 4.2,4.4,4.6,4.8,5.0	6mg/ml e 8mg/ml



**Figura 4.7.** Cristais de  $\beta$ -xilosidase crescidos em 15% de PEG 3350, acetato de sódio 50mM pH 4.0 e 8mg/ml de proteína.

Foram feitas tentativas trocando o tampão acetato de sódio por fosfato de sódio (tabela 4.3) mas não foram observadas melhoras na qualidade da difração para os cristais obtidos nestas tentativas.

**Tabela 4.3.** Ensaio de cristalização para a  $\beta$ -xilosidase trocando o tampão acetato de sódio por fosfato de sódio.

Tampão	Agente precipitante	pH	Concentração da proteína
NaPh 10mM	PEG 3350 15 %	5.5,5.8,6.0,6.2,6.4	6mg/ml e 8mg/ml
NaPh 20mM	PEG 3350 15%	5.5,5.8,6.0,6.2,6.4	6mg/ml e 8mg/ml
NaPh 30mM	PEG 3350 15 %	5.5,5.8,6.0,6.2,6.4	6mg/ml e 8mg/ml
NaPh 40mM	PEG 3350 15 %	5.5,5.8,6.0,6.2,6.4	6mg/ml e 8mg/ml
NaPh 50mM	PEG 3350 15 %	5.5,5.8,6.0,6.2,6.4	6mg/ml e 8mg/ml

Considerou-se o uso da proteína sem passar pela coluna de troca iônica, o que permitia atingir concentrações de até de 12.5mg/ml, porém num grau de pureza muito menor, as condições testadas com esta concentração se mostram na tabela 4.4. Apesar de terem sido obtidos cristais nestas condições, eles são menores que os obtidos nas mesmas condições com a proteína purificada.

**Tabela 4.4.** Ensaio de cristalização para a  $\beta$ -xilosidase, usando a proteína parcialmente pura.

Tampão	Precipitante	pH	[ ] proteína
NaAc 50mM	PEG 3350 15 %	5.0, 5.5,6.0	10mg/ml e 12.5mg/ml
NaAC 50mM	PEG 3350 18%	5.0, 5.5,6.0	10mg/ml e 12.5mg/ml

Visando estabilizar a proteína, adicionamos xilose na solução de cristalização, (tabela 4.5) já que foi reportado que  $\beta$ -xilosidase de *Thermoanaerobacterium saccharolyticum*<sup>15</sup> foi cristalizada em complexo com xilose e os cristais difrataram até 2.4 Å de resolução. Infelizmente, em nosso caso, a adição da xilose não alterou a forma macroscópica dos cristais (sempre se observaram placas) nem a qualidade da difração.

**Tabela 4.5.** Ensaio de cristalização para a  $\beta$ -xilosidase com xilose.

Tampão	Precipitante	ligante	pH	[ ] proteína
NaAc 50mM	PEG 3350 15 %	Xilose 1mM	5.0, 5.5,6.0	10mg/ml e 12.5 mg/ml
NaAC 50mM	PEG 3350 18%	Xilose 1mM	5.0, 5.5,6.0	10mg/ml e 12.55 mg/ml

Em uma última tentativa se decidiu mudar o agente precipitante. Assim, utilizou-se no lugar de PEG 3350, sulfato de amônio em diferentes tampões. Neste caso não apareceram cristais em nenhuma condição (tabela 4.6).

**Tabela 4.6.** Ensaio de cristalização para a  $\beta$ -xilosidase com sulfato de amônio.

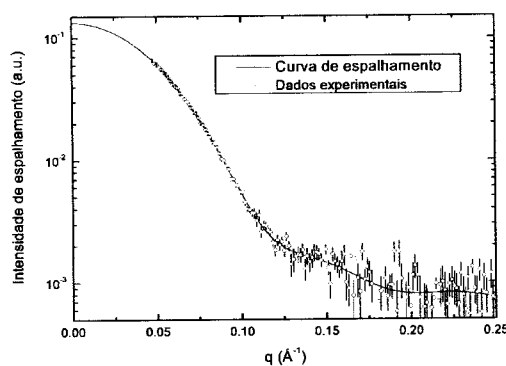
Tampão	Agente precipitante (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	pH	Concentração da proteína
NaAc 0.1M	0.25M,0.5M,0.62M 0.7M	4.0,4.2,4.4,4.6,4.8,5.0	10mg/ml
KPh 0.1M	0.25M,0.5M,0.62M 0.7M	5.0,5.4,5.8,6.2,6.4,6.8	10mg/ml
Tris 0.1M	0.25M,0.5M,0.62M 0.7M	7.0,7.2,7.4,7.6,7.8,8.0	10mg/ml

Em vista dos resultados negativos nas diferentes tentativas de reproduzir cristais que difratassem à alta resolução, considerou-se a possibilidade de trabalhar com proteína deglicosilada. Estes ensaios foram feitos no laboratório do Dr. Golubev, contudo a proteína deglicosilada precipita sugerindo que os carboidratos ajudam na estabilidade da mesma (comunicação pessoal do Dr. Golubev).

Na tentativa de separar as populações que apresentam grau de glicosilação diferente foi feita uma cromatografia de afinidade. A proteína concentrada foi aplicada numa coluna de sefarose com uma lectina (concanavalina A) ligada covalentemente. A concanavalina A é uma proteína que tem afinidade por glucose e manose. As tentativas de eluir a  $\beta$ -xilosidase num gradiente de manose ou glucose não foram bem sucedidas. Em geral não foi possível recuperar a proteína mesmo em altas concentrações (1M) de glucose ou manose.

#### 4.3.1. Determinação do envelope estrutural.

Usando o programa GNOM<sup>16,17</sup> a função de distribuição de distâncias,  $P(r)$  (Figura 4.9), foi calculada a partir da curva de espalhamento de raios-X a baixo ângulo. (Figura 4.8). Assim, determinou-se que a dimensão máxima é 90 Å e o raio de giro é 36.9 Å. O envelope foi modelado usando o programa GASBOR<sup>18,19</sup>. Dez modelos independentes foram comparados entre si e a figura 4.10 mostra o modelo que apresentou menor índice de discrepância com os dados experimentais ( $\chi^2 = 1.42$ ).



**Figura 4.8.** Curva de espalhamento a baixo ângulo da  $\beta$ -xilosidase de *Trichoderma reesei* em um tampão 20 mM TrisHCl, pH 6.



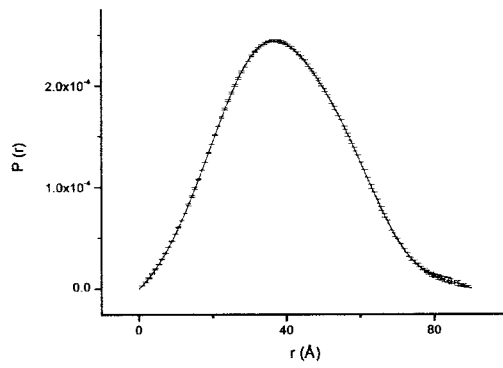


Figura 4.9. Distribuição de distâncias  $P(r)$  da  $\beta$ -xilosidase de *Trichoderma reesei* calculada usando o programa GNOM.

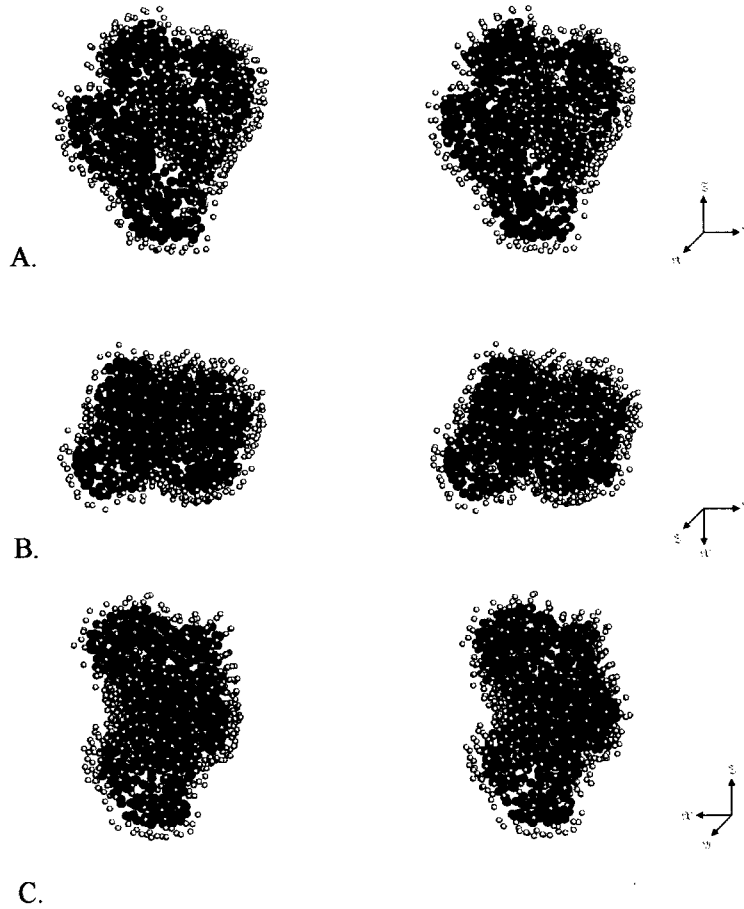
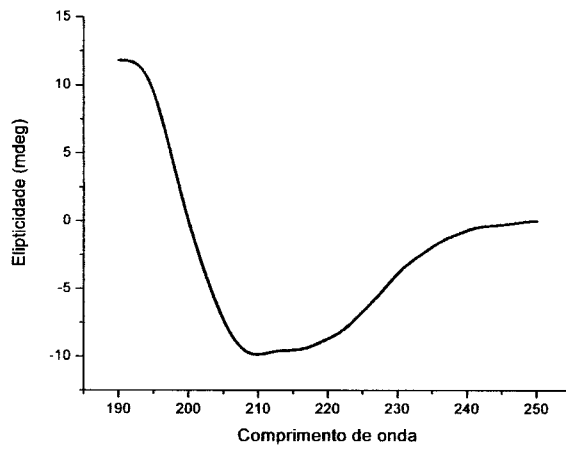


Figure 4.10. Representação estereográfica de três diferentes vistas do envelope da  $\beta$ -xilosidase de *T. reesei*. Em b. e c. as figuras foram rodadas  $90^\circ$  ao redor de Y e Z. Com respeito a figura a, respectivamente.

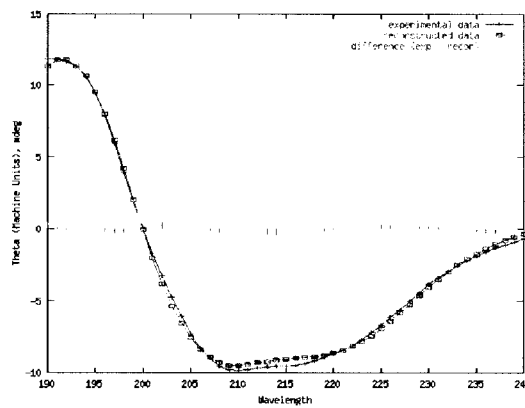
4.3.2 Estrutura secundária da  $\beta$ -xilosidase de *T. reesei*

Visando determinar o conteúdo de estrutura secundária da  $\beta$ -xilosidase de *T. reesei* foram conduzidos experimentos de dicroísmo circular como descrito na seção 4.2.3. O espectro da proteína é apresentado na figura 4.11a e os ajustes da deconvolução do espectro usando os programas CONTILL e CDSSTR são mostradas nas figuras 4.11b e figuras 4.11c, respectivamente.

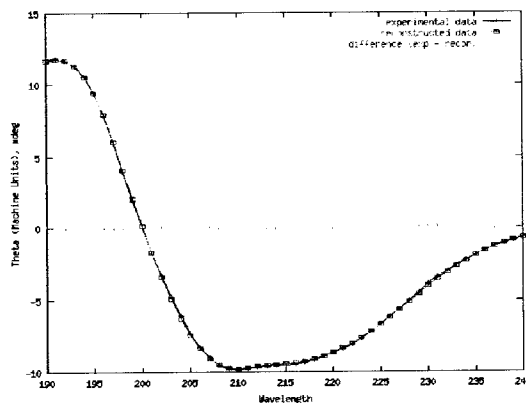
O melhor ajuste entre os dados experimentais e o espectro reconstruído foi obtido com o programa CDSSTR, mas os resultados obtidos com CONTINLL também são apresentados. Isto se deve ao fato de ter sido reportado que com o uso de CDSSTR os valores de NRMSD tendem a ser mais baixos, embora isto não signifique necessariamente um melhor resultado<sup>20</sup>.



a. CD espectro



b. CONTINLL



c.CDSSTR

Figure 4.11. a. Espectro de dicroísmo circular da  $\beta$ -xiliosidase *T. reesei*. b. Ajuste dos dados experimentais e a curva estimada pelo programa CONTINLL c. Ajuste dos dados experimentais e a curva estimada pelo programa CDSSTR.

O espectro é característico de uma proteína  $\alpha/\beta$ , com linhas espectrais atribuíveis as transições dos grupos amida da cadeia principal da proteína. A transição  $n\pi^*$  gera um mínimo ao redor dos 222 nm e indica a presença de estruturas em forma de  $\alpha$ -hélice<sup>21</sup>. O mínimo local em 210nm é uma média entre o mínimo a 208nm proveniente da polarização paralela  $\pi\pi^*$  do eixo da hélice e do mínimo a 215nm proveniente da transição  $n\pi^*$  das fitas beta. Assim, isto sugere que a estrutura secundária da proteína é  $\alpha/\beta$ . Os valores estimados do conteúdo de estrutura em hélice, beta, voltas e desordenada são apresentados na tabela 4.1.

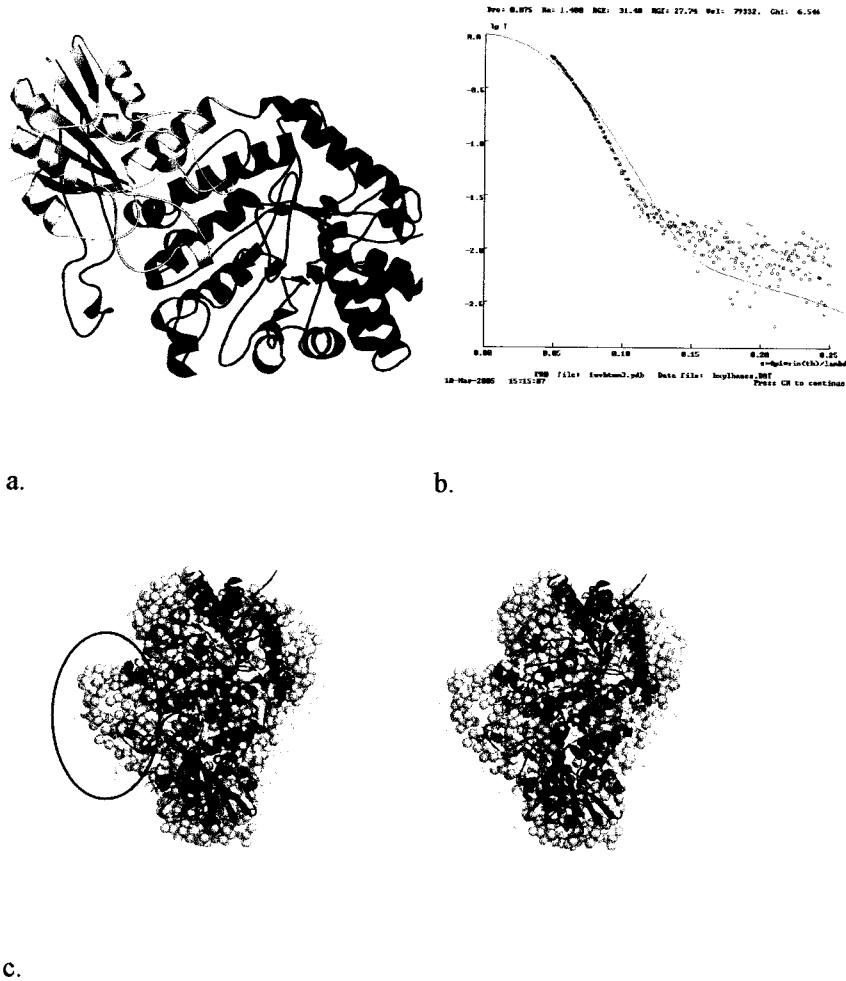
Tabela 4.1. Conteúdo de estrutura secundária da  $\beta$ -xiliosidase *T. reesei*, determinado a partir do espectro de dicroísmo circular.

Programa	Hélice	Beta	Voltas	desordenada	NMRSD
CDSSTR	23.0%	27.0%	20.0%	30.0%	0.013
CONTINLL	24.1%	25.5%	21.7%	28.6%	0.038

#### 4.3.3. Enovelamento da $\beta$ -xiliosidase de *T. reesei*.

A  $\beta$ -D-glucanase de cevada é um membro da família GHF3 e a sua estrutura tridimensional foi determinada por estudos cristalográficos (Figura 4.12.a). Baseados na hipótese de que proteínas que pertencem a uma mesma família das GHs apresentam morfologias parecidas, supõe-se que esta estrutura representa um modelo para todos os membros da família. Usando o programa CRY SOL<sup>22</sup> foi calculada a curva teórica de espalhamento de raios-X a baixos ângulos para a  $\beta$ -D-glucanase de

cevada. Ao contrário do que se esperava, a curva teórica mostrou uma grande diferença com os dados experimentais da  $\beta$ -xilosidase de *T. reesei* ( $\chi = 6.5$ ) (figura 4.12.b), o que indica diferença entre as duas estruturas. A superposição da estrutura cristalográfica da  $\beta$ -D-glucanase de cevada com o envelope da  $\beta$ -xilosidase, feita com o programa SUPCOMB<sup>23</sup>, mostra que o envelope da  $\beta$ -xilosidase de *T. reesei* é muito maior que o modelo da  $\beta$ -D-glucanase de cevada (Figura 4.12.c), sugerindo a presença de um domínio adicional na primeira, que poderia ser ajustado dentro da região assinalada com um círculo na figura 4.12c.



**Figure 4.12.** a. Estrutura cristalográfica da  $\beta$ -D-glucanase de cevada membro da família GHF3. b. Comparação da curva de espalhamento de raios-X a baixos ângulos da  $\beta$ -xilosidase *T. reesei* (pontos) com a curva teórica da  $\beta$ -D-glucanase de cevada (linha sólida) c. Superposição do envelope da  $\beta$ -xilosidase *T. reesei* e a estrutura da  $\beta$ -D-glucanase

Com o intuito de determinar os possíveis domínios da  $\beta$ -xilosidase de *T. reesei* foi feita uma análise da seqüência de aminoácidos desta enzima com a ajuda do servidor <http://supfam.org/SUPERFAMILY>. Assim, em acordo com a classificação desta enzima na família GHF3, foi predito um enovelamento tipo barril TIM para os resíduos do N-terminal (61-413) seguido de um domínio  $\alpha/\beta$  (resíduos 414-643). Os restantes 146 resíduos do C-terminal não apresentaram homologia com nenhum enovelamento descrito até o momento.

Usando o programa JUFO<sup>24</sup> foi determinada a possível estrutura secundária da  $\beta$ -xilosidase de *T. reesei*. Esta foi alinhada com a estrutura da  $\beta$ -D-glucanase de cevada usando o programa CLUSTALW<sup>25</sup>. O alinhamento sugere que a região C-terminal da  $\beta$ -xilosidase pode formar um terceiro domínio (Figura 4.13), devido à presença de um grande *loop* no final do segundo domínio e à seqüência de várias fitas betas que poderiam formar um pequeno  $\beta$ -sandwich, característico dos domínios CBMs presentes nas glicosil hidrolases. Assumindo que de fato existe um terceiro domínio na  $\beta$ -xilosidase de *T. reesei* e que os dois primeiros têm a mesma composição em estrutura secundária da  $\beta$ -D-glucanase, o terceiro domínio deve ser um domínio formado por folhas beta para poder explicar a porcentagem de estrutura beta nos espectros de dicroísmo circular. Enquanto 232 resíduos (29.1%) da  $\beta$ -xilosidase de *T. reesei* adotariam uma conformação tipo hélice e 129 (16.1%) uma tipo beta, na  $\beta$ -D-glucanase 225 resíduos (37.2%) encontram-se em estruturas tipo hélice e somente 64 (10.5%) formam fitas beta. Isto nos faz supor que algumas regiões que foram preditas com estrutura desordenada na  $\beta$ -xilosidase podem ser fitas beta. De fato a presença de um ou mais domínios que ligam carboidratos (CBMs) diferentes ao domínio catalítico nas glicosil hidrolases tem sido reportada<sup>26</sup>. Particularmente vários CBMs que ligam xilano já foram reportados como domínios com enovelamento tipo  $\beta$ -sandwich<sup>27</sup>.

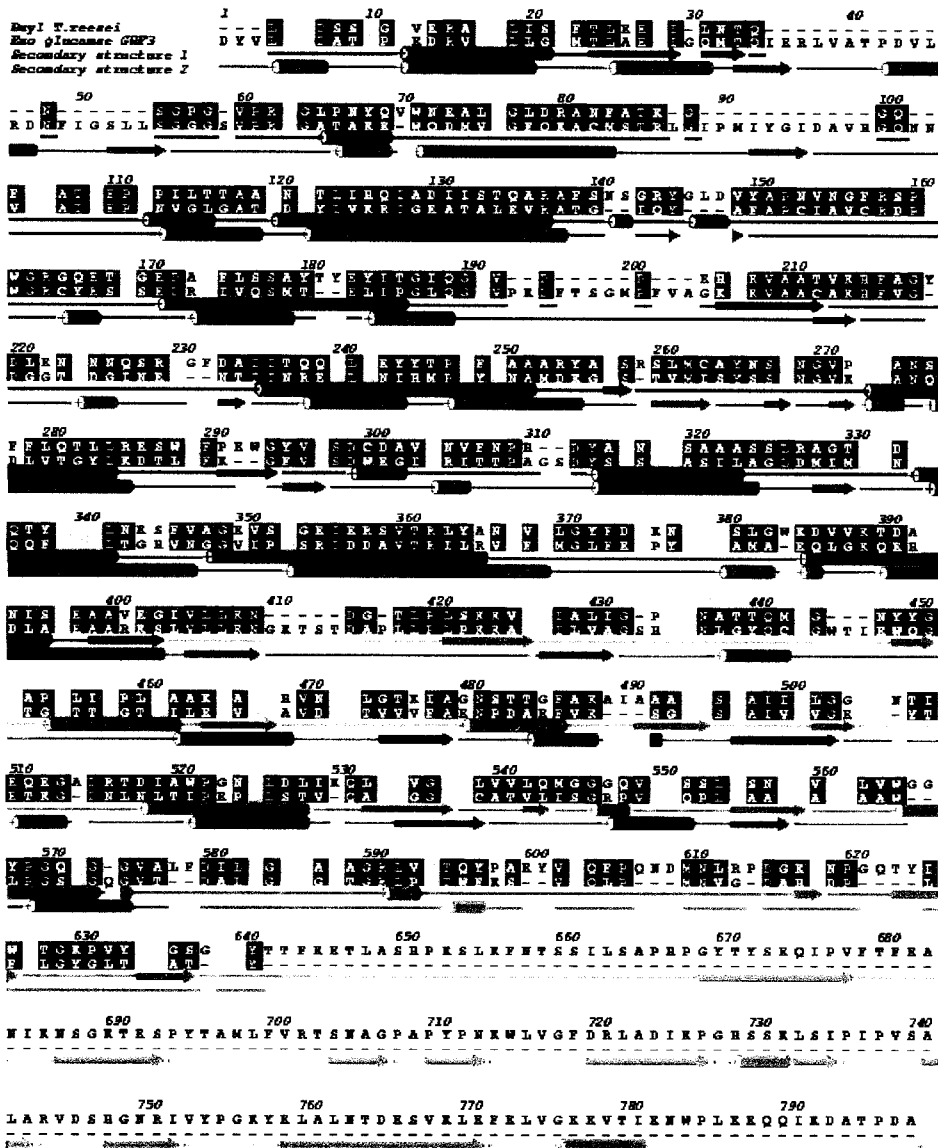


Figure 4.13. Alinhamento da seqüência da  $\beta$ -Xilosidase de *T. reesei* (797aa) e  $\beta$ -D-glucanase de cevada (605aa). A similaridade entra as seqüências foi calculada com ajuda do programa ALSCRIPT<sup>28</sup> e é apresentada em diferentes tons de azul. Azul escuro corresponde a uma alta similaridade e o mais claro a uma baixa. Os domínios 1-3 da  $\beta$ -xilosidase de *T. reesei* são apresentados em vermelho, verde claro e ciano, respectivamente. Os domínios 1-2 da  $\beta$ -D-glucanase são apresentados em laranja e verde, respectivamente.

## 4.5 Conclusões

- O envelope da  $\beta$ -xilosidase de *T. reesei* foi determinado a partir dos dados de espalhamento de raios-X a baixos ângulos da proteína em solução a 20 Å de resolução, mostrando que a enzima é um monômero e que apresenta três domínios. Usando algoritmos de reconhecimento de possíveis domínios foi determinado que esta proteína apresenta, além dos dois domínios característicos da família GHF3, um barril TIM e um domínio  $\alpha/\beta$ , um terceiro domínio. A predição da estrutura secundária e os dados de dicroísmo circular indicam que este terceiro domínio apresentaria um enovelamento formado por folhas beta. Com base nesta evidência é proposto que este terceiro domínio seja um módulo de ligação de carboidratos (CBM) e, portanto participe no reconhecimento do substrato.

## 4.6 Referências

- <sup>1</sup> Surnakki A, Tenkanen M, Buchert J and Viikari L. (1997), Hemicellulases in the bleaching of chemical pulps, *Adv Biochem Eng Biotechnol.*, **57**:261-287.
- <sup>2</sup> Kuhand RC, Singh A and Eriksson KL (1997), Microorganisms and enzymes involved in the degradation of plant fiber cell wall, *Biochem.Eng.Biotechnol.*,**57**:45-125
- <sup>3</sup> Biely P., (1985) Microbial xylanolytic system, *Trends Biotechnol.* **3** 286-290.
- <sup>4</sup> A. M. Golubev, J. R. Brandão Neto, E. V. Eneyskaya, A. V. Kulminskaya, M. A. Kerzhner, K. N. Neustroev and I. Polikarpov (2000) Purification, crystallization and preliminary X-ray study of  $\beta$ -xylosidase from *Trichoderma reesei* *Acta Cryst.* **D56**, 1058-1060.
- <sup>5</sup> Margolles-Clark E., Tenkanen M., Nakari-Setälä T. and Penttilä M., (1996) Cloning of genes encoding  $\alpha$ -L-arabinofuranosidase and  $\beta$ -xilosidase from *Trichoderma reesei* by expression in *Saccharomyces cerevisiae*, *Appl. Environ. Microbiol.* **62**:3840-3846.
- <sup>6</sup> Gómez, M., Isorna, P., Rojo, M. and Estrada, P. (2001), Kinetic mechanism of  $\beta$ -xylosidase from *Trichoderma reesei* QM 9414. *J. Mol. Cat. B: Enzymatic* **16** : 7-15.
- <sup>7</sup> Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) Three dimensional structure of a barley  $\beta$ -D-glucan exohydrolase, a family 3 glycosyl hydrolase, *Structure* **7**, 179-190.
- <sup>8</sup> Manavalan, P. and Johnson, W. C. Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra, *Anal. Biochem.* **167**, 76-85.
- <sup>9</sup> Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set, *Anal. Biochem.* **287**(2), 252-260.
- <sup>10</sup> Provencher, S.W. and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37.
- <sup>11</sup> Andrade, M. A., Chacón, P., Merelo, J. J. and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network, *Prot. Engineering.* **6**, 383-390.
- <sup>12</sup> Sreerema, N. and Woody, R.W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism, *Anal. Biochem.* **209**, 32-44.



- <sup>13</sup> Lobley, A. and Wallace, B.A. (2001) DICHROWEB: A Website for the analysis of protein secondary structure from circular dichroism spectra, *Biophysical J.* **80**, 373a.
- <sup>14</sup> Mao, D., Wachter, E. and Wallace, B.A. (1982) Folding of the H<sup>+</sup>-ATPase Proteolipid in Phospholipid Vesicles, *Biochemistry* **21**, 4960-4968.
- <sup>15</sup> Yang J.Y., Yoon H. J., Ahn H.J., Lee B., Kim H.W., Laivenieks M., Vieille C., Gregory J. Z., and Suh S.W. (2002) Crystallization and preliminary X-ray crystallographic analysis of  $\beta$ -xylosidase from *Thermoanaerobacterium saccharolyticum*, a thermophilic anaerobe *Acta Cryst* **D58**:531-532
- <sup>16</sup> Svergun, D. I., Semenyuk, A. V. and Feigin, L. A. (1988) Small-Angle-Scattering-Data Treatment by the Regularization Method, *Acta Cryst.* **A44**, 244-250.
- <sup>17</sup> Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr.* **25**, 495-503.
- <sup>18</sup> Svergun, D. I., Petoukhov, M. V. and Koch, M. H. (2001). Determination of Domain Structure of Proteins from X-Ray Solution Scattering, *Biophys. J.* **80**, 2946-2953.
- <sup>19</sup> Petoukhov, M. V. and Svergun, D. I. (2003) New methods for domain structure determination of proteins from solution scattering data, *J. Appl. Cryst.* **36**, 540-544.
- <sup>20</sup> Whitmore, L. and Wallace, B.A. (2004) DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Research* **32W**, 668-673.
- <sup>21</sup> Woody, R. W. (1996). Theory of Circular Dichroism of Proteins, in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., Ed), pp. 25-67, Plenum, New York.
- <sup>22</sup> Svergun, D., Barberato, C. and Koch, M. H. J. (1995) CRY SOL: a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *J. Appl. Cryst.* **28**, 768-773.
- <sup>23</sup> Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J. Appl. Crystallogr.* **34**, 33-41.
- <sup>24</sup> Meiler, J., Mueller, M., Zeidler, A., Schmaeschke, F., (2001) Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks, *Journal of Molecular Modelling* **7(9)**, 360-369.
- <sup>25</sup> Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence

---

alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673-4680.

<sup>26</sup> Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781.

<sup>27</sup> Fernandes, A. C., Fontes, C. M. G. A., Gilbert, H. J., Hazlewood, G. P. and Fernandes, T. H. (1999) Homologous xylanases from *Clostridium thermocellum* : evidence for bifunctional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes, *Biochem. J.* **342**, 105-110.

<sup>28</sup> Barton G. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

# Anexo 3

Adriana L. Rojas, Hannes Fischer, Elena V. Eneiskaya, Anna A. Kulminskaya, Konstantin A. Shabalin, Kirill N. Neustroev, Aldo F. Craievich, Alexander M. Golubev e Igor Polikarpov.(2005). Structural insights into the beta-Xylosidase from *Trichoderma reesei* obtained by Synchrotron Small-Angle X-ray Scattering and Circular Dichroism Spectroscopy. *Biochemistry*, *aceito*.

Structural insights into the  $\beta$ -Xylosidase from  
*Trichoderma reesei* obtained by Synchrotron Small-  
Angle X-ray Scattering and Circular Dichroism  
Spectroscopy.

*Adriana L. Rojas*<sup>1</sup>, *Hannes Fischer*<sup>1</sup>, *Elena V. Eneiskaya*<sup>2</sup>, *Anna A. Kulminskaya*<sup>2</sup>, *Konstantin A. Shabalin*<sup>2</sup>,  
*Kirill N. Neustroev*<sup>2</sup>, *Aldo F. Craievich*<sup>3</sup>, *Alexander M. Golubev*<sup>2</sup> and *Igor Polikarpov*.<sup>1\*</sup>

<sup>1</sup>Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, CEP 13560-970, São Carlos, SP, Brazil.

<sup>2</sup>Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg, 188300, Rússia.

<sup>3</sup>Instituto de Física, Universidade de São Paulo, São Paulo, SP, Brazil.

**Running title:** Molecular shape and domain organization of *Trichoderma reesei*  $\beta$ -Xylosidase

\*To whom correspondence should be addressed. Email: [ipolikarpov@ifsc.usp.br](mailto:ipolikarpov@ifsc.usp.br), Phone: +55(16) 3373-9874, fax: +55(16) 3373-9881

## SUMMARY

The enzyme  $\beta$ -xylosidase from *Trichoderma reesei*, a member of glycosyl hydrolase family 3 (GHF-3) is a glycoside hydrolase which acts at the glycosidic linkages of 1,4- $\beta$ - xylooligosaccharides and that also exhibits  $\alpha$ -L-arabinofuranosidase activity on 4-nitrophenyl  $\alpha$ -L-arabinofuranoside. In this work, we show that the enzyme forms monomers in solution and derive the low-resolution molecular envelope of the  $\beta$ -xylosidase from small-angle X-ray scattering (SAXS) data using *ab initio* simulated annealing algorithm. The radius of gyration and the maximum dimension of the  $\beta$ -xylosidase are 36.9 Å and 90 Å, respectively. In contrast to the fold of the only structurally characterized member of GHF-3, the  $\beta$ -D-glucan exohydrolase, which has two distinct domains, the shape of the  $\beta$ -xylosidase indicates the presence of three domains located in the same plane. Domain recognition algorithms were used to show that the C-terminal part of the amino acid sequence of the protein forms the third domain. Circular dichroism spectroscopy and secondary structure prediction programs demonstrate that this additional domain adopts predominantly  $\beta$ -conformation.

**Keywords:**  $\beta$ -Xylosidase; *Trichoderma reesei*; glycosyl hydrolase; small-angle X-ray scattering; circular dichroism.

## 1. INTRODUCTION

*Trichoderma reesei*  $\beta$ -xylosidase is a glycoside hydrolase that acts toward unbranched xylans, glucuronoxylans and  $\beta$ -1,4- xylooligosaccharides with different degrees of polymerization. The only resulting product of all these reactions is D-xylose, which characterizes *T. reesei*  $\beta$ -xylosidase as an exoglycosidase (1). Xylan is a major structural polysaccharide in plant cells, and is the second most abundant polysaccharide in nature, accounting for approximately one-third of all renewable organic carbon on earth (2). Xylan together with cellulose (1,4- $\beta$ -glucan) and lignin (a complex polyphenolic compound) make up the major polymeric constituents of plant cell walls (3). Recently, there was a significant industrial interest in xylan and its hydrolytic enzymatic complex, as a supplement in animal feed, for the manufacture of bread, food and drinks, textiles, bleaching of cellulose pulp, ethanol and xylitol production (4).

Xylan degradation is a multistep process involving a number of enzymatic activities (5). Xylanases (1,4- $\beta$ -D-xylan xylanohydrolases; E.C. 3.2.1.8) hydrolyze the internal  $\beta$ -1,4-xylosidic linkages of the xylan backbone. Xylosidases (1,4- $\beta$ -D-xylan xylohydrolases; E.C. 3.2.1.37) in turn catalyze hydrolysis of 1,4- $\beta$ -D-xylans, removing successive D-xylose residues from the non-reducing termini, whereas arabinofuranosidases ( $\beta$ -L-arabinofuranoside arabinofuranohydrolase; E.C. 3.2.1.55) hydrolyze terminal non-reducing  $\beta$ -L-arabinofuranoside residues in  $\beta$ -L-arabinosides.

$\beta$ -Xylosidases are produced by fungi, bacteria, yeast, marine algae, protozoans, snails, crustaceans, insect and seeds, but the principal commercial source are filamentous fungi.  $\beta$ -Xylosidases from various organisms have been grouped into glycoside hydrolase (GH) families 3, 39, 43, 52, and 54 in the CAZY database (<http://afmb.cnrs-mrs.fr/CAZY/>) (6).  $\beta$ -Xylosidase from *T. reesei* belongs to GHF 3. There are more than 180 known members of this family, most of which are classified as  $\beta$ -D-glucosidases,  $\beta$ -D-xylosidases, or N-acetyl  $\beta$ -D-glucosaminidases.

Barley  $\beta$ -D-glucan exohydrolase is one of the best-characterized member of GHF3, both in functional and structural terms, and the only representative member of the family, which crystal structure has been determined at 2.2 Å resolution (7). It folds into two distinct domains. The first domain consists of an  $(\alpha/\beta)_8$  TIM-barrel domain, whereas the second domain consists in a six-stranded  $\beta$  sandwich, which contains a  $\beta$  sheet of five parallel  $\beta$  strands and one antiparallel  $\beta$  strand, with three  $\alpha$  helices on either side of the sheet. Although members of each glycoside hydrolase family have similar three-dimensional (3D) conformations, comparison of the small angle X-ray scattering (SAXS) data of the *T. reesei*  $\beta$ -xylosidase with the theoretical scattering curve for the  $\beta$ -D-glucan exohydrolase shows poor agreement, indicating gross differences between the two structures. A search for possible domains of the protein was performed using the Superfamily HMM library and genome assignments server (8) and the primary structure of *T. reesei*  $\beta$ -xylosidase. Thus two known domains, characteristic for family GHF3, comprising the first 643 residues of *T. reesei*  $\beta$ -xylosidase could be identified, but the last 146 residues of the C-terminus form additional, third, domain of an unknown fold. Circular dichroism (CD) spectroscopy and secondary structure prediction algorithms indicate that the third domain has predominantly  $\beta$  conformation. Docking of the  $\beta$ -D-glucan exohydrolase structure into molecular envelope of the *T. reesei*  $\beta$ -xylosidase demonstrate that all three domains are positioned in the same plane forming heart-shaped structure.

## 2. RESULTS AND DISCUSSION

### 2.1. Shape of *T. reesei* $\beta$ -xylosidase

To obtain information about *T. reesei*  $\beta$ -xylosidase tertiary structure and the molecular shape we submitted this protein to SAXS analysis. The X-ray scattering curve of  $\beta$ -xylosidase obtained as described in Materials and Methods is shown in Figure 1. The distance distribution functions,  $p(r)$ , evaluated by the indirect Fourier transform with the program GNOM (9,10), yielded the radius of gyration and the maximum dimension of the  $\beta$ -xylosidase of 36.9 Å and 90 Å respectively (Figure 2). The shape of the protein was determined from the X-ray scattering data with the program GASBOR (11,12), using a real-space algorithm. The results of ten independent simulated annealing computations were compared with each other; the best model is given in Figure 3. Molecular envelope of the protein is sculpted in the form of a heart with the longitudinal dimension of about 90Å and its transversal dimension of approximately 55Å.

### 2.2. Secondary structure of *T. reesei* $\beta$ -xylosidase

In search for secondary structure information of *T. reesei*  $\beta$ -xylosidase, we collected its circular dichroism spectroscopy (CD) data. The CD spectrum of the protein in solution obtained as described in Materials and Methods is shown in Figure 4a. The secondary structure content was estimated using the programs K2D (13), SELCON3 (14), CONTINLL (15) and CDSSTR (16,17). The best fit between experimental data and reconstructed data was found using CDSSTR (Figure 4b). However it is reported that CDSSTR tends to give the lowest NRMSD values, although this does not necessarily mean that those results are the most correct (18). For this reason the fit to experimental data obtained with CONTINLL is also given in Figure 4c.

The spectra are strongly characteristic of an  $\alpha/\beta$  protein, with spectral bands attributable to electron



transitions in the amide groups of the protein backbone. The carbonyl oxygen lone-pair rotational transition  $n\pi^*$  gives rise to the minimum at ~222 nm, indicative of a right-handed  $\alpha$ -helix (19). The local minimum at ~210 is an average between the minimum at ~208 nm consequential from  $\pi\pi^*$  polarized parallel to the helical axis and the minimum at ~215 nm provide from  $n\pi^*$  transition in  $\beta$ -strands. This is indicative of  $\alpha/\beta$  structure of the protein. The secondary structure content estimated with both programs is given in Table 1.

### 2.3. The putative fold of $\beta$ -xylosidase from *T. reesei*.

Barley  $\beta$ -D-glucan exohydrolase is the only member of GHF3 with known three-dimensional structure (Figure 5a). It is tempting, therefore, to use this model as a template for *T. reesei*  $\beta$ -xylosidase structural model. Following this reasoning we calculated theoretical scattering curve for the  $\beta$ -D-glucan exohydrolase using the program CRY SOL (20). To our surprise comparison of this theoretical curve with the experimental SAXS data of the *T. reesei*  $\beta$ -xylosidase gives  $\chi = 6.5$  (Figure 5b), which clearly indicate gross difference between the two structures. Superposition of the crystallographic structure of  $\beta$ -D-glucan exohydrolase with  $\beta$ -xylosidase free atom model, performed by the program SUPCOMB (21) (Figure 5c), shows that the  $\beta$ -xylosidase shape is significantly larger than the  $\beta$ -D-glucan exohydrolase, indicating a presence of additional domain in  $\beta$ -xylosidase structure that could fill the region inside the circle in Figure 5c. In order to determine the amino acid sequence extension and organization of possible domains, we conducted analysis of the  $\beta$ -xylosidase primary structure using the server <http://supfam.org/SUPERFAMILY>. As a result N-terminal part of the protein (residues 61-413) was predicted to fold as a TIM barrel domain followed by an  $\alpha/\beta$  domain (residues 414-643), in agreement with the classification of this enzyme as a member of GHF3. The last 146 residues remained unclassified.

2D sequence alignment of *T. reesei*  $\beta$ -xylosidase and barley  $\beta$ -D-glucan exohydrolase, another

member of CHF3 also indicate that C-terminal region of  $\beta$ -xylosidase could form the third domain (Figure 6). This is not surprising since some fungal  $\beta$ -xylosidases from other families have already been reported to possess three structural domains (22). Furthermore, the secondary prediction performed using *JUFO: Secondary Structure Prediction for Proteins* server (23) show that the third domain should be in the most part composed of  $\beta$ -strands. This notion was confirmed by CD analysis of *T. reesei*  $\beta$ -xylosidase. While in total 232 amino acid residues of *T. reesei*  $\beta$ -xylosidase adopts helical conformation and 129 form  $\beta$ -structures. 225 residues form helices and 64 form  $\beta$ -strands in  $\beta$ -D-glucan exohydrolase X-ray structure. Assuming an existence of the additional structural domain in the former structure and taking into account that the first two domains of these proteins have the same fold and similar secondary structure composition, the third domain of *T. reesei*  $\beta$ -xylosidase should be a  $\beta$ -fold domain. This fact is in accord with the general observation that GHs consist of a catalytic domain joined by flexible linker sequences to one or more carbohydrate binding modules (CBM). The function of the CBMs is largely to attach the enzyme to its substrate, and therefore enhance the rate of catalysis by increasing the probability of enzyme/substrate interaction (24). This observation opens the possibility that the two auxiliary domains of *T. reesei*  $\beta$ -xylosidase play a role in recognizing the substrate xylo-oligosaccharides. Moreover, in line with our studies, a number of xylan-binding domains found in xylanases adopt  $\beta$ - (predominantly  $\beta$ -sandwich) folds (25).

### 3. MATERIALS AND METHODS

#### 3.1. Protein purification, Sample preparation and SAXS data collection

The  $\beta$ -xylosidase was purified from a culture of *T. reesei* after 48-60 h of cultivation as previously described (26,27). The resulting purified enzyme was dialyzed against water and lyophilized. This lyophilized enzyme was dissolved in 90 mM Tris buffer at pH 7 at a concentration of 7.3 mg/mL. The sample was centrifuged for 3 minutes at 20000g.

The small-angle X-ray scattering experiments were carried out at LNLS, Campinas, Brazil, using a

synchrotron SAXS beamline (Kellermann *et al.*, 1997). The wavelength of the income monochromatic X-ray beam was  $\lambda=1.608 \text{ \AA}$ . A 1D X-ray position sensitive detector (PSD) was utilized to record the scattering intensity as a function of the modulus of the scattering vector  $q = (4\pi/\lambda)\sin\theta$ ,  $\theta$  being half the scattering angle. The parasitic scattering from air and beamline windows were subtracted from the total measured intensities. Desmearing of the experimental results was performed since the entrance window of the 1D PSD was 8mm height.

The sample-to-detector distance (702 mm) was adjusted in order to record the scattering intensity for  $q$  values ranging from 0.01 to  $0.25 \text{ \AA}^{-1}$ . The parasitic scattering intensity produced by air and slits was subtracted from the total measured SAXS intensity. The samples were encapsulated inside a cell with two thin parallel mica windows. Data analysis was performed using the GNOM program package (9,10). Model simulations were done using the program GASBOR (11,12). At least ten independent reconstructions starting from random approximations yielded reproducible results.

The superposition of crystallographic structures and free atom models was done using the program SUPCOMB (21). To keep the original low-resolution limit of the SAXS data, molecular envelopes were computed from coordinate files output by GASBOR (11,12).

### 3.2. Circular dichroism

Circular dichroism spectra were acquired in-house on a J715 spectropolarimeter (Jasco Corp., Japan). The protein at a concentration of 0.1 mg/mL was placed in a 10-mm path-length cuvette. Spectra were obtained in 10 mM Tris buffer at pH 7, and spectra for the buffer were subtracted from those for the *T. reesei*  $\beta$ -xylosidase.

Secondary structure content was analyzed using the programs CDSSTR (16,17) and CONTINLL (15). Data were submitted online to the DichroWeb website (<http://www.cryst.bbk.ac.uk/cdweb/html/home.html>) (28) via the online submission procedure. Data over the range 190-240 nm was used. The NRMSD (29) of Fit parameter was found to be 0.013 with

CDSSTR and 0.038 with CONTIN in both approaches were used the reference database number 7.

## CONCLUSIONS

The shape of  $\beta$ -xylosidase in solution was restored at 20 Å resolution from synchrotron x-ray scattering data. It demonstrates that the enzyme forms monomers in solution and indicates that it is composed of three distinct domains. Fold recognition algorithms and 2D sequence alignment show that the first two domains fold, respectively, as a TIM barrel and an  $\alpha/\beta$  domain and, furthermore, confirm the existence of the third, N-terminal, domain. CD data and the secondary structure prediction analysis provide evidence that this putative domain should have predominantly  $\beta$ -fold. Based on this evidence, we propose that first two domains of *T. reesei*  $\beta$ -xylosidase have a molecular fold similar to that of barley  $\beta$ -D-glucan exohydrolase, another member of GHF 3, and, in addition, this monomeric enzyme has additional, predominantly  $\beta$  domain, that could possibly function as a carbohydrate binding module involved in the recognizing of the substrate.

**ACKNOWLEDGMENT.** This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, via grants 01/07014-0 and 99/03387-4, and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and grant 05-04-49216 from Russian Federation for Basic Research, Russia.

## FIGURE CAPTIONS

**Figure 1.** X-ray small angle scattering curve of the  $\beta$ -xylosidase in 20 mM TRIS at pH 6 buffer superimposed with the theoretical scattering curve from the restored low-resolution model.

**Figure 2.** The distance distribution functions,  $p(r)$ , of *T. reesei*  $\beta$ -xylosidase computed by the indirect

Fourier transform with the program GNOM (9,10).

**Figure 3.** The shape of  $\beta$ -Xylosidase in solution at 20 Å resolution was restored from synchrotron x-ray scattering data using an *ab initio* simulated annealing algorithm implemented in program GNOM (9,10). Only the best model (normalized spatial discrepancy (NSD) is equal to 1.27, Chi against experimental data equal to 1.42) is shown. Three different stereoviews of the protein shape are given. The middle and bottom rows are rotated counterclockwise by 90° around Y and Z axes, respectively.

**Figure 4.** a. Circular dichroism spectra of *T. reesei*  $\beta$ -xylosidase. b. Fit to experimental data obtained with program CONTINLL (15) c. Fit obtained with program CDSSTR (16,17).

**Figure 5.** a. Crystallographic structure of barley  $\beta$ -D-glucan exohydrolase as a template for GHF3. b. Comparison of the SAXS data of the *T. reesei*  $\beta$ -xylosidase (dots) with the theoretical scattering curve of the template for GHF3 (solid line) shows poor agreement ( $\chi = 6.5$ ), indicating gross differences between these two structures c. Superposition of the *T. reesei*  $\beta$ -xylosidase molecular envelope with 3D structure of barley  $\beta$ -D-glucan exohydrolase indicates that  $\beta$ -xylosidase could have additional. third. domain.

**Figure 6.** Amino acid sequence alignment of the *T. reesei*  $\beta$ -xylosidase (797aa) and barley  $\beta$ -D-glucan exohydrolase (605aa). The amino acid similarity between primary structures as calculated by the program ALSCRIPT is shown in three different shades of grey. The darker color represents higher sequence similarity. Secondary structure elements of the *T. reesei*  $\beta$ -xylosidase domains 1–3, predicted by the JUFO server (23), are depicted in black, grey and light grey, respectively.

Figure 1.

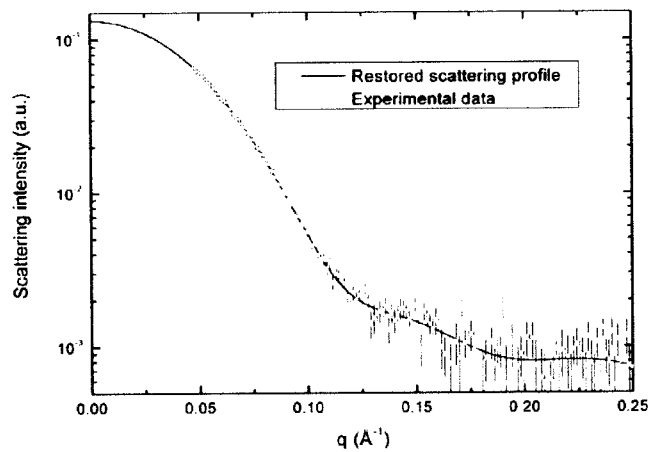


Figure 2.

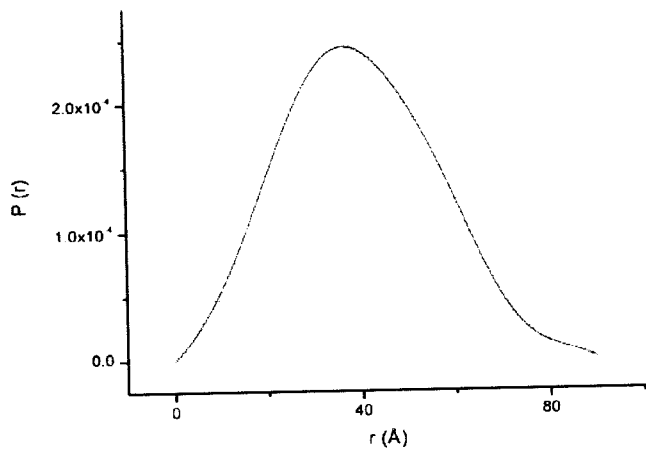
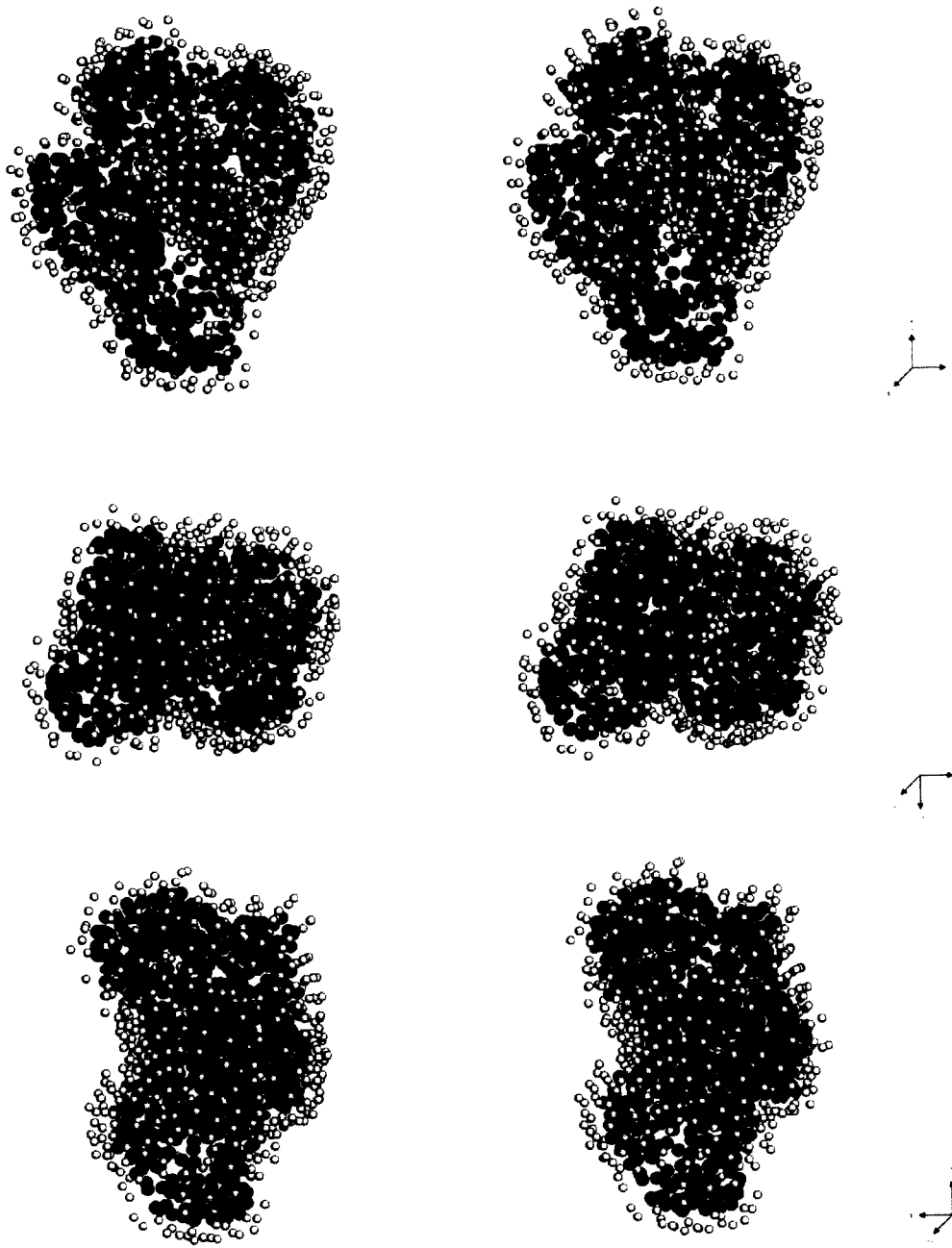
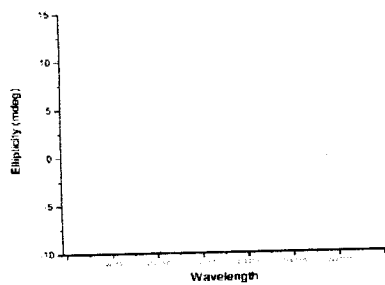


Figure 3.

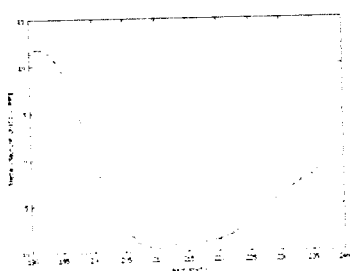




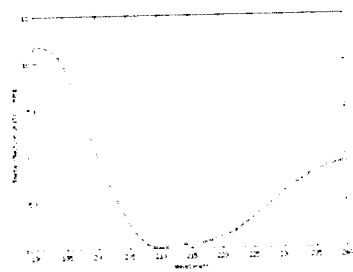
**Figure 4.**



**a. CD spectra**



**b. CONTINLL**

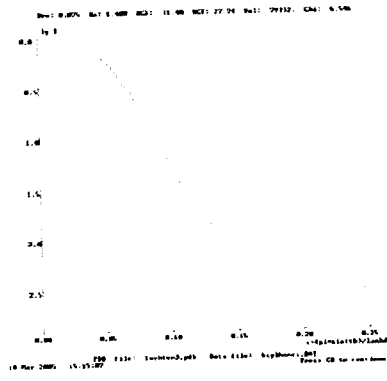


**c. CDSSTR**

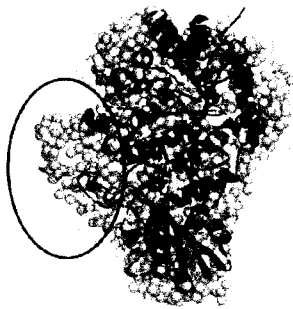
Figure 5.



a.



b.



c.

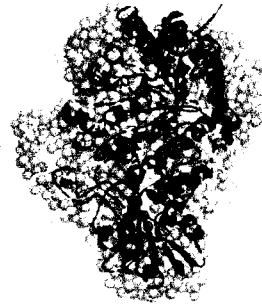
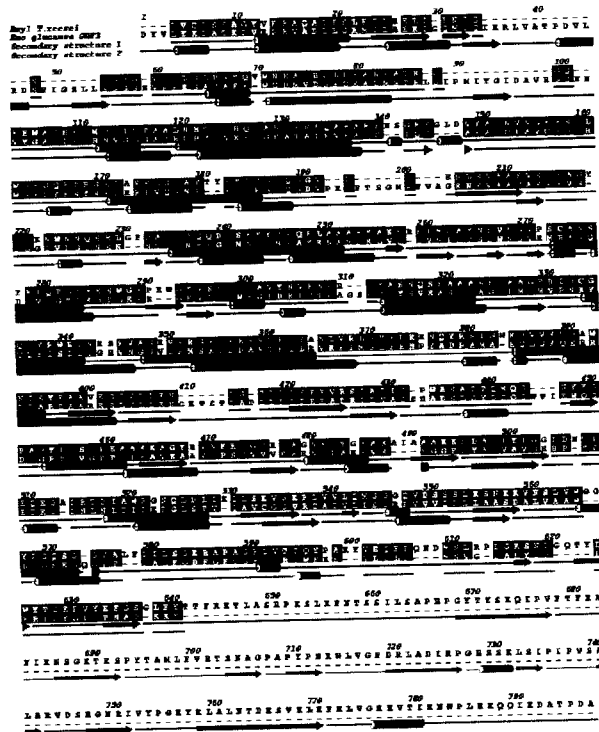


Figure 6.



**TABLES.****Table1.** Secondary structural content of the *T. reesei*  $\beta$ -xylosidase evaluated from CD spectra using CONTINLL and CDSSTR.

<b>Program</b>	<b>Helix</b>	<b>Beta</b>	<b>Turn</b>	<b>Random</b>	<b>NMRSD</b>
<b>CDSSTR</b>	23.0%	27.0%	20.0%	30.0%	0.013
<b>CONTINLL</b>	24.1%	25.5%	21.7%	28.6%	0.038

## REFERENCES

1. Herrmann, M. C., Vrsanska, M., Jurickova, M., Hirsch, J., Bierly, P. and Kubicek, C. P. (1997) The beta-D-xylosidase of *Trichoderma reesei* is a multifunctional beta-D-xylan xylohydrolase, *Biochem. J.* **321**, 375-381.
2. Prade, R. A. (1995) Xylanases: from biology to biotechnology, *Biotech. Genet. Eng. Rev.* **13**, 100-131.
3. Kulkarni, N., Shendye, A. and Rao, M. (1999) Molecular and biotechnological aspects of xylanases, *FEMS Microbiol. Rev.* **23**, 411-456.
4. Joseleau, J. P., Comtat, J. and Ruel, K. (1992) Chemical structure of xylans and their interaction in the plant cell walls, *In Xylans and Xylanases*, (Visser, J., Beldman, G., Someren, M. A. K. v. AND Voragen, A. G.J., eds), pp. 1-15, Elsevier, Amsterdam.
5. Dekker, R. F. H. and Richards, G. N. (1976) Hemicellulases: their occurrence, purification, properties, and mode of action, *Adv. Carbohydr. Chem. Biochem.* **32**, 277-352.
6. Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach, *In Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G., Henrissat, B. and Svensson, B., eds), pp. 3-12, The Royal Society of Chemistry, Cambridge.
7. Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) Three dimensional structure of a barley  $\beta$ -D-glucan exohydrolase, a family 3 glycosyl hydrolase, *Structure* **7**, 179-190.
8. Gough, J., Karplus, K., Hughey, R. and Chothia, C. 2001 Assignment of Homology to Genome Sequences using a Library of Hidden Markov Models that Represent all Proteins of Known Structure, *J. Mol. Biol.* **313(4)**, 903-919.
9. Svergun, D. I., Semenyuk, A. V. and Feigin, L. A. (1988) Small-Angle-Scattering-Data Treatment by

the Regularization Method, *Acta Cryst. A44*, 244-250.

10. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr.* 25, 495–503.
11. Svergun, D. I., Petoukhov, M. V. and Koch, M. H. (2001). Determination of Domain Structure of Proteins from X-Ray Solution Scattering, *Biophys. J.* 80, 2946–2953.
12. Petoukhov, M. V. and Svergun, D. I. (2003) New methods for domain structure determination of proteins from solution scattering data, *J. Appl. Cryst.* 36, 540–544.
13. Andrade, M. A., Chacón, P., Merelo, J. J. and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network, *Prot. Engineering.* 6, 383-390.
14. Sreerema, N. and Woody, R.W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism, *Anal. Biochem.* 209, 32-44.
15. Provencher, S.W. and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20, 33-37.
16. Manavalan, P. and Johnson, W. C. Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra, *Anal. Biochem.* 167, 76-85.
17. Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set, *Anal. Biochem.* 287(2), 252-260.
18. Whitmore, L. and Wallace, B.A. (2004) DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Research* 32W, 668-673.
19. Woody, R. W. (1996). Theory of Circular Dichroism of Proteins, in *Circular Dichroism and the*

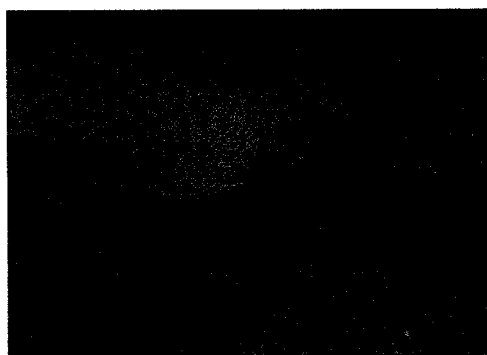
20. Svergun, D., Barberato, C. and Koch, M. H. J. (1995) CRY SOL: a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *J. Appl. Cryst.* 28, 768–773.
21. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J. Appl. Crystallogr.* 34, 33–41.
22. Sunna, A. and Antranikian, G. (1997). Xylanolytic enzymes from fungi and bacteria. *Crit. Rev. Biotechnol.* 17(1):39–67.
23. Meiler, J., Mueller, M., Zeidler, A., Schmaeschke, F., (2001) Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks, *Journal of Molecular Modelling* 7(9), 360-369.
24. Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S. J., Simpson, P. J., Williamson, M. P., Hazlewood, G. P., and Gilbert, H. J. (1999) The type II and X cellulose-binding domains of *Pseudomonas xylanase A* potentiate catalytic activity against complex substrates by a common mechanism, *Biochem. J.* 342, 473–480.
25. Fernandes, A. C., Fontes, C. M. G. A., Gilbert, H. J., Hazlewood, G. P. and Fernandes, T. H. (1999) Homologous xylanases from *Clostridium thermocellum* : evidence for bifunctional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes, *Biochem. J.* 342, 105–110.
26. Savel'ev, A. N., Eneyskaya, E. V., Isaeva-Ivanova, L. S., Shabalin, K. A., Golubev, A. M. and Neustroev, K. N. (1997). The carbohydrate moiety of  $\alpha$ -galactosidase from *Trichoderma reesei*. *Glycoconjugate J.* 14, 897–905.

27. Golubev, A.M., Brandão-Neto, J. R., Eneyskaya, E. V., Kulminkaya, A. V., Kerzhner, A. V., Neustroev, K. N., and Polikarpov, I. (2000) Purification, crystallization and preliminary X-ray study of  $\beta$ -xylosidase from *Trichoderma reesei*, *Acta Cryst. D56*, 1058-1060.
28. Lobley, A. and Wallace, B.A. (2001) DICHROWEB: A Website for the analysis of protein secondary structure from circular dichroism spectra, *Biophysical J.* 80, 373a.
29. Mao, D., Wachter, E. and Wallace, B.A. (1982) Folding of the H<sup>+</sup>-ATPase Proteolipid in Phospholipid Vesicles, *Biochemistry* 21, 4960-4968.



# **CAPÍTULO 5.**

## **Laminarinases**



As beta-glucanas são polissacarídeos lineares, não ramificados, compostos por unidades de beta-D-glicopiranosil com ligações 1,4 (aprox. 70%) e 1,3 (aprox. 30%), sendo componentes estruturais das paredes celulares de grãos, principalmente aveia e cevada. Essas enzimas se destacam por suas propriedades funcionais, como a capacidade de formar soluções viscosas e géis em presença de água e aquecimento, apresentando-se estáveis na presença de sais e açúcares, podendo assim serem utilizadas como espessantes e/ou viscosantes em alimentos.

Para a indústria cervejeira, as beta-glucanas da cevada são consideradas um fator negativo, de maneira que a sua degradação é de grande importância na produção de malte e obtenção de cerveja de qualidade. Se as beta-glucanas não forem degradadas totalmente, impedem a passagem das outras enzimas dos grãos, produzindo uma modificação irregular do endosperma. Isto resultaria em baixo rendimento do mosto e possível turvação da cerveja, assim como em um processamento menos eficiente devido à obstrução dos filtros.

As  $\beta$ -1,3 glucanases estão amplamente distribuídas entre bactérias, fungos e plantas superiores, sendo classificadas como exo- $\beta$ -1,3 glucanases (EC 3.2.1.58) e endo- $\beta$ -1,3 glucanases (EC 3.2.1.6 e EC 3.2.1.39). Diferentes papéis fisiológicos já foram propostos para estas enzimas. Em plantas, acredita-se que estejam envolvidas na diferenciação celular e na defesa contra fungos patógenos<sup>1</sup>. Em fungos, as  $\beta$ -1,3 glucanases parecem ter diferentes funções, estando envolvidas em processos morfológicos, na mobilização de  $\beta$ -glucano no processo de reconhecimento patógeno-planta<sup>2</sup>. Também foi descrita uma  $\beta$ -1,3 glucanase que pode estar envolvida no começo da embriogênese<sup>3</sup>. Em bactérias, funções metabólicas já foram reportadas para endo- $\beta$ -1,3 glucanases e endo- $\beta$ -1,3 -1,4-glucanases<sup>4</sup> tendo estes dois tipos de enzimas bacterianas especificidades muito próximas.<sup>5</sup>

### 5.1. Endo-glucanase de *Rhodothermus marinus*

*Rhodothermus marinus* é uma bactéria halofílica isolada das águas termais marinhas finlandesas<sup>6</sup>. Esta bactéria produz algumas enzimas que degradam

polissacarídeos tais como xilanases<sup>7,8</sup>, celulasas<sup>9</sup> e  $\beta$ -glucanases que agem em lichenina e laminarina<sup>10,11</sup>. A laminarinase de *Rhodotermus marinus* foi caracterizada por Krahl e colaboradores<sup>11</sup> como uma endo-glucanase monomérica de 32 kDa que hidrolisa ligações  $\beta(1-3)$ - $\beta(1-4)$  glucanas. Esta enzima faz a hidrólise com retenção da configuração anomérica possuindo homologia com as glucanases da família GH16, que inclui 1,3-1,4 $\beta$  glucanases (lichenases) que hidrolisam especificamente ligações 1,4 $\beta$  na posição do oxigênio da glucose<sup>12,13</sup>.

Estudos de dicroísmo circular revelaram que sua estrutura tridimensional está composta principalmente por folhas  $\beta$ <sup>22</sup>. Análises utilizando oligossacarídeos como substrato indicaram a presença de dois subsítios (+1 e +2) para a ligação do substrato na região reduzida e três ou quatro subsítios (21-24) na região não reduzida (nomenclatura em acordo com Davies e colaboradores<sup>25</sup>).

A substituição dos resíduos Glu129, Asp131 e Gln134, altamente conservados na família 16 das GHs, acarretaram uma drástica redução na hidrólise de  $\beta$ -glucanos, sugerindo a sua importância na hidrólise enzimática.

No presente trabalho obtiveram-se cristais nativos da laminarinase de *Rhodotermus marinus* que difrataram até 2.1Å de resolução, não tendo sido possível, porém, a resolução da estrutura.

### **5.1.1. Metodologia**

#### **5.1.1.1. Cristalização**

A laminarinase de *Rhodotermus marinus* foi cristalizada pelo método da gota pendurada. Para determinar as condições iniciais dos ensaios de cristalização foi feito um teste de precipitação, conforme sugerido pelo Dr Alexander Golubev. Este teste consistiu em misturar PEG 3350 à proteína em determinadas concentrações e verificar a presença de precipitado na solução, após 30min - 1 hora. A proteína purificada e liofilizada pelo grupo do Dr Golubev foi cedida para os ensaios de cristalização. O liofilizado foi resuspendido em uma solução 5mM de fosfato de potássio pH 6.0 na concentração de 8 mg/ml. Gotas de 10 $\mu$ l contendo a solução de proteína e uma solução tampão com PEG foram dispostas em lamínulas sobre um recipiente fechado para minimizar a evaporação durante o tempo de espera. Assim, a solução de proteína foi misturada com uma solução de 40% PEG 3350, 10mM fosfato

de potássio pH 5.0. As diferentes proporções entre as soluções de proteína e tampão utilizadas neste teste estão apresentadas na tabela 5.1.

**Tabela 5.1.** Concentrações de PEG e de proteína usados no teste de precipitação para a laminarinase de *R. marinus*.

Volume (µl)		Concentração inicial	
tampão	proteína	PEG 3350 (%)	Proteína (mg/ml)
1	4	8	6.4
2	3	16	4.8
3	2	24	3.2
4	1	32	2.0

Após 40 minutos as gotas com concentrações iniciais de PEG maiores de 24% apresentaram precipitados, enquanto as gotas com concentrações menores ainda estavam claras. Este valor foi usado nos primeiros ensaios de cristalização, como limite máximo da concentração inicial de PEG. Assim, foi desenhado um *screening* variando a molaridade do fosfato de potássio e pH da solução, mantendo-se a concentração final do PEG fixa em 22%. O objetivo deste teste é facilitar a escolha de um caminho a seguir nos primeiros ensaios de cristalização.

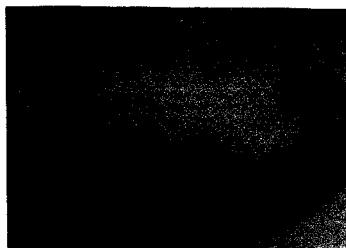
### 5.1.1.2 Coleta e processamento de dados

Foram coletados vários conjuntos de dados no Laboratório de Cristalografia de Proteínas do Instituto de Física de São Carlos, contudo devido ao tamanho dos cristais, o valor médio da razão  $I/\sigma$  para as reflexões medidas estava abaixo de 2 nas últimas faixas de resolução (3.0-2.8Å). Como o padrão de difração apresentava um bom perfil, tentamos melhorar as medidas coletando novos conjuntos de dados no síncroton em Campinas. Dois conjuntos foram coletados (um nativo e um derivado) na linha de Cristalografia de Proteínas do LNLS (Laboratório Nacional de Luz Síncroton)<sup>26</sup>, usando um detector tipo Marccd. Os cristais foram congelados com nitrogênio líquido depois de serem mergulhados em soluções crioprotetoras. O derivado de césio foi preparado com o método do *quick cryo-soaking*<sup>27,28</sup>. As imagens de difração foram processadas com os programas MOSFLM<sup>29</sup> e SCALA.

## 5.1.2. Resultados e discussão

### 5.1.2.1 Cristalização

Os ensaios de cristalização foram feitos como descrito na seção 5.1.1.1. Cristais em forma de agulhas apareceram em gotas que continham 5 µl da solução da proteína e 5 µl de um tampão 10mM fosfato de potássio pH 5.0 com 6% PEG 3350. Estas gotas foram equilibradas contra 800µl de uma solução 25% PEG 3350 50mM de fosfato de potássio pH 6.0 a 18°C. Os cristais aparecem depois de 1 semana (figura 5.1)



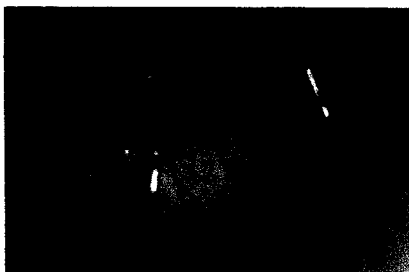
**Figura 5.1.** Agulhas de laminarinase de *Rhodotermus marinus*. Gota: 6% PEG 3350 em 10mM fosfato de potássio a pH 5.0. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0

A partir destas condições, novos ensaios foram feitos variando o pH e a concentração do fosfato de potássio (tabela 5.2). Observou-se que o aumento de pH resultou no crescimento de cristais mais separados e as variações na concentração do fosfato influenciaram o tamanho dos mesmos.

**Tabela 5.2.** Ensaios de cristalização da laminarinase de *Rhodotermus marinus*

Tampão	Agente precipitante	pH	Concentração da proteína
KH <sub>2</sub> PO <sub>4</sub> 5mM	PEG 3350 6%	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 10mM	PEG 3350 6%	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 25mM	PEG 3350 6 %	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 50mM	PEG 3350 6 %	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 60mM	PEG 3350 6 %	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 70mM	PEG 3350 6 %	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 80mM	PEG 3350 6 %	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml
KH <sub>2</sub> PO <sub>4</sub> 90mM	PEG 3350 6 %	5.0, 5.5, 6.0, 6.5, 7.0	8mg/ml

Os melhores resultados nesta etapa foram obtidos para pH 6.0- 6.5 e 90 mM de fosfato de potássio. Embora houvesse a presença constante de cristais pequenos, foi possível encontrar alguns de maior tamanho (figura 5.2).

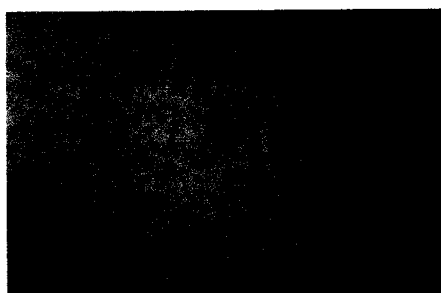


**Figura 5.2.** Cristais da laminarinase de *Rhodotermus marinus*. Gota: 6% PEG 3350 em 90mM fosfato de potássio a pH 6.5. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0

Tentativas para aumentar o tamanho dos cristais usando aditivos (tabela 5.3), assim como *macroseeding* não foram bem sucedidas. Só obtivemos cristais um pouco maiores na condição com glucose (figura 5.3).

**Tabela 5.3.** Ensaio de cristalização da laminarinase de *Rhodotermus marinus*, adicionando aditivos.

Aditivo	[] PEG3350	Tampão KH <sub>2</sub> PO <sub>4</sub>	pH	[] proteína
Glucose 5mM	15 %	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Sulfato de cádmio 0.5mM	6%	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Sulfato de Manganês 0.5mM	6 %	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Sulfato de Magnésio 0.5mM	6 %	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Cloreto de cádmio 0.5mM	6%	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Cloreto de cálcio 0.5mM	6 %	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Cloreto de Magnésio 0.5mM	6 %	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml
Sulfato de Manganês 0.5mM	6 %	50mM, 60mM, 70mM, 80mM, 90mM, 100mM	6.5	8mg/ml



**Figura 5.3.** Cristais da laminarinase do *Rhodotermus marinus* (*macroseeding*). Gota: 6% PEG 3350 em 90mM fosfato de potássio a pH 6.5, glucose 5mM. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0

Novas tentativas foram feitas adicionando diferentes concentrações de etanol no poço, tendo sido os melhores resultados obtidos com etanol 2% (figura 5.4), em diferentes concentrações de fosfato. Assim, foi feito um novo *screening* variando a concentração de fosfato na gota e usando no poço uma solução de PEG 25% 50mM de fosfato de potássio e etanol 2%.



**Figura 5.4.** Cristais da laminarinase de *Rhodotermus marinus*. Gota: 6% PEG 3350 em 65mM fosfato de potássio a pH 6.5. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0 e etanol 2%.

Os melhores resultados foram obtidos numa concentração de fosfato de potássio 120mM pH6.0 (figura 5.5). Os cristais cresceram após uma semana e atingiram seu tamanho máximo em duas semanas. Embora no começo de seu crescimento apresentassem uma aparência de monocristais, com o tempo cresceram formações de *clusters* ligados ao cristal.



**Figura 5.5.** Cristal da laminarinase de *Rhodotermus marinus*, usados para a coleta de dados de raios-x. Gota: 6% PEG 3350 em 120mM fosfato de potássio a pH 6.0. Poço: 25% PEG 3350 50mM de fosfato de potássio pH 6.0 e etanol 2%.

O cristal da figura 5.5 foi usado durante a coleta dos dados de difração de raios-X no LNLS. Após limpar cuidadosamente o *cluster* ligado ao cristal, ele foi dividido em duas partes: Uma foi usada como cristal nativo e a outra foi mergulhada numa solução com césio 0.5M e coletada como derivado.

### 5.1.2.2. Coleta e processamento de dados.

O cristal nativo difratou até 2.1Å de resolução e pertence ao sistema monoclinico C2 com parâmetros  $a=52.94$ , Å  $b=119.71$  Å,  $c=109.94$  Å,  $\alpha=90.00$ ,  $\beta=89.89$ ,  $\gamma=90.00$ . Devido ao tamanho do parâmetro de rede, em algumas orientações as reflexões aparecem muito próximas (figura 5.6), e usando o programa MOSFLM não foi possível separar estas reflexões. Assim foi coletado um conjunto a baixa resolução visando separar estas reflexões. Embora as reflexões ficassem mais afastadas o programa não conseguiu predizer todas as reflexões (figura 5.7).

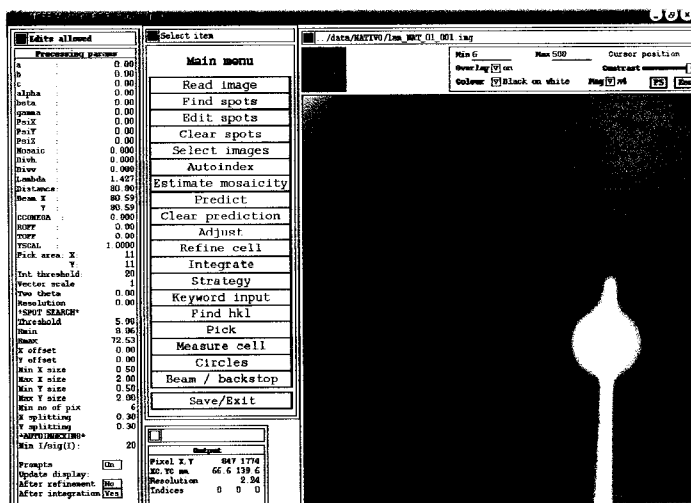
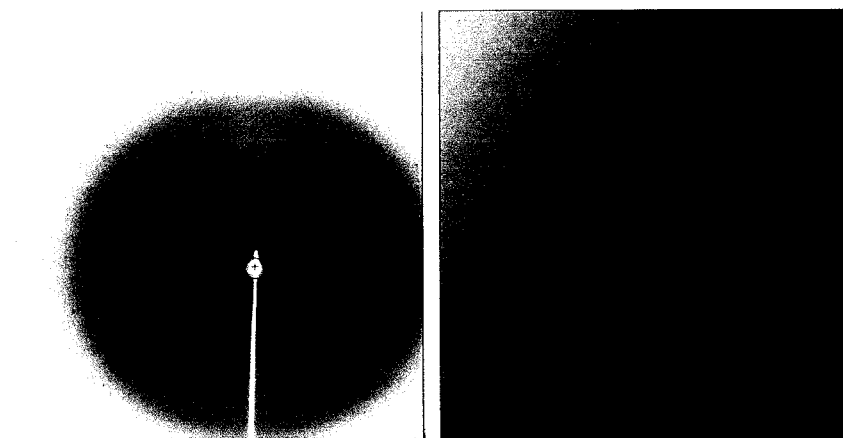


Figura 5.6. Imagem de difração do cristal nativo da laminarinase de *Rhodotermus marinus*. Devido ao parâmetro da rede as reflexões aparecem muito próximas.

O cristal que foi mergulhado na solução de césio difratou até 2.9 Å, mas embora para algumas imagens o perfil das reflexões seja bem determinado, em outras apareceram reflexões muito próximas, o que dificulta a indexação e a integração das imagens de difração. Alguns dados estatísticos do processamento são mostradas na tabela 5.4.





a. b.  
**Figura 5.7.** Imagens de difração do cristal derivado da laminarinase de *Rhodotermus marinus*. a. imagem a  $\text{PHI}=0^\circ$  b. Zoom da imagem a  $\text{PHI}=90^\circ$ , apresenta reflexões duplas.

**Tabela 5.4.** Estatística do processamento dos cristais nativo e derivado da laminarinase de *Rhodotermus marinus*.

	Nativo	Derivado
Graus coletados	261	180
Grupo espacial	$C_2$	$C_2$
Dimensões da célula (Å)	a=52.94Å, b=119.71 Å c=109.94Å, $\alpha=90.00$ $\beta=89.89$ , $\gamma=90.00$ .	a=52.37Å, b=119.08Å c=109.33Å, $\alpha=90.00$ $\beta=89.73$ , $\gamma=90.00$
Resolução (Å)	30.0-2.1	30.0-2.9
Multiplicidade	4.1 (2.3)	4.4(4.1)
$R_{\text{merge}}^1$ (%)	12 (29.1)	25.0(55.0)
Completeza (%)	84.5 (84.5)	95.5(95.5)
$I/\sigma(I)$	5.9 (1.6)	2.7(2.3)

As tentativas de fazer substituição molecular com os 8 membros da família GH16 com estrutura tridimensional determinada não foram bem sucedidas. Apesar de de um cristal derivado de Césio ter sido preparado, os resultados não permitem afirmar que de fato o Césio tenha se ligado na estrutura cristalina, já que as tentativas de achar as posições destes íons com o programa SHELXD não foram bem sucedidas. Neste caso não seria correto de chamar este cristal de derivado, mas mantivemos esta denominação durante a discussão apenas para ressaltar a diferença de tratamento a que foram submetidos os cristais.

## 5.2. Exo-glucanase de *Trichoderma viride*

A exo-1,3- $\beta$ -glucanase de *Trichoderma viride* é uma enzima de aproximadamente 61 kDa que mostra atividade com laminarina e laminari-oligossacarídeos. Esta enzima realiza a hidrólise com inversão da configuração anomérica, não tendo a seqüência N-terminal homologia com nenhuma das glucanases conhecidas<sup>30</sup>. No presente estudo foi feita a caracterização da exo-1,3- $\beta$ -glucanase de *Trichoderma viride* usando a técnica de CD, a qual mostro que esta enzima apresenta uma predominância de estrutura secundaria tipo folhas beta.

### 5.2.2. Metodologia

#### 5.2.2.1. Purificação

A proteína parcialmente pura foi gentilmente cedida pelo grupo do Dr Golubev, tendo sido entao implementado um ultimo passo de purificação com a colunade troca anionica MonoQ HR5/5 (Amershan). A proteína ressuspendida em agua foi aplicada na coluna previamente equilibrada com tampão Tris 20 mM pH 7,2 e eluída com um gradiente linear de NaCl de 0 a 0.5 M em fluxo de 1 ml/min. O conjunto das frações correspondentes ao peso molecular da laminarinase foi dialisado por 12 horas a 4o C contra água destilada e concentrado com Microcom.

As amostras foram analisadas em SDS-PAGE (concentração de poliacrilamida de 15 %). As bandas de proteínas foram visualizadas através de coloração com *Coomassie Blue*.

A dosagem da concentração da proteína foi determinada pelo reagente de Bradford<sup>31</sup> (Bio-Rad) conforme especificações do fabricante.

#### 5.2.2.2. Dicroísmo Circular

O espectro de dicroísmo circular foi obtido em um espectropolarimeto J715 (Jasco corp., Japão). A proteína, solubilizada em tampão 3mM TrisHcl pH 7.2, a uma concentração de 0.1mg/ml foi colocada em uma cuveta de 10 mm de caminho ótico. Um total de 16 varreduras, no intervalo de comprimento de onda de 250 a 190 nm,

foram efetuadas para determinação de uma média. O espectro do tampão sem proteína foi medido e subtraído do espectro da exo-1,3-β-glucanase do *Trichoderma viride*. O conteúdo de estrutura secundária foi estimado mediante a deconvolução do espectro, usando os programas CDSSTR<sup>32,33</sup>, CONTINLL<sup>34</sup>, K2D<sup>35</sup> e SELCON3<sup>36</sup> no website DichroWeb (<http://www.cryst.bbk.ac.uk/cdweb/html/home.html>)<sup>38</sup>. Os menores valores de NRMSD<sup>39</sup> foram obtidos com os programas CDSSTR (0.023) e com CONTIN (0.185). Nos dois casos o conjunto 4 da base de dados do DichroWeb foi usado como referência.

### 5.2.3. Resultados e Discussão

#### 5.2.3.1. Purificação

Na figura 5.8 apresenta-se um cromatograma com um perfil típico desta cromatografia e na figura 5.9 o gel SDS correspondente.

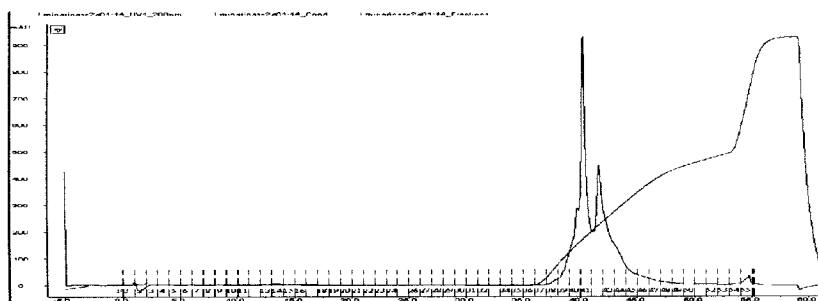


Figura 5.8. Perfil da cromatografia de troca iônica (MonoQ 5/5) para a exo-1,3-β-glucanase do *Trichoderma viride*.

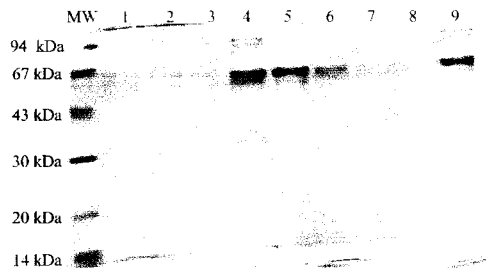
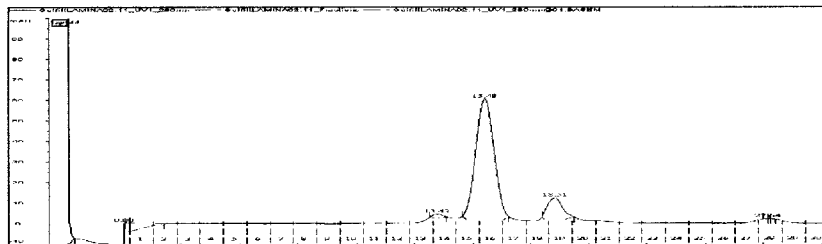


Figura 5.9. Eletroforese em gel de poliacrilamida (SDS-PAGE) das amostras coletadas depois da MonoQ, no poço 4 verifica-se a amostra eluída no pico 1 e nos poços 5 e 6 a proteína eluída no segundo pico. A amostra que foi injetada na coluna se observa no poço 9.

As frações correspondentes ao segundo pico da cromatografia foram reunidas, dialisadas e concentradas como descrito na seção 5.1.1 antes de serem usadas nos ensaios de cristalização.

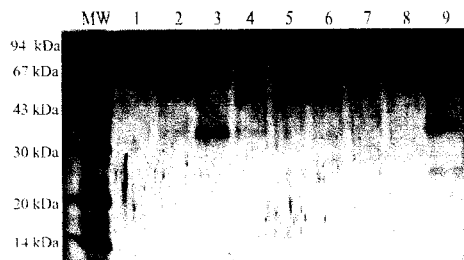
Os ensaios de cristalização foram conduzidos pelo método da gota pendurada. Utilizou-se o *kit* Cristal Screen I-II (Hampton Research) para os ensaios iniciais empregando-se amostras obtidas por coluna de troca iônica MonoQ a temperatura de 18 °C, em concentrações de 5 mg/ml e 10 mg/ml. Não verificou-se o surgimento de cristais ou de precipitados cristalinos.

Para verificar o estado de oligomerização da proteína, utilizou-se uma coluna de gel filtração. Na figura 5.10 mostra-se o perfil de eluição desta cromatografia enquanto na figura 5.11 o gel correspondente é apresentado.



**Figura 5.10.** Perfil da cromatografia de gel filtração (Superdex 200 HR10/30) para a exo-1,3-β-glucanase de *Trichoderma viride*.

O volume de eluição ( $V_e$ ) foi de 15.48ml o que corresponde a uma massa de 62.9 kDa e um raio de Stokes de 33.83 Å. Comparando-se com o SDS-PAGE pode-se concluir que a proteína é um monômero em solução.



**Figura 5.11.** Eletroforese em gel de poliacrilamida (SDS-PAGE), corado com prata. Poços 2 e 3: frações correspondentes ao segundo pico da cromatografia, que corresponde a um peso molecular de 67 kDa. No poço 9 a amostra que foi injetada na coluna.

5.2.3.2. Dicroísmo circular

Analisando o espectro ilustrado na figura 5.12, verificamos a existência de um mínimo no intervalo de  $\lambda = 210 \text{ nm}$  à  $\lambda = 215 \text{ nm}$ , sugerindo a predominância de um tipo de estrutura secundária, uma vez que espectros de dicroísmo circular são resultantes de uma média de estruturas em solução. A deconvolução mostrou uma predominância de estrutura em folhas beta (39%), pouca contribuição de estrutura em hélice (9%) e uma considerável porcentagem de estrutura desordenada (31%) além de voltas (21%). Os melhores resultados foram achados com o programa CDSSTR, NMRSD = 0.023 (tabela 5.5).

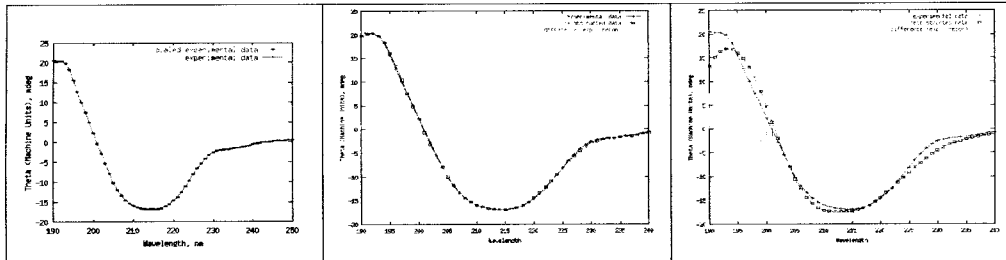


Figure 5.12. a. Espectro de dicroísmo circular da Exo-1,3- $\beta$ -glucanase do *Trichoderma viride*. b. Ajuste dos dados experimentais e a curva estimada pelo programa CONTINLL c. Ajuste dos dados experimentais e a curva estimada pelo programa CDSSTR.

Tabela 5.5. Conteúdo de estrutura secundaria da exo-1,3- $\beta$ -glucanase de *Trichoderma viride*, determinado a partir do espectro de dicroísmo circular.

Programa	Hélice	Beta	Voltas	Desordenada	NMRSD
CONTINLL	6.6%	39.4%	23.5%	30.6%	0.162
CDSSTR	9.0%	39.0%	21.0%	31.0%	0.023

### 5.3. Conclusões

- Foi cristalizada a laminarinase de *Rhodotermus marinus* pelo método da gota pendurada usando diferentes concentrações de PEG na gota e no poço, tendo sido utilizada na gota uma solução tampão 120mM fosfato de potássio pH 6.0 e 6% PEG 3350 e no poço uma solução 50mM de fosfato de potássio pH 6.0 25% PEG 3350 e etanol 2%. Os cristais difrataram até 2.1Å de resolução mostrando pertencer ao sistema monoclinico C2 com parâmetros  $a=52.94\text{\AA}$ ,  $b=119.71\text{\AA}$ ,  $c=109.94\text{\AA}$ ,  $\alpha=90.00$ ,  $\beta=89.89$ ,  $\gamma=90.00$ . Devido a qualidade do padrão de difração não foi possível a resolução da estrutura.
- O espectro de dicroísmo circular mostrou que a estrutura secundária da exo-1,3- $\beta$ -glucanase de *Trichoderma viride* apresenta uma predominância de estrutura em folhas beta (39%), pouca contribuição de estrutura em hélice (9%) e uma considerável porcentagem de estrutura desordenada (31%) além de voltas (21%). As tentativas de cristalização desta enzima não foram bem sucedidas.

## 5.4. Referências

- <sup>1</sup> Grenier, J., Potvin, C., and Asselin, A. (1993) *Plant Physiol.* **103**, 1277-1283
- <sup>2</sup> Batchman, E. S., and McClay, D. R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6808-6813
- <sup>3</sup> Watanabe, T., Kasahara, N., Aida, K., and Tanaka, H. (1992) *J. Bacteriol.* **174**, 186-190
- <sup>4</sup> Chen, L., Fincher, G. B., and Hoj, P. B. (1993) *J. Biol. Chem.* **268**, 13318-13326
- <sup>5</sup> Varghese, J. N., Garrett, T. P., Colman, P. M., Chen, L., Hoj, P. B., and Fincher, G. B. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2785-2789
- <sup>6</sup> Alfredson, G.A., Kristjánsson, J.K., Hjørleifsdóttir, S. and Stetter, K.O. (1988) *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland. *J. Gen. Microbiol.* **134**, 299-306.
- <sup>7</sup> Dahlberg, L., Holst, O. and Kristjánsson, J.K. (1993) Thermostable xylanolytic enzymes from *Rhodothermus marinus* grown on xylan. *Appl. Microbiol. Biotechnol.* **40**, 63-68.
- <sup>8</sup> Karlsson, E.N., Bartonek-Roxa, E. and Holst, O. (1997) Cloning and sequence of a thermostable multidomain xylanase from the bacterium *Rhodothermus marinus*. *Biochim. Biophys. Acta* **1353**, 118-124.
- <sup>9</sup> Halldórsdóttir, S., Þórólfssdóttir, E.T., Spilliaert, R., Johansson, M., Thorbjarnardóttir, S.H., Palsdóttir, A., Hreggvidsson, G.Ó., Kristjánsson, J.K., Holst, O. and Eggertsson, G. (1998) Cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. *Appl. Microbiol. Biotechnol.* **49**, 277-284.
- <sup>10</sup> Spilliaert, R., Hreggvidsson, G.O., Kristjánsson, J.K., Eggertsson, G. and Palsdóttir, A. (1994) Cloning and quencing of a *Rhodothermus marinus* gene, *bglA*, coding for a thermostable b-glucanase and its expression in *Escherichia coli*. *Eur. J. Biochem.* **224**, 923-930.
- <sup>11</sup> Krah, M., Misselwitz, R., Politz, O., Thomsen, K.K., Welfle, H. and Borriss, R. (1998) The laminarinase from thermophilic eubacterium *Rhodothermus marinus*. Conformation, stability, and identification of active site carboxylic residues by site-directed mutagenesis. *Eur. J. Biochem.* **257**, 101-111.
- <sup>12</sup> Anderson, M.A. and Stone, B.A. (1975) A new substrate for investigating the specificity of b-glucan hydrolases. *FEBS Lett.* **52**, 202-207.
- <sup>13</sup> Parrish, F.W., Perlin, A.S. and Reese, E.T. (1960) Selective enzymolysis of poly-b-d-glucans, and structure of the polymers. *Can. J. Chem.* **38**, 2094-2104.



- <sup>25</sup> Davis, G.J., Wilson, K.S. & Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557-559.
- <sup>26</sup> Polikarpov, I., Perles, L. A., *et al.* & Craievich, A. (1997). Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *J. Synchrotron Rad.* **5**, 72-76
- <sup>27</sup> Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst.* **D56**, 232-237.
- <sup>28</sup> Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta Cryst.* **D57**, 996-1002.
- <sup>29</sup> Leslie, A.G.W., (1992), Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
- <sup>30</sup> Kulminkaya A., Thomsen K., Shabalin K., Sidorenko I., Eneyskaya E., Savelev A. and Neutroev K., (2001), Isolation, enzymatic properties, and mode of action of an exo-1,3-b-glucanase from *Trichoderma viride*, *Eur. J. Biochem.* **268**, 6123-6131.
- <sup>31</sup> Bradford, M. (1976), A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248 - 254.
- <sup>32</sup> Manavalan, P. and Johnson, W. C. Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra, *Anal. Biochem.* **167**, 76-85.
- <sup>33</sup> Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set, *Anal. Biochem.* **287**(2), 252-260.
- <sup>34</sup> Provencher, S.W. and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37.
- <sup>35</sup> Andrade, M. A., Chacón, P., Merelo, J. J. and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network, *Prot. Engineering.* **6**, 383-390.
- <sup>36</sup> Sreerema, N. and Woody, R.W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism, *Anal. Biochem.* **209**, 32-44.
- <sup>37</sup> Lobley, A. and Wallace, B.A. (2001) DICHROWEB: A Website for the analysis of protein secondary structure from circular dichroism spectra, *Biophysical J.* **80**, 373a.





<sup>38</sup> Lobley, A. and Wallace, B.A. (2001) DICHROWEB: A Website for the analysis of protein secondary structure from circular dichroism spectra, *Biophysical J.* **80**, 373a.

<sup>39</sup> Mao, D., Wachter, E. and Wallace, B.A. (1982) Folding of the H<sup>+</sup>-ATPase Proteolipid in Phospholipid Vesicles, *Biochemistry* **21**, 4960-4968.

# Anexo 4

(Outras Publicações)

Nagem, R. P. A., Ambrosio A. L. B., Rojas, A. L., Navarro M. V. A. S., Golubev A. M., Garratt, R. and Polikarpov I. (2005). Getting the most of X-ray home sources. *Acta Crystallographica D*. Aceito.

## Getting the most out of X-ray home sources

R. A. P. Nagem,<sup>a,†</sup>  
 A. L. B. Ambrosio,<sup>a</sup> A. L. Rojas,<sup>a</sup>  
 M. V. A. S. Navarro,<sup>a</sup>  
 A. M. Golubev,<sup>b</sup> R. C. Garratt<sup>a</sup>  
 and I. Polikarpov<sup>a\*</sup>

<sup>a</sup>Institute of Physics of São Carlos, University of São Paulo, Avenida Trabalhador São-carlense 400, CEP 13560-970, São Carlos, SP, Brazil, and <sup>b</sup>Petersburg Nuclear Physics Institute, Gatchina, St Petersburg, 188300, Russia

<sup>†</sup> Current address: Biochemistry and Immunology Department, Institute of Biological Sciences, Federal University of Minas Gerais, Avenida Antônio Carlos 6627, Caixa Postal 486, CEP 31270-901, Belo Horizonte, MG, Brazil.

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

Received 2 March 2005  
 Accepted 22 April 2005

The structures of a 14 kDa phospholipase, an 18 kDa proteinase inhibitor and a novel glycoside hydrolase with molecular weight 60 kDa were solved using the SAD technique and the effects of the amount of anomalous signal, completeness and redundancy of data on heavy-atom substructure determination, phasing and model building were analyzed. All diffraction data sets were collected on a Cu-anode X-ray home source. The structure of the phospholipase was obtained using the anomalous scattering contribution from its 16 S atoms. Three-dimensional models for the other two macromolecules were obtained using the anomalous contribution of 1 atoms rapidly incorporated into the crystal through the quick cryo-soaking method of derivatization. These results were used to discuss the application of sulfur- and iodine-SAD approaches in combination with X-ray home sources for high-throughput protein crystal structure solution. The estimates of the anomalous signal from S atoms in the gene products of four genomes are given and the prospects for increasing the anomalous contribution using longer wavelengths (*e.g.* from a chromium home source) and quick cryo-soaking derivatization are discussed. The possibility of rapidly preparing tangible home-source isomorphous derivatives suggests that this approach might become a valuable tool in the future of post-genomic projects.

### 1. Introduction

The electron-density map required to construct an atomic model of a macromolecule by means of X-ray crystallography can only be obtained if the phase and the amplitude of the structure factors are available. However, during a typical X-ray diffraction experiment only the structure-factor amplitudes can be obtained. This is the so-called 'phase problem' and has been discussed in detail in several textbooks (Blundell & Johnson, 1976; Drenth, 1999).

The first largely successful technique developed to estimate the phases and consequently to obtain an initial electron-density map of a macromolecular crystal was pioneered by Perutz and coworkers in the 1950s (Green *et al.*, 1954). This approach resulted in the elucidation of the first three-dimensional structure of a protein, haemoglobin. This technique is based on the isomorphous differences observed in the diffraction pattern owing to the incorporation of a few heavy atoms into the ordered structure of native crystals and is termed the multiple/single isomorphous replacement (MIR/SIR) method. The anomalous scattering of X-rays can also be used to yield independent phase information (Laue, 1916; Bijvoet, 1949). In practice, its use in phase determination was reported in the 1980s (Hendrickson & Teeter, 1981; Wang,

1000 1107 50907444905012989

1985). The anomalous signal, which is observed as differences between Friedel-related reflections (Bijvoet differences), is usually small and depends on both the atom type and the wavelength of the incident X-rays. The combined use of such approaches in a SIRAS/MIRAS (single/multiple isomorphous replacement with anomalous scattering) experiment has also been used for phase determination of several macromolecule crystals.

More recently, with the advent of tunable synchrotron beamlines devoted to protein crystallography experiments, the use of the anomalous diffraction techniques with either single-wavelength (SAD) or a multi-wavelength (MAD; Hendrickson, 1991) data has become routine. The MAD method uses only the wavelength-dependence of  $f'$  and  $f''$  (dispersive and anomalous components of the atomic structure factor) of the anomalously scattering atoms for phasing. Usually, three or four data sets are collected at various wavelengths in order to optimize the differences between the  $f'$  and  $f''$  contributions. Nowadays, the MAD approach in conjunction with selenomethionine-derivatized proteins has become the method of choice for solving new protein crystal structures. Nevertheless, several recent experiments have demonstrated that even when anomalous signals are weak, the SAD approach can be used for phasing and structure determination if based on high-quality diffraction data (Dauter *et al.*, 1999, 2002; Liu *et al.*, 2000; Bond *et al.*, 2001; Dauter & Adamiak, 2001; Gordon *et al.*, 2001; Lemke *et al.*, 2002; Micossi *et al.*, 2002; Nagem *et al.*, 2003; Debreczeni, Bunkóczi, Ma *et al.*, 2003).

Now that a number of genomes have been completely elucidated and many more are being sequenced, the availability of thousands and thousands of genes has created the challenge of high-throughput structure elucidation of gene products. This in turn has stimulated an increasing interest in the development of less expensive equipment for automatic crystal growth (Stevens, 2000; Watanabe *et al.*, 2002), techniques that require little time and effort for successful phasing (Dauter *et al.*, 2000; Nagem *et al.*, 2001; Evans & Bricogne, 2002) as well as robust software for heavy-atom location (Schneider & Sheldrick, 2002; Burla *et al.*, 2003), density-modification protocols (Terwilliger, 2002; Sheldrick, 2002) and automatic electron-density interpretation (Perrakis *et al.*, 1999; Terwilliger, 2002; Joergers & Sacchettini, 2002). Massive investment has been made in the construction, maintenance and upgrade of advanced dedicated protein crystallography beamlines on synchrotrons worldwide. Training of specialized personnel to help synchrotron users to collect diffraction data from crystals of macromolecules ranging from small proteins to viruses or ribosomes also plays an important role in the competitive world of synchrotron beam time. We therefore may conclude that within the last 50 y drastic changes have influenced the way protein crystallography is being performed.

All these technological innovations and advances developed within high-throughput structure-solution projects pose a few questions. How can laboratories with limited access to synchrotron sources still contribute to such projects? How can we make the most of an X-ray home source in such a way that it can be used not only to test crystals prior to a synchrotron

data collection but also for structure solution of novel proteins?

In the present work, we describe the use of X-rays from a rotating Cu-anode home source to solve the structures of a 14 kDa sulfur-rich enzyme, an 18 kDa proteinase inhibitor and a novel 60 kDa glycosidase by the SAD method. The first protein was solved using the anomalous scattering contribution of its 16 S atoms, while the structures of the latter two macromolecules were obtained using the anomalous contribution of I atoms rapidly incorporated into the crystal through the quick cryo-soaking approach (Dauter *et al.*, 2000; Nagem *et al.*, 2001, 2003). The use of sulfur phasing, which is associated mostly with synchrotron data collection, but also with data sets collected on X-ray home sources, has been widely exploited by many crystallographers in recent years (Dauter *et al.*, 1999; Liu *et al.*, 2000; Bond *et al.*, 2001; Dauter & Adamiak, 2001; Gordon *et al.*, 2001; Lemke *et al.*, 2002; Micossi *et al.*, 2002; Debreczeni, Bunkóczi, Girmann *et al.*, 2003; Debreczeni, Girmann *et al.*, 2003; Olsen, Flensburg, Olsen, Bricogne *et al.*, 2004; Olsen, Flensburg, Olsen, Seibold *et al.*, 2004). Owing to the natural occurrence of S atoms in proteins, it has been proposed to use these anomalous scatterers for phasing despite their small anomalous contribution within the practically available X-ray range ( $f'' = 0.56 e^-$  at the Cu  $K\alpha$  wavelength or  $f' = 1.14 e^-$  at the Cr  $K\alpha$  wavelength). Alternatively, an elevated anomalous signal can be obtained by derivatization of protein crystals using the quick cryo-soaking approach with appropriate high atomic weight ions (halides or alkali metals). The quick cryo-soaking approach has mostly been used in combination with synchrotron sources (Nagem *et al.*, 2003). Here, we report our results on the application of this derivatization technique in conjunction with diffraction data collected at a Cu rotating-anode home source. The crystal structure determination of macromolecules with molecular weights of up to 60 kDa on an in-house X-ray source carry the promise that these methods can be routinely applied for protein structure solution as a complement to modern synchrotron-radiation-based techniques.

## 2. Materials and methods

### 2.1. Crystal preparation and data collection

Three different proteins were purified to homogeneity and crystallized using the hanging-drop method as described elsewhere.

The first protein is a class II Lys49-PLA<sub>2</sub> from the venom of *Agkistrodon contortrix laticinctus*. It is a sulfur-rich enzyme (14 kDa) belonging to a large well studied family of phospholipases A<sub>2</sub> (Acl-PLA<sub>2</sub>; PDB code 1s8i; Ambrosio *et al.*, 2005). Well diffracting crystals were grown in hanging drops after mixing equal volumes of protein solution at 10 mg ml<sup>-1</sup> and mother-liquor solution. The mother-liquor solution consisted of 2.0 M ammonium sulfate, 0.1 M Tris-HCl pH 8.5. A suitable crystal for data acquisition was obtained after transferring a single crystal from the crystallization drop to a cryogenic solution containing the mother liquor and 15%

**Table 1**  
Data statistics for Nat-Acl-PLA<sub>2</sub>, I-Aaw-Ghy, Nat-Bba-Pin and I-Bba-Pin data sets.

Values in parentheses refer to the highest resolution shell in each data set. All data sets were collected with Cu K $\alpha$  radiation from a Rigaku Ultra-X 18 rotating-anode generator.

	Nat-Acl-PLA <sub>2</sub>				I-Aaw-Ghy				Nat-Bba-Pin	I-Bba-Pin	
Space group	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2				<i>P</i> 2 <sub>1</sub>				<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
Unit-cell parameters (Å, °)	<i>a</i>	<i>b</i>	70.57, <i>c</i>	57.45	<i>a</i>	49.96, <i>b</i>	93.35, <i>c</i>	67.83, $\beta$	106.92	<i>a</i> = 46.70, <i>b</i> = 64.14, <i>c</i> = 59.24	<i>a</i> = 47.05, <i>b</i> = 63.12, <i>c</i> = 59.78
Mosaicity (°)	0.45				0.9				0.5	0.5	
$\Delta\phi$ rotation (°)	1.0				1.0				0.7	1.0	
Resolution range (Å)	26.63–1.61 (1.65–1.61)				26.63–1.87 (1.97–1.87)				26.43–1.87 (1.97–1.87)	25.27–2.20 (2.32–2.20)	
Data-collection time (h)	21				47				15	23	
Total $\phi$ rotation (°)	90	180	270	359	102	205	308	411	213	250	
Total reflections	131349	263493	395738	526515	95202	180160	272594	360302	124202	88825	
Unique reflections	35899	35928	35942	35959	37859	44936	45075	45205	14580	16658	
Redundancy	3.6 (3.4)	7.3 (6.8)	11.0 (10.2)	14.6 (14.1)	2.5 (2.3)	4.0 (3.8)	6.0 (5.7)	8.0 (7.5)	8.5 (8.4)	5.3 (5.6)	
Overall $R_{\text{sym}}$ (%)	5.7 (27.5)	6.3 (30.9)	6.2 (31.4)	6.1 (31.4)	5.9 (17.9)	4.3 (12.0)	5.1 (13.5)	5.3 (13.8)	5.0 (21.0)	10.2 (36.8)	
Completeness (%)	99.6 (97.2)	99.7 (97.9)	99.8 (99.3)	99.8 (99.8)	64.1 (61.5)	84.4 (72.6)	88.3 (77.9)	91.0 (81.6)	95.8 (95.8)	95.5 (95.5)	
$I/\sigma(I)$	23.8 (4.0)	33.9 (5.7)	41.6 (7.0)	49.5 (8.1)	7.0 (4.2)	10.9 (6.0)	9.2 (5.4)	9.1 (5.2)	11.0 (3.6)	6.7 (2.0)	

ethylene glycol for a few seconds. This crystal (Nat-Acl-PLA<sub>2</sub>) was then flash-cooled to 100 K in a cold nitrogen stream and used for data collection. A redundant data set was collected on a MAR345 image-plate detector using Cu K $\alpha$  radiation generated from a Rigaku Ultra-X 18 rotating-anode generator operating at 80 mA and 50 kV and focused using Osmic mirrors. Diffraction images were processed with *DENZO* (Otwinowski & Minor, 1997) and scaled with *SCALEPACK* (Otwinowski & Minor, 1997) in four different batches of total  $\phi$  rotation. As a result, four different data sets of total  $\phi$  rotation equal to 90, 180, 270 and 359° were obtained.

The second macromolecule is a proteinase inhibitor with 164 amino-acid residues extracted from the *Bauhinia bahinoides* tree (Bba-Pin). Protein crystallization was achieved using 8% PEG 4000, 0.1 M sodium acetate trihydrate pH 4.6 as the mother-liquor solution. Two crystals were prepared for data collection: a native and an iodine derivative. The first crystal (Nat-Bba-Pin) was obtained in the same fashion as Nat-Acl-PLA<sub>2</sub>. However, the second crystal (I-Bba-Pin) was prepared according to the quick cryo-soaking procedure for derivatization in which native crystals are rapidly soaked in cryogenic solutions containing a high concentration of an appropriate salt; in this case the mother liquor with 15% ethylene glycol and 0.5 M NaI. Both crystals were rapidly frozen in a nitrogen stream and the same technical aspects of data collection used for the first protein were applied here. Diffraction images for each data set were processed with *MOSFLM* (Leslie, 1992) and scaled with *SCALA* (Evans, 1997) in a single batch of total  $\phi$  rotation.

The last protein is a novel extra cellular glycoside hydrolase (537 amino acids) from the filamentous fungus *Aspergillus awamori* (Aaw-Ghy; PDB codes 1y4w and 1y9g; Nagem *et al.*, 2004). Only a single iodine derivative (I-Aaw-Ghy) was prepared for this protein using the quick cryo-soaking approach. One native crystal was transferred from the crystallization drop to a new solution drop containing the mother liquor (15% PEG 3350, 0.1 M sodium acetate pH 4.2), 20% glycerol and 0.5 M NaI. The crystal was immersed for a few

minutes and then flash-cooled in the nitrogen stream. More than 400 diffraction images ( $\Delta\phi = 1.0^\circ$ ) were collected from this derivative and processed with *MOSFLM*. Following the procedure adopted with the Nat-Acl-PLA<sub>2</sub> data set, intensities were scaled with *SCALA* in four different batches of total  $\phi$  rotation equal to 102, 205, 308 and 411°.

Further details of data acquisition and data statistics for all crystals used in this work are shown in Table 1.

## 2.2. Anomalous scatterer search, phase calculation and model building

The location of S atoms in the Nat-Acl-PLA<sub>2</sub> crystal and iodide anions in the I-Bba-Pin and I-Aaw-Ghy derivatives was performed using the dual-space recycling algorithm (Miller *et al.*, 1993) implemented in the *SHELXD* program (Sheldrick *et al.*, 2001). During anomalous substructure identification *SHELXD* default parameters were used. No resolution cutoff was imposed. Anomalous differences obtained with *SHELXPRO* (Sheldrick & Schneider, 1997) were separately chosen by the *SHELXD* algorithm for 500 multi-substructure solution trials.

A total of four different runs of *SHELXD* were performed with Nat-Acl-PLA<sub>2</sub>; each run corresponded to data sets with different total  $\phi$  rotation (90, 180, 270 and 359°). In all four cases *SHELXD* was set to search for 16 sulfur sites. A similar approach was adopted for the location of 12 partially occupied iodide anions in the I-Aaw-Ghy data sets (102, 205, 308 and 411°). As a consequence, the effect of data-set redundancy in the location of heavy-atom sites could be studied by analyzing the outcome from each run. In the proteinase-inhibitor case, *SHELXD* was set to search for eight iodine sites in a single run, corresponding to a data set of total  $\phi$  rotation equal to 250° (I-Bba-Pin).

Data sets that exhibited potentially correct results for the anomalous scatterer searches, as indicated by the large difference between the CCs from the best solution and all

**Table 2**

Detailed statistics obtained at each step from heavy-atom search to model building using Nat-Acl-PLA<sub>2</sub>, I-Bba-Pin and I-Aaw-Ghy data sets.

Phase calculation using SAD or SIRAS methods and model building were carried out only with promising data sets.

Parameters	Nat-Acl-PLA <sub>2</sub>				I-Bba-Pin		I-Aaw-Ghy		
	90	180	270	359	250	102	205	308	411
Total $\varphi$ rotation (°)	16	16	16	16	8	12	12	12	12
No. heavy atoms to search	2115	2127	2102	2098	697	3842	4899	5196	5418
No. $\Delta F$ s used in search <sup>†</sup>	0/500	0/500	13/500	14/500	66/500	3/500	21/500	42/500	87/500
No. solutions/total trials	16.6	14.4	38.4	42.4	37.2	16.0	20.1	27.1	30.4
CC of best trial (%)	13.1	10.7	12.1	12.2	17.2	8.0	10.9	11.0	13.9
(CC) of other trials (%)	—	—	17	17	8	10	10	10	10
No. sites used for phasing	—	—	0.53	0.54	0.52 (0.63)§	0.43	0.44	0.48	0.49
(FOM) after phasing (lr)‡	—	—	0.42	0.45	0.39 (0.46)	0.30	0.36	0.40	0.41
(FOM) after phasing (hr)*	—	—	0.92	0.92	0.90 (0.94)	0.71	0.92	0.93	0.93
(FOM) after dm†† (lr)‡	—	—	0.94	0.94	0.86 (0.90)	0.74	0.89	0.90	0.89
(FOM) after dm†† (hr)*	—	—	0.86	0.87	0.71 (0.85)	0.32	0.69	0.78	0.81
Electron-density correlation‡‡	—	—	108/121	108/121	123/163 (154/163)	0/537	499/537	500/537	496/537
No. residues built/in total	—	—	1	1	9 (3)	0	11	10	9
No. polypeptide chains	—	—	1	1	9 (3)	0	11	10	9

<sup>†</sup> Anomalous differences used for triplet generation. <sup>‡</sup> lr, low-resolution range (25.0–4.0 Å). <sup>§</sup> Values in parentheses refer to phase calculation using the SIRAS method. <sup>\*</sup> hr, high-resolution range (25.0–2.5 Å). <sup>††</sup> dm, density-modification protocol. <sup>‡‡</sup> Correlation coefficient between experimental (after density modification) and refined electron-density maps.

other trials, were used in phase calculation. Only major anomalous scatterer sites from the best *SHELXD* solution were input into *SHARP* (de La Fortelle & Bricogne, 1997) with their respective data set for SAD phase calculation. Whenever necessary, anomalous scatterer coordinates were inverted to assign the correct enantiomorph substructure. Density modification was performed with the program *SOLOMON* (Abrahams & Leslie, 1996) using the user-friendly interface implemented in *SHARP*. The resulting electron-density maps were used for automatic model building and iterative refinement with the programs *ARP/wARP* (Perrakis *et al.*, 1999) and *REFMAC* (Murshudov *et al.*, 1997), respectively. In the proteinase-inhibitor case, improved experimental phases were obtained when native and iodine-derivative data sets were combined in SIRAS phase calculation with *SHARP*. As a result of this, model building was continued with two distinct maps: the first at 2.20 Å resolution with SAD-derived phases and the second at 1.87 Å resolution with extended SIRAS-derived phases.

Detailed statistics obtained for each step from heavy-atom search to model building in all cases analyzed are shown in Table 2.

### 2.3. Anomalous signal estimates

The anomalous signal contribution from S atoms in the Nat-Acl-PLA<sub>2</sub> crystal and iodide ions in the I-Aaw-Ghy and I-Bba-Pin derivatives was assessed by calculating the ratio  $\langle \Delta F_{\text{ANO}} \rangle / \langle F \rangle$  for several resolution shells in each data set. Only non-centric reflections were used in these calculations. Whenever necessary, theoretical estimation of anomalous signal was obtained using the formula  $\langle \Delta F_{\text{ANO}} \rangle / \langle F \rangle = 2^{1/2} N_A^{1/2} f'' / Z_{\text{eff}} N_P^{1/2}$  suggested by Hendrickson & Teeter (1981). Here,  $N_A$  is the number of anomalous scatterers in the asymmetric unit with anomalous contribution  $f''$ . As  $f''$  depends on the atom type and on the wavelength of the incident X-rays, the program *CROSSEC* (Cromer, 1983) was used to estimate its value for each anomalous scatterer at the

corresponding wavelength.  $N_P$  is the number of non-H atoms in the asymmetric unit and  $Z_{\text{eff}}$  is the mean normal scattering for atoms present in proteins, which assumes the value of  $6.7 \text{ e}^-$ .

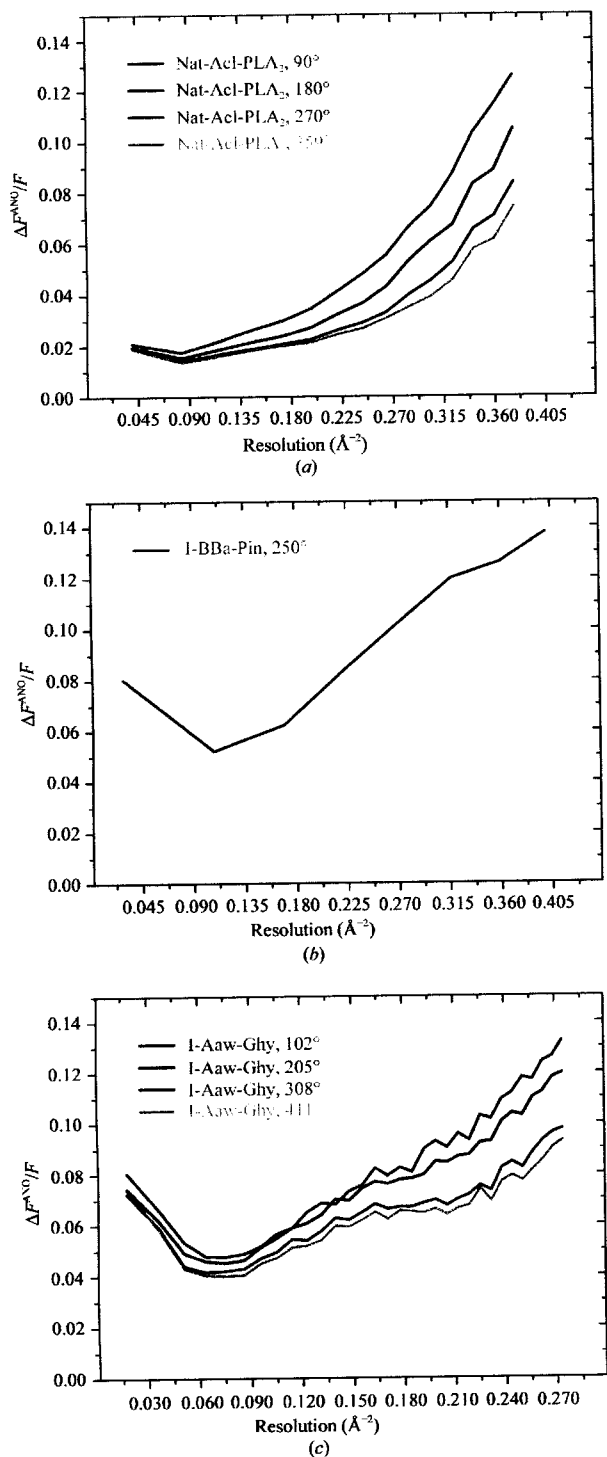
The anomalous contribution from S atoms in hypothetical gene-product crystals was calculated using the same formula. The amino-acid sequences of annotated genes in *Bacillus subtilis*, *Escherichia coli* K12, *Vibrio cholerae* and *Xylella fastidiosa* genomes were obtained from the National Center of Biotechnology Information (NCBI) through the FTP site <ftp://ftp.ncbi.nih.gov/genbank/genomes/>.

## 3. Results and discussion

### 3.1. Checking for the presence of and locating anomalous scatterers

The scattering of X-rays arising from the anomalous scatterers present in protein crystals results in breakage of Friedel's law and the emergence of non-zero Bijvoet differences ( $\Delta F_{\text{ANO}} = |F_{hkl} - F_{\bar{h}\bar{k}l}|$ ). As a result, the ratio  $\langle \Delta F_{\text{ANO}} \rangle / \langle F \rangle$  obtained for several resolution shells can indicate not only the presence of anomalous scatterers but also the amount of anomalous signal in each data set. Moreover, Bijvoet differences can be used with direct methods to determine the complete anomalous scatterer substructure within the asymmetric unit of a crystal, as implemented in *SHELXD*.

For Nat-Acl-PLA<sub>2</sub>, in which naturally occurring S atoms are the only measurable source of anomalous scattering, the total anomalous signal is lower than 2% (Fig. 1). As data-set redundancy increases, the symmetry-related reflections are measured several times and the precision of the Bijvoet differences improves. Therefore, the anomalous signal, which is overestimated at the beginning of data collection, approaches its theoretical value of 1.5% (corresponding to 16 fully occupied S atoms among 121 amino acids). Such redundancy-dependent behaviour has previously been observed (Dauter &



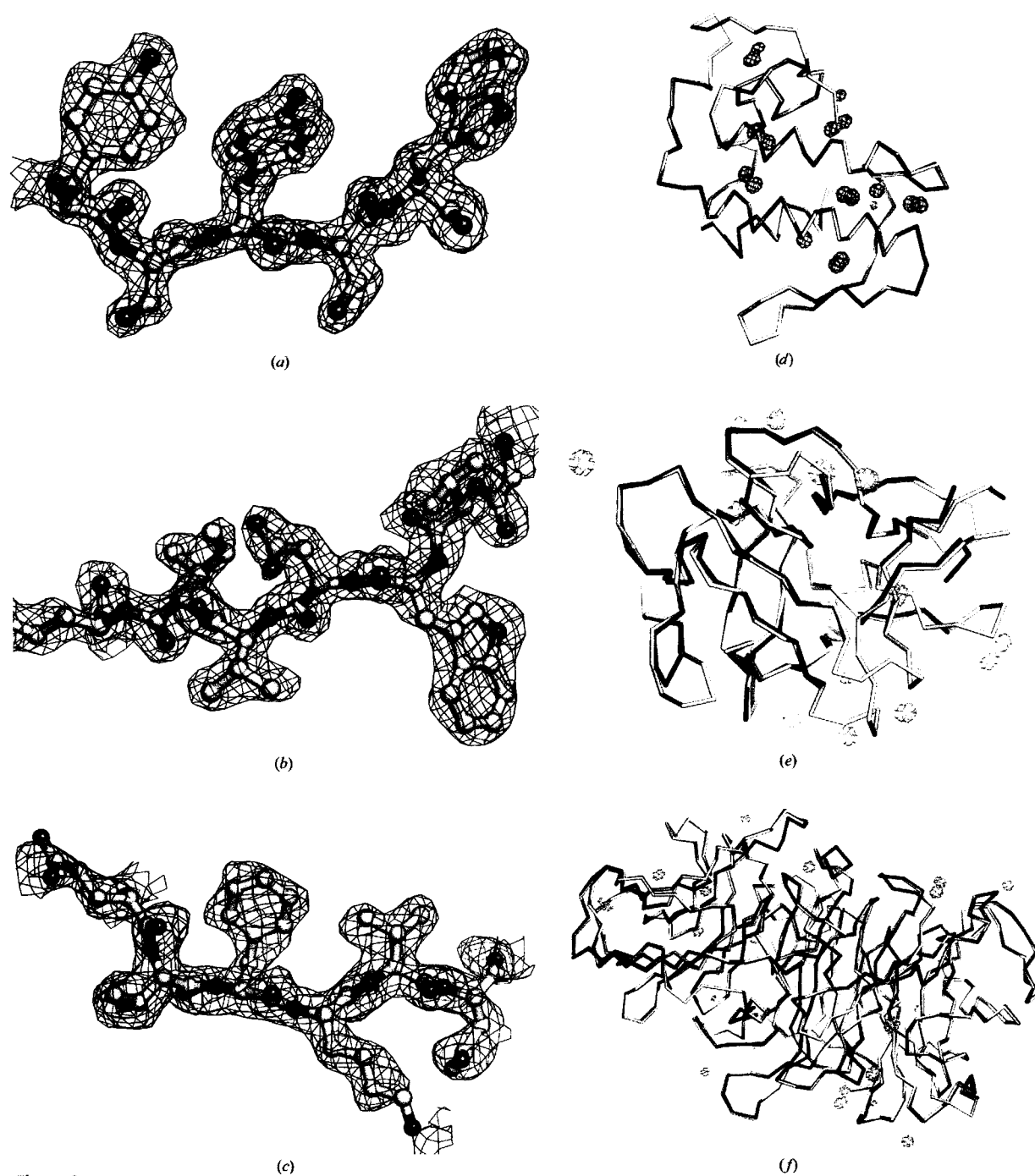
**Figure 1**  
Anomalous signal estimates as a function of resolution for the Nat-Acl-PLA<sub>2</sub>, I-Bba-Pin and I-Aaw-Ghy data sets. As the redundancy increases, the anomalous signal obtained for each data set can be measured with greater accuracy.

Nagem, 2002) and, particularly for data sets with small anomalous signal, plays an important role during anomalous substructure identification, phasing and structure determination. This fact can be readily confirmed by analyzing the four Nat-Acl-PLA<sub>2</sub> subsets integrated through 90, 180, 270 and 359°. Even though all four subsets show very similar completeness and almost the same number of anomalous difference  $\Delta F$  values could be obtained for each of them (Table 2), only the two most redundant subsets (270 and 359°) gave correct solutions for the anomalous scatterer substructure with *SHELXD*. Nevertheless, if appropriate resolution cutoffs were imposed, it might in principle be possible to locate major sulfur sites with the 180° or even 90° subsets. Surprisingly, in addition to the 16 sites expected for the 14 cysteine and two methionine residues of Acl-PLA<sub>2</sub>, *SHELXD* was able to identify two extra sites. It was subsequently verified that the two extra sites correspond to a second conformation of one of the methionine residues and a sulfate ion which was derived from the crystallization solution.

When the Cu  $K\alpha$  wavelength is used for data acquisition, S atoms display an anomalous contribution to the atomic scattering factor ( $f''$ ) of only 0.56 e<sup>-</sup>. As a result of this, the theoretical anomalous contribution from naturally occurring S atoms in both iodine derivatives tested is less than 0.5% and hence most of the anomalous contribution is provided by I atoms. The total anomalous signal of the I-Bba-Pin and I-Aaw-Ghy derivatives is plotted as a function of resolution in Fig. 1.

It is interesting to note that the I-Bba-Pin derivative, which has an asymmetric unit content slightly larger than that of Nat-Acl-PLA<sub>2</sub>, exhibits almost a fourfold greater anomalous signal. This fact can be explained not only by the incorporation of several iodide anions into the crystallographic structure, but also by the fact that I atoms show an anomalous scattering contribution that is 14 times stronger than that of S atoms at the same wavelength ( $f'' = 6.8$  e<sup>-</sup> at 1.54 Å). The amount of anomalous signal present in this data set could in principle suggest that no more than two I atoms were fully incorporated into the crystallographic structure. However, in practice, eight partially occupied major iodine sites were correctly identified in the I-Bba-Pin asymmetric unit by *SHELXD* (Table 2).

In the case of I-Aaw-Ghy, identification of the iodine substructure was also straightforward, as suggested by the contrast in correlation coefficients presented in Table 2. Careful inspection of the results obtained from each of the *SHELXD* jobs (102, 205, 308 and 411°) indicates that a higher contrast in CC values is observed within the last two subsets. A combination of increasing completeness and redundancy is most likely to be the main reason for these results. On the basis of the final refined structure, it was verified that all top ten iodine sites in each *SHELXD* solution were correct sites, with the exception of the first subset (102°), where a mixture of correct (top six) and incorrect (last four) sites was obtained. This result indicates that even though the data set demonstrates a high anomalous signal, high completeness and also



**Figure 2**  
 Automatic interpretation of experimental electron-density maps. (a) SAD-derived electron-density map from the Nat-Acl-PLA<sub>2</sub> most redundant data set after the density-modification protocol. (b) Electron-density map for Bba-Pin after density modification using SIRAS phases at 1.87 Å resolution. (c) Map obtained for the I-Aaw-Ghy data set after density modification using 411° of data. In all cases representative parts of refined models are shown. (d) C $\alpha$  trace of Nat-Acl-PLA<sub>2</sub> hybrid model (red) and the refined model (grey) after manual construction of the C-terminal portion and positional refinement. (e) C $\alpha$  trace of the Bba-Pin hybrid model (red) obtained with the experimental map at 2.20 Å resolution superimposed on the final model (grey) refined against the high-resolution data set. (f) C $\alpha$  representation of the I-Aaw-Ghy hybrid model (red) obtained using 205° diffraction data and the final model (grey) refined against the most redundant data set (411°). Anomalous maps obtained with phases generated on the basis of final models are shown in green (S atoms) and gold (I atoms). All figures were prepared using the program *PyMol* (DeLano, 2002).



moderate to high redundancy are essential for the location of anomalous scatterers with low occupancy.

### 3.2. Structure determination

An initial Nat-AcI-PLA<sub>2</sub> hybrid model (polypeptide fragments and dummy atoms) was obtained with the program *ARP/wARP* using either of the last two data sets (270 and 359°) with its respective experimental electron-density map. In the two cases tested, very similar qualitative and quantitative results were obtained. However, only the 359° subset was used for subsequent steps of manual construction, side-chain docking, water picking and positional refinement, aiming towards further structural studies.

In the proteinase-inhibitor case, the resulting set of experimental phases obtained both by SAD and SIRAS methods were submitted to the *ARP/wARP* program in two different jobs. As a result of this, initial hybrid models of the proteinase inhibitor were built within two distinct maps: the first at 2.20 Å resolution and the second at 1.87 Å resolution. Automatic interpretation of the lower resolution map took less than 30 min, rendering a hybrid model consisting of 123 amino-acid residues distributed into nine polypeptide fragments. This partial model represents 75% of the total structure. Further automatic model building under the same initial conditions did not lead to the improvement of the hybrid model. Nevertheless, better results were obtained using the higher resolution map. When SIRAS-derived phases were used, automatic model building was extremely fast (a few minutes) and around 95% of the total structure was built within the first four building cycles.

Determination of 1-Aaw-Ghy hybrid models was also performed automatically. However, it is interesting to note that even though all four data sets (102, 205, 308 and 411°) showed reasonable results during the heavy-atom search and phasing, the first data set, with the lowest completeness and

redundancy, was the only one which did not render sufficient information for convergent automatic model building and refinement. In the last three cases, practically the same number of amino-acid residues were built into each experimental electron-density map into a similar number of polypeptide chains. The basic difference between the results obtained for the last three data sets was the time required for model building. While automatic model building of this 60 kDa protein structure using the most redundant data set (411°) took less than 5 min, similar results were obtained within hours when the second data set (205°) was used. The main reason for this difference is the quality of each experimental electron-density map used for model building.

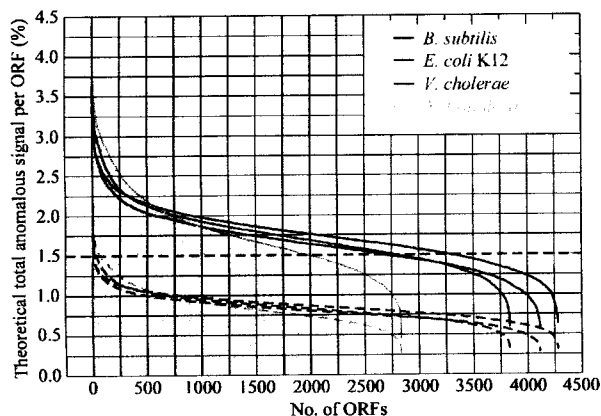
In Fig. 2, representative parts of the refined models together with the experimental maps used for automatic model building are given (Figs. 2a, 2b and 2c). A comparison between automatically built hybrid models and refined models for all three proteins tested are also shown (Figs. 2d, 2e and 2f).

### 3.3. Genomic scale analysis

The choice of the incident X-ray wavelength is one of the major decisions required for data acquisition. The decision relies on the evaluation of several wavelength-dependent factors, such as absorption, radiation damage, photon flux, resolution limit and anomalous signal. In fact, the incident X-ray wavelength is chosen as a compromise between all these factors and the purpose of data acquisition (Polikarpov *et al.*, 1997; Teplyakov *et al.*, 1998). However, as X-ray home sources produce monochromatic radiation, optimization of these parameters by changing the wavelength is not an option and alternative procedures should be applied. In cases where anomalous signal is essential for structure determination, an increase in the anomalous scattering contribution is achieved by selecting appropriate anomalous scatterers for derivatization that match the monochromatic X-ray home-source radiation and by collecting a highly complete and redundant data.

It has been suggested for a long time that S atoms could be used as the only source of anomalous signal in protein crystallography (Wang, 1985). From the biochemical point of view, this approach has two major advantages: firstly, S atoms are naturally present in proteins, which turns native crystals into derivatives, and secondly, the occupancy of sulfur sites is equal to 1 (although in some cases disordered/flexible S atoms may behave as if their occupancy was lower than 1). On the other hand, this approach has a clear disadvantage: S atoms provide a small anomalous contribution at the Cu K $\alpha$  wavelength (0.56 e<sup>-</sup>). Nevertheless, it is known that a few proteins display a high sulfur content which increases the total anomalous contribution to a level sufficient for structure elucidation, provided high-resolution diffraction data with adequate completeness and redundancy are available.

In this work, we demonstrate that an anomalous contribution of 1.5% is sufficient for structure elucidation of well diffracting crystals using the sulfur-SAD technique with data sets collected in-house. Similar results, sometimes with a



**Figure 3**  
Theoretical anomalous signal from S atoms present in expected gene products from *B. subtilis*, *E. coli* K12, *V. cholerae* and *X. fastidiosa* genomes. Anomalous contributions calculated at the Cu K $\alpha$  (dashed lines) and Cr K $\alpha$  (filled lines) wavelengths are shown. Filled areas illustrate the number of ORFs amenable to S-SAD structure determination using Cu K $\alpha$  (in brown) and Cr K $\alpha$  (in orange).

somewhat lower or a slightly higher anomalous signal, have been obtained by other crystallographers (Ramagopal *et al.*, 2003 and references therein). However, how many proteins or gene products present in elucidated genomes display a theoretical anomalous contribution above 1.5%?

In Fig. 3, we show theoretical calculations of the anomalous contribution from S atoms at the Cu  $K\alpha$  wavelength for each open-reading frame (ORF) in four different bacterial genomes. Unfortunately, the number of expected gene products with anomalous signal higher than 1.5% does not exceed 2.5% of the total number of ORFs in all four genomes evaluated in this work. Examining, for instance, all 2832 *X. fastidiosa* ORFs, approximately 97.5% of them (2760) show a theoretical anomalous contribution of less than 1.5%, while only 72 gene products display higher values. It is important to mention that our estimate does not take into account the fact that a number of proteins will have metal ions (such as Zn, Fe, Cu *etc.*) bound to them which would significantly increase their anomalous signal. In addition, considering that in some special cases an anomalous signal of around 1% has been successfully used for structure determination, more gene products could be solved by the sulfur-SAD approach, at least from the point of view of required anomalous signal. However, taking into consideration that this level of anomalous contribution requires high or even atomic resolution data, our estimates most probably represent an upper limit of proteins in the genome amenable to sulfur-SAD structure determination.

The possibility of using soft X-ray radiation obtained from specially designed chromium systems (home sources with Cr rotating anode, confocal multilayer optics designed for Cr  $K\alpha$  radiation, helium beam path and large-aperture detector; Yang *et al.*, 2003) can increase the applicability of this approach. At the Cr  $K\alpha$  wavelength, S atoms display twice the anomalous signal ( $f'' = 1.14 e^-$ ) than at the Cu  $K\alpha$  wavelength. Surprisingly, however, the number of gene products with total anomalous signal superior than 1.5% is increased by a factor that can reach two orders of magnitude (Fig. 3). Under these conditions, around 1982 *X. fastidiosa* ORFs would naturally display an anomalous signal higher than 1.5%. Our purely theoretical analysis may give an overoptimistic estimate of the benefits of soft X-ray radiation. Nevertheless, the results allow one to consider the possibility of phasing and determining the structure of a considerable number of proteins using the S-SAD approach with copper and/or chromium rotating-anode home sources.

#### 3.4. Obtaining more from rotating anodes

Apart from the potential success of the sulfur-SAD approach, it is clear that a considerable number of macromolecules will still require preparation of derivatives, either because they do not display a high anomalous signal or because in-house data quality is not good enough for phase determination. Preparation of isomorphous derivatives through the quick cryo-soaking approach using halides seems to be an appropriate option in such cases. The introduction of

negatively charged anomalous scatterers (such as iodide) into the crystal structure is achieved during a short soak (30–300 s) of the native crystal in a cryogenic solution containing in addition a high concentration of an appropriate salt. The procedure is performed during preparation of the crystal for data collection. It is cheap and not time-consuming.

The use of iodine compounds (such as sodium iodide and potassium iodide) for fast derivatization has some advantages in comparison with traditional transition-metal reagents (potassium tetrachloroplatinum, potassium dicyanoaurate and mercury acetate). Although iodide anions interact with water molecules through hydrogen bonds, they do not form strong complexes or bind covalently to certain chemical functions of the protein. As a result, some protein crystals, including Bba-Pin and Aaw-Ghy, can undergo derivatization soaks at high concentration of such reagents without damaging their crystalline order. Additionally, as diffusion through the crystal channels takes place very rapidly, a great number of sites around the protein surface are filled with iodide anions (Fig. 2). Most frequently, these ions replace ordered water molecules and their occupancy results from the competition between them. Since I atoms possess a strong anomalous signal at the copper  $K\alpha$  wavelength, comparable to that of platinum ( $6.9 e^-$ ), gold ( $7.3 e^-$ ) and mercury ( $7.7 e^-$ ), the contribution of all partially occupied sites can dramatically increase the total anomalous signal of the crystal, as observed in Fig. 1.

In application to structural genomics, the total anomalous signal of both iodine derivatives obtained in the present study is much higher (above 5%) than the theoretical anomalous signal from S atoms in all gene products analyzed, which makes it instrumental for structure determination of protein crystals that diffract to medium and low resolution. For example, as observed for the I-Bba-Pin data set, a strong anomalous signal facilitated the search for anomalous scatterers and phase determination.

In principle, the use of an in-house chromium radiation source ( $\lambda = 2.29 \text{ \AA}$ ) would further increase the strength of anomalous scattering of I atoms. At this wavelength, they display an  $f''$  of  $12.8 e^-$ ! However, special attention must be taken during data acquisition and data processing in order to avoid radiation damage and to include absorption corrections from the mounting loop and the solvent. This is especially true for quick cryo-soaked crystals, where a high concentration of heavy atoms is found in the non-crystalline phase.

The quick cryo-soaking approach can also be used with positively charged anomalous scatterers (Nagem *et al.*, 2001) such as alkali metals and lanthanides. Indeed, during the course of this work, the structure of a  $2 \times 56$  kDa enzyme was determined in our laboratory using such an approach (Mantovani *et al.*, manuscript in preparation). This provides additional flexibility in the process of derivatization and increases the chances of success in phase determination, since anions and cations naturally occupy different interaction sites on protein surfaces and display different affinities which depend on the protein surface-charge distribution and the pH of the cryo-derivatization solution. Additionally, different heavy-ion sites permit the combined use of derivatives in

MIR(AS) phasing, even when none of them are strong enough to produce interpretable electron-density maps via the SAD or SIR(AS) methods.

In this work, a simple consecutive data-acquisition strategy has been applied. It has been suggested, however, that synchrotron data involving anomalous signal should be collected in such a way that Friedel mates are measured close in time (Hendrickson & Ogata, 1997). This approach takes into account the fact that small anomalous differences cannot be precisely determined if errors from radiation damage accumulate in measured intensities. This strategy can be applied to home-source data acquisition, particularly when the crystals belong to a low-symmetry space group or suffer from considerable radiation damage.

## References

- Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.
- Ambrosio, A. L. B., Nonato, M. C., Selistre-de-Araujo, H. S., Arni, R., Ward, R. J., Ownby, C. L., de Souza, D. H. F. & Garratt, R. C. (2005). *J. Biol. Chem.* **280**, 7326–7335.
- Bijvoet, J. M. (1949). *Nature (London)*, **173**, 888–891.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
- Bond, C. S., Shaw, M. P., Alphey, M. S. & Hunter, W. N. (2001). *Acta Cryst.* **D57**, 755–758.
- Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2003). *Acta Cryst.* **D59**, 662–669.
- Cromer, D. T. (1983). *J. Appl. Cryst.* **16**, 437–438.
- Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* **D57**, 990–995.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
- Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* **D58**, 494–506.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
- Dauter, Z. & Nagem, R. A. P. (2002). *Z. Kristallogr.* **217**, 694–702.
- Debreczeni, J. É., Bunkóczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 393–395.
- Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 688–696.
- Debreczeni, J. É., Girmann, B., Zeeck, A., Krätzner, R. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 2125–2132.
- DeLano, W. L. (2002). *The PyMol Molecular Graphics System*. Delano Scientific, San Carlos, CA, USA.
- Drenth, J. (1999). *Principles of Protein X-ray Crystallography*, 2nd ed. Heidelberg: Springer.
- Evans, G. & Bricogne, G. (2002). *Acta Cryst.* **D58**, 976–991.
- Evans, P. R. (1997). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
- Gordon, E. J., Leonard, G. A., McSweeney, S. M. & Zagalsky, P. F. (2001). *Acta Cryst.* **D57**, 1230–1237.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Ioerger, T. R. & Sacchettini, J. C. (2002). *Acta Cryst.* **D58**, 2043–2054.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Laue, M. von (1916). *Ann. Phys.* **50**, 433–446.
- Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* **D58**, 2096–2101.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Liu, Z.-J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
- Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* **D58**, 21–28.
- Miller, R., De Titta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). *Acta Cryst.* **D57**, 996–1002.
- Nagem, R. A. P., Polikarpov, I. & Dauter, Z. (2003). *Methods Enzymol.* **374**, 120–137.
- Nagem, R. A. P., Rojas, A. L., Golubev, A. M., Korneeva, O. S., Eneyskaya, E. V., Kulminskaya, A. A., Neustroev, K. N. & Polikarpov, I. (2004). *J. Mol. Biol.* **344**, 471–480.
- Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. (2004). *Acta Cryst.* **D60**, 250–255.
- Olsen, J. G., Flensburg, C., Olsen, O., Seibold, M., Bricogne, G. & Henriksen, A. (2004). *Acta Cryst.* **D60**, 618.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Polikarpov, I., Teplyakov, A. & Oliva, G. (1997). *Acta Cryst.* **D53**, 734–737.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* **D59**, 1020–1027.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
- Sheldrick, G. M., Hauptmann, H. A., Weeks, C. M., Miller, M. & Usón, I. (2001). *International Tables For Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 333–351. Dordrecht: Kluwer Academic Publishers.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–341.
- Stevens, R. C. (2000). *Curr. Opin. Struct. Biol.* **10**, 558–563.
- Teplyakov, A., Oliva, G. & Polikarpov, I. (1998). *Acta Cryst.* **D54**, 610–614.
- Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1937–1940.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Watanabe, N., Murai, H. & Tanaka, I. (2002). *Acta Cryst.* **D58**, 1527–1530.
- Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* **D59**, 1943–1957.