

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

Amanda Abdalla Valério

**“Estudos estruturais e funcionais de duas glicosídeo hidrolases: a
celulase putativa XF0810 de *Xylella fastidiosa* e a lisozima digestiva 1
de *Musca domestica*”**

Tese apresentada ao Instituto de
Biologia para obtenção do Título de
Doutor em Biologia Funcional e
Molecular, na área de Bioquímica.

Orientador: Prof. Dr. João Alexandre Ribeiro Gonçalves Barbosa

Campinas, 2007

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

V237e

Valério, Amanda Abdalla

Estudos estruturais e funcionais de duas glicosídeo hidrolases: a celulase putativa XF0810 de *Xylella fastidiosa* e a lisozima digestiva 1 de *Musca domestica* / Amanda Abdalla Valério. – Campinas, SP: [s.n.], 2007.

Orientador: João Alexandre Ribeiro Gonçalves Barbosa.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Glicosídeo hidrolases. 2. Celulase. 3. *Xylella fastidiosa*. 4. Lisozima. 5. *Musca domestica*. I. Barbosa, João Alexandre Ribeiro Gonçalves. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

Título em inglês: Structural and functional studies of two glycosyl hydrolases: the putative cellulase XF0810 of *Xylella fastidiosa* and the digestive lysozyme 1 of *Musca domestica*.

Palavras-chave em inglês: Glycoside hydrolases; Cellulase; *Xylella fastidiosa*; Lysozyme; *Musca domestica*.

Área de concentração: Bioquímica.

Titulação: Doutora em Biologia Funcional e Molecular.

Banca examinadora: João Alexandre Ribeiro Gonçalves Barbosa, Celso Eduardo Benedetti, Jörg Kobarg; Richard Charles Garratt; Sônia Maria de Freitas.

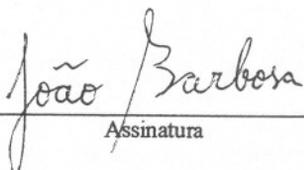
Data da defesa: 25/07/2007.

Programa de Pós-Graduação: Biologia Funcional e Molecular.

Campinas, 25 de julho de 2007

BANCA EXAMINADORA

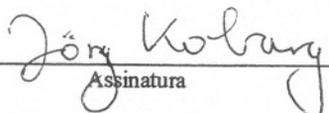
Prof. Dr. João Alexandre R.G. Barbosa (Orientador)


Assinatura

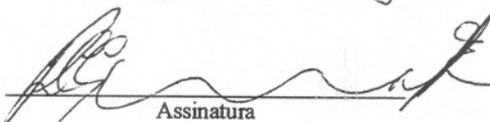
Prof. Dr. Celso Eduardo Benedetti


Assinatura

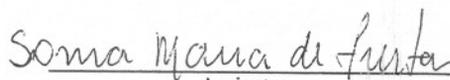
Prof. Dr. Jörg Kobarg


Assinatura

Prof. Dr. Richard Charles Garratt


Assinatura

Profa. Dra. Sônia Maria de Freitas


Assinatura

Profa. Dra. Paula Regina Kuser Falcão

Assinatura

Prof. Dr. Marcos Roberto de Mattos Fontes

Assinatura

Profa. Dra. Ana Paula Ulian de Araújo

Assinatura

Aos Meus Pais, Regina e José Adilson;
Ao meu irmão, Júnior (In memoriam);
Ao meu marido, Álvaro.

**Aos Meus Pais, Regina e José Adilson;
Ao meu irmão, Júnior (*in memoriam*);
Ao meu marido, Álvaro.**

Agradecimentos

Á Deus pelo dom da vida e por me proporcionar minha evolução terrena;

Á FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pela bolsa de doutorado.

Ao CNPq, LNLS (Laboratório Nacional de Luz Síncroton) pela infra-estrutura onde este trabalho de doutorado foi desenvolvido.

Ao Dr. João Alexandre Ribeiro Gonçalves Barbosa pela orientação e confiança.

Ao Prof. Dr. Sandro Marana e Fabiane Cançado pela oportunidade de realizar a colaboração deste trabalho.

À coordenadora da SCPG/Biologia Funcional e Molecular/IB/Unicamp, Profa. Dra. Helena Coutinho Franco de Oliveira pela acessoria.

Á secretária da SCPG/Biologia Funcional e Molecular/IB/Unicamp da Andréia Vigilato pelo apoio e acessoria nos trâmites burocráticos.

Aos membros da banca do exame de qualificação Prof. Dr. Hiroshi Aoyama, Prof. Dr. Ricardo Aparício e Prof. Dr. Jörg Kobarg pela avaliação realizada.

Aos membros titulares da banca Prof. Dr. Celso Eduardo Benedetti, Prof. Dr. Jörg Kobarg, Prof. Dr. Richard Charles Garratt, Profa. Dra. Sônia Maria de Freitas e aos suplentes Profa. Dra. Paula Regina Kuser Falcão, Prof. Dr. Marcos Roberto de Mattos Fontes e Profa. Dra. Ana Paula Ulian de Araújo, por terem aceitado o convite.

Aos amigos do grupo: Fábio, Priscila, Camila, Juliana, Tatiana, Carol, Andréia, Mário, Marcos e Dra. Beatriz Guimarães pela convivência e apoio científico.

Ao apoio técnico das técnicas do LNLS e especialmente Andréia nos experimentos.

A todos os amigos do LNLS especialmente a Givanil, Zildene, Daniel Trindade, Alexandre, Marcos Alborguetti, Natália, Andrés, Rosi, Beatriz, Sandra, Amadeu e aos amigos que conheci no LNLS: Valéria, Dra. Kelly, Aline, Rafael, Karine, Bruno, Dra. Patrícia Ribeiro, Dra. Leandra, Dra. Letícia, Leandro, Dr. Júlio pela amizade e apoio científico.

A acunpultura e o Dr. Gláucio Oliveria por me ajudar a manter o equilíbrio físico.

Aos amigos de Valinhos: Dr. João, Maria Inês, Regina, Irene, Iara, Moacir, Ana, Maura, Rita, Carlos, Vitor por proporcionarem meu crescimento espiritual, pelo apoio e amizade.

A minha maravilhosa família especialmente a Ana, Zezo, Bete, Gabriel, Valéria, D. Josefa, Vó Nair e Aneide e todos que torcem por mim, pelas orações e preces.

Ao meu querido marido Álvaro pelo apoio, carinho, incentivo e paciência nesta jornada científica.

Aos meus pais queridos que sempre me apoiaram e me deram forças e carinho para continuidade na área científica e ao meu irmão Junior (*in memoriam*) que foi o principal responsável pela minha escolha profissional nesta caminhada científica.

**“Comece fazendo o que é necessário, depois o que é possível,
e de repente você estará fazendo o impossível.”**

São Francisco de Assis

ÍNDICE

LISTA DE FIGURAS	x
LISTA DE TABELAS	xii
LISTA DE ABREVIACÕES	xiii
RESUMO	xv
ABSTRACT	xvii
1. INTRODUÇÃO	1
1.1. Glicosídeo hidrolases	1
1.2. Mecanismo de ação da glicosídeo hidrolases	3
1.3. Sítios de ligação do substrato da glicosídeo hidrolases	5
1.4. Celulases (E.C.3.2.1.4) (família 5)	7
1.5. Celulase de <i>Xylella fastidiosa</i> : XF0810	12
1.6. Lizozimas (E.C.3.2.1.17) (família 22)	13
1.7. Lizozimas digestivas	15
1.8. Lizozimas digestivas de <i>Musca domestica</i>	16
1.9. Estrutura terciária de lizozimas	16
2. OBJETIVOS	18
3. MATERIAIS E MÉTODOS	19
3.1. Materiais	19
3.1.1. Bactérias	19
3.1.2. Oligonucleotídeos	20
3.1.3. Vetores	20
3.1.4. Meios e soluções para o crescimento bacteriano	21
3.1.5. Soluções para extração de DNA plasmidial (Mini-prep)	22
3.1.6. Soluções para eletroforese	22
3.1.6.1. Eletroforese de DNA: Gel de agarose	22
3.1.6.2. Eletroforese de proteínas: SDS-PAGE	22
3.1.7. Tampão utilizado para ensaio enzimático	23
3.2. Métodos	24
3.2.1. Biologia Molecular	24
3.2.1.1. Reação em cadeia da polimerase (PCR)	24
3.2.1.2. Eletroforese em gel de agarose	25
3.2.1.3. Purificação de fragmentos de DNA e de produtos de PCR	26
3.2.1.4. Reação de ligação de DNA - Clonagem	26
3.2.1.5. Digestão de DNA com enzimas de restrição	26
3.2.1.6. Células competentes	27

3.2.1.7. Transformação bactérias competentes	27
3.2.1.8. PCR de colônia	28
3.2.1.9. Extração de DNA plasmideal	28
3.2.1.10. Sequenciamento de DNA	29
3.2.2. Expressão de proteínas	30
3.2.2.1. Teste de expressão de proteínas	30
3.2.2.2. Eletroforese de proteína (SDS-PAGE)	31
3.2.2.3. Expressão de proteínas	31
3.2.3. Purificação de proteínas	32
3.2.3.1. Purificação de proteínas por cromatografia	32
3.2.3.2. Diálise	33
3.2.3.4. Concentração de proteína	34
3.2.4. Análises espectroscópicas e espectrométricas	34
3.2.4.1. Determinação da concentração de proteína	34
3.2.4.2. Espectrometria de massa	34
3.2.4.3. DLS	36
3.2.4.4. Filtração em gel analítica	36
3.2.4.5. Dicroísmo circular	37
3.2.4.6. Ensaio enzimático – atividade exocelulase	38
3.2.5. Cristalização de proteínas	39
3.2.5.1. Ensaio de cristalização	39
3.2.5.2. Ensaio de cristalização com robôs	39
3.2.5.3. Soaking	40
3.2.6. Modelagem Molecular	41
3.2.7. Cristalografia de proteínas	42
3.2.7.1. Difração de raios X	42
3.2.7.2. Processamento dos dados	42
3.2.7.3. Substituição molecular	43
3.2.7.4. Refinamento	43
4. RESULTADOS E DISCUSSÃO	45
4.1. Celulase	46
4.1.1. Amplificação, clonagem e subclonagem	46
4.1.2. Expressão	49
4.1.3. Purificação	51
4.1.4. Espalhamento dinâmico de luz (DLS)	55
4.1.5. Gel filtração analítica	56
4.1.6. Dicroísmo circular	59
4.1.7. Ensaio enzimático: exocelulase	61
4.1.8. Ensaio de cristalização	63
4.1.9. Modelagem Molecular	64
4.2. Lisozima	73
4.2.1. Cristalização e soaking	73
4.2.2. Difração de raios-X e processamento de dados	74
4.2.3. Substituição molecular e refinamento	78

6. CONCLUSÕES	93
6.1. Celulase	93
6.2. Lisozima	94
7. REFERÊNCIAS BIBLIOGRÁFICAS	96
8. ANEXOS	103
8.1. Clorose variegada de citros (CVC) e <i>Xylella fastidiosa</i>	103
8.2 Artigos: publicado, aceito e em preparação	108

Lista de Figuras

Figura 1. Os dois principais mecanismos da hidrólise enzimática de ligações glicosídicas.	4
Figura 2. Desenho esquemático dos subsítios de ligação de açúcares em diversas glicosil-hidrolases.	6
Figura 3. Mecanismo de ação celulasas	9
Figura 4. Enovelamento barril (α/β) ₈ e a estrutura do barril-TIM da celulase de <i>Acidothermus cellulolyticus</i> .	12
Figura 5. Dois possíveis mecanismos catalíticos para HEWL	14
Figura 6. Estrutura da lisozima 1HEW	17
Figura 7. Seqüência XF0810c.	47
Figura 8 Gel de agarose 1 %. Amplificação do gene da celulase XF0810c.	48
Figura 9. Gel de agarose 1 %. Digestão do plasmídeo pGEM-XF0810c.	49
Figura 10. Gel de agarose 1 %. PCR colônia do plasmídeo pET29a-XF0810c.	49
Figura 11. SDS-PAGE 12% da expressão XF0810c.	50
Figura 12. Seqüência XF0810c, seqüência do corte da tripsina.	50
Figura 13. Cromatograma da purificação por cromatografia de troca iônica da XF0180c, coluna HiTrap QFF.	52
Figura 14. SDS-PAGE 12% da purificação por cromatografia de troca iônica XF0810c.	53
Figura 15. Cromatograma da purificação por cromatografia de filtração em gel (fração lavagem da QFF) da XF0180cel, coluna Superdex 16/60	54
Figura 16. SDS-PAGE 12% da purificação por cromatografia de filtração em gel XF0180c.	55
Figura 17. DLS dos picos resultantes da purificação em gel filtração XF810c, (a) Pico 1 e (b) Pico 2.	56
Figura 18. Cromatograma da purificação de filtração em gel analítica (amostra fração lavagem da QFF) da XF0180c, coluna Superdex 16/60.	57
Figura 19. Regressão linear e curva de calibração geradas a partir da filtração em gel analítica.	58
Figura 20. Espectro do dicróismo circular da proteína Xf0810cel medida no UV distante.	60
Figura 21. Resultado do teste de atividade enzimática para XF0810c.	62
Figura 22. Resultado do teste de atividade enzimática para o controle positivo da celulase de <i>Tricoderma reesei</i> .	62
Figura 23. Fotos dos cristais da proteína XF0810cel nas gotas do kit JCSG (Molecular Dimensions) (a) microcristais - condição D4 e (b) agulhas - condição E1.	64
Figura 24. Alinhamento das celulasas XF0810c e XF0818 com a celulase 1ECE (<i>Acidothermus cellulolyticus</i>)	65
Figura 25. Gráfico verify plot do modelo XF0810c	67
Figura 26. Plot de Ramachandran do modelo XF0810c	69
Figura 27. Stereoview do modelo da XF0810c com o ligante	70
Figura 28. Stereoview da 1ECE	71

Figura 29: Stereoview do modelo da XF0810c	71
Figura 30: Sobreposição dos resíduos catalíticos	73
Figura 31. Cristal da lisozima 1 de <i>Musca domestica</i> .	74
Figura 32. Padrão de difração da lisozima 1 complexada com (a) Q3 1,85 Å, (b) Q4 1,92 Å, (c) Q6 2,25 Å.	76
Figura 33. Representação esquemática da estrutura de lisozima 1 de <i>Musca domestica</i> .	81
Figura 34. Monômeros MdL1 com açúcar	81
Figura 35. Superfície da molécula MdL1: zoom na cavidade para entrada do ligante Q4.	82
Figura 36. Alinhamento múltiplo de estruturas representativas de lisozimas c: 1JSF, 1LMN, 1HEW, 1GD6, 1IIZ e 2FDB.	83
Figura 37. Stereoview do enovelamento da lisozima (família 22)	85
Figura 38. Zoom no loop (98-100) da lisozima c com o ligante.	87
Figura 39. Esquema átomos de açúcar e das interações com o ligante	90
Figura 40. Superposição da MdL1 e HEWL mostrando as diferenças na ligação com o ligante.	91
Figura 41. Stereoview da MdL1 das interações do ligante Q4 no sítio ativo da enzima	92
Figura 42. Sintomas da doença clorose variegada de citros (CVC).	104
Figura 43. Microscopia eletrônica de <i>Xylella fastidiosa</i> em vaso do xilema.	106

Lista de Tabelas

Tabela 1. Seqüência de oligonucleotídeos iniciadores desenhados para amplificação e clonagem.	20
Tabela 2. Meios para o crescimento bacteriano.	21
Tabela 3. Soluções para o crescimento bacteriano.	22
Tabela 4. Estimativa de estrutura secundária da proteína XF0810cel, programa CDNN.	60
Tabela 5. Estimativa de estrutura secundária da proteína XF0810c pelos programas Dicroprot	60
Tabela 6. Aminoácidos conservados celulasas 1ECE e XF0810c.	66
Tabela 7. Estatística dos dados obtidos da coleta dos cristais de lisozima 1 de <i>Musca domestica</i> com os ligantes Q3, Q4 e Q6.	77
Tabela 8. Superposição dos representantes das lisozimas da família	86
Tabela 9. pKa teórico dos resíduos catalíticos da MdL1 calculados pelo programa PROPKA.	88
Tabela 10. Lista dos genes selecionados de <i>X. fastidiosa</i> .	107

Lista de abreviações

ϵ	Coeficiente de extinção molar
λ	Comprimento de onda
θ	Elipticidade
Å	Ångström (1×10^{-10} metros)
3D	Tridimensional
ACN	Acetonitrila
APS	Persulfato de amônio
BLAST	<i>Basic Local Alignment Search Tool</i>
C	Concentração
°C	Graus Celsius
CTO	Triacetilchitotriose
DLS	Espalhamento dinâmico de luz
DNA	Ácido desoxirribonucleico
DTT	1,4 ditiotreitól
E.C.	<i>Enzyme Commission Number</i>
EDTA	Ácido etilenodiaminotetracético
FPLC	<i>Fast Performance Liquid Chromatography</i>
HEWL	<i>Hen Egg White Lysozyme</i>
IPTG	Isopropil- β -D-tiogalactopiranosídeo
kDA	Quilo Dalton
l	Comprimento do caminho ótico
LB	Meio de cultura Luria Bertani
LNLS	Laboratório Nacional de Luz Síncroton
mM	Mili Molar (massa molecular)
MdL1	Lisozima 1 de <i>Musca domestica</i>
MÊS	Morfolinoetanosulfônico
MRW	Peso médio do resíduo
MW	Peso molecular
NAM	Ácido n-acetilmurâmico
NAG	N-acetilglucosamina
(NAG)₃	Tri-n-acetilglucosamina
Nm	Nanômetro (10^{-9} metros)
OD_{600nm}	Densidade ótica a 600nm
Pb	Pares de bases
PCR	Reação da polymerase em cadeia (<i>Polimerase Chain Reaction</i>)
PDB	Protein Data Bank
PEG	Polietilenoglicol
pH	Potencial hidrogeniônico
pKa	Constante
Pnpc	P-nitrofenilfosfato - celobiosídeo
Q3	Metilumbeliferilchitotriosídeo
Q4	Chitotetraose
Q5	Chitopentaose

Q6	Chitohexaose
rpm	Rotações por minuto
SDS	Dodecil-sulfato de sódio
SDS-PAGE	<i>SDS polyacrilamide gel electrophoresis</i>
TAE	Tampão tris ácido-acético EDTA
TB	Meio de cultura Terrific Broth
TEMED	N,n,n',n' tetrametil etilenodiamina
TFA	Ácido triflúor acético
TIM	Triose fosfato isomerase
Tm	Temperatura média de transição
Tris	Tris-hidroximetil aminometano
U	Unidade enzimática
UV	Ultravioleta
XF	<i>Xylella fastidiosa</i>
X-gal	5-bromo-4cloro-3-indolil- β -D-galactosídeo

Resumo

Glicosídeo hidrolases (EC 3.2.1.-) são enzimas que hidrolisam ligações glicosídicas. Neste trabalho realizamos estudos funcionais e estruturais de duas glicosídeo hidrolases: a celulase XF0810 de *Xylella fastidiosa* e a lisozima digestiva 1 de *Musca domestica* (MdL1). A celulase XF0810 está anotada no genoma da *X. fastidiosa* como membro da família 5 das glicosídeo hidrolases (EC 3.2.1.4). Após a amplificação, clonagem, expressão e purificação da mesma; prosseguimos com os experimentos de filtração em gel analítica, dicroísmo circular, DLS, ensaio enzimático e cristalização desta proteína. Foi feito um estudo de modelagem onde ficou evidenciado que a XF0810 não pertenceria à família 5 pois quatro dos sete resíduos conservados que caracterizam esta família foram substituídos, incluindo os dois resíduos catalíticos de glutamato essenciais para o mecanismo de clivagem de retenção. Isto foi comprovado através da ausência de atividade nos ensaios feitos com a proteína purificada. A MdL1 pertence à família 22 das glicosídeo hidrolases (EC 3.2.1.17) e foi cristalizada com o ligante N-acetilquitotetraosídeo para a difração de raios X. A resolução da estrutura (2H5Z no PDB) foi realizada por meio do método de substituição molecular tendo como modelo a estrutura nativa. A análise comparativa da MdL1 com outras lisozimas de quatro classes diferentes de animais mostrou grande semelhança e pequenas diferenças apenas na região das voltas. Estas diferenças foram utilizadas para explicar as características especiais de uma lisozima com função digestiva. A volta na região definida pelos resíduos 98-100 apresenta uma deleção na MdL1,

tornado-a menos exposta ao solvente, podendo justificar a resistência à proteólise. O resíduo Gln100 participa de uma interação com o ligante. Os resíduos Thr107 e Asn46 são apontados como responsáveis pelo decréscimo dos pK_a dos grupos carboxilas dos resíduos catalíticos Glu32 e Asp50, respectivamente. A diminuição dos pK_a explica o pH ótimo mais ácido característico de lisozimas digestivas.

Abstract

Glycoside hydrolases (EC 3.2.1.-) are enzymes that hydrolyze glycoside bonds. In this work we studied functional and structural features of two glycoside hydrolases: the cellulase XF0810 of *Xylella fastidiosa* and the digestive lysozyme 1 of *Musca domestica* (MdL1). The cellulase XF0810 is annotated in the genome of *X. fastidiosa* as member of family 5 of glycoside hydrolases (EC 3.2.1.4). After the amplification, cloning, expression and purification of XF0810; we continued with the experiments of analytical gel filtration, circular dichroism, DLS, enzymatic assay and crystallization of this protein. A homology model was built which showed that XF0810 did not belong to family 5 because four of seven conserved residues that characterize the family were substituted, including the two catalytic residues of glutamate that are essential for the retention hydrolysis mechanism. This was further confirmed by the absence of activity in the assays (exocellulase with PNPc) performed with the purified protein. The MdL1 belongs to the family 22 of glycoside hydrolases (EC 3.2.1.17) and was crystallized with the ligand N-acetylchitotetraose for X-ray diffraction. The resolution of the structure (2H5Z in PDB) was accomplished by molecular replacement with the native structure as the searching model. The comparative analysis of MdL1 with other lysozymes of four different classes of animals showed a high similarity and few differences appeared only in the loops' regions. These differences were used to explain the special characteristics of the lysozyme with a digestive function. The loop in the region defined by residues 98-100 presents one deletion in the MdL1, becoming less

exposed to the solvent, this might justify the proteolysis resistance. The residue Gln100 participates directly in an interaction with the ligand. The residues Thr107 and Asn46 are pointed out as responsible for a reduction in the pK_a of the carboxyl groups of catalytic residues Glu32 e Asp50, respectively. The reduction in pK_a explains the more acidic pH optimum that characterizes the digestive lysozymes.

1. Introdução

1.1. Glicosídeo hidrolases

Os carboidratos nas formas de mono-, di-, oligo- e polissacarídeos desempenham um papel único na natureza. O número de possíveis combinações para um pequeno oligossacarídeo resulta em uma diversidade estrutural e funcional para estes compostos. Oligo- e polissacarídeos desempenham um papel central na cadeia de processos biológicos tais como estoque e utilização de alimento, invasão viral e eventos altamente seletivos de sinalização celular. As enzimas que hidrolisam estes açúcares - glicosídeo hidrolases - também estão implicadas em um amplo espectro de processos biológicos. As interações entre carboidrato e proteínas tiveram um significativo avanço em particular na área clínica, várias estruturas de cristais foram resolvidas de enzimas relacionadas ao metabolismo de carboidratos proporcionando assim, um ponto de partida para o desenho racional de drogas. Como exemplo relacionado com a área médica podemos citar os defeitos nos genes que codificam glicosídeo hidrolases que são responsáveis por muitas doenças genéticas como desordens de estoque lisossomal. Para esses casos, inibidores específicos de glicosídeo hidrolases podem ser investigados quanto ao potencial agente terapêutico (von Itzstein & Colman, 1996; Davies & Henrissat, 1995).

Vários sistemas de classificação existem para glicosídeo hidrolases baseados na especificidade do produto ou do substrato, modo de ataque (exo/endo) e mecanismos estereoquímicos. Em 1991, a classificação de

glicosídeo hidrolases baseando-se nas similaridades de seqüências de aminoácidos foi introduzida (Henrissat, 1991). Esta classificação é atualizada frequentemente (Henrissat & Bairoch, 1993) (Henrissat & Bairoch, 1996), e fornece cerca de 60 famílias de glicosídeo hidrolases baseada na similaridade seqüencial tendo 27 famílias com estrutura conhecida.

As glicosídeo hidrolases (E.C 3.2.1.x) são a chave para o metabolismo de carboidratos e são encontradas nos três maiores reinos (archeabacteria, eubactéria e eucariotos). A nomenclatura proposta pela União Internacional de Bioquímica e Biologia Molecular é baseada no tipo de reação que a enzima catalisa ou em sua especificidade pelo substrato. Para glicosídeo hidrolases, EC 3.2.1.x, os três primeiros dígitos indicam que estas enzimas hidrolisam ligações o-glicosil, enquanto o último número indica a especificidade do substrato e as vezes reflete o mecanismo molecular. Henrissat & Davies, em 1997, mostraram as limitações da utilização deste sistema de classificação, pois este não reflete as características estruturais da enzima, o substrato e seus produtos estereoquímicos. O termo exo/endo providencia uma terminologia adicional, mas freqüentemente é difícil de determinar. A estereoquímica da hidrólise é de extrema importância, mas não é uma base para a classificação global. A classificação de glicosídeo hidrolases em famílias baseadas na similaridade de seqüência de aminoácidos é complementar a outras classificações e juntamente com a determinação de estrutura 3D proporciona uma ferramenta de predição para os cientistas modernos (Henrissat & Davies, 1997).

1.2. Mecanismo de ação das glicosídeo hidrolases

A hidrólise enzimática de ligações glicosídicas ocorre por meio da catálise ácida que requer dois resíduos de aminoácidos: um doador de prótons e um nucleófilo/base (Koshland, 1953; Sinnott, 1990). Essa hidrólise é usualmente promovida por resíduos de aspartato e/ou glutamato, encontrados em sítios ativos e altamente conservados durante a evolução das glicosídeos hidrolases (Gilkes et al., 1991). A hidrólise ocorre via dois mecanismos principais, retenção ou inversão, na configuração anomérica do átomo de carbono do glicosídeo hidrolisado (Koshland, 1953). Em ambos os mecanismos, retenção ou inversão, a posição do doador de prótons é idêntica. Nas enzimas que utilizam o mecanismo de retenção, a base catalítica nucleófilica é próxima do carbono anomérico do açúcar. Esta base, por outro lado, é mais distante nas enzimas que utilizam o mecanismo de ação de inversão o qual deve acomodar uma molécula de água entre a base e o açúcar. Essa diferença resulta em uma distância média entre os dois resíduos catalíticos de $\sim 5,5 \text{ \AA}$ nas enzimas de retenção é de $\sim 10 \text{ \AA}$ nas enzimas de inversão (figura 1).

As enzimas estudadas neste trabalho (celulase e lisozima) utilizam somente o mecanismo de retenção que se inicia com a ligação inicial do substrato a enzima, seguido do ataque catalisado pelo ácido do grupo nucleófilo da enzima ao centro anomérico do substrato, formando o intermediário glicosil-enzima (McCarter & Withers, 1994, Sinnott 1990). Este intermediário é então hidrolisado por um ataque catalisado por base, feito por uma molécula de água ao centro anomérico

do substrato formando o produto e retornando a enzima ao estado de protonação original (McCarter & Withers, 1994, Sinnott 1990).

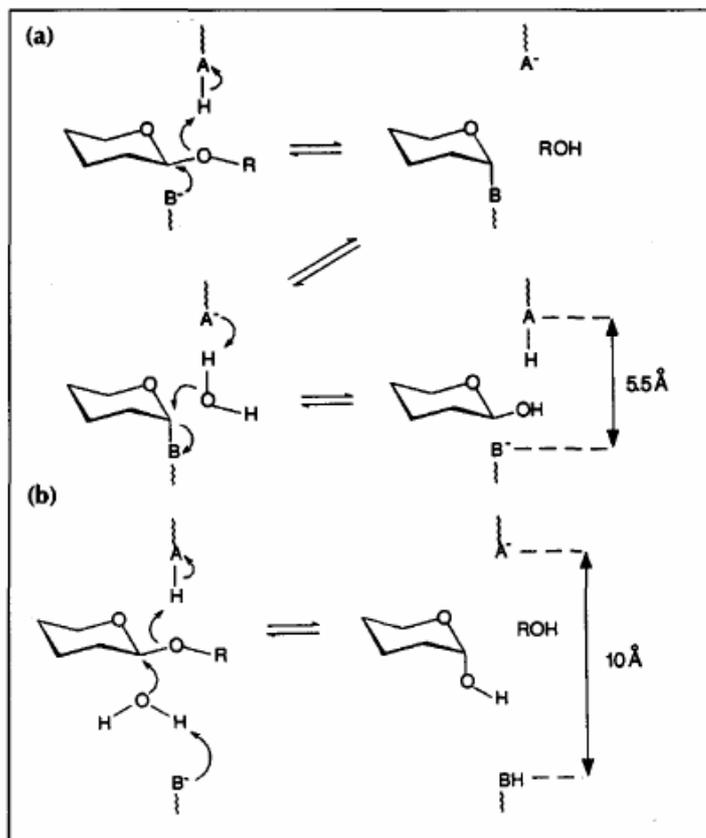


Figura 1: Os dois principais mecanismos da hidrólise enzimática de ligações glicosídicas que foram primeiramente propostos por Koshland. (a) No mecanismo de retenção o oxigênio glicosídico é protonado por uma catálise ácida (AH) e a assistência nucleofílica para o aglicon é providenciada pela base B⁻. O resultado glicosil enzima é hidrolisada por uma molécula de água e sua segunda substituição nucleofílica no carbono anomérico gera um produto com a mesma estereoquímica do substrato. (b) O mecanismo de inversão, no qual a protonação do oxigênio glicosídico e a liberação do aglicon é acompanhada por ataque concomitante de uma molécula de água que é ativada pelo resíduo base (B⁻). Esta simples substituição nucleofílica resulta em um produto com estereoquímica oposta ao substrato (Davies & Henrissat, 1995).

1.3. Sítios de ligação do substrato das glicosídeo hidrolases

O mapeamento dos subsítios de ligação de açúcares nas glicosídeo hidrolases iniciou-se em 1960 com estudos de enzimas amilolíticas. Os subsítios das enzimas eram marcados por i , $i+1$, $i+2$ até $i+n$, com os números aumentando positivamente na direção da extremidade redutora do substrato. A razão foi que em química de açúcares, os oligômeros são desenhados por convenção com a extremidade não redutora no lado esquerdo e a redutora no direito. A figura 2a ilustra uma enzima exo-amilolítica marcada com os subsítios 1, 2, 3, 4, 5 e 6 (da extremidade não redutora a redutora) com a clivagem entre subsítios 1 e 2. Este sistema é inapropriado para exoamilases que clivam a terminação não redutora do açúcar. Além disso, a descoberta de subsítios anteriores do subsítio 1 obrigou alguns autores a incluírem subsítios negativos.

Em 1978 Suganuma e colaboradores continuaram utilizando a nomenclatura 1, 2, 3...n e simultaneamente introduziram o conceito de marcação de subsítios relativos ao ponto de clivagem enzimática. Subsítios são marcados como $-n$ representando a extremidade não redutora e n a redutora e a clivagem ocorre entre os subsítios $+1$ e -1 . Então os subsítios na extremidade redutora são marcados $+1$, $+2$, $+3$ até $+n$ e na extremidade não redutora -1 , -2 , -3 até $-n$ (figura 2c-g).

A nomenclatura alfabética dos subsítios de glicosídeo-hidrolases foi introduzida por cristalógrafos Phillips e colaboradores (1967) por meio da análise da estrutura da HEWL (Lysozyme Hen Egg White), onde foram identificados seis subsítios de ligação para sacarídeos marcados de A-F, no qual A é a extremidade

não redutora e F é a extremidade redutora do substrato. O ponto de clivagem é entre os subsítios D e E (figura 2b).

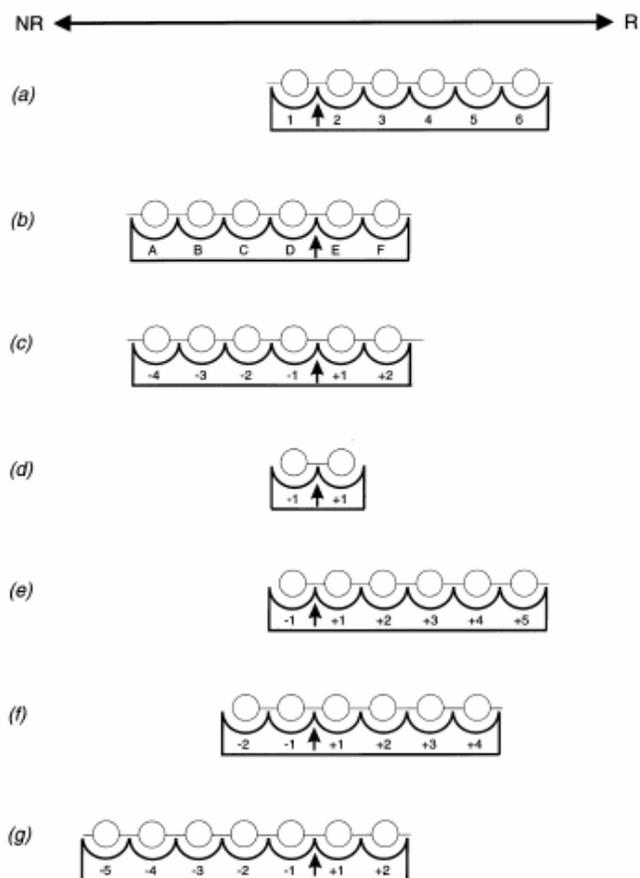


Figura 2: Desenho esquemático dos subsítios de ligação de açúcares em diversas glicosil-hidrolases. Por convenção as extremidades não redutoras são desenhadas na esquerda e as extremidades redutoras na direita. Os pontos de clivagem estão indicados pelas setas. (a) subsítio marcado por números aplicados à glucoamilase; (b) subsítios marcados por letras aplicados a HEWL; (c) subsítios da HEWL marcados pelo esquema proposto $-n$ para $+n$; (d) o sistema $-n$, $+n$ aplicado para monoglicosidades não específicas e dissacaridases; (e) o sistema $-n$, $+n$ aplicados para enzimas que clivam monossacarídeos da extremidade não redutora do substrato tais como glucoamilase; (f) o sistema $-n$, $+n$ aplicado para enzimas que clivam unidades de dissacarídeos da extremidade não redutora do substrato tais como β -amilase; (g) o sistema $-n$, $+n$ aplicado para enzimas que clivam unidades de dissacarídeo da extremidade redutora do substrato proposta para celobiohidrolase I de *T. reesei* (Davies *et al.*, 1997).

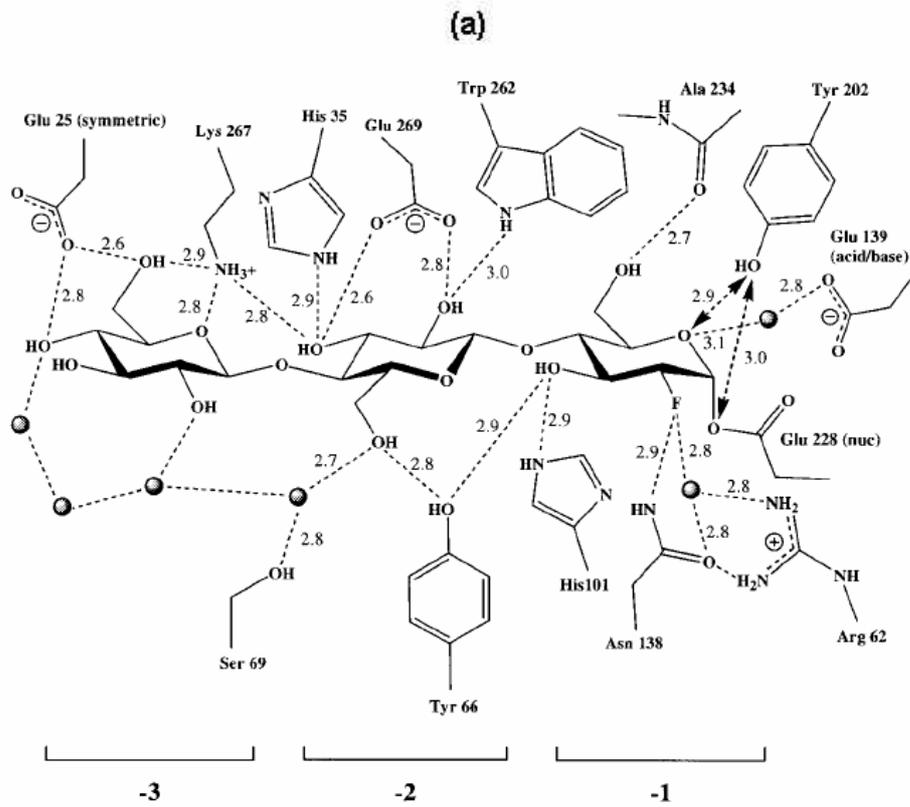
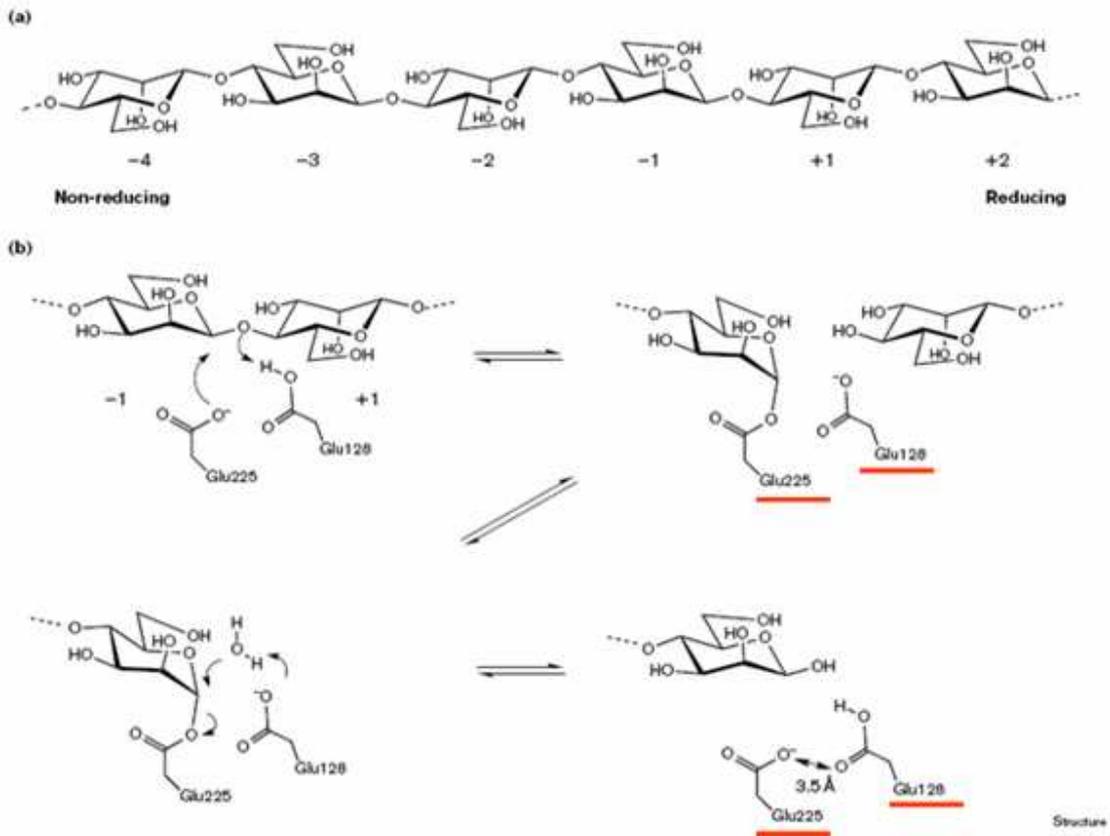
1.4. Celulases (E.C.3.2.1.4) (família 5)

As celulases são enzimas que catalisam a hidrólise de ligações β -1,4 glicosídicas entre os resíduos de glicose da celulose. A família 5 é uma das maiores famílias de β -glicohidrolase (Sakon *et al.*, 1996). A seqüência dos membros da família 5 é diversa e relatada por manter sete resíduos conservados próximos ao sítio catalítico (Wang *et al.*, 1993) equivalentes a Arg62, His116, Asn161, Glu162, His238, Tyr240 e Glu282, numeração de *Acidothermus cellulolyticus* (Sakon *et al.*, 1996). Essas famílias apresentam mecanismo de clivagem de retenção (Davies & Henrissat, 1995) proposto inicialmente por Koshland (1953). Dois resíduos conservados da família 5 estão diretamente envolvidos na reação: o Glu282 foi identificado como nucleófilo (Tull *et al.*, 1991) e o Glu162 foi denominado como doador de próton, base B⁻ e ácido AH respectivamente (figura 1 e 3).

Na figura 3a podemos ver um exemplo do mecanismo de ação catalisado pela enzima β -mananase do organismo *Thermomonospora fusca* no substrato específico para hidrólise de ligações O-glicosídicas. Neste exemplo é mostrado o mecanismo de retenção da configuração anomérica do substrato, no qual o oxigênio glicosídico (resíduo +1) é protonado pelo doador de prótons (Glu128) e o átomo de carbono anomérico (resíduo -1) é atacado pelo nucleófilo (Glu225), resultando num intermediário (manosil-mananase) que é então hidrolisado por uma molécula de água gerando o produto com a mesma configuração anomérica do substrato (Hilge *et al.*, 1998). A figura 3b ilustra a representação esquemática

da rede de ligações de hidrogênio enzima-oligossacarídeo no complexo covalente intermediário da celulase de *Bacillus agaradharens* que adquire a conformação de cadeira, as moléculas de água que são importantes para a ligação com o substrato e para manutenção do mecanismo de clivagem estão representadas pelas esferas, (Varrot *et al.*, 2000). Na figura 3c podemos observar os estados de transição do mecanismo catalítico que foi esquematizado assumindo o mínimo movimento do substrato e a não movimentação dos átomos da proteína da celulase de *Acidothermus cellulolyticus* (Sakon *et al.*, 1996).

No Protein Data Bank (PDB) (Bernstein *et al.*, 1997) estão depositadas estruturas de celulasas de vários organismos e famílias. Especificamente na família 5 das celulasas encontramos estruturas dos organismos: 1ECE (*Acidothermus cellulolyticus*), 1H1N (*Thermoascus aurantiacus*), 1BQC (*Thermomonospora fusca*), 1H5V (*Bacillus agaradharens*) e 1CEC (*Clostridium thermocellum*). Essas famílias apresentam o enovelamento comum de barril (β/α)₈, a estrutura de um barril-TIM, na figura 4 podemos observar a estrutura celulase 1ECE.



(b)

Legenda na próxima página.

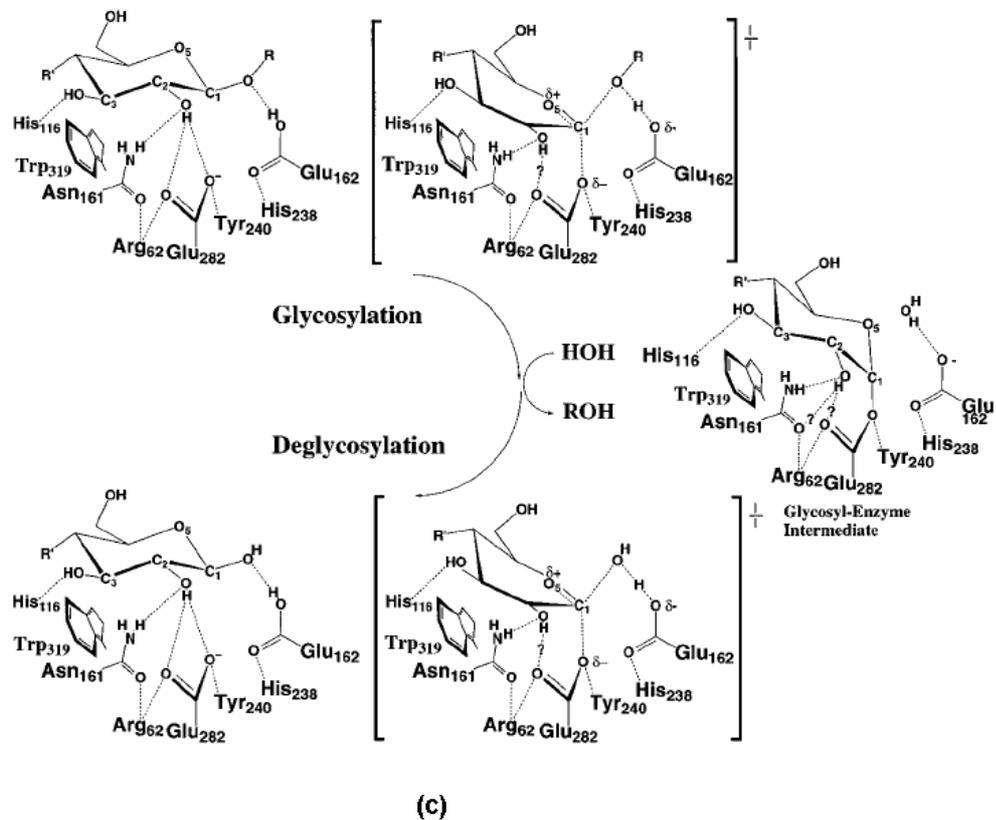


Figura 3: (a) Reação catalisada da enzima β -mananase do organismo *Thermomonospora fusca* (Hilge *et al.*, 1998), os resíduos catalíticos estão sublinhados em vermelho (b) Representação esquemática da rede de ligações de hidrogênio enzima-oligossacarídeo da celulase de *Bacillus agaradharens* (Varrot *et al.*, 2000). (c) Estados de transição da reação da celulase de *Acidothermus cellulolyticus* (Sakon *et al.*, 1996).

Nas estruturas de proteínas, em geral, os domínios mais freqüentes são os domínios alfa-beta (α/β) que consistem em um núcleo de fitas β (paralelas ou torcidas) envoltas por hélices α . A maioria das enzimas glicolíticas são estruturas α/β . Nestes domínios a região formada por loops participa na ligação e ação catalítica. Existem três principais classes de proteínas α/β , na primeira classe (onde está inserida a celulase da família 5) é descrita como um núcleo de fitas β torcidas e paralelas organizadas fechadas juntas na forma de barril, as hélices α que conectam as fitas β estão fora deste barril. Este domínio estrutural é chamado de barril-TIM devido a estrutura da enzima triosefosfato isomerase, onde foi primeiramente observado. A estrutura do barril (α/β)₈ é encontrada em muitas proteínas diferentes, na maioria enzimas com seqüências de aminoácidos e funções completamente diferentes. Nestas estruturas as fitas β e alfa hélices α formam um “esqueleto” estrutural da enzima enquanto os loops/voltas contém os aminoácidos responsáveis pela química catalítica. No barril as cadeias laterais hidrofóbicas das hélices α estão empacotadas contra as cadeias laterais hidrofóbicas das fitas β . As hélices α são antiparalelas e adjacentes às fitas β que as conectam. Assim o barril é como se fosse uma concha de resíduos hidrofóbicos vindos da hélice α e folha β . O outro lado das cadeias das fitas β apontam para o interior do barril formando um núcleo hidrofóbico, este núcleo então consiste exclusivamente dos resíduos das cadeias laterais das fitas β . O empacotamento das interações entre hélices α e fitas β são dominados por resíduos de aminoácidos hidrofóbicos. O sítio ativo em todo barril α/β é um bolsão formado pela região do loop que conectam as carboxilas terminais das fitas β com os aminos terminais das hélices α .

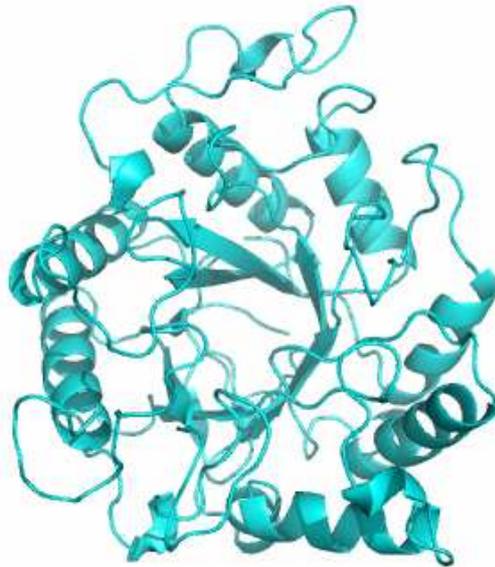


Figura 4: Enovelamento barril TIM (α/β)₈ da estrutura da celulase de *Acidothermus cellulolyticus* (Sakon *et al.*, 1996).

1.5. Celulase de *Xylella fastidiosa*: XF0810

No ano de 2000 o genoma da *X. fastidiosa* foi seqüenciado (Simpson *et al.*, 2000), sendo este o primeiro genoma de um fitopatógeno, representando destaque da ciência brasileira no cenário mundial. Para o estudo do projeto foram selecionados genes da *X. fastidiosa* que codificam hidrolases (celulases, glicosidases e peptidases), vide anexo I. Devido às dificuldades encontradas para a realização do projeto inicial tais como a expressão das proteínas na forma insolúvel, concentramos nossas atividades na celulase XF0810. As celulases são importantes para a *X. fastidiosa*, pois estão associadas à obtenção de glicose e colonização do sistema vascular da planta, afetando, desta forma o desenvolvimento das plantas infectadas. Especula-se que as celulases são

utilizadas para romper as paredes da membrana primária das pontuações, permitindo a migração das bactérias de um vaso a outro.

Como foi anotado no genoma da *X. fastidiosa*, o gene XF0810 codificaria a enzima celulase que pertenceria a família 5 das celulases. Fizemos um alinhamento estrutural da celulase XF0810 e ao analisarmos o alinhamento, verificamos a não conservação de 4 dos 7 resíduos importantes que participam do sítio ativo da família 5 de celulases .

1.6. Lisozimas (E.C.3.2.1.17) (família 22)

As lisozimas são enzimas que catalisam a hidrólise de ligações glicosídicas β -1,4 entre o ácido n-acetilmurâmico (NAM) e resíduos de n-acetilglucosamina (NAG) de peptidoglicanas da parede celular bacteriana. Baseados na seqüência e na similaridade estrutural, estas enzimas são classificadas nas famílias 22, 23, 24, 25 e 73 de glicosídeo hidrolases (Coutinho & Henrissat, 1999). Na família 22, onde existem 219 lisozimas encontradas em vertebrados, invertebrados e fungos, a lisozima da clara de ovo de galinha (Hen Egg White: HEWL) é o membro melhor caracterizado desta família.

A lisozima HEWL foi a primeira estrutura de uma enzima resolvida pelo método de difração de raios X (Blake *et al.*, 1965), a estrutura cristalográfica do complexo co-cristalizado com tri-N-acetilglucosamina (NAG)₃ resultou em propostas para o mecanismo de ação catalítico com a indicação de dois resíduos catalíticos, Glu35 e Asp52. Analisando o mecanismo de ação proposto, foi verificado que o resíduo catalítico Glu35 age como ácido protonando o oxigênio glicosídeo enquanto o outro resíduo catalítico Asp52 age como grupo estabilizador

eletrostático do íon carbonium intermediário (Phillips, 1967). Recentemente este mecanismo foi revisado e uma catálise via um intermediário covalente (glicosil-enzima) formado com o resíduo catalítico aspartato foi proposto (Vocadlo et al., 2001). A carboxila do Glu35 atua como ácido e protona o oxigênio da ligação glicosídica, o Asp52 efetua um ataque nucleofílico ao carbono do substrato formando um intermediário “glicosil-enzima”. Posteriormente o Glu35 desprotonado atua como base ativando uma molécula de água que realiza um ataque nucleofílico sobre o intermediário covalente regenerando o Asp52 e liberando o segundo produto (figura 5).

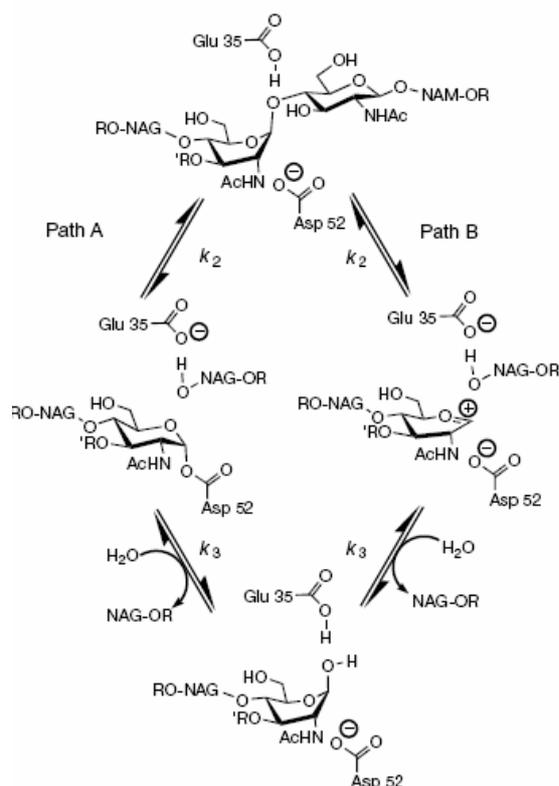


Figura 5: Dois possíveis mecanismos catalíticos para HEWL. Via A o mecanismo de Koshland proposto por Vocadlo e via B o mecanismo de Phillips. R, cadeia oligossacarídeo, R', cadeia lateral peptídeo (Vocadlo *et al.*, 2001).

O mecanismo de catálise das lisozimas da família 22 depende da ação dos dois resíduos catalíticos com grupos carboxila em estados de ionização diferentes, o Glu35 apresenta um pKa em torno de 6,5 enquanto o Asp 52 apresenta um pKa próximo de 3,5 (Yang & Hamaguchi, 1980; Mckenzie & White, 1991), logo no pH ótimo da enzima (em torno de 5,0), o Asp52 encontra-se desprotonado e o Glu35 está protonado.

1.7. Lisozimas digestivas

Lisozimas da família 22 geralmente fazem parte do sistema de defesa contra bactéria (Jollès & Jollés, 1984), entretanto, algumas delas estão também envolvidas na digestão de bactérias e são encontradas nos vertebrados que tem uma câmara de fermentação em seus intestinos (por exemplo, ruminantes como *Bos taurus*, Dobson *et al.*, 1984) e em insetos que se alimentam de material orgânico em decomposição (por exemplo, a mosca doméstica *Musca domestica* e mosca de fruta *Drosophila melanogaster*, Lemos & Terra, 1991; Prager, 1996). Estas lisozimas apresentam diversas adaptações à função digestiva tais como um alto nível de expressão no intestino, resistência à hidrólise por proteinases presentes no trato digestivo e atividade bacteriolítica com pH ótimo ácido (Dobson *et al.*, 1984). Resistência à inativação pelas proteinases ácidas foi descrita para lisozima de vaca, a qual é aproximadamente seis vezes mais resistente à pepsina do que a HEWL (Dobson *et al.*, 1984). Alinhamentos de seqüências de lisozimas digestivas têm sido usados para propor uma base molecular destas adaptações onde foi verificada a redução do número de ligações peptídicas lábeis a ácidos e uma diminuição no número de aminoácidos básicos como arginina (Prager, 1996).

1.8. Lisoziimas digestivas de *Musca domestica*

Duas lisoziimas digestivas são encontradas no intestino médio da *Musca domestica*: lisoziima 1, AAQ20048 e lisoziima 2, AAQ20047 (Cançado *et al*, 2006). Baseado na similaridade seqüencial com a HEWL, as lisoziimas 1 e 2 de *Musca domestica* foram classificadas como lisoziimas da família 22.

A lisoziima 1 apresenta 122 aminoácidos e massa molecular de 13816 Da apresentando 55,4% de similaridade e 38,5% identidade com a HEWL, esta exibe um pH ótimo de 4,5 sob *Micrococcus lysodeikticus* e é resistente à catepsina D (uma protease ácida). A lisoziima 2, apresenta 122 aminoácidos e massa molecular de 13890 Da, sendo muito similar à lisoziima 1 (81% similaridade e 70% identidade) (Cançado *et al*, 2006).

1.9. Estrutura terciária de lisoziimas

A estrutura terciária das lisoziimas digestivas não é conhecida, embora 17 diferentes lisoziimas da família 22 já terem sido resolvidas, correspondendo a 492 estruturas no Protein Data Bank (PDB) (Bernstein *et al.*, 1997). As lisoziimas da família 22 apresentam o mesmo enovelamento com poucas diferenças em suas estruturas. A estrutura da HEWL (figura 6) foi elucidada por Phillips em 1965, apresenta um enovelamento α/β e dependendo do critério, consiste em 5 ou 7 hélices α e 3 fitas β antiparalelas, a enzima é aproximadamente elipsoidal em sua forma com uma ampla cavidade em um lado formando o sítio ativo.

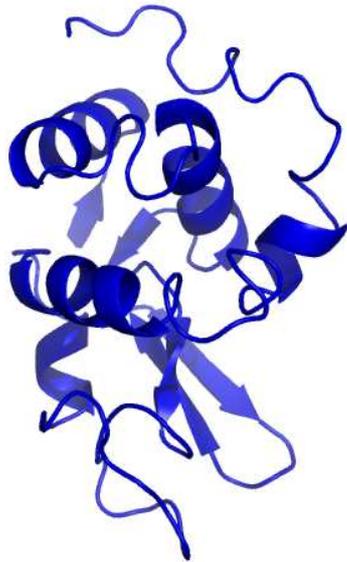


Figura 6: Estrutura da lisozima 1HEWL (Cheetham et al., 1992).

As estruturas terciárias de duas lisozimas do sistema imune de insetos são conhecidas. O núcleo da estrutura da lisozima do bicho-da-seda *Tasar *Antheraea mylitta** (AmL) (Jain et al., 2001) e da lisozima de bicho da seda *Bombix mori* (BmL) (Ito et al., 1995) são similares ao da HEWL, embora diferenças foram observadas nos resíduos de aminoácidos que formam o sítio ativo e em algumas voltas externas como a deleção de alguns resíduos de aminoácidos em várias voltas destas estruturas.

Existem quatro classes de animais com lisozimas da família 22 representadas no PDB: mamíferos, aves, peixes e insetos onde todas estas classes estão representadas por estruturas associadas à função imune.

2. Objetivos

Inicialmente os objetivos foram de estudar as propriedades estruturais e funcionais das hidrolases de *X. fastidiosa* com a finalidade de proporcionar um melhor entendimento dos mecanismos de patogenicidade da bactéria. Devido às dificuldades encontradas no decorrer do trabalho, focamos nossas atividades para o estudo funcional e estrutural da glicosídeo hidrolase celulase putativa XF0810. Com a colaboração com o Prof. Dr. Sandro Marana, ocorreu a possibilidade de realizarmos estudos estruturais da glicosídeo hidrolase lisozima 1 de *Musca domestica*, para podermos contribuir na compreensão das bases moleculares de adaptação desta lisozima à função digestiva.

Os objetivos específicos foram:

- Clonagem dos genes XF0810.
- Expressão da proteína codificada pelo gene XF0810 em sistema de expressão procariótico (*Escherichia coli*).
- Purificação da celulase expressa na forma solúvel.
- Realização de estudos estruturais, funcionais, ensaios de cristalização, difração e resolução da estrutura da celulase XF0810.
- Cristalização da lisozima 1 de *Musca domestica* com os ligantes Q3, Q4, Q5 e Q6.
- Coleta de dados de difração de raios X e processamento de dados da lisozima 1 com os ligantes Q3, Q4, Q5 e Q6.
- Resolução da estrutura terciária da lisozima 1 complexada com o ligante.

3. Materiais e Métodos

Celulase

3.1. Materiais

Os oligonucleotídeos iniciadores para PCR foram obtidos da empresa Invitrogen Life Technologies. As endonucleases de restrição, T4 DNA ligase, Taq DNA polimerase, nucleotídeos, reagentes para PCR, marcadores de massa molecular de DNA e de proteína, foram obtidos do fabricante Fermentas. Os kits para purificação do produto de PCR (QIAquick PCR purification kit), e de extração de gel de agarose (QIAquick gel extraction kit) foram obtidos do fabricante Qiagen. As colunas de cromatografia de troca iônica, gel filtração foram obtidas do fabricante Amersham Pharmacia Biotech. Os demais reagentes químicos que foram utilizados para preparação de soluções foram obtidos das empresas Sigma, Merck, Fluka e Aldrich.

3.1.1. Bactérias

As linhagens de *E. coli* utilizadas foram:

- DH5 α : para manutenção e multiplicação de vetores plasmidiais.
- BL21(DE3) Δ SlyD: para expressar proteínas de interesse. Esta cepa possui baixos níveis de produção de proteases, que poderiam degradar a proteína alvo e carrega uma cópia cromossômica do gene I que codifica a enzima T7 RNA polimerase controlado pelo promotor lacUV5 e ativado pela adição de IPTG e também possui uma deleção no gene SlyD, este gene codifica uma proteína que contém 15 histidinas no C-terminal.

3.1.2. Oligonucleotídeos

De acordo com a seqüência do gene selecionado no banco de dados da *X. fastidiosa* (<http://aeg.lbi.ic.unicamp.br/xf/>), foram desenhados os primers *forward* que são aqueles que irão parear com a extremidade 5' da fita do DNA genômico e os primers *reverse* que irão parear com a extremidade 3' da fita de DNA genômico complementar. Estes primers foram desenhados para uma temperatura de *melting* (Tm) ao redor de 70°C. Aos primers desenhados, foram adicionados os sítios de restrição das enzimas: NdeI, CATATG, no início do primer *forward* e XhoI, CTCGAG, no início do primer *reverse*.

Tabela 1 – Seqüência de oligonucleotídeos iniciadores desenhados para amplificação e clonagem. Os sítios das enzimas de restrição estão em vermelho.

ORF	Bp	Sítios enzimas restrição	Seqüência Forward (F) Reverse (R)
XF0810	1698	NdeI/XhoI	5' -CATATgTACggTgAAACgATggATgC-3' 5' -CTCgAgTTATTCggAAAACtggACATTTCC-3
XF0810c	1080	NdeI/XhoI	5' -CATATgTACggTgAAACgATggATgC-3' 5' -CTCgAgCCAACCATTgAgCgggACgCTgTTATC-3'
XF0180ct	618	NdeI/XhoI	5' -CATATgTTgCTAgTgATAACAgCgTCCCgCTC-3' 5' -CTCgAgTTATTCggAAAACtggACATTTCC-3

3.1.3. Vetores

Para a clonagem e expressão de proteínas foram utilizados os plasmídeos:

- pGEM-T-Easy (Promega) – possui o promotor T7 e SP6, o operon *lac* que pode ser utilizado na seleção de colônias transformadas, duas extremidades 3'-T livres onde o produto do PCR pode ser ligado

diretamente e a marca de seleção para ampicilina.

- pET-29a (Novagem) – possui o promotor T7, o operon *lac*, os sítios da enzimas de restrição NdeI /XhoI e a marca de seleção para canamicina.

3.1.4. Meios e soluções para o crescimento bacteriano

Os meios e soluções para o crescimento bacteriano estão nas tabelas 2 e 3 respectivamente.

Tabela 2 - Meios para o crescimento bacteriano

Meio	Reagentes
LB ¹ líquido	1 % de triptona, 0,5 % de extrato de levedura, 0,5 % de NaCl
LB sólido	1 % de triptona, 0,5 % de extrato de levedura, 0,5 % de NaCl, 1,5 % de agar
TB ² líquido	1,2 % de triptona, 2,4 % de extrato de levedura, 0,4 % de glicerol, 1 % de solução salina – 0,17 M de KH ₂ PO ₄ , 0,72 M de K ₂ HPO ₄
TB sólido	1,2 % de triptona, 2,4 % de extrato de levedura, 0,4 % de glicerol, 1 % de solução salina: 0,17 M de KH ₂ PO ₄ , 0,72 M de K ₂ HPO ₄ , 1,5 % de agar
TYM	2 % de triptona, 0,5 % de extrato de levedura, 0,58 % de NaCl, 0,25 % de MgSO ₄ .7H ₂ O

¹LB Luria Bertani, ²TB Terrific broth

Tabela 3 - Soluções para o crescimento bacteriano

Solução	Reagentes
TBF1	0,03 M de acetato potássico, 0,05 M de MnCl ₂ , 0,10 M de KCl, 0,01 M de CaCl ₂ , 15 % de glicerol
TBF2	0,01 M de ácido 3-(N-morfolino) propanosulfônico (MOPS) pH 7,0, 0,075 M de CaCl ₂ , 0,01 M de KCl, 15 % de glicerol

3.1.5. Soluções para extração de DNA plasmidial (Mini-prep)

Solução I (TEG) - 50 mM de glicose; 25 mM de Tris-HCl pH 8,0; 10 mM de EDTA pH 8,0.

Solução II - 0,2 N de NaOH; 1 % de SDS.

Solução III – 3 M de acetato potássio.

3.1.6. Soluções para eletroforese

3.1.6.1. Eletroforese de DNA: Gel de agarose

Gel (1%): 1 mg/mL de agarose dissolvida em tampão 45 mM de Tris-acetato (TAE) contendo 1 mM de EDTA (ácido etilenodiaminotetracético) pH 8,0. Os géis foram acrescidos de 0,5 µg/mL brometo de etídio, para marcação das bandas de DNA. Utilizamos o marcador de massa molecular de DNA (fermentas) para determinação do número de pares de bases presente nas amostras de DNA.

3.1.6.2. Eletroforese de proteínas: gel de poliacrilamida com dodecil sulfato de sódio (SDS-PAGE - sodium dodecyl sulfate-polyacrylamide gel electrophoresis)

- Gel de resolução (15 %): 2,5 mL de solução de acrilamida/bis-acrilamida (30/0,8 %), 1,88 mL de tampão (1,5 M de Tris-HCl, pH 8,8), 0,57 mL de água MilliQ e 25 µL de solução de SDS 20 %. Os catalisadores da

polimerização foram 25 µL de persulfato de amônio (APS) 10 % e 4 µL de N,N,N',N'- tetrametiletilenodiamino (TEMED).

- Gel de empacotamento (5%): 3,84 mL de água; 0,83 mL da solução de acrilamida/bis-acrilamida (30/0,8 %); 0,25 mL de tampão (1,5 M de Tris-HCl pH 6,8); 25 µL de SDS 20 %; 50 µL de APS (10 %) e 4 µL de TEMED.
- Tampão de amostra para eletroforese: 62 mM de Tris-HCl pH 6,8; 5% β-mercaptoetanol; 2 % (p/v) de SDS; 0,001 % (p/v) de azul de bromofenol e 10 % de glicerol.
- Tampão de corrida: Tris-glicina (25 mM de Tris; 250 mM de glicina; 0,1 % p/v de SDS).
- Corante para SDS-PAGE: 25 % (v/v) de etanol, 10 % (v/v) de ácido acético, 0,25 % de Coomassie Brilliant Blue R (Bio-Rad).
- Descorante para SDS-PAGE: 7,5 % de ácido acético; 25 % de etanol.

3.1.7. Tampão utilizado para ensaio enzimático

Tampão Carbonato: 53 g de carbonato de sódio, 42 g de bicarbonato de sódio para 1 litro de solução.

3.2. Métodos

Celulase

3.2.1. Biologia Molecular

3.2.1.1. Reação em cadeia da polimerase (PCR)

A reação em cadeia da polimerase foi utilizada para a amplificação do gene selecionado, a seqüência gênica foi obtida através do banco de dados da *X. fastidiosa* (<http://www.lbi.ic.unicamp.br/xf>). Para realizarmos uma reação de 25 µL utilizamos: 2,0 µL de DNA genômico (26 mg/µL) da bactéria que foi cedido pelo grupo da Prof^a Dra. Anete P. de Souza do Centro de Biologia Molecular e Engenharia Genética da UNICAMP, 8,05 µL de água MillQ, 1,25 µL de MgCl₂ 50 mM, 2,5 µL de tampão (1X) Taq. polimerase.(10X), 1,0 µL de DNTP, 0,2 µL de Taq polimerase (5u/µL), 2,5 µL do primer F (forward) (1pmol/µL) e 2,5 µL do primer R (reverse) (1pmol/µL) e 5 µL de betaína (1M). Foi adicionado à reação de PCR o agente adjuvante betaína (N,N,N-trimetilglicina), que auxilia na desestabilização de estruturas secundárias de DNA e na amplificação de seqüências ricas em GC (Mytelka & Chamberlin, 1996 e Rees *et al.*, 1993). A reação foi realizada em um termociclador Gene Amp PCR System 9700 - Applied Biosystem. O programa utilizado foi constituído de 30 ciclos de amplificações em seis etapas:

1^a. Etapa - desnaturação inicial do DNA molde em uma temperatura de 95°C por 4 minutos;

2^a. Etapa – desenovelamento do DNA molde por 30 segundos 94°C;

3^a. Etapa – anelamento dos oligonucleotídeos por 30 segundos a 55°C;

4^a. Etapa – extensão (polimerização) do DNA, em temperatura de 72°C por

1 minuto;

5ª. Etapa - amplificação do DNA realizado pela repetição das etapas 2 – 4 por 30 vezes;

6ª. Etapa - polimerização final em temperatura de 72°C por 5 minutos.

A última etapa foi o armazenamento da reação à 4°C por um tempo indefinido. Os produtos da PCR foram analisados em gel de agarose 1 % e purificados utilizando o kit QIAquick gel extraction da Qiagen conforme instrução do fabricante.

3.2.1.2. Eletroforese em gel de agarose

As análises de DNA foram realizadas por meio de eletroforese em gel de agarose 1 %. As amostras de DNA (plasmídeos, produtos de amplificação por reação de PCR e produtos de digestão com enzimas de restrição) foram preparadas contendo o tampão de corrida (Fermentas) para serem aplicadas ao gel. Os géis foram acrescidos de 0,5 µg/mL de brometo de etídio, para visualização dos fragmentos. O tampão usado para a eletroforese foi o TAE 1X. (40 mM de Tris-HCl, 1 mM de EDTA, pH 8,0). Após a aplicação das amostras de DNA no gel, este foi submetido a uma corrente elétrica de 75 mA, por aproximadamente duas horas. A visualização dos fragmentos foi realizada através do transiluminador UV TI202 (Sigma). Foi utilizada uma câmera digital DC120 (Kodak) acoplada a um computador para o registro das imagens dos géis. A concentração de DNA foi estimada pela da comparação entre a intensidade da banda de DNA marcada com brometo de etídeo e a intensidade de bandas padrão de DNA.

3.2.1.3. Purificação de fragmentos de DNA e de produtos de PCR

Os produtos de amplificação por PCR foram purificados com o QIAquick Extraction Gel Kit (Qiagen) de acordo com o protocolo fornecido pelo fabricante, as bandas de interesse presentes no gel de agarose foram cortadas e o DNA de interesse foi eluído.

3.2.1.4. Reação de ligação de DNA - Clonagem

As ligações dos fragmentos de DNA aos vetores de clonagem (inicialmente pGEM e posteriormente a subclonagem no vetor de expressão pET29a) previamente digeridos e purificados foram realizadas de acordo com a metodologia descrita em Ausubel *et al.* (1995). Foi utilizada a enzima T4 DNA ligase (Fermentas). Os plasmídeos recombinantes obtidos foram utilizados para a transformação de bactérias competentes.

3.2.1.5. Digestão de DNA com enzimas de restrição

Os plasmídeos foram digeridos com as enzimas de restrição específicas para cada clonagem de acordo com a metodologia descrita em Sambrook *et al.* (1989). Para cada μg de DNA utilizou-se 1 U (unidade) de enzima para digestão dos plasmídeos, as reações prosseguiram em banho termostático a 37°C por um intervalo de 3 horas. A digestão é realizada para conferência por meio da liberação do inserto no tamanho específico levando-se em consideração as regiões dos plasmídeos de cada gene de interesse. Após a digestão é feita uma corrida eletroforética em gel de agarose 1 %, para verificar a liberação do inserto de DNA.

3.2.1.6. Células competentes

Para a obtenção de células competentes, as bactérias foram preparadas segundo o protocolo Sambrook *et. al.* (1989) onde uma colônia isolada é inoculada por um período de 16 horas (37°C, 200 rpm) em 5 mL de TYM. Após diluição (1:100) do pré-inóculo, a solução foi crescida em 10 mL de TYM até alcançar OD₆₀₀ entre 0.2 – 0.6. Depois deste período, acrescentou-se 40 mL de TYM, deixando crescer até alcançar OD₆₀₀ entre 0.5 – 0.9 e novamente acrescentou-se 200 mL de TYM, esperando alcançar OD₆₀₀ 0.6. Alcançada a densidade ótica, a amostra foi deixada em banho de gelo por um período curto de tempo e as células foram centrifugadas (2100 g 15 min, 4°C). O sobrenadante foi descartado, as células foram ressuspensas em 100 mL de TBF1 e centrifugadas (2100 g, 8 min, 4°C). Novamente o sobrenadante foi descartado, as células ressuspensas em 10 mL de TBF2, alíquotas de 100 µL foram transferidas para tubos eppendorf e estocadas a uma temperatura de 80°C.

3.2.1.7. Transformação de bactérias competentes

As bactérias competentes foram transformadas segundo o protocolo Sambrook *et. al.* (1989) com cada vetor contendo o inserto de interesse de cada gene e as linhagens de células competentes de *E. coli* DH5α/BL21(DE3)ΔSlyD preparadas de acordo com o protocolo descrito no item 3.2.1.6. Cerca de 10 µL do vetor foi incubado com 80 µL de tampão de transformação (10 mL = 1 mL de 10xKCM, 1,5 mL de 10% PEG, 7,5 ml de água) e foram colocadas em gelo, após foram adicionados 50 µL de células das bactérias competentes e estas foram

levadas ao gelo por 30 minutos, após foram levadas a temperatura ambiente por 10 minutos. As células transformadas foram plaqueadas em meio LB/TB sólido contendo o antibiótico de resistência específico para cada plasmídeo e colocadas em estufa a 37°C para o crescimento durante 18 horas. As colônias de bactérias resistentes ao antibiótico foram testadas por PCR de colônia.

3.2.1.8. PCR de colônia

As colônias de bactérias transformadas com vetores recombinantes foram selecionadas em meio sólido LB/TB contendo o antibiótico específico. Para verificar se o vetor continha a seqüência de DNA correspondente as proteínas de interesse, as colônias de bactérias foram ressuspensas, uma a uma, em 100 µL de água, e aquecidas a 95°C por 5 min antes de prosseguir com os ciclos de amplificação utilizando primers da seqüência específica. A reação foi feita conforme descrito no item 3.2.1.1.

3.2.1.9. Extração de DNA plasmideal

Na extração do DNA plasmideal, o método utilizado foi o de lise alcalina modificado (Le Gouill *et al.*, 1994). Uma colônia isolada de bactéria foi crescida durante 16 horas (37°C, 200 rpm) em meio LB/TB, contendo antibiótico conforme a marca de resistência do plasmídeo codificador da proteína de interesse. Após crescimento desta, retirou-se uma alíquota adicionou-se 30 % de glicerol, para manutenção de estoque à -80°C. O restante das células foi coletado por centrifugação a 4°C. A seguir as células foram solubilizadas em TEG: 50 mM de glicose; 25 mM de Tris-HCl pH 8,0; 10 mM de EDTA e submetidas à lise por meio

da adição de uma solução, recém preparada contendo: 0,2 N de NaOH, 1% de SDS. A seguir adicionou-se clorofórmio 100% v/v, para solubilização de fosfolipídeos e proteínas hidrofóbicas, agitaram-se os tubos por inversão e incubou-se à temperatura ambiente por 1 minuto. Adicionou-se 3 mL da solução 3 M de potássio e 5 M de acetato, pH 5,5, para diminuir o pH e para favorecer a precipitação de proteínas. Recolheu-se então a fase aquosa em um novo frasco e adicionou-se duas vezes o volume da amostra, de etanol absoluto gelado, para precipitação dos plasmídeos, misturou-se a amostra por inversão. As amostras foram centrifugadas por 10 min a 18000 g e descartou-se o sobrenadante. Após total evaporação do álcool, o DNA plasmideal foi solubilizado em tampão 0,1 M de Tris-HCl pH 8,0 e adicionou-se 10 µg/mL RNase de pâncreas bovino (81,4 unidades/mg – USB) para eliminar RNA. A purificação dos plasmídeos produzidos na Miniprep foi feita utilizando o kit QIAquick PCR purification (Quiagen), de acordo com o protocolo do fabricante .

3.2.1.10. Sequenciamento de DNA

Foi utilizado o seqüenciador automático ABI PRISM 377 DNA Sequencer Applied Biosystem para confirmar a integridade dos genes clonados. A quantidade de DNA por reação das amostras enviadas para plasmídeo foi de 300-800 ng (300-600 vetores alta cópia / 500-800 vetores baixa cópia), e 5 pmol/µL de primers dos respectivos genes. A análise das seqüências geradas foi feita com os programas Chromas e BLAST A operação do seqüenciador foi executada pela técnica Zildene Correa.

3.2.2. Expressão de proteínas

3.2.2.1. Teste de expressão de proteínas

Os testes de expressão foram realizados com a inoculação das colônias de interesse em meio LB/TB líquido com 50 mg/mL de canamicina por cerca de 16h a 37°C e 200 rpm. Alíquotas desta cultura foram estocadas à -80°C em uma solução de glicerol 80 %. Outra alíquota desta cultura (pré-inóculo) é inoculada em meio LB/TB líquido com 50 mg/mL de canamicina em tubos de ensaio, à 37°C, 200 rpm, até que a densidade óptica a 600 nm (OD_{600nm}) atingisse um valor entre 0,6 e 1,0. As medidas foram feitas no espectrofotômetro Ultrospec 3000 Pharmacia Biotech. Após atingir a OD desejada, retira-se uma alíquota do cultivo para o controle não induzido e se induz a produção das proteínas adicionando isopropil- β -D-tiogalactosídeo (IPTG) 0,4 M. Manteve-se o cultivo por 2 horas no shaker e retirou-se uma alíquota. A cultura foi mantida até 16 horas para retirada de outra alíquota.

As amostras (induzidas e não induzidas) foram preparadas por centrifugação 10000 g por 10 minutos, resuspensão dos pellets em água e adição do tampão de amostra para o gel de poliacrilamida. Ferveu-se as amostras. Centrifugou-se à 10000 g por 10 minutos, a amostra do sobrenadante foi considerada a fração solúvel e o precipitado a fração insolúvel. A verificação do nível de expressão e a massa molecular aproximada das proteínas codificadas por cada gene, foi acompanhada através da aplicação das amostras em gel de poliacrilamida 15%, no sistema SDS-PAGE onde foi aplicado um marcador de massa molecular.

Para tentativa de obtenção de um maior rendimento das proteínas expressas na forma solúvel foram feitas variações na linhagem de células de expressão de *E. coli* BL21(DE3) (Rosetta, Star, C41, C43, Δ SlyD) e BL21 pLys-s, na concentração dos indutores lactose / IPTG (Isopropil β -D-tiogalactosídeo), na temperatura e no tempo de indução.

3.2.2.2. Eletroforese de proteína (SDS-PAGE)

Para análise qualitativa das proteínas, foi utilizada a eletroforese de proteínas em géis de SDS-PAGE (15%) de acordo com o método descrito por Laemmli et. al. (1970). As amostras foram diluídas na proporção 1:1 com tampão de amostra para SDS-PAGE e desnaturadas por cinco minutos a 95°C para então serem aplicadas ao gel de empacotamento e as proteínas serem separadas no gel de resolução. A eletroforese foi conduzida com a utilização de um aparato Mini-Protean II Dual Slab Cell (Bio-Rad) a 100 mA por aproximadamente duas horas. A seguir os géis foram colocados na solução corante e a remoção do excesso de corante foi realizada utilizando a solução de descorante sob agitação leve.

3.2.2.3. Expressão de proteínas

De acordo com os resultados dos testes de expressão o protocolo que ficou estabelecido devido ao melhor desempenho nos testes foi: a linhagem de *E.coli* BL21(DE3) Δ SlyD, com a indução de IPTG 0,4 mM na temperatura de 20°C por 6 h em agitação constante de 200 rpm para celulase XF0810c.

Após a determinação das condições de expressão na forma solúvel, a mesma foi submetida à expressão em grande escala. Foi feito um pré-inóculo de 3

mL em meio TB/LB líquido por cerca de 16 horas com canamicina (50 mg/mL) e uma alíquota das células BL21(DE3) Δ SlyD estocadas com agitação constante de 200 rpm a uma temperatura de 37°C. O processo realizado foi o mesmo dos testes de expressão utilizando 500 mL de meio de cultura LB/TB com 50 μ L de canamicina (50 mg/mL) e IPTG 0,4 mM em uma temperatura de 20°C por 6h. Após o cultivo, as células foram centrifugadas e o pellet resuspendido em gelo. A lise celular foi realizada pelo processo de sonicação em 6 ciclos de 30 segundos com amplitude de 40 % intercalados com 30 segundos de descanso. Centrifugou-se esta amostra e o sobrenadante foi submetido à precipitação do DNA com sulfato de estreptomicina 2 % (p/v) por 45 minutos na câmara fria. Realizou-se nova centrifugação e recuperou-se o sobrenadante. A expressão da proteína foi analisada pelo sistema SDS-PAGE.

3.2.3. Purificação de proteínas

3.2.3.1. Purificação de proteínas por cromatografia

A expressão em larga escala de proteínas foi realizada, o “pool” da proteína expressa foi submetido à purificação utilizando o aparelho de cromatografia líquida de performance rápida (FPLC) da AKTA (Amershan Pharmacia Biotech). Todo o processo foi realizado com monitoramento constante da absorvância a 280 nm, da condutância e da pressão. Antes de cada corrida cromatográfica cada coluna foi equilibrada com 2 volumes de coluna do tampão adequado e após as corridas cada coluna utilizada foi lavada com água (2 volumes de coluna) e então etanol 20% (2 volumes de coluna). Todas as soluções foram filtradas e deaeradas antes

de iniciar as cromatografias. Inicialmente foi utilizada a cromatografia de troca iônica (aniônica), em uma coluna HiTrap Q FF 5 mL (Amersham-Biosciences), a escolha desta coluna para troca aniônica foi devido à pesquisa utilizando a seqüência de aminoácidos da proteína truncada e o banco de dados ExPASy-ProtParam/tool que mostrava uma carga total negativa para XF0810c. Utilizamos o tampão A: 20 mM de Tris-HCl pH 8,5 para equilibrar a coluna, a amostra foi aplicada e em seguida foi coletado o flow-through. Em seguida os picos foram eluídos em um gradiente linear progressivo (0-100%) de tampão B (tampão A e 1 M de NaCl) em 20 volumes de coluna no fluxo de 1 mL/min. Posteriormente foi realizada a cromatografia de filtração em gel na coluna Superdex (200) 16/60, volume 120 mL utilizando o tampão 20 mM de Tris-HCl pH 8.5 e 100 mM de NaCl. A proteína de interesse foi detectada pelo SDS-PAGE.

3.2.3.2. Diálise

O processo de diálise foi realizado na amostra de proteína eluída na purificação de troca iônica com o objetivo de diminuirmos a concentração de sal da amostra para a aplicação da mesma na cromatografia de filtração em gel e outros experimentos. A diálise foi conduzida a 4°C contra um volume de 2 litros no tampão 10 mM de Tris-HCl pH 8,5 por 2 horas (tempo mínimo) com a realização de três trocas com volume de dois litros.

3.2.3.4. Concentração de proteína

Para a obtenção das proteínas em uma concentração desejável para as análises espectrocópicas, espectrométricas e a cristalização, foi utilizado o sistema Amicon Ultra 10.000 MW conforme as instruções do fabricante.

3.2.4. Análises espectroscópicas e espectrométricas

Celulase

3.2.4.1. Determinação da concentração de proteína

A concentração da proteína foi determinada por meio do espectrofotômetro DU 640 (Beckman Coulter) utilizando a medida de absorção da solução a 280 nm, em uma cubeta de quartzo de um 1 cm de caminho óptico. Para estimativa da concentração das proteínas foi utilizado o coeficiente de extinção (ϵ) calculado especificamente para proteína através do programa ProtParam (<http://us.expasy.org/tools/protparam.html>). Os valores de ϵ calculados foram utilizados na equação de Beer-Lambert para determinar a concentração das proteínas:

$$A = \epsilon \cdot C \cdot l$$

Onde A é a absorvância a 280 nm; ϵ é o coeficiente de extinção molar em 280 nm ($M^{-1} \text{ cm}^{-1}$); l é o caminho óptico em centímetros (cm) e C - concentração em mol/L.

3.2.4.2. Espectrometria de massa

As amostras da proteína foram preparadas utilizando o protocolo de digestão em gel (redução e alquilação) dividido em três etapas:

Etapa 1: extração do gel

As amostras de proteínas foram submetidas a uma corrida no sistema SDS-PAGE, as bandas contendo as proteínas de interesse foram excisadas do gel e transferidas cada uma para um tubo. Em seguida foram lavadas três vezes com 200 μL de uma solução 25 mM de NH_4HCO_3 em 50 % de acetonitrila (ACN) pH 8.0. Após, foram novamente lavadas com 200 μL de água milli-Q e foi adicionado 25 μL da solução de 10 mM de ditioneitol (DTT) em 25 mM NH_4HCO_3 . A solução foi agitada em vortex e os tubos ficaram a 56 $^\circ\text{C}$ durante uma hora. O sobrenadante então foi descartado e foi adicionado 25 μL de 55 mM de iodoacetamida. Os tubos foram submetidos à agitação ficaram quarenta e cinco minutos em temperatura ambiente sem a interferência de luz. Descartou-se então o sobrenadante e as amostras foram lavadas três vezes com 200 μL da solução de ACN 50% em 25 mM de NH_4HCO_3 e uma vez com 200 μL de água milli-Q . Foi adicionado 100 μL de ACN 100% para desidratação do gel e as amostras foram incubadas durante cinco minutos a temperatura ambiente, o excesso de líquido foi retirado e as amostras foram submetidas ao Speed Vac.

Etapa 2: digestão enzimática

As amostras dos géis foram reidratadas com uma solução gelada de tripsina 15 $\mu\text{g}/\text{mL}$ em NH_4HCO_3 . Esperou-se dez minutos a 4 $^\circ\text{C}$ e adicionou-se às amostras a solução 25 mM de NH_4HCO_3 em uma quantidade necessária para que os géis ficassem cobertos, após incubou-se à 37 $^\circ\text{C}$ por vinte e quatro horas.

Etapa 3: extração de peptídeos

Os sobrenadantes resultantes da digestão foram transferidos para novos tubos, adicionou-se 20-50 μL de ACN 50%, ácido trifluor acético (TFA) 5%, agitou-se suavemente os tubos durante trinta minutos e o sobrenadante foi retirado. Novamente, adicionou-se mais 25-50 μL de ACN 50% e TFA 5% por 30 minutos, retirou-se o sobrenadante, após o volume das soluções foram reduzidos para 10 μL através do Speed Vac.

A proteína foi analisada através do aparelho ESI Q-Tof Ultima (Micromass, UK) acoplado a um sistema de cromatografia líquida HPLC onde foi inserida uma amostra aquosa da solução contendo os peptídeos da proteína digeridos pela tripsina.

3.2.4.3. DLS

Os experimentos de espalhamento dinâmico de luz foram realizados em um equipamento DynaPro-99-E-15 (Protein Solutions). Todas as medidas foram realizadas a 25°C. Utilizou-se concentração de 1 mg/mL de proteína em tampão 10 mM Tris-HCl pH 8,5, a amostra foi centrifugada por 30 minutos a 4620g antes de ser analisada. Os dados foram tratados com o programa Dynamics V6 fornecido pelo fabricante do aparelho.

3.2.4.4. Filtração em gel analítica

Com a finalidade de determinarmos a massa da proteína em estudo e seus estados de agregação realizamos a filtração em gel analítica, que se baseia no princípio da cromatografia por exclusão de tamanho da amostra. A coluna

Superdex 200 16/60 (Amersham Biosciences) montada no aparelho AKTA Purifier foi usada. Para determinação do volume vazio da coluna foi aplicado na coluna o blueextran na concentração de 1 mg/mL. Em seguida, foram aplicados os marcadores de proteínas com diferentes massas moleculares conhecidos (aldolase, quimotripsina, ovoalbumina e albumina) nas concentrações de 3 mg/mL e finalmente foi aplicada a amostra contendo a proteína de interesse. Para normalização dos dados obtidos na cromatografia utilizou-se a fórmula da constante $K_{av} = (V_e - V_o) / (V_t - V_o)$ onde V_e = volume de eluição da proteína, V_o = volume vazio da coluna e V_t = volume total da coluna empacotada. Para determinação da massa molecular da proteína de interesse foi utilizada uma curva de calibração, construída a partir da regressão linear dos valores de K_{av} e logaritmos das massas moleculares das proteínas conhecidas. A partir desta curva se extrapolou o valor desejado a partir do K_{av} da proteína de interesse.

3.2.4.5. Dicroísmo circular

Os experimentos de dicroísmo circular foram realizados no espectropolarímetro JASCO J-810 a uma temperatura de 25°C. Os espectros de dicroísmo circular no UV distante (200 - 260 nm) foram obtidos utilizando-se uma cubeta de quartzo de 0,2 cm de caminho ótico, a uma concentração de 0,16 mg/mL de proteína no tampão 10 mM de fosfato de sódio pH 7,4. Foram acumuladas 20 curvas consecutivas e a média dos espectros foi armazenada. Todos os espectros de dicroísmo circular foram corrigidos pela contribuição do tampão utilizado e as elipticidades observadas foram transformadas em elipticidade molar por resíduo $[\theta]$, de acordo com a equação:

$$[\theta]_{\lambda} = \theta_{\text{obs}} \times \text{MRW} / 10 \times d \times c$$

Onde λ = comprimento de onda em nanômetros; θ_{obs} = a elipticidade medida em graus; MRW = peso médio do resíduo (utilizando um peso molecular médio do resíduo de 110 Da), d = comprimento do caminho óptico (cm) e c = concentração da proteína em g/mL (Adler *et al.*, 1973). A estrutura secundária da seqüência de aminoácidos das proteínas foi estimada a partir de espectros de UV-distante usando-se o programa CDNN Deconvolution (Bohm *et al.*, 1992) e Dicroprot (Deléage & Geourjon, 1993). Os programas utilizados para registro e tratamento dos dados foram Spectra manager (JASCO) e Origin versão 7.0 (Microcal).

3.2.4.6. Ensaio enzimático – atividade exocelulase

Em um tubo foi colocado 80 μL P-nitrofenilfosfato-celobiosídeo (PNPc) (1 mg/mL em 50 mM de tampão acetato de sódio pH 5,8) mais 200 μL da enzima e incubado a 45°C por 20 min. Em seguida, foi adicionado 0,5 mL de tampão carbonato e a absorbância foi lida à 420 nm. Uma unidade de atividade da enzima é igual à quantidade de enzima necessária para produzir 1 μmol de PNPc por min. Nos experimentos realizados, a quantidade utilizada de substrato e enzima foi de 80 μL e 20 μL respectivamente, o tempo de incubação foi de 0, 5, 30, 60 e 240 min, na temperatura de 45 °C. Os controles utilizados foram o branco do substrato (sem a enzima) e o branco da enzima (sem o substrato). O experimento para obtenção da curva padrão de p-nitrofenolato foi realizado utilizando a solução 1

mM de PNPC. As amostras foram medidas em triplicata. Foi realizado também o controle positivo da enzima utilizando a celulase de *Trichoderma reesei* 2,5 KU – 6,9 unidades/mg.

3.2.5. Cristalização de proteínas

Lisozima

3.2.5.1. Ensaio de cristalização

Os ensaios de cristalização iniciais foram realizados pelo método de difusão de vapor em gota sentada (sitting drop). Nas placas com 24 poços, foram colocados 300 µL da solução do kit no reservatório e 1 µL desta solução com 1 µL da solução da proteína no suporte da gota de cristalização. Foram utilizados os kits de cristalização Crystal Screen e Crystal Screen 2 da Hampton Research e os 1 a 10 da Jena Bioscience.

Celulase

3.2.5.2. Ensaio de cristalização com robôs

Os ensaios de cristalização foram realizados no laboratório robotizado de cristalização de proteínas onde se encontram os robôs:

- Matrix Maker Workstation/ Emerald Biosystems - um robô que prepara soluções a partir de diferentes soluções estoque. Neste caso ele prepara as soluções de 6 kits de cristalização:

Kit 1: Crystal Screen e Crystal Screen 2 (96 condições) da Hampton Reserach reestruturado,

Kit 2: Wizard I e II (96 condições) da Emerald BioSystems reestruturado,

Kit 3: PACT (96 condições) da Nextal/Quiagen reestruturado,

Kit 4: Joint Center for Structural Genomics (96 condições) da Nextal/Quiagen reestruturado,

Kit 5: Salt RX (96 condições) da Hampton Research,

Kit 6: Precipitant Synergy (64 condições) da Emerald BioSystems.

- Honey Bee 961 System Genomic Solution/ Havard Bioscience Company – Protein Crystallization System - um robô que prepara as gotas de cristalização contendo a solução dos kits de cristalização e a proteína de interesse utilizando o método de gota sentada em placa CrystalQuick da Grenier bio-one. Esta placa contém 96 reservatórios, cada um com 3 posições de gotas.

Lisozima

3.2.5.3. Soaking

Para a incorporação de ligantes específicos no cristal da proteína de interesse foi utilizado o método de soaking, onde foi adicionado 1 μ L da solução de cada ligante diluído em água à cada gota que continha o cristal da proteína nas placas de cristalização. A concentração da proteína (lisozima 1 de *Musca domestica*) foi de 0,7 mM e de cada ligante foi de:

- 4- metilumbeliferil-N-acetil quitotriosídeo Q3 (0,2 mM)
- N-acetil quitotetraosídeo Q4 (1,0 mM)
- N-acetil quitopentaosídeo Q5 (1,0 mM)
- N-acetil quitohexaosídeo Q6 (0,5 mM)

As placas de cristalização contendo os cristais e ligantes permaneceram à temperatura de 18°C por uma semana.

Celulase

3.2.6. Modelagem Molecular

O princípio do método utilizado para modelagem molecular de proteínas feito neste trabalho é o baseado na homologia entre a seqüência de aminoácidos da proteína com estrutura 3D determinada (1ECE) e da proteína de interesse (XF0810c) sem estrutura 3D determinada. O programa utilizado para a construção do modelo foi o MODELLER (Sali and Blundell, 1993) que utilizou arquivos contendo o alinhamento das proteínas, as instruções do MODELLER e as coordenadas atômicas depositadas no Protein Data Bank (Bernstein *et al.*, 1977) da proteína de estrutura conhecida. Foi utilizado para os “input files” um arquivo contendo valores padrão propostos pelo programa. Para a visualização gráfica dos modelos gerados, utilizamos o programa O (Jones, 1991), para a construção da figuras utilizamos o programa PyMOL (Delano, 2002) e para avaliação da qualidade dos modelos utilizamos o programa PROCHECK (Laskowski; *et al.*, 1993) que analisa a qualidade estereoquímica do modelo e VERIFY 3D (Luthy, *et al.*, 1992) que indica as regiões mal construídas do modelo.

3.2.7. Cristalografia de proteínas

Lisozima

3.2.7.1. Difração de raios X

Os dados cristalográficos foram coletados na linha de luz de cristalografia de proteínas D03B-MX1 do LNLS, esta linha é equipada com um detector MARCCD que possui uma superfície circular de 165 mm de diâmetro combinado com um goniômetro MAR DTB. Cada cristal foi pego diretamente da gota com um loop (Hampton Research) e este foi fixado no goniômetro, o cristal foi resfriado com um fluxo direto de gás nitrogênio a 100 K com o objetivo de minimizar os danos causados pela radiação. Os dados foram coletados utilizando o método de rotação com 1° por imagem durante o tempo de exposição de 60 segundos, o comprimento de onda da radiação foi de aproximadamente 1,43 Å. As soluções nas quais os cristais cresceram propiciaram uma proteção parcial contra a formação de gelo.

3.2.7.2. Processamento dos dados

Os dados coletados resultaram em um conjunto de fotos digitalizadas onde cada foto contém os pontos de difração, estes foram processados com o programa HKL2000 que utiliza o DENZO e SCALEPACK (Otwinowski & Minor, 1997). Através do programa DENZO, primeiramente foi identificada a célula unitária e o grupo espacial, então foi realizado a autoindexação onde consiste em se obter os melhores parâmetros de rede a , b , c , α , β e γ possíveis para uma das 14 redes de Bravais e da orientação do cristal (rotação em x , y , z) em relação ao feixe de raios

X e do eixo de rotação do cristal. Depois a indexação foi feita com um refinamento de parâmetros da célula unitária onde se associam os índices de Miller h k l para cada uma das reflexões e a integração onde os pixels de cada ponto são utilizados para calcular a intensidade total. Após o agrupamento e escalonamento dos dados foram realizados pelo programa SCALEPACK que lê e organiza todos os arquivos das reflexões e faz comparações e estatísticas, fazendo uma média ponderada das intensidades das reflexões equivalentes produzindo arquivos de estatísticas do agrupamento, das reflexões que foram rejeitadas e das reflexões independentes aceitas.

3.2.7.3. Substituição molecular

A substituição molecular foi realizada com o programa MOLREP (Collaborative Computacional Project, Number 4, 1994; Vagin & Teplyakov, 1997) utilizando como modelo um monômero das coordenadas da estrutura de lisozima 1 de *M. domestica* depositada no PDB (Bernstein *et al.*, 1997) com o código 2FBD.

3.2.7.4. Refinamento

Para o refinamento da estrutura resolvida por substituição molecular foram utilizados os programas CNS (Brunger *et al.*, 1998) versão 1.1 e REFMAC (Collaborative Computacional Project, Number 4, 1994; Murshudov *et al.*, 1997). Inicialmente realizamos um ciclo de refinamento “simulated annealing” utilizando o programa CNS versão 1.1 e após esta primeira etapa, foram realizados vários ciclos de refinamento utilizando o programa REFMAC. No final de cada ciclo eram observados os valores dos fatores R (R_{factor} e R_{free}). As modificações manuais

introduzidas no modelo atômico durante os ciclos de refinamento foram feitas através do programa O (Jones *et al.*, 1991) e COOT (Emsley & Cowtan, 2004) pela inspeção dos mapas de densidade eletrônica 3Fo-2Fc e Fo-Fc com contornos de aproximadamente 1 e 3 σ , respectivamente. As moléculas de água foram adicionadas pelo procedimento automático do ARP/WARP (Lamzin & Wilson, 1997) e verificadas individualmente. As coordenadas do ligante foram adicionadas à molécula após alguns ciclos de refinamento, para isto, utilizamos as coordenadas do ligante triacetilchitotriose (CTO) e os parâmetros de topologia obtidos através do banco de dados Hic-Up (Kleywegt & Jones, 1998) como ponto de partida. Como a estrutura da lisozima de *M. domestica* com o ligante Q4 apresenta um dímero na unidade assimétrica, a modelagem do ligante no segundo monômero foi feita pela aplicação da matriz de rotação e translação obtidas pela superposição da molécula B na molécula A. Foi verificado que o ligante ocupava somente a molécula B. A superposição dos monômeros A e B foi feita com o programa COOT (Emsley & Cowtan, 2004).

4. Resultados e Discussão

Inicialmente para o estudo deste trabalho, foram selecionados genes da *X. fastidiosa* que codificam hidrolases (vide anexo I), cujos objetivos seriam de estudar propriedades funcionais e estruturais destas enzimas para um melhor entendimento dos mecanismos de patogenicidade da bactéria. Devido às dificuldades encontradas na realização do trabalho (expressão das proteínas na forma insolúvel), as atividades foram concentradas em uma enzima, a glicosídeo hidrolase celulase putativa XF0810, pelo fato da expressão desta ser na forma solúvel. Com a colaboração do Prof. Dr. Sandro Marana surgiu a oportunidade de estudo de mais uma glicosídeo hidrolase, a lisozima 1 de *M. domestica* com o intuito de contribuir na compreensão das bases moleculares de adaptação desta lisozima à função digestiva.

As glicosídeos hidrolases, são enzimas que hidrolizam ligações glicosídicas e estão envolvidas em um amplo espectro de processos biológicos. Neste trabalho foram estudadas duas destas enzimas, as celulases que hidrolizam ligações β (1-4) da celulose e as lisozimas que catalisam a hidrólise ligações 1-4 entre resíduos de peptídeoglicans do ácido acetil-n-muramico e n- acetilglucosamina da parede celular bacteriana.

4.1. Celulase de *X. fastidiosa* (XF0810)

4.1.1. Amplificação, clonagem e subclonagem

Inicialmente o projeto foi desenvolvido visando a obtenção da celulase de *X. fastidiosa* codificada pelo gene XF0810. Para isto, foi feita a amplificação do gene que codifica a mesma, clonagem em vetor pGEM e suclonagem em vetor de expressão pET. Entretanto, ao realizarmos os testes de expressão, a proteína se apresentou na forma insolúvel. Várias tentativas foram realizadas através de novos testes de expressão para a obtenção da XF0810 na forma solúvel utilizando algumas variáveis tais como: linhagens de *E. coli*, diferentes vetores, temperatura, concentração do indutor, tempo de indução, porém todas estas tentativas não apresentaram os resultados esperados, a proteína ainda se apresentava na forma insolúvel (dados não demonstrados). Paralelamente a estes experimentos de laboratório iniciamos o estudo de modelagem molecular para a XF0810 e analisando o alinhamento estrutural da celulase de estrutura resolvida 1ECE do organismo *Acidothermus cellulolyticus* comparada com a celulase XF0810 observamos que a 1ECE não possui o C terminal adicional que a celulase XF0810 apresenta (figura 7). Através desta análise, resolvemos desenhar um novo primer para a celulase de *X. fastidiosa* para a retirada da cauda adicional C terminal para então podermos realizar uma nova amplificação, clonagem e expressão do gene da celulase truncada (domínio celulase sem o C-terminal) XF0810c na tentativa da produção de proteínas solúveis.

>XF0810

```
MYGETMDAQNVAVSDTHFVEPLHGWNWRGLETQAQHLLQGLDQRPWREVLDO
MQSLGINAIRLPLCSDTLHGTMPTNLDLVRNPDLKGRALQIADAIIDEA
GKRGMRVLLAYHGVECPDGNPLLRVDESEHQWISDVQFITSHYRAQQK
VVMGVLDLADMANHRPFQSGGDSTPDWNRVVERAAAAAILAMTPEWLGWQP
VGLNPPCLDASAPISDDNIQSQHCVPLRIPARNLLLMPHFAGTDIDTEAA
LGA FSGKQTVLPNSLDATDAEQLAHRIDALLAFGIRQGFYGSWMTSAQMP
FGMLDNDGRTPRTALIAQLHRWVGVS RVDVA SENAATKNQTTTDTINGCVA
SDNSVPLNGWDTSFSGVATYTYTGYKGGALMLDPIQSHAQITALNPIQLN
LGGIPAAAMAGAYLRVQGPKGSTTVYVTD RYPTGSSGALDLSNPAFASIGN
MAQGRIPVQWKVVSAPVSGNLIYRVKKGSSGWWAAIQVREHRYPVLKLEI
CQDGTWLNLPKRNMYFVVGTRLGNQPLSMRMTDIRGQTLIDTLPALPKRA
SSKAYSVNGNVQFSE
```

Figura 7: Seqüência XF0810c, a cauda C-terminal adicional está marcada em azul.

Com a obtenção do novo primer desenhado para celulase XF0810c, iniciamos os experimentos de biologia molecular com o gene que foi amplificado por PCR a partir do DNA genômico de *X. fastidiosa*, o produto obtido foi um fragmento de 1080 pb correspondente ao tamanho do gene (figura 8). O fragmento do gene amplificado foi clonado em vetor pGEM-T-Easy resultando no plasmídeo pGEM-XF810c, então foi realizada então a transformação deste plasmídeo em células competentes de *E. coli* DH5 α . As colônias transformantes foram selecionadas em meio LB sólido contendo ampicilina (marca de seleção do plasmídeo pGEM) e X-gal. Foi feita a extração dos plasmídeos através do método de mini-prep, após a extração a seqüência do plasmídeo foi confirmada através do sequenciamento de DNA.

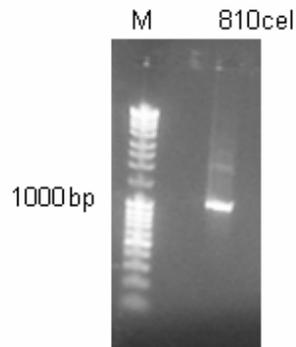


Figura 8: Gel de agarose 1 %. Amplificação do gene da celulase XF0810c (1080 pb). Marcador de peso molecular DNA Ladder Mix (M), gene amplificado da celulase (810c).

A digestão do plasmídeo pGEM-XF810c foi realizada para que ocorresse a liberação do gene, para isto foram utilizadas as endonucleases de restrição NdeI e XhoI (figura 9). Com o gene da XF0810c realizamos uma subclonagem no vetor de expressão pET29a (previamente digerido com as mesmas enzimas) que resultou no plasmídeo pET29a-XF0810c, este foi usado para transformar de células de *E. coli* da linhagem BL21(DE3) Δ SlyD. As colônias foram crescidas em meio LB/TB sólido contendo canamicina (marca de seleção do plasmídeo pET29a).

Foi realizado o teste de PCR de colônia (figura 10) para a confirmação do plasmídeo pET29a-XF0810c e seleção dos clones contendo o inserto de interesse. As colônias selecionadas foram submetidas a uma midi-prep para extração dos plasmídeos e um novo sequenciamento confirmou a integridade dos genes.

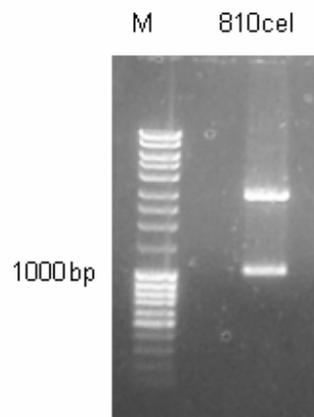


Figura 9: Gel de agarose 1 %. Digestão do plasmídeo pGEM-XF0810c. Marcador de peso molecular DNA Ladder Mix (M), plasmídeo pGEM-XF0810c digerido.

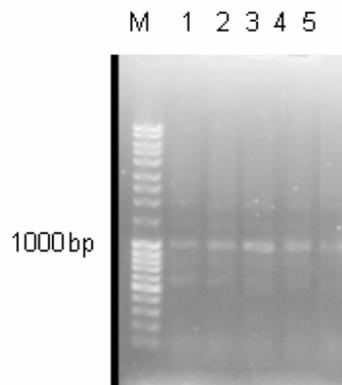


Figura 10: Gel de agarose 1 %. PCR de colônia do plasmídeo pET29a-XF0810c (1080 bp). Marcador de peso molecular DNA Ladder Mix (M), cinco colônias positivas numeradas respectivamente (1,2,3,4 e 5).

4.1.2. Expressão

A expressão da celulase XF810c foi realizada em 1 litro de LB com indução de 0,4 mM de IPTG por um período de 6 horas à 20°C. (figura 11).

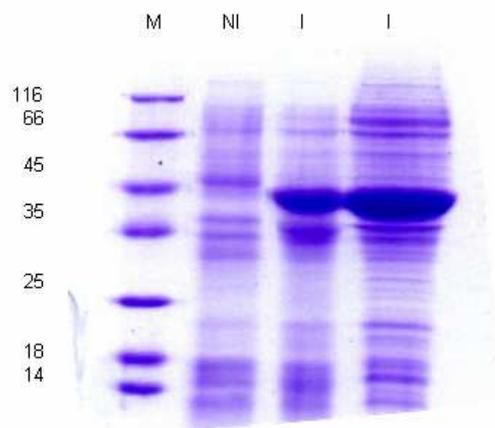


Figura 11: SDS-PAGE 12% da expressão XF0810c (39,4 kDa). Marcador de peso molecular (M), fração não induzida (NI), fração induzida (I).

O resultado obtido foi o esperado, a proteína se apresentou na forma solúvel. Analisando o resultado do gel e com a finalidade de certificarmos que a banda expressa correspondia à proteína de interesse, submetemos esta a um protocolo de digestão da banda recortada do gel com tripsina para leitura de sua massa através do aparelho Q-Tof Ultima (Micromass, UK), o resultado desta análise confirmou que a banda expressa seria endoglucanase de *X. fastidiosa* codificada pelo gene XF0810c, pois a tripsina havia clivado as seguintes seqüências de peptídeos **TALQIADAIIDEAGK** e **VDVASENAATK** presentes na seqüência XF0810c, conforme demonstrado na figura 12.

```

MYGETMDAQNNAVSDTHFVEPLHGVNWRGLETAQHLLQGLDQRPWREVLDQMQSLGINAIRLPLCSDTLHGTMPT
TNLDLVRNPDLKGRTALQIADAIIDEAGKRGMRVLLAYHGVECPTDGNPLLRVDESEHQWISDVQFITSHYR
AQQKVVMGVDLADMANHRPFQSGGDSTPDWNRVVERAAAAAILAMTPEWLIQVQPVGLNPPCLDASAPISDDNI
QSQHCVPLRIPARNLLLMPHFAGTDIDTEAALGAFSGKQTVLPNSLDATDAEQLAHRIDALLAFGIRQGFYGS
WMTSAQMPFGMLDNDGRTPTALIAQLHRWVGVSVDVASENAATKNQTTTDTNGCVA
SDNSVPLNGW

```

Figura 12: Seqüência XF0810c, em vermelho está a seqüência do corte com tripsina.

4.1.3. Purificação

Com a proteína solúvel, realizamos a expressão em larga escala da celulase XF0810c com objetivo de termos grande quantidade para estabelecer um protocolo de purificação da mesma, com o intuito de obter a proteína pura numa concentração desejável para realizarmos ensaios de cristalização. Iniciamos através da técnica de troca iônica (aniônica) em uma coluna HiTrap Q FF 5 mL (Amersham Pharmacia Biotech), utilizando o tampão A 20 mM Tris-HCl pH 8,5. Os picos foram eluídos através do gradiente progressivo 0-100% de tampão B (20 mM Tris-HCl pH 8,5 + 1 M NaCl) em 20 volumes de coluna em um fluxo de 1 mL/min (figuras 13 e 14). A escolha desta coluna para troca aniônica foi devido à pesquisa utilizando a seqüência de aminoácidos da proteína e o banco de dados ExPASy-ProtParam/tool que mostrou um pI calculado para a proteína truncada de 5.05 e um número total de resíduos carregados negativamente (Asp +Glu) de 42%.

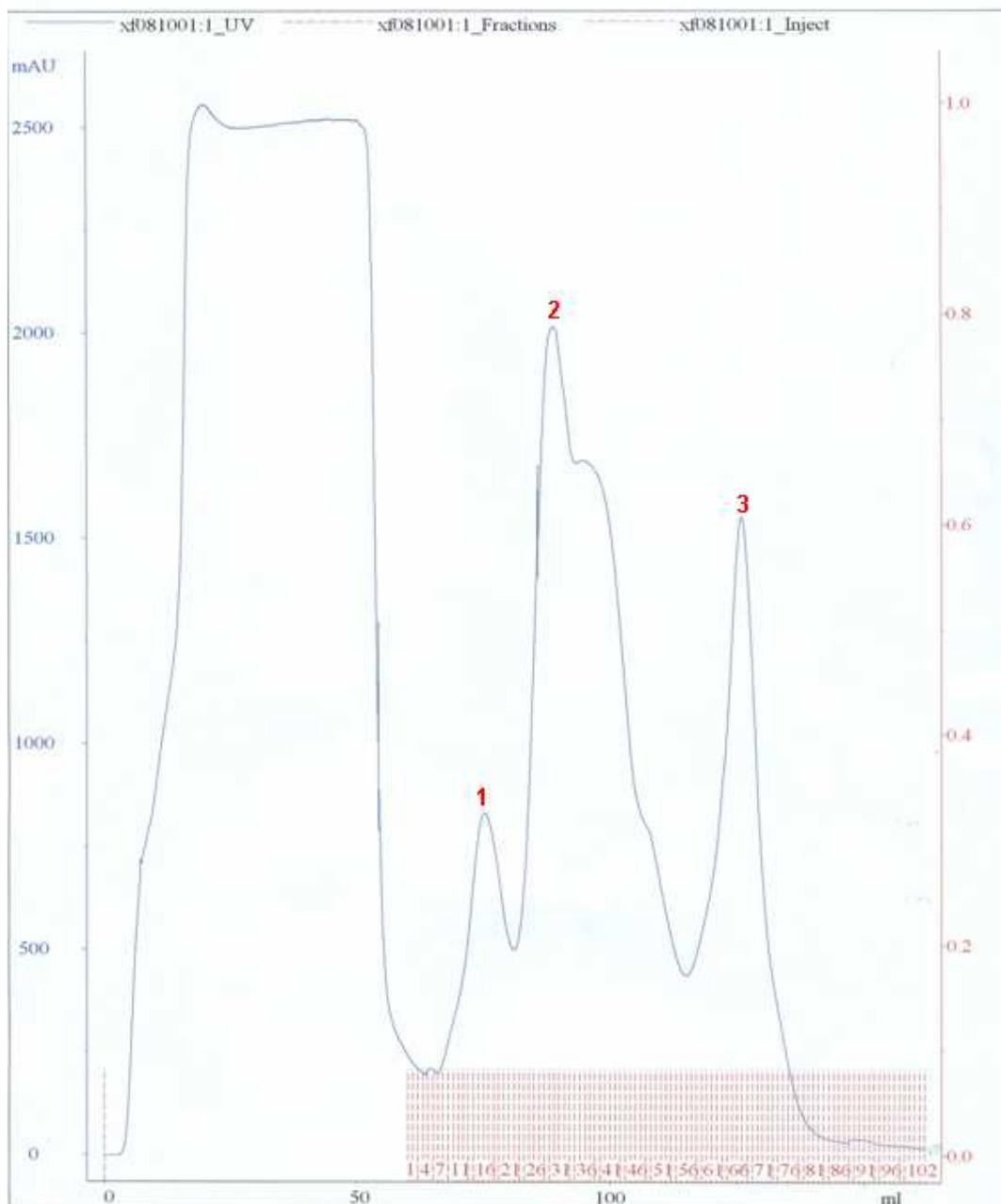


Figura 13: Cromatograma da purificação da XF0180c em coluna de troca iônica HiTrap QFF. Os picos eluídos estão marcados respectivamente pelos números 1, 2 e 3.

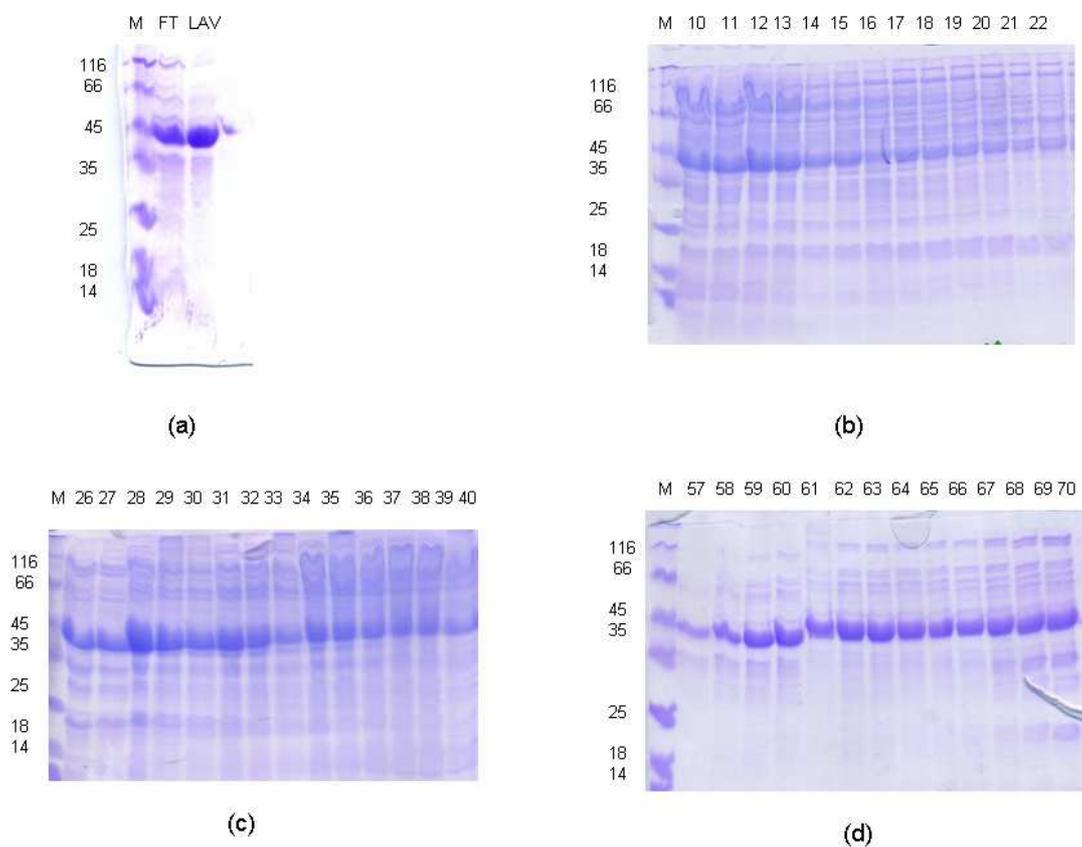


Figura 14: SDS-PAGE 12% da purificação por cromatografia de troca iônica da XF0810c (39,4 kDa). Marcador de peso molecular (M), **(a)** flowthrough (FT) e lavagem (Lav), **(b)** frações 10-22 (primeiro pico), **(c)** frações 26-40 (segundo pico), **(d)** frações 57-70 (terceiro pico).

Analisando os géis resultantes da purificação de troca iônica, verificamos a eluição da proteína XF0810c em vários picos da cromatografia (flowthrough, lavagem, pico 1, pico 2 e pico 3). Observamos que na fração “lavagem”, a proteína se encontrava com um maior grau de pureza, prosseguimos os experimentos submetendo esta fração a uma purificação de filtração em gel na coluna Superdex

(200) 16/60, volume 120 mL utilizando o tampão 20 mM Tris-HCl pH 8.5 + 100 mM NaCl (figuras 15 e 16) com o objetivo de remover as contaminações da fração tornando a proteína pura.

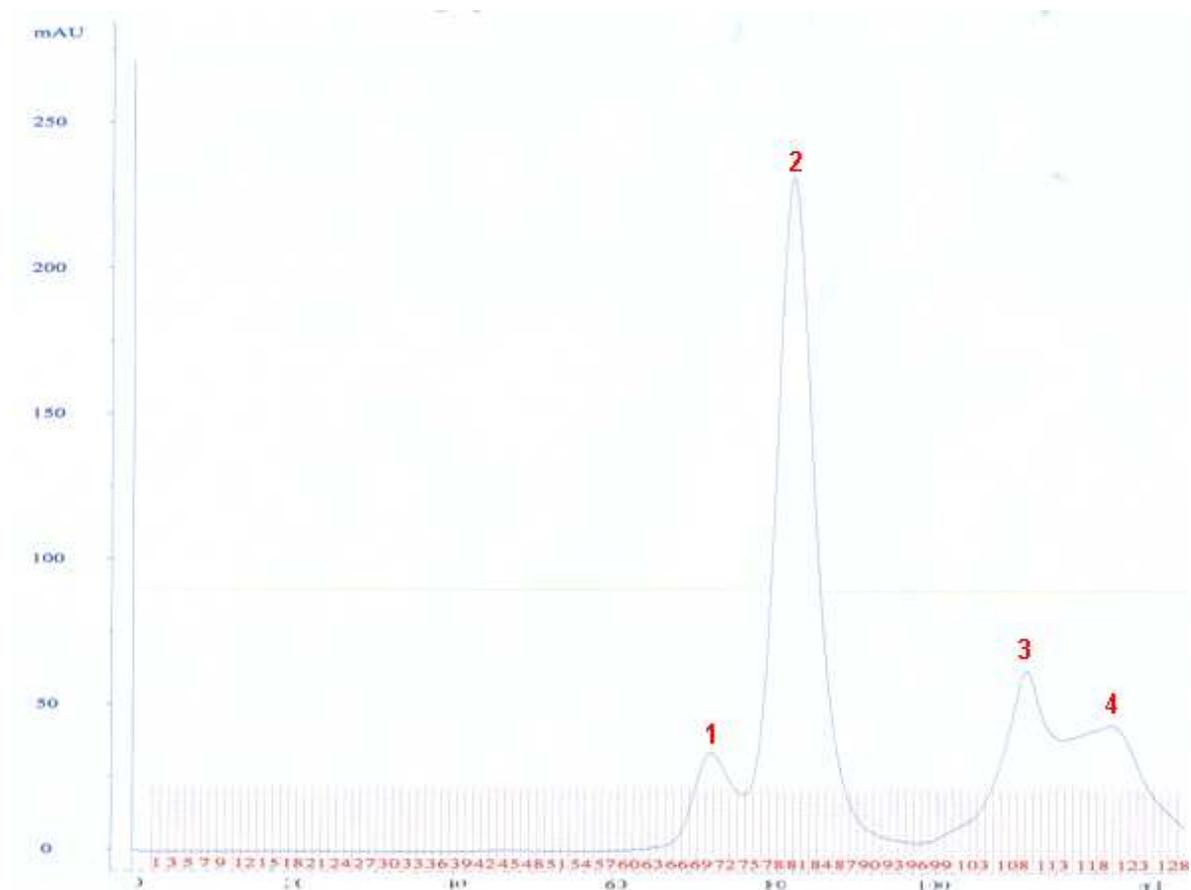


Figura 15: Cromatograma da purificação da XF0180cel (fração lavagem da QFF), em coluna de filtração em gel Superdex 16/60. Os picos eluídos estão marcados respectivamente pelos números 1, 2, 3 e 4.



Figura 16: SDS-PAGE 12% da purificação por cromatografia de filtração em gel da XF0180c (39.4 kDa). Marcador de peso molecular (M), frações 67-74 (pico 1), frações 75-87 (pico 2), frações 105-113 (pico 3), frações 114-123 (pico 4).

Verificamos que a proteína XF0810c havia sido eluída pura, porém em dois picos da cromatografia. Com a finalidade de verificarmos a massa molecular da proteína e seus estados de agregação submetemos a amostra dos picos 1 e 2 resultantes da filtração em gel ao experimento de DLS e gel filtração analítica.

4.1.4. Espalhamento dinâmico de luz (DLS)

Ao observarmos o resultado da purificação por cromatografia de filtração em gel, verificamos que a proteína XF0810c estava pura, porém se encontrava em dois picos, 1 e 2. Com a finalidade de analisarmos a agregação da proteína, submetemos as amostras dos picos 1 e 2 (em uma concentração de 1 mg/mL) ao experimento de DLS (Dynamic Light Scattering) (figura 17).

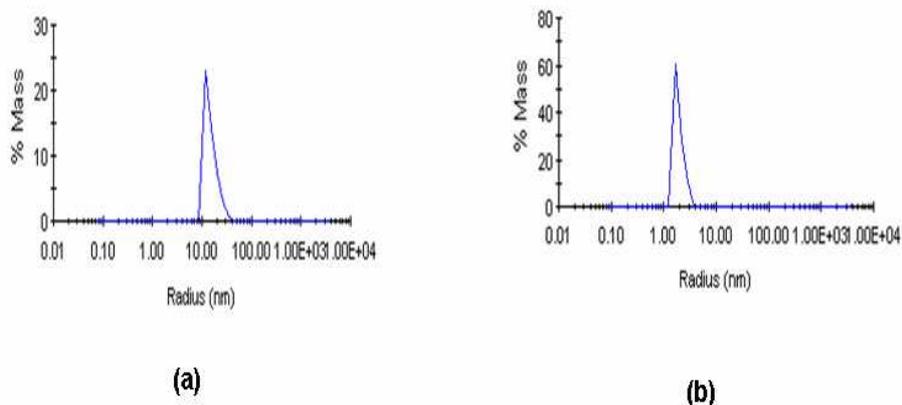


Figura 17: DLS dos picos resultantes da purificação filtração em gel XF810c, **(a)** pico 1 e **(b)** pico 2.

Neste resultado, verificamos que o pico 1 apresenta um raio de giro de 15,1 nm e uma massa relativa de 1928 kDa e o pico 2 um raio de giro de 1,9 nm e uma massa relativa de 16 kDa, demonstrando a presença de agregados no pico 1.

4.1.5. Gel filtração analítica

Com o intuito de confirmar a massa molecular da proteína XF0810c e verificar o estado de agregação da mesma, submetemos a amostra da fração da lavagem vinda da troca iônica na coluna QFF a uma cromatografia de filtração em gel analítica na coluna Superdex (200) 16/60 (figura 18). Para determinação da massa molecular da proteína, realizamos os cálculos das constantes K_{av} , fizemos uma curva de calibração e uma regressão linear com os valores de K_{av} e os logaritmos das massas moleculares (figura 19).

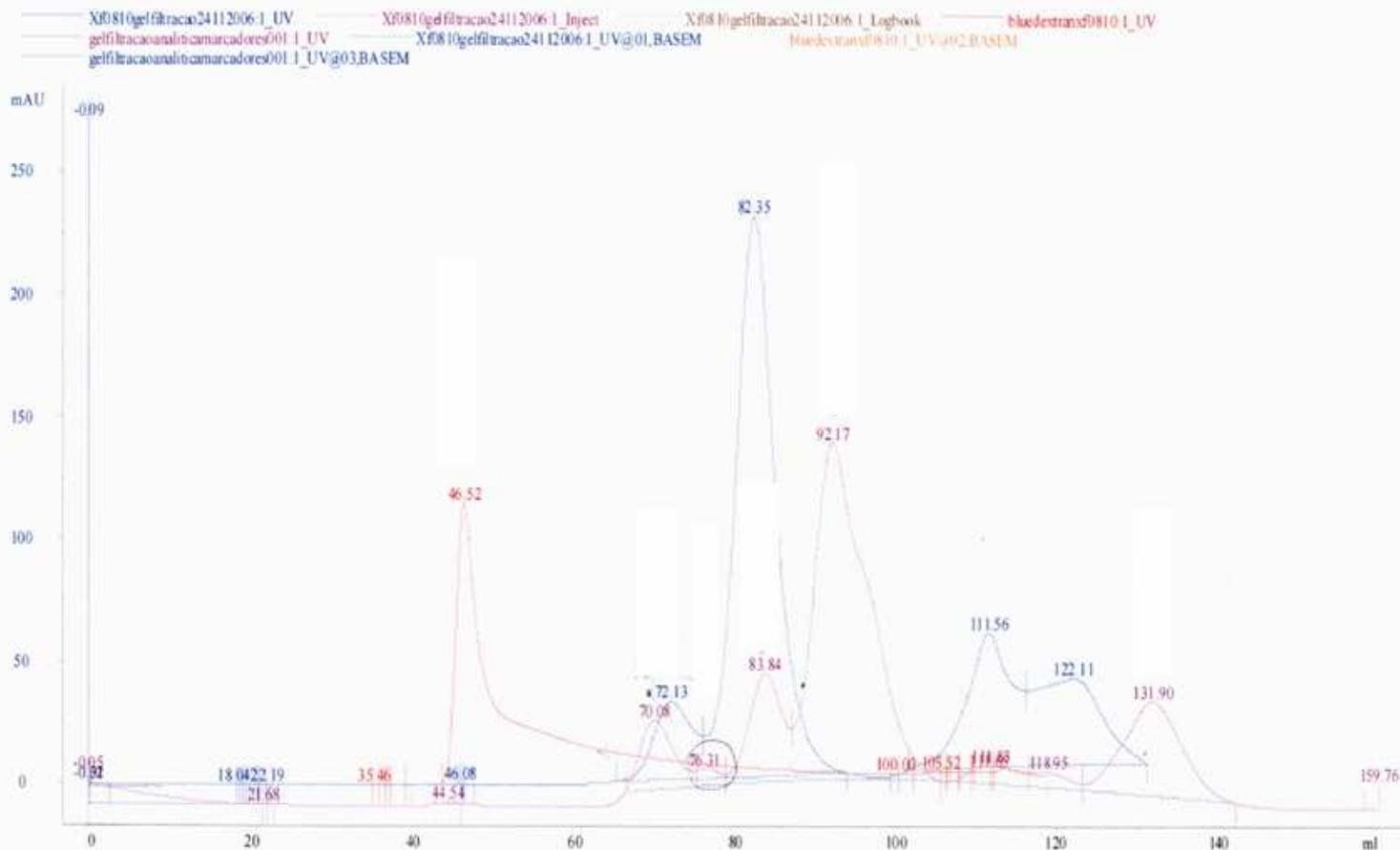
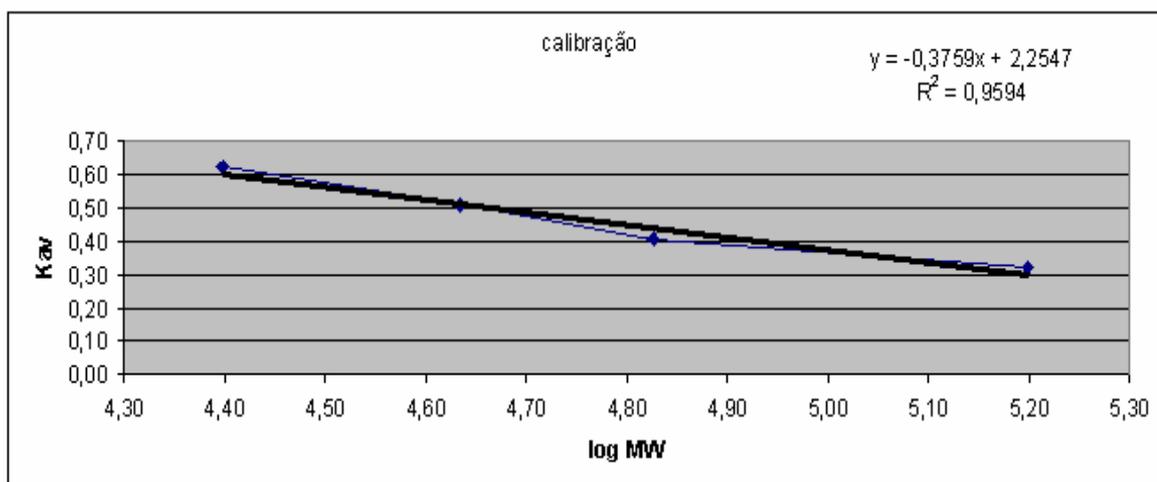


Figura 18: Cromatograma da filtração em gel analítica (amostra fração lavagem da QFF) da XF0180c, coluna Superdex 16/60. Bluedextran (pico 46,52) está em vermelho. Os marcadores em roxo: aldolase (pico 70,08), albumina (pico 76,31), ovoalbumina (pico 83,84), quimiotripsina (pico 92,17) e a amostra da proteína XF0180c em azul (pico1: 72,13 e pico 2: 82,35).

proteína	MW	Ve	Kav	LOG MW
quimiotripsina	25000	92,17	0,62	4,40
ovalbumina	43000	83,84	0,51	4,63
albumina	67000	76,31	0,41	4,83
aldolase	158000	70,08	0,32	5,20
bluedextran		46,52		
volume coluna		120		



XF0810	Ve	Kav	log MW	MW
pico 1	72,13	0,35	5,07	117747 Da
pico 2	82,35	0,49	4,70	50228 Da

Figura 19: Regressão linear (linha preta) e curva de calibração (linha azul) geradas a partir da filtração em gel analítica.

Com o resultado da purificação de filtração em gel analítica, constatamos que a proteína XF0810c (39,5 kDa) provavelmente se encontrava na forma de trímero (117,7 kDa) no pico 1 e na forma de monômero (50,2 kDa) no pico 2. Posteriormente a este experimento, realizamos a concentração da proteína submetendo a amostra do pico 2 (monômero) a uma centrifugação de 4620 g em temperatura de 4°C no filtro concentrador Amicon (10 MW). Em seguida foi realizada uma diálise em 20 mM de Tris-HCl pH 8,5.

4.1.6. Dicroísmo circular

Com a proteína XF0810c pura e na concentração de 2 mg/mL prosseguimos para realização do experimento de dicroísmo circular, porém antes da realização do mesmo, submetemos a amostra que se destinaria a este experimento a uma diálise contra o tampão 10 mM fosfato de sódio pH 7,4, isto foi necessário porque o tampão em que a amostra se encontrava anteriormente não era recomendado para o experimento pois aumentava a condutividade do aparelho. Analisando o espectro resultante do dicroísmo circular no UV distante (figura 20), observamos mínimos entre 205 e 222 nm característicos de estruturas secundárias (Woody, 1994; Kelly e Price, 1997). Utilizando os dados do experimento, realizamos as estimativas da estrutura secundária da proteína XF0810c, através do programa CDNN (tabela 4) e DICROPROT (tabela 5). Os resultados indicaram que a proteína apresentaria uma mistura de hélices α e fitas β , mostrando que estes estariam de acordo com os membros da família 5 de celulases, onde as estruturas são compostas por um barril-TIM.

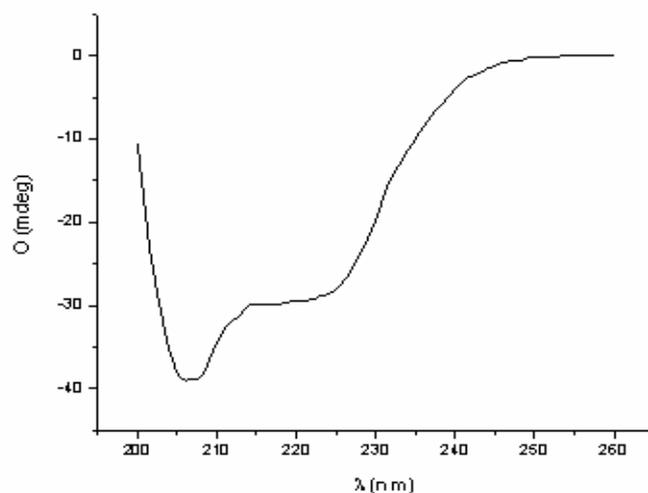


Figura 20: Espectro do dicroísmo circular da proteína XF0810c medida no UV distante.

Tabela 4: Estimativa de estrutura secundária da proteína XF0810c pelo programa CDNN.

CDNN	200-260 nm	205-260 nm	210-260 nm
Hélice α	32,7 %	36,6 %	33,2 %
Fitas β	17,4 %	15,3 %	16,9 %
Outros	47,0 %	47,0 %	50,1 %

Tabela 5: Estimativa de estrutura secundária da proteína XF0810c pelo programa Dicroprot.

DICROPROT	Hélice α	Fitas β
Fasman	0,37	-0,07
Chen	0,23	0,16
Bolotina	0,35	0,09
Chang	0,21	0,29
Yang	0,27	0,06
Self Consist Meth	0,37	-0,07
Self Selcon3 Meth	0,371	0,102
K2D Method	0,29	0,15
CONTIN Method	0,15	0,0
Varselec	24,15	0,91

4.1.7. Ensaio enzimático: exocelulase

Realizamos com a proteína XF0810c purificada do pico 2 na concentração de 2 mg/mL o ensaio enzimático para exocelulase medindo a absorvância à 420 nm nos tempos: 0, 5, 30, 60 e 240 minutos. Os resultados deste ensaio deram negativo para atividade de celulase (figura 21), estando de acordo os estudos de modelagem realizados para XF810c, onde verificamos a substituição de 4 dos 7 resíduos importantes que participam do sítio catalítico das celulases, indicando que XF0810 não faz parte da família 5 de celulases, pois os resíduos chaves estariam ausentes não permitindo a manutenção do mecanismo de clivagem na reação catalítica de retenção. Realizamos o mesmo ensaio enzimático utilizando a celulase de *Trichoderma reesei* (comercialmente adquirida através da empresa Sigma) em uma concentração de 1 mg/mL com a finalidade de termos um controle positivo para a atividade da enzima (figura 22). Em todos os ensaios foi realizada também a leitura das absorvâncias do branco da enzima (tampão do substrato + enzima) e branco do substrato (substrato + tampão da enzima) medidos nos mesmos tempos resultando em valores de 0,00 até 0,02 UA.

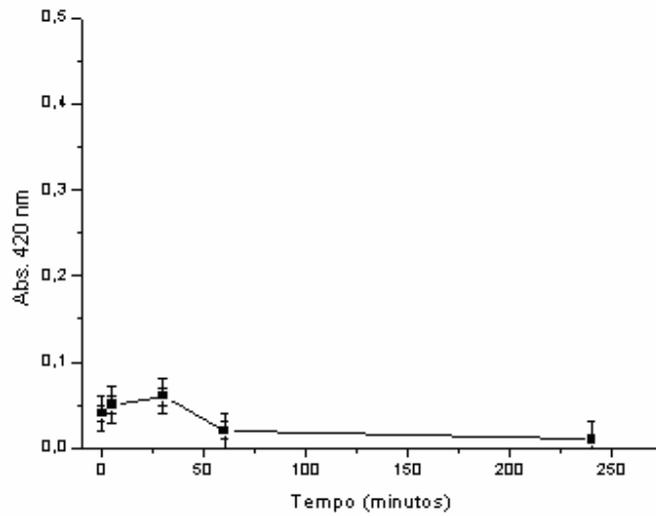


Figura 21: Resultado do teste de atividade enzimática para XF0810c, as medidas foram feitas em triplicata (amostras independentes) nos tempos de 0, 5, 30, 60 e 240 minutos. As medidas foram obtidas através do espectro de absorção a 420 nm.

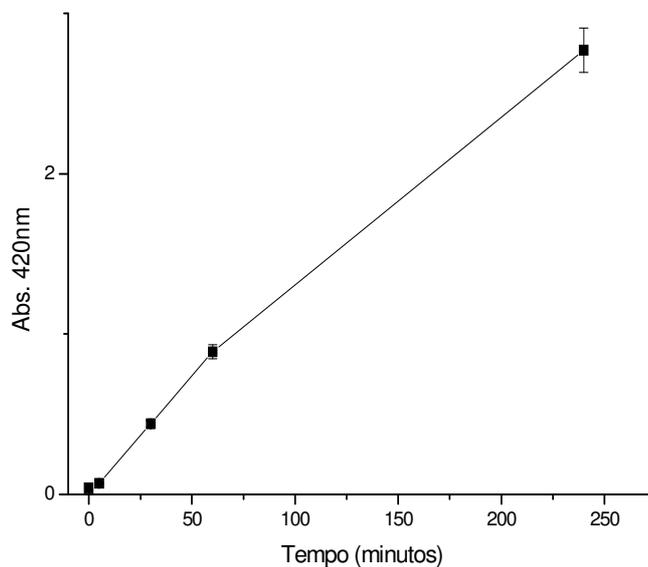


Figura 22: Resultado do teste de atividade enzimática para o controle positivo, a celulase de *Trichoderma reesei*, as medidas foram feitas em triplicata (amostras independentes) nos tempos de 0, 5, 30, 60 e 240 minutos. As medidas foram obtidas através do espectro de absorção a 420 nm.

4.1.8. Ensaios de cristalização

Para a realização dos ensaios de cristalização da proteína XF0810c, tivemos que aumentar a concentração da mesma submetendo a amostra contendo a proteína à centrifugação no filtro amicon (10 MW). A proteína atingiu a concentração de 3,6 mg/mL (uma concentração razoável para os testes de cristalização) e então através dos robôs Matrix Maker e Honeybee fizemos os screenings iniciais de cristalização: Crystal Screen/ Crystal Screen 2, Wizard I/ II, PACT, JCSG, Salt RX, Precipitant Synergy, utilizando 0,4 µL da amostra da proteína e 0,4 µL de amostra do tampão de cristalização em cada gota. Após uma semana do experimento, as gotas foram visualizadas e foi observado cerca de 30% de precipitado.

Dois meses após os ensaios de cristalização, visualizamos as gotas e notamos a presença de microcristais (figura 23a) da proteína XF0810c na condição D4 do kit JCSG (Molecular Dimensions) que contem 100 mM de HEPES de sódio pH 7,5, 10% p/v PEG 8000 e a presença de agulhas na condição E1 do mesmo kit que contem 100 mM de Tris pH 8,5, 40% p/v PEG 400 e 200 mM de sulfato de lítio (figura 23b).



(a)



(b)

Figura 23: Fotos dos cristais da proteína XF0810c nas gotas do kit JCSG (Molecular Dimensions) **(a)** microcristais - condição D4 e **(b)** agulhas - condição E1.

4.1.9. Modelagem Molecular

Foi realizada a modelagem molecular da celulase XF0810c, foi utilizado o alinhamento de sua seqüência de aminoácidos com a celulase de estrutura 3D conhecida: 1ECE (*Acidothermus cellulolyticus*) (figura 24), ao analisarmos o alinhamento, verificamos a falta de 4 dos 7 resíduos importantes que participam do sítio catalítico de celulases na XF0810c. Somente 3 resíduos estão conservados (Arg60, His112, His239 que equivalem a Arg62, His116 e His 238 na 1ECE – tabela 6), os resíduos catalíticos (Glu162 e Glu282) e os resíduos Asn161 e Try240, estão ausentes em XF0810 (Asp159, Met160, Phe289 e Ala241 respectivamente).

Figura 24: Alinhamento das celulases XF0810c e XF0818 com as celulase 1ECE (*Acidothermus cellulolyticus*) os números correspondem a XF0810c. O alinhamento foi produzido pelo programa CLUSTALX (Thompson et al., 1997). Asteriscos (*) representam os resíduos estritamente conservados, dois pontos (:) e ponto (.) representam resíduos parcialmente conservados. Os resíduos marcados em negrito pertencem a família 5 de celulases. A linha inferior mostra a estrutura secundária da 1ECE com as setas para fitas β e cilindros amarelos para hélices α .

O princípio do método utilizado para a construção do modelo é que a estrutura terciária/quaternária (coordenadas atômicas) é mais conservada do que a estrutura primária das proteínas, logo como a celulase XF0810c apresenta homologia entre sua seqüência de aminoácidos e a da proteína com estrutura 3D determinada (1ECE), podemos modelar a XF0810c. Um alinhamento com a ferramenta BLAST (Basic Local Alignment Search Tool) resultou em 39% de identidade e 58% de similaridade da XF0810c com a 1ECE.

Tabela 6 - Aminoácidos conservados nas celulases da família 5 em 1ECE e aminoácidos correspondentes na XF0810c (em vermelho estão marcados os aminoácidos que não são conservados).

1ECE	XF0810c
R62	R60
H116	H112
N161	D159
E162	M160
H238	H239
Y240	A241
E282	F289

O programa utilizado para a construção do modelo foi o MODELLER que utiliza arquivos de entrada: alinhamento das proteínas (.pir), as instruções do job (.top, foram utilizados valores padrão) e coordenadas atômicas depositadas no PDB da proteína de estrutura conhecida considerando parâmetros de dinâmica molecular e impedimento espacial dos átomos. Foram gerados cinco modelos para celulase XF0810c. O melhor modelo foi escolhido de acordo com a avaliação da qualidade de cada um utilizando os programas VERIFY 3D (figura 25) que analisa regiões mal construídas e PROCHECK (figura 26) que checa a qualidade estereoquímica.

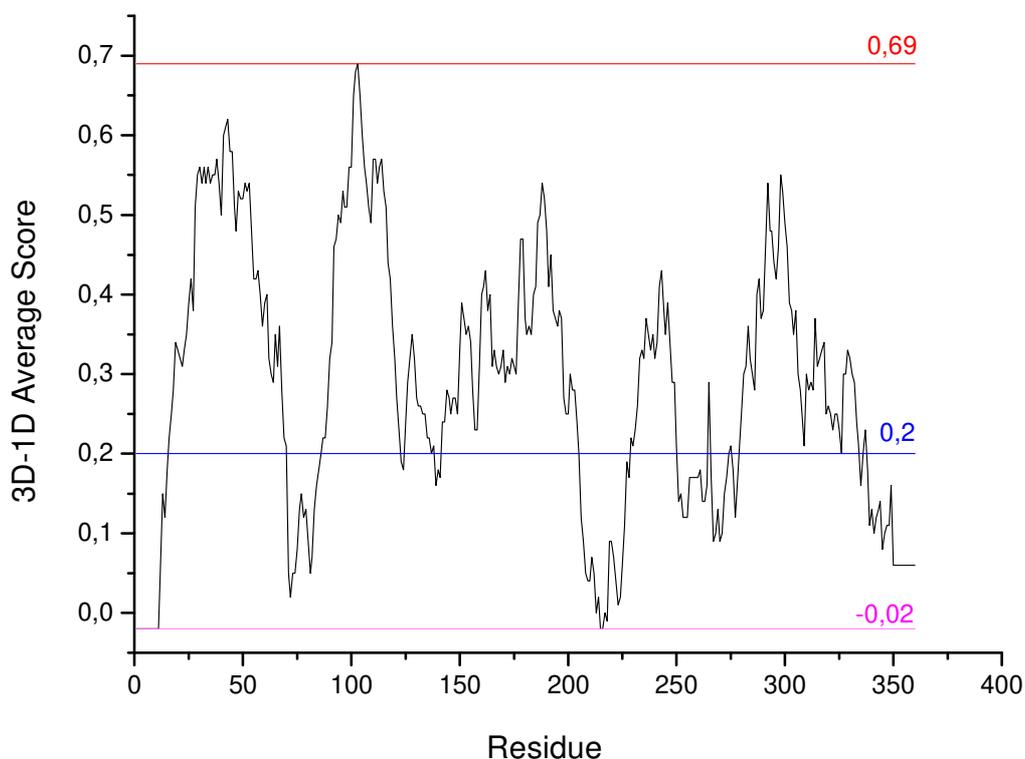


Figura 25: Gráfico do valor médio 3D-1D por resíduo do modelo XF0810c obtido com o programa VERIFY 3D .

O programa VERIFY 3D trabalha com o inverso do enovelamento de proteína, encontrando na seqüência de aminoácidos o que é mais compatível com o ambiente dos resíduos na estrutura tridimensional. O ambiente é descrito pela área do resíduo enterrado na proteína e inacessível ao solvente, a fração da área da cadeia lateral que é coberta por átomos polares (O e N) e a estrutura secundária local. De acordo com o gráfico gerado para o modelo da XF0810c, o score total para 360 resíduos de aminoácidos foi de 103,78 e verificamos que o resultado da média 3D-1D de resíduos acima de 0,2 é de 70% demonstrando que a seqüência é compatível com a estrutura 3D e que o modelo somente apresenta problemas em algumas regiões (valor < 0,2) em torno dos resíduos 70-80, no N e C terminal pelo fato destas regiões na seqüência de aminoácidos da XF0810c apresentarem deleções e nas regiões em torno dos resíduos 205-210, 250-260 pelo fato da seqüência apresentar inserções.

Analisando o gráfico de Ramachandran observamos que dois resíduos (Thr313 e Lys338) se encontram na região generosamente permitida, uma justificativa para este fato seria que Thr313 equivale a Arg304 na estrutura 1ECE utilizada para o modelo e que já ocupava uma região generosamente permitida e a Lys338 ocupa o mesmo lugar que uma glicina na estrutura da 1ECE.

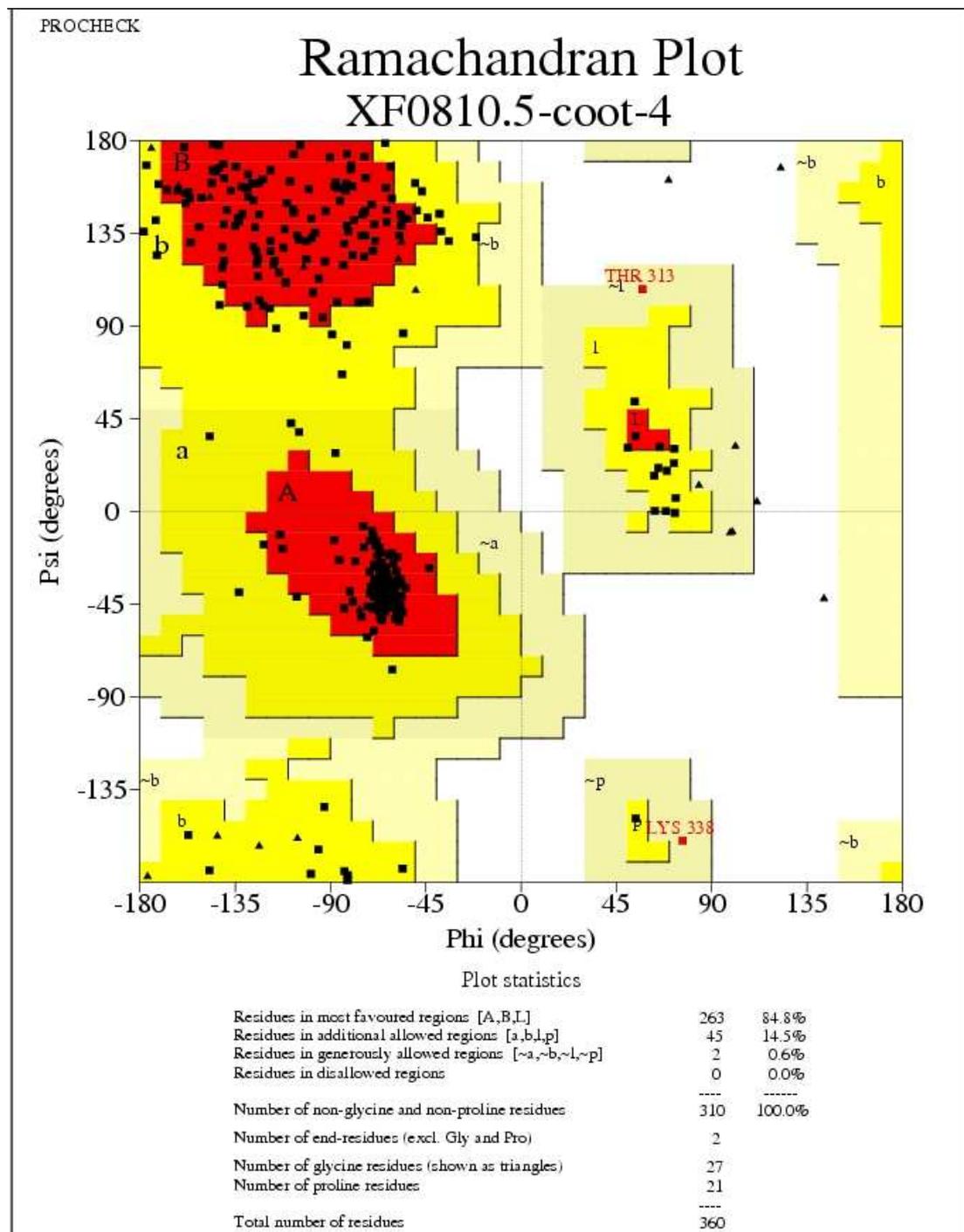


Figura 26: Gráfico de Ramachandran do modelo XF0810c.

A figura 27 ilustra o modelo gerado da XF0810c mostrando o enovelamento de barril TIM característico da família 5 de celulase.

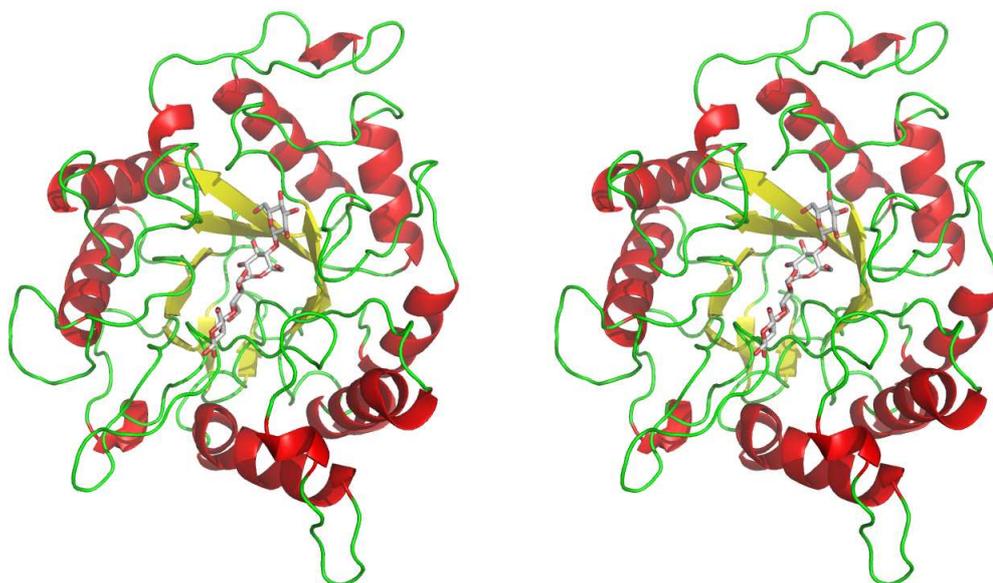


Figura 27: Vista estéreo do modelo da XF0810c com o ligante. O ligante foi modelado usando as coordenadas do ligante presente na estrutura resolvida 1ECE.

Na figura 28, podemos observar os sete resíduos conservados característicos da família 5 na 1ECE e na figura 29 os resíduos equivalentes para o modelo construído XF0810c com presença de um ligante (açúcar). As interações entre os resíduos com o ligante podem ser vistas como linhas tracejadas.

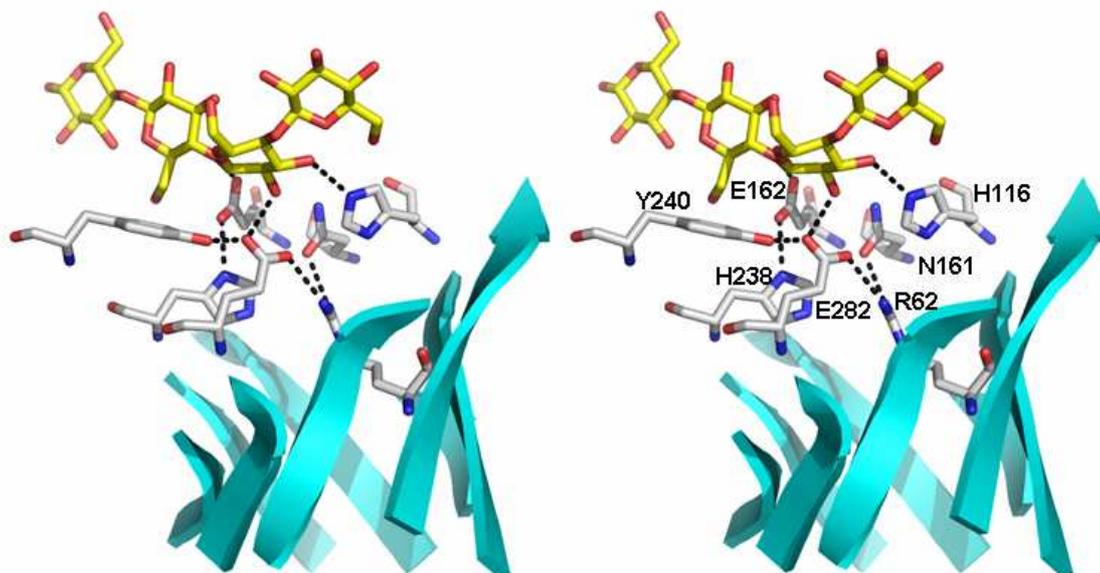


Figura 28: Vista estéreo da 1ECE com os resíduos conservados característicos na região catalítica da família 5 das celulasas interagindo com análogo do substrato.

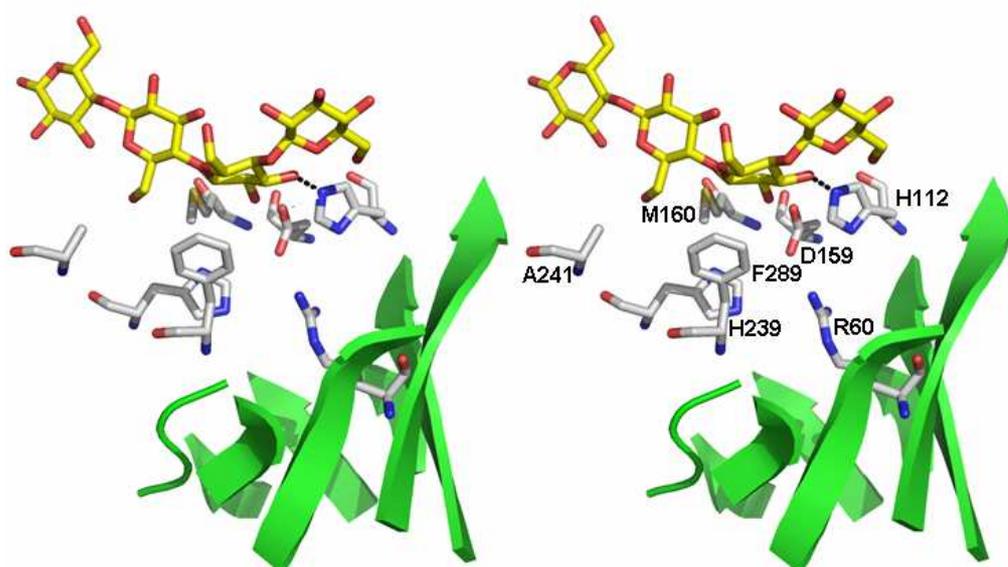


Figura 29: Vista estéreo do modelo da XF0810c com os resíduos equivalentes aos resíduos conservados na 1ECE na região catalítica interagindo com um análogo do substrato.

Analisado o modelo da XF0810c notamos que os resíduos ausentes Asn161 e Try240 na 1ECE substituídos por Asp159 e Ala241 e principalmente os catalíticos Glu162 e Glu282 na 1ECE substituídos por Met160 e Phe289 não podem ser compensados por outros nas proximidades e não podem exercer o mesmo papel que na 1ECE. Comparando as figuras 28 e 29 podemos observar que a única ligação de hidrogênio conservada entre a 1ECE e o modelo XF0810c é a da H116/H112 com o substrato, todas as demais foram desfeitas.

Na figura 28 podemos observar a importância das ligações de hidrogênio para manter a conformação e estabilização dos grupos carboxilas dos resíduos catalíticos Glu162 e Glu282. Temos como exemplos o NH1 da R62 que interage com OE2 do Glu162 e o OH da Tyr240 faz interação com o OE1 do Glu282.

Verificamos que não seria possível ocorrer o mecanismo de clivagem na XF0810c, pois a Met160 não atuaria como doador de prótons como o Glu162 e a Phe289 não atuaria como nucleófilo como atua o Glu282 não permitindo assim a catálise, estes fatos nos levam a propor que a XF0810c não faz parte da família 5 de celulases. Na figura 30 podemos observar a sobreposição da estrutura 1ECE com o modelo da XF0810c destacando a diferença de três resíduos substituídos.

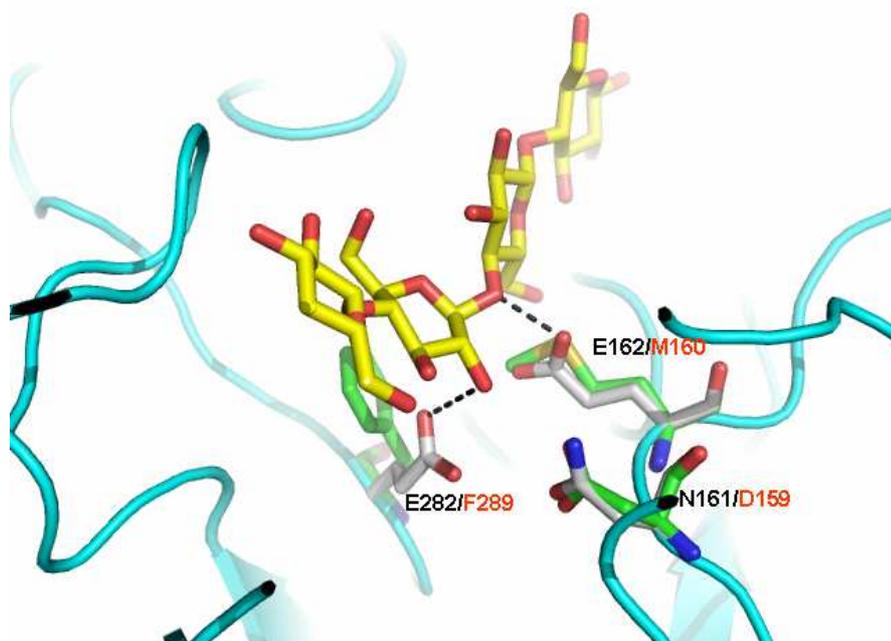


Figura 30: Sobreposição dos resíduos catalíticos da 1ECE (cinza) e dos equivalentes na XF0810c (verde).

4.2. Lisozima 1 de *M. domestica* (MdL1)

Em colaboração com o Prof. Dr. Sandro Marana e a aluna de doutorado Fabiane Cançado do Instituto de Química da Universidade de São Paulo foi possível a obtenção do conhecimento teórico e prático das técnicas de cristalização e cristalografia de proteínas através das atividades realizadas com a proteína lisozima 1 de *Musca domestica* (MdL1).

4.2.1. Cristalização e soaking

Amostras da proteína lisozima 1 de *Musca domestica* (MdL1) purificada de acordo com o artigo Cançado *et. al.* 2007, na concentração de 9,6 mg/mL em

tampão de 10 mM de ácido 2-morfolinoetanosulfônico (MES) foram submetidas à cristalização utilizando o método de difusão de vapor de gota sentada (sitting drop) colocando 1 μ L de proteína e 1 μ L de solução do reservatório das placas de cristalização na melhor condição de cristalização encontrada (Marana *et al.*, 2006) (figura 31): 1,4 M de sulfato de amônio, 0,1 M de sal ácido de sódio 4-(2-hidroxietil) piperazina-1-etanosulfônico (HEPES sódio) pH 7,5 e 1% de PEG 400. Os cristais cresceram em uma ou duas semanas em uma temperatura de 18°C. Para incorporação dos ligantes específicos no cristal da proteína foi utilizado o método de soaking que consistiu em colocar 1 μ L da solução de cada ligante Q3 (0,2 mM), Q4 (1,0 mM), Q5 (1,0 mM) e Q6 (0,5 mM) nas gotas com cristais da proteína, estes cristais foram deixados a 18°C no período de uma semana para a incorporação do ligante.

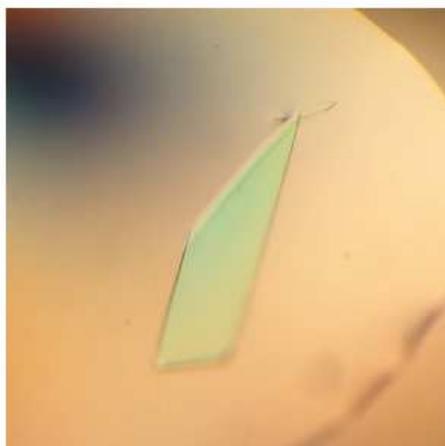


Figura 31: Cristal da lisozima 1 de *Musca domestica*.

4.2.2. Difração de raios X e processamento de dados

Os cristais da proteína MdL1 com os ligantes foram levados à linha de cristalografia de proteínas do Laboratório Nacional de Luz Síncrotron (LNLS) D03B-MX1 onde foram testados para o experimento de difração de raios X. Em cada cristal foi colocada uma solução crioprotetora de 20% de glicerol para evitar a formação de anéis de gelo. Cristais contendo os ligantes Q3, Q4 e Q6 difrataram até 1,85 Å (figura 32a), 1,92 Å (figura 32b) e 2,25 Å (figura 32c) de resolução, respectivamente. Como podemos observar na figura 32b e c os padrões de difração apresentam anéis de gelo que são caracterizados pela formação de círculos com difração bem intensas, a formação de gelo no cristal foi devida ineficiência da solução crioprotetora. O processamento dos dados coletados dos três complexos foi realizado através do programa HKL2000 utilizando Denzo e Scalepack, foi feita a indexação, refinamento da célula unitária, integração, escalonamento e agrupamento. Os cristais dos complexos Q3, Q4 e Q6 pertencem ao grupo espacial monoclinico $P2_1$, a tabela 7 sumariza as estatísticas dos dados coletados e processados.

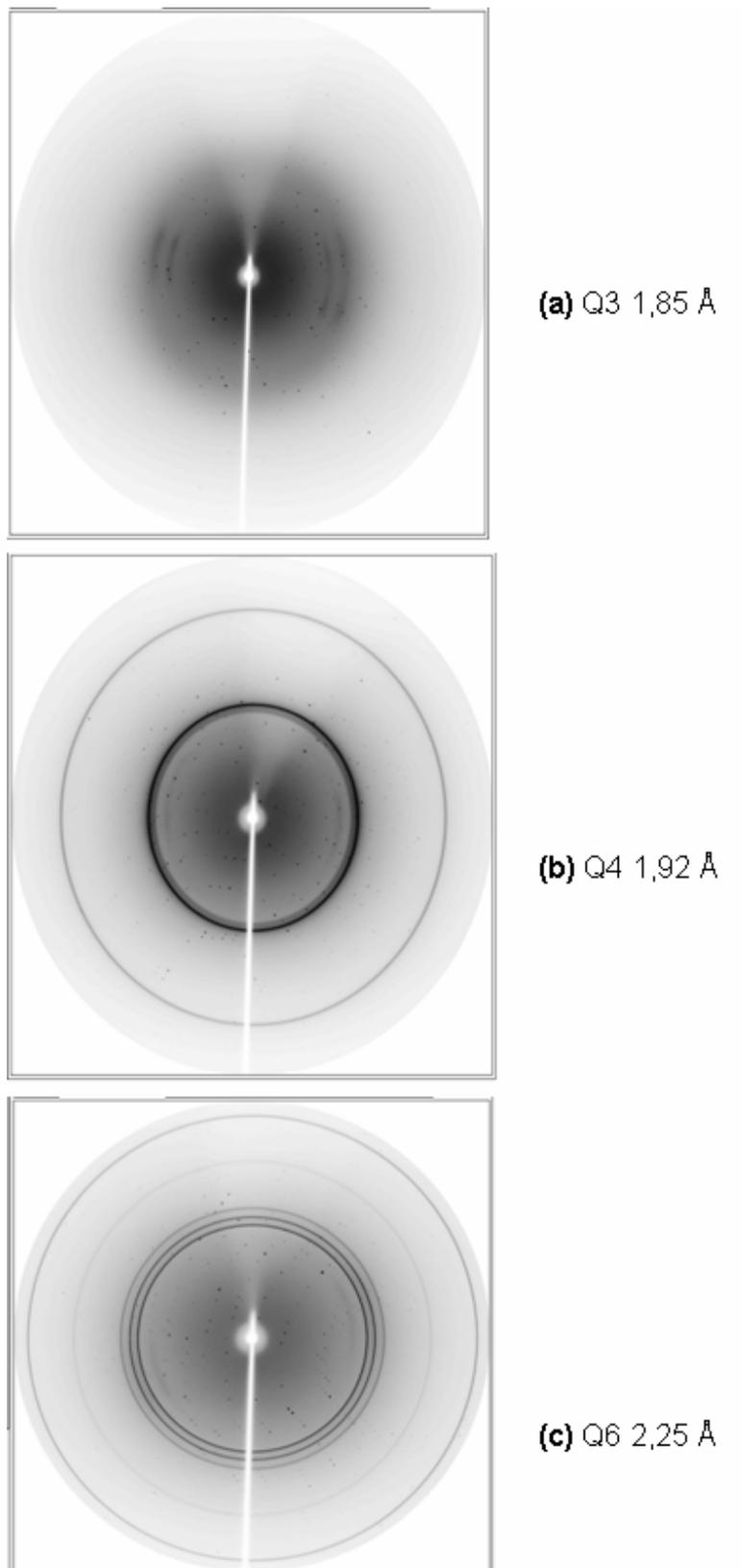


Figura 32: Padrão de difração da lisozima 1 complexada.

Tabela 7: Estatística dos dados obtidos da coleta dos cristais de lisozima 1 de *Musca domestica* com os ligantes Q3, Q4 e Q6. Os valores entre parênteses correspondem ao da última camada de resolução.

	Q3	Q4	Q6
Grupo espacial	P2 ₁	P2 ₁	P2 ₁
Cél. unitária(Å; °)	a = 36.1, b = 78.9, c = 44.9; $\alpha = \gamma = 90.0 \beta = 103.1$	a = 35.9, b = 78.8, c = 44.8; $\alpha = \gamma = 90.0 \beta = 102.2$	a = 35.7, b = 79.0, c = 44.5; $\alpha = \gamma = 90.0 \beta = 102.0$
Mosaicidade (°)	0.5	1.0	1.5
Temperatura (K)	100	100	100
Comp. de onda (Å)	1.427	1.427	1.427
Rotação (°)	1.0	1.0	1.0
Distância cristal-detector (mm)	80.0	83.5	100.0
No. Imagens	140	201	208
Limite resolução (Å)	30.00-1.85 (1.92-1.85)	30.00-1.92 (1.99-1.92)	30.00- 2.25 (2.33- 2.25)
I/σ(I)	16.5 (2.3)	16.2 (2.5)	8.0 (1.9)
Completeza (%)	97.3 (80.6)	88.1 (82.0)	93.5 (88.4)
Multiplicidade	2.8 (2.2)	3.8 (2.8)	3.7 (2.7)
R_{sym}	0.063 (0.285)	0.078 (0.307)	0.183 (0.534)
No. de reflexões	57117	62924	39775
No. de reflexões únicas	20442 (1693)	16455 (1521)	10870 (1007)

4.2.3. Substituição molecular e refinamento

Para resolução das estruturas cristalográficas dos dados coletados e processados é necessário o conhecimento das fases. O método utilizado para obtenção deste dado foi o de substituição molecular tendo como modelo um monômero da estrutura do cristal nativo de MdL1 (este pertence ao mesmo grupo espacial dos derivados, monoclinico $P2_1$) que foi resolvida pelo método de substituição molecular tendo como modelo a estrutura da lisozima de clara de ovo de galinha (1HEWL) apresenta 38,5% de identidade e 55,4% de similaridade. As coordenadas da MdL1 nativa foram depositadas no PDB (Bernstein *et al.*, 1997) com o código 2FBD. Retiraram-se as moléculas de solvente e manteve-se somente uma das conformações dos átomos com dupla conformação. O coeficiente de Matthews (Matthews, 1968) foi calculado para lisozima 1 (Marana *et al.*, 2006) para duas moléculas na unidade assimétrica como sendo $2,3 \text{ \AA}^3 \text{ Da}^{-1}$, tendo uma porcentagem de solvente de 45%. A substituição molecular foi realizada com o programa MOLREP, a melhor solução encontrada pelo programa apresentou um R_{factor} de 0,369 e um coeficiente de correlação de 0,662. As fases obtidas foram utilizadas para calcular os mapas de diferença de cada complexo e a visualização foi feita pelo programa O e COOT.

Ao analisarmos os mapas das estruturas dos complexos, escolhemos o complexo Q4 para o refinamento por este apresentar um mapa com melhor definição da densidade eletrônica do ligante. Primeiramente realizamos um refinamento pelo do programa CNS utilizando o método “simulated annealing”

com o intuito de melhorar a eficiência do refinamento. Esse procedimento possibilita minimização de energia do sistema com aumento de temperatura acarretando o movimento molecular do sistema com a finalidade de reduzir a influência da estrutura utilizada como modelo na substituição molecular. Os demais ciclos posteriores de refinamento foram realizados pelo programa REFMAC do pacote de programas CCP4. Os valores dos fatores R do último ciclo de refinamento foram de $R_{\text{factor}}=0,178$ e $R_{\text{free}}=0,242$.

A validação da estrutura do complexo lisozima 1 e o ligante Q4 foi feita utilizando a interface do programa COOT (Emsley & Cowtan, 2004). Foram analisados os parâmetros estereoquímicos pela análise do gráfico de Ramachandran, gráficos de distorção geométrica, variância do fator de temperatura, distorção do ângulo ômega, análise de rotâmeros, ajuste da densidade eletrônica e outros. Foi realizado o depósito estrutura no PDB segundo as instruções do banco de dados (Bernstein *et al.*, 1997). A estrutura recebeu o código 2H5Z.

A estrutura da lisozima 1 de *Musca domestica* com o ligante apresenta um dímero na unidade assimétrica, para verificação da possibilidade da existência do mesmo em solução com significância biológica realizamos o experimento de DLS (dynamic light scattering) da amostra da proteína em solução que resultou em um padrão monodisperso com uma massa estimada de 13 kDa (a massa calculada da lisozima é de 14.200 Da). Também foram realizados experimentos de cromatografia de filtração em gel que indicaram uma massa molecular estimada de 6,5 kDa para MdL1 e na eletroforese em condições não desnaturantes mostrou uma massa relativa de 14.000 kDa para MdL1 (dados não demonstrados obtidos

por Cançado F.C. e Marana S.R.). Uma análise *in silico* foi feita com o programa PISA http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html (Krissinel and Henrick, 2005) que calcula a área da interface do dímero de 490,9 Å² (menos que 8% da superfície do monômero), apresentando valor de complexação de zero. Todos estes dados indicam que o dímero não ocorre naturalmente e é resultante do empacotamento cristalino.

Os monômeros pertencem ao enovelamento da família 22 de lisozimas, consistindo em um “core” central composto por três hélices α longas e uma pequena (Thr4-Leu15, Glu21-Ser34, Ile86-Gly101 e Thr107 e Ser112) e uma folha β antiparalela com duas fitas (As40-Try51 e Ile56-Asn57) e quatro 3_{10} hélices. A folha beta difere de outras lisozimas que são compostas geralmente por três fitas β . A estrutura tridimensional do monômero da enzima MdL1 está representada na figura 33.

Na unidade assimétrica, a molécula A não é ocupada pelo açúcar apesar do seu sítio ativo estar exposto a uma mesma região do solvente que o sítio ativo ocupado da molécula B. Uma possível razão para isso é que se o trissacarídeo estiver presente na molécula A na mesma configuração que na molécula B, o sítio -4 do sacarídeo na molécula A tem um impedimento estérico com a cadeia principal da Gln100 da molécula B conforme ilustrado na figura 34.

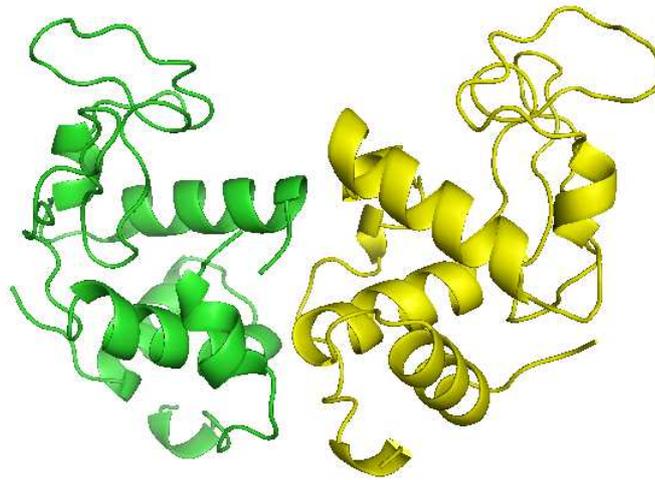


Figura 33: Representação esquemática da estrutura de lisozima 1 de *Musca domestica* mostrando um dímero na unidade assimétrica. Figura gerada com o programa PyMOL (Delano, 2002).

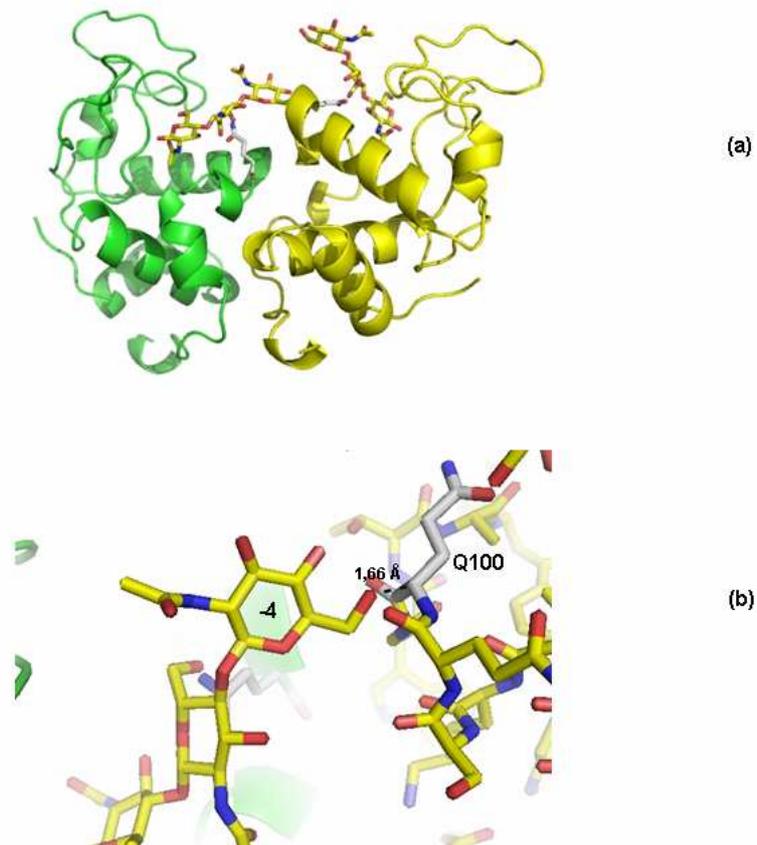


Figura 34: (a) Monômeros da Mdl1 com açúcar, em verde o monômero A e em amarelo monômero B. (b) Zoom no sítio -4 do açúcar e o resíduo Q100 do monômero B mostrando a distância de 1,66 Å.

Na figura 35 observamos a cavidade do monômero B da MdL1 para entrada do ligante tetra-N-acetilglucosamina com o mapa da densidade eletrônica [Fo-Fc], podemos observar que a densidade está fraca no sítio -4 do açúcar devido a este não fazer interações com a proteína, porém discutiremos este ponto mais adiante.

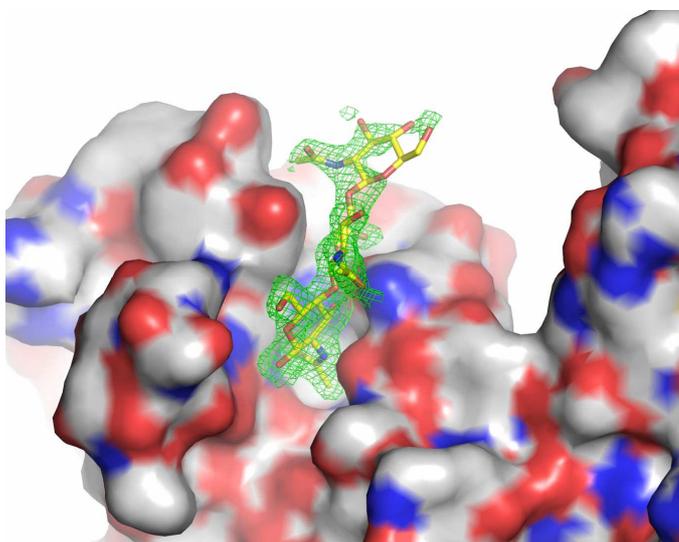


Figura 35: Superfície da molécula MdL1: zoom na cavidade para entrada do ligante Q4. A cor azul representa os átomos de nitrogênio e a vermelha os átomos de oxigênio. O mapa da densidade eletrônica para o ligante [Fo-Fc] está representado em verde com um nível contorno de 3σ . A figura foi desenhada com o programa PyMOL (Delano, 2002).

Foi realizado o alinhamento estrutural da MdL1 com lisozimas da família 22 de quatro classes de animais diferentes: 1JSF de mamífero *Homo sapiens* (Harata et al., 1998), 1LMN de peixe *Oncorhynchus mykiss* (Karlsen et al., 1995), 1HEWL de ave *Gallus gallus* egg white (Cheetham et al., 1992), 1GD6 de inseto *Bombyx*

mori (Matsura et al., 2002) e 1IIZ de inseto *Antheraea mylitta* (Jain et al., 2001) (figura 36). Todas estas estruturas correspondem a lisozimas associadas a função imune e nenhuma estrutura de lisozima digestiva havia sido depositada no PDB até o presente trabalho. Foram escolhidas duas estruturas de lisozimas de insetos (1GD6 e 1IIZ) por estarem mais próximas da de mosca podendo nos propiciar maiores informações.

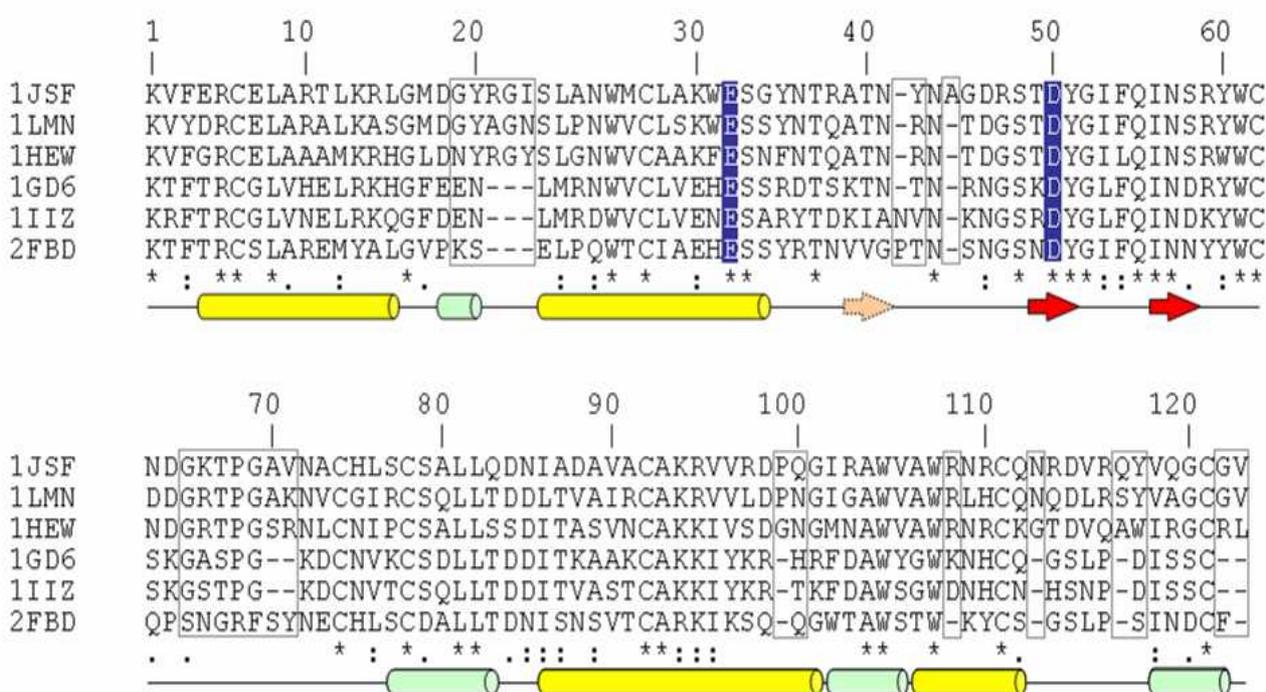


Figura 36: Alinhamento múltiplo de estruturas representativas de lisozimas da família 22: 1JSF, 1LMN, 1HEWL, 1GD6, 1IIZ e 2FDB. Os resíduos catalíticos estão marcados em azul. Nos boxes estão as regiões que não estão estruturalmente alinhadas, a numeração se refere à estrutura 2FBD. Este alinhamento é uma versão modificada (foram feitos alinhamentos estruturais) da inicialmente produzida pelo programa CLUSTALX (Thompson et al., 1997). Asteriscos representam os resíduos conservados, e os dois pontos e ponto representam os resíduos parcialmente conservados. A linha inferior mostra a estrutura secundária da 2FBD com as setas para fitas β e cilindros amarelos para hélices α e verde para hélices 3_{10} .

O alinhamento estrutural mostra nove regiões de inserções ou deleções, com cinco delas na região C-terminal. A deleção de 3 aminoácidos na região do resíduo Ser20, onde os resíduos adotam a conformação de hélice 3_{10} parece ser uma característica de lisozimas de insetos. A segunda região que contém gaps no alinhamento é entre a Val40 e o catalítico Asp50 mostrando diferenças no modo de acomodar e manter os resíduos Asn44, Asn/Asp46 e Ser48, os quais junto com Asn57 são responsáveis pelo posicionamento da cadeia lateral do Asp50. A maior região não alinhada é em torno do resíduo 70, outra deleção é próxima ao resíduo Gln100 que é uma característica de lisozimas de insetos. Uma análise estrutural destas estruturas sobrepostas através do programa COOT (Emsley & Cowtan, 2004) o que observamos foi a similaridade do core estrutural, verificando algumas diferenças significativas na região dos loops/voltas como mostra a figura 37.

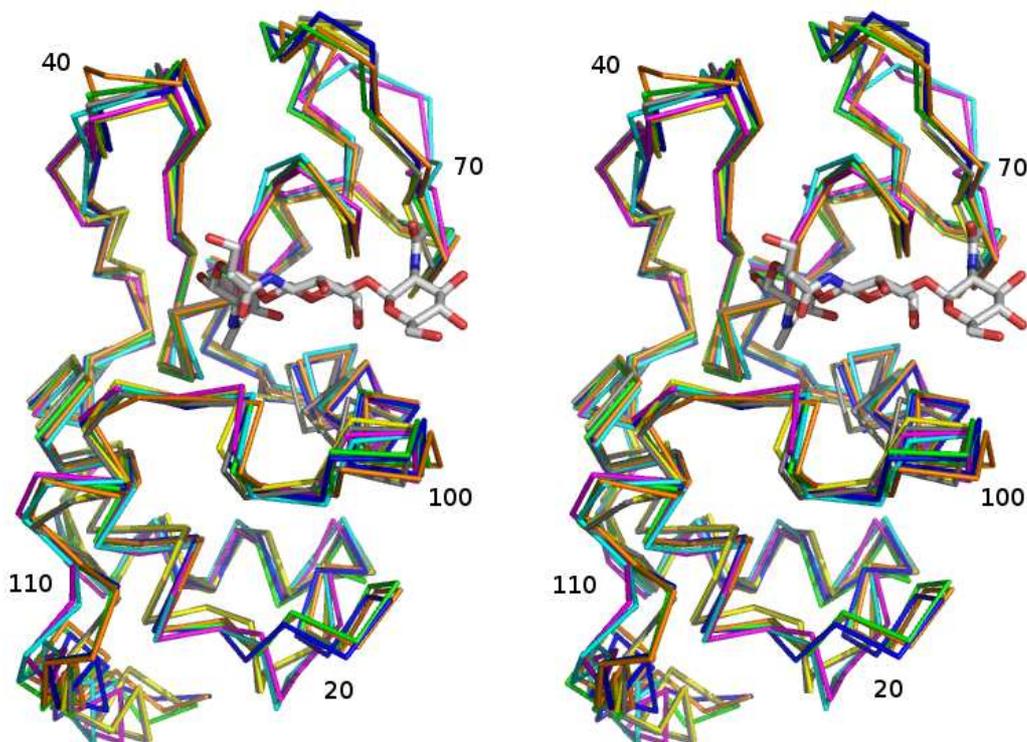


Figura 37: Vista estéreo do enovelamento da lisozima (família 22), superposição de carbonos-alfa de todas as classes de animais representadas no PDB: 2H5Z em amarelo, 2FBD em cinza, 1IIZ em magenta, 1GD6 em cian, 1HEWL em azul, 1JSF em laranja e 1LMN em verde. As regiões que apresentam diferenças entre as estruturas estão indicadas pelos números que mostram os resíduos de aminoácidos destas regiões.

O RMSD entre as estruturas superpostas demonstrou que MdL1 é apenas um pouco mais próxima das outras lisozimas de insetos (média RMSD de 1,09 Å) do que das outras lisozimas representadas por aves, mamíferos e peixes (média RMSD de 1,19 Å) (tabela 8). Este fato contrasta com a grande similaridade entre as estruturas de insetos (RMSD de 0,66 Å) e indica que as divergências para a função digestiva exigiram uma maior adaptação da estrutura.

Tabela 8: Superposição dos representantes das lisozimas da família 22 de insetos: 2FBD_A, 2FBD_B, 2H5Z_B (*Musca domestica*), 1GD6 (*Bombyx mori*) e IIZ (*Anthera mylitta*); aves: 1HEWL (*Gallus gallus*); mamíferos: 1JSF (*Homo sapiens*) e peixes: 1LMN (*Oncorhynchus mykiss*). O _A e _B no final dos códigos do PDB denotam o monômero. Os números na última coluna representam a média dos valores dos números marcados na escala cinza respectivamente.

	Resolução (Å)	Identidade (%)	RMSD (Å)							mean
			2FBD_B	2FBD_A	2H5Z_B	1GD6	1IIZ	1HEWL	1JSF	
2FBD_A	1.90	100	0.24							0.27
2H5Z_B	1.92	100	0.34	0.24						1.09
1GD6	2.50	46	1.12	1.08	1.14					0.66
1IIZ	2.40	43	1.11	1.05	1.06	0.66				1.19
1HEWL	1.75	43	1.21	1.21	1.22	1.04	1.22			1.02
1JSF	1.15	46	1.17	1.20	1.19	0.98	1.20	0.66		1.20
1LMN	1.80	42	1.10	1.23	1.18	1.03	1.17	0.56	0.47	0.56

As adaptações evolutivas das lisozimas foram matéria de revisão em Prager, 1996, tais como, um menor número de resíduos de arginina, menos ligações lábeis a ácidos (Asp, Asn e Gln), especificamente a falta da ligação mais lábil a hidrólise ácida Asp99-Pro99A (101 e 102 na numeração do esquema de Prager). Nesta região em MdL1, definida pelos resíduos 98-100, há uma deleção antes do resíduo Gln100. Esta região forma uma volta de conformação “mais fechada” (menos exposta ao solvente) em relação a outras estruturas (figura 38), este fato poderá contribuir para a resistência desta enzima ao ataque proteolítico, uma característica das lisozimas digestivas.

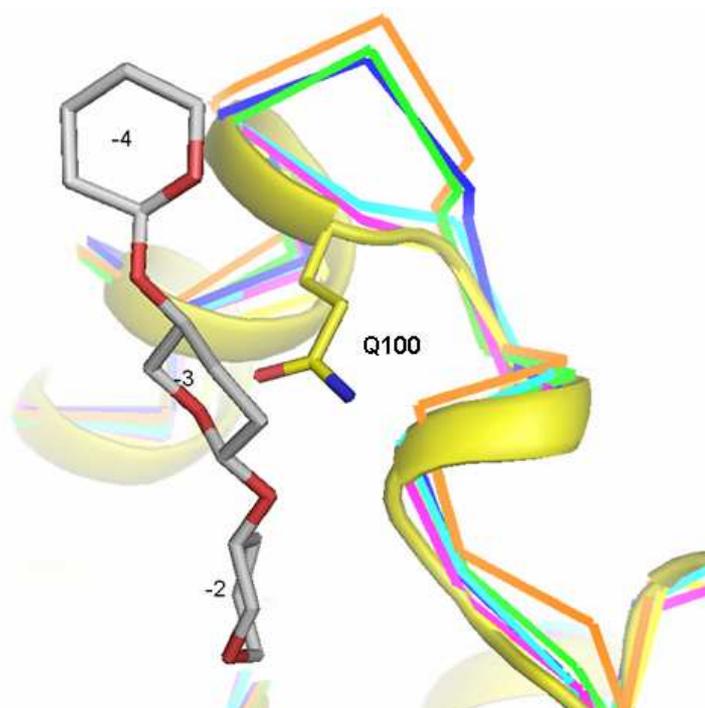


Figura 38: Zoom na volta (98-100) das lisozimas da família 22 com o ligante representado apenas pelos anéis. Superposição de carbonos-alfa de todas as classes de animais representadas no PDB: 2H5Z em amarelo, 1IIZ em magenta, 1GD6 em cyan, 1HEWL em azul escuro, 1JSF em laranja e 1LMN em verde.

Outra adaptação das lisozimas digestivas estudada é um pH ótimo mais ácido comparado com outras lisozimas (Dobson *et al.*, 1984 e Prager 1996). A diminuição do pH ótimo é um reflexo direto da diminuição nos valores de pK_a s dos grupos carboxila das cadeias laterais dos resíduos catalíticos: o nucleófilo Asp50 e o doador de prótons Glu32. No pH ótimo da enzima ($\approx 5,0$), o Asp50 está desprotonado e o Glu32 protonado. Considerando que a atividade da enzima depende destes dois grupos ionizáveis, o pH ótimo da enzima é calculado pela média aritmética entre os pK_a s destes grupos (Segel, 1993). Buscamos então

fatores na estrutura que justificassem uma redução nos pK_a s dos grupos catalíticos e conseqüentemente implicassem num pH ótimo mais ácido para MdL1.

O fato do pH ótimo ser mais ácido é parcialmente justificado na estrutura de MdL1 pelo fato da existência de uma ligação de hidrogênio entre o O γ da Thr107 com o O ϵ 2 do Glu32. Comparado com outras estruturas do alinhamento (figura 36), esta treonina é substituída por uma alanina ou glicina em outras seqüências de lisozimas que não podem formar a ligação de hidrogênio. Mudanças da Thr107 para glicina ou alanina foram feitas manualmente pela edição nas coordenadas do PDB de MdL1 e cálculos teóricos do efeito produzido pela mudança sugerem que a presença desta ligação de hidrogênio é responsável pelo decréscimo de 0.6 unidades no pK_a do resíduo catalítico Glu32 (tabela 9).

Tabela 9: pK_a teórico dos resíduos catalíticos da MdL1 calculados pelo programa PROPKA (Li *et al.*, 2005). Os números separados pelas barras correspondem a valores do pK_a para o monômero A e B respectivamente.

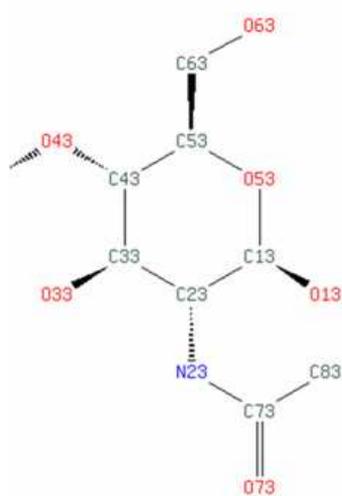
	Glu32	Asp50
Nativa (2FBD)	4.96/5.18	3.94/3.66
Complexo (2H5Z)	4.99/5.15	3.83/4.00
Complexo T107A	5.63/5.79	3.81/3.98
Complexo T107G	5.56/5.73	3.80/3.97

Realizamos uma comparação das interações da enzima MdL1 com o ligante tetra-N-acetilglucosamina com a estrutura 1HEWL resolvida com o ligante tri-N-acetilglucosamina, a densidade eletrônica encontrada para o ligante na MdL1 mostra uma triose que corresponde ao (NAG)₃ nos seis subsítios de ligação de substrato classificados -4 à +2, originalmente nomeado A à F, um por

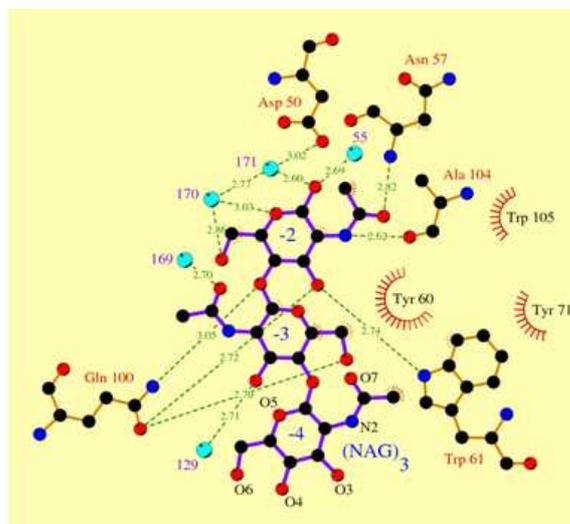
monossacarídeo (Philips, 1967; Blake *et al.*, 1967; Davies *et al.*, 1997). A configuração é similar a da estrutura 1HEWL (Cheetham *et al.*, 1992), os sítios -2 e -3 apresentam boas densidades, enquanto o -4 é apresenta uma densidade pobre devido a baixa ocupância e/ou desordem, este fato pode estar associado a hidrólise do ligante em tri-,di-e monossacarídeo e também pelo fato desta unidade de açúcar não apresentar interações com a proteína e com moléculas de água. A mesma razão da baixa ocupância/desordem devido a hidrólise pode explicar o fato que nós encontramos um trissacarídeo ao invés to tetrassacarídeo que foi incorporado por soaking (NAG)₄.

Notamos que existem cinco ligações de hidrogênio conservadas feitas com o açúcar do sítio -2, o O1 com W55 e W171 (águas 74 e 76 na 1HEWL); N2 com Ala104; O3 com Trp61; O7 com Asn57 são as mesmas nas duas estruturas. A exceção é uma ligação de hidrogênio entre o O3 e Gln100 que é exclusiva da MdL1. Este mesmo resíduo faz outras ligações de hidrogênio com os átomos do ligante O4 e O6 nas posições -2 e -3 respectivamente. Na 1HEWL estas interações são substituídas por ligações de hidrogênio entre os átomos OD2 do Asp101 e OD1 Asn103 (numeração da 1HEWL) e O6 do sacarídeo na posição -3 e -4 respectivamente. Fazendo uma comparação com as outras estruturas de insetos, notamos que na região onde existe a deleção próxima do resíduo 100 é similar a MdL1, mas nem AmL ou BmL tem um resíduo de Gln na posição 100, sendo este ocupado por uma Thr ou His, respectivamente. Ambos aminoácidos não são capazes de formar a mesma interação com o substrato como a Gln da MdL1.

O ligante na MdL1 faz seis ligações de hidrogênio com as moléculas de água (figura 39). Duas destas águas, W55 e W171, juntas com águas W172 e W18 formam uma rede na região do sacarídeo na posição -1, como visto na 1H6M (Vocadlo 2001). Esta rede está presente na HEWL, com exceção para W18 (figura 40). Estas águas estruturais mostram claramente o espaço reservado para o sacarídeo na posição -1 e ajuda manter a cadeia lateral dos resíduos catalíticos na posição.



(a)



(b)

Figura 39: (a) Nomenclatura dos átomos de açúcar; (b) Esquema das interações do ligante com suas respectivas distâncias, as águas estão em ciano. Produzido pelo LIGPLOT (Wallace et al., 1995).

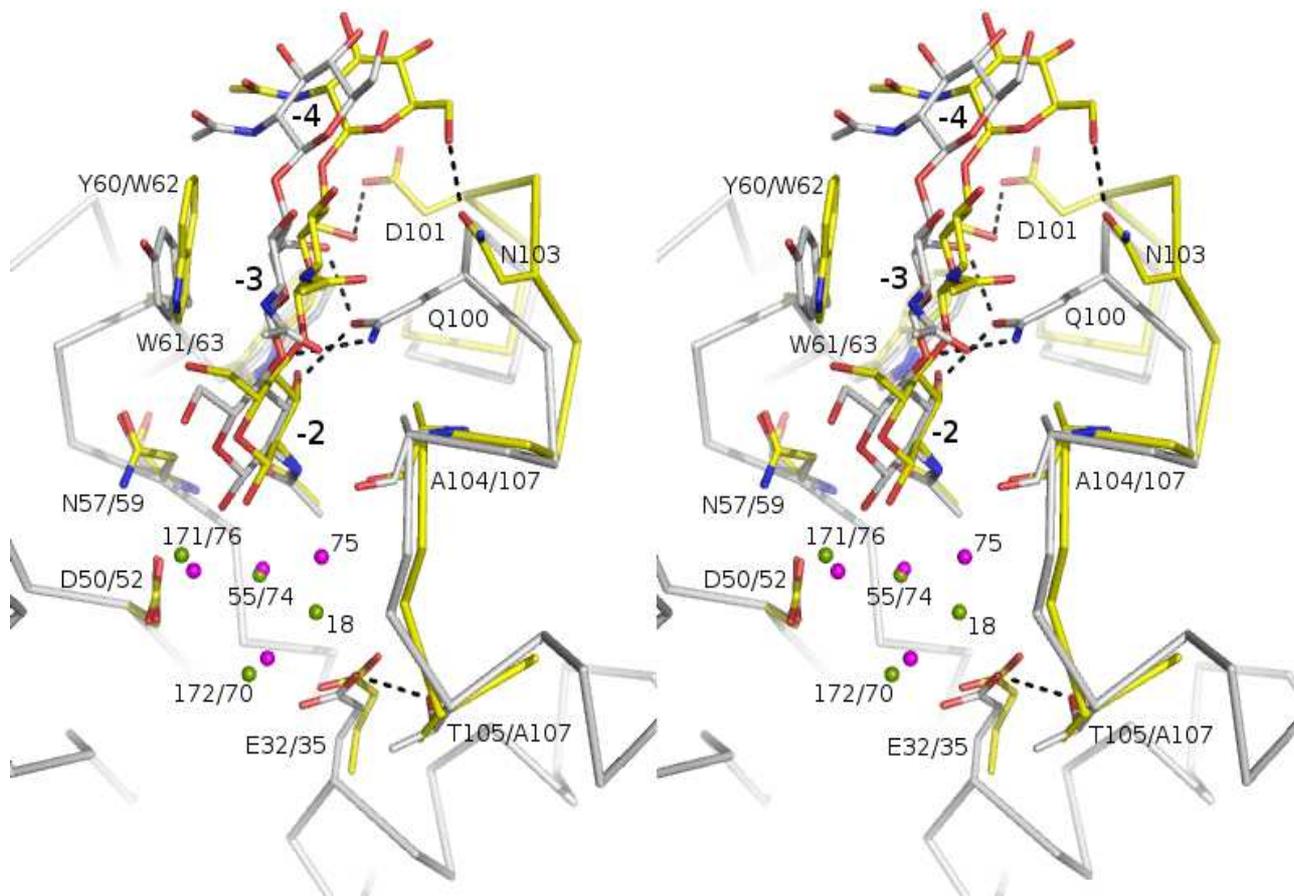


Figura 40: Vista estéreo da superposição da MdL1 (cinza e as águas em verde) e HEWL (amarelo e as águas em magenta) mostrando as diferenças na interação com o ligante. Algumas ligações de hidrogênio estão representadas pelas linhas tracejadas.

A figura 41 mostra a interação do sítio ativo com o ligante. Verificamos o ligante Q4 envolto no mapa de densidade eletrônica [Fo-Fc] gerado pelo protocolo do programa CNS “simulated annealing omit map”. Este método consiste em fazer um mapa “annealing” com uma região “omitida” (neste caso o ligante), com as principais interações observadas no sítio ativo da MdL1 representadas por linhas tracejadas. Notamos que o resíduo Gln 100 é essencial para interação com o ligante.

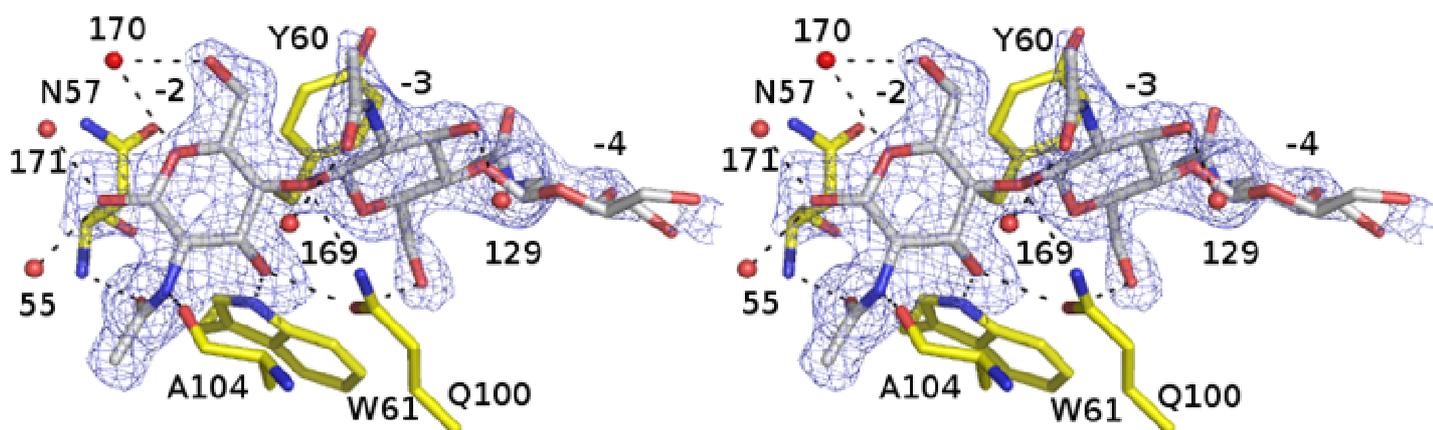


Figura 41: Vista estéreo das interações do ligante Q4 no sítio ativo da enzima MdL1 representadas por linhas tracejadas. O mapa de densidade eletrônica [F_o-F_c] mostrado na figura foi gerado a partir do programa CNS (Brunger *et al.*, 1998) utilizando o procedimento de simulated annealing omit map para o resíduo do ligante, o nível contorno do mapa é 1 σ . A figura foi desenhada com o programa PyMOL (Delano, 2002).

6. Conclusões

6.1. Celulase

- A celulase XF0810 truncada (sem o C-terminal) XF0810c domínio celulase, foi expressa e purificada da fração solúvel do lisado bacteriano (*E.coli*);
- A identidade da XF0810c foi confirmada através de espectrometria de massa;
- A purificação da celulase XF0810c foi realizada em duas etapas de cromatografia: troca iônica e filtração em gel;
- Foram realizados estudos sobre os estados de agregação da XF0810c (DLS e filtração em gel analítica) mostrando que a proteína estava provavelmente na forma de trímero e monômero;
- O modelo da estrutura da celulase XF0810c foi gerado pelo programa MODELLER e é composto por um barril TIM. Neste modelo foi verificada a substituição de 4 resíduos importantes que participam do sítio catalítico da família 5 de celulasas;
- O espectro de dicroísmo circular mostrou que a XF0180c apresenta uma característica de hélices α e fitas β coerente com o enovelamento de barril TIM;
- XF0180c não demonstrou qualquer atividade no teste de atividade enzimática para exocelulase como era esperado a partir da modelagem;
- Através dos ensaios de cristalização da XF0810c, foi detectada a presença de cristais em duas condições de cristalização.

6.2. Lisozima

- A lisozima 1 de *Musca domestica* (MdL1) foi cristalizada em uma condição previamente conhecida e através da técnica de soaking foi obtido o complexo com o inibidor tetra-N-acetilglucosamina;
- O complexo da proteína MdL1 com o inibidor difratou até 1,9 Å;
- O processamento de dados, do cristal difratado do complexo da proteína MdL1 foi realizado com o programa HKL2000 resultando em um R_{sym} de 0,078 e uma completeza de 88,1% ;
- A estrutura da proteína MdL1 com inibidor foi resolvida pelo método de substituição molecular utilizando o programa MOLREP e a estrutura nativa (código de 2FBD) como modelo;
- O refinamento do complexo da MdL1 foi realizado com os programas CNS e REFMAC convergindo para um $R = 0,178$ e $R_{\text{free}} = 0,242$;
- A estrutura da MdL1 com o inibidor foi depositada no PDB onde recebeu o código 2H5Z;
- Foi realizado uma comparação estrutural MdL1 com outras lisozimas de 4 diferentes classes de animais onde foi demonstrado que o core estrutural das estruturas é semelhante e que somente existem diferenças na região do loops/voltas;
- Foi feito um estudo comparativo entre a lisozima de galinha resolvida com o ligante tri-N-acetilglucosamina e a MdL1 com o ligante tetra-N-acetilglucosamina e foi demonstrado que existem cinco ligações de

hidrogênio que são conservadas entre as duas estruturas na interação com o ligante. Todas estas interações acontecem no sub-sítio -2 ;

- Na MdL1 o resíduo Gln100 interage com o ligante de forma única;
- Nas lisozimas de insetos através de análises de alinhamento estrutural, verificamos no loop dos resíduos 98-100 existe uma deleção que o torna menos exposto ao solvente, este fato poderia justificar a resistência destas lisozimas a proteólise;
- Foi verificado que MdL1 a presença de uma ligação de hidrogênio entre o O γ da Thr107 e o resíduo catalítico Glu32. Em cálculos teóricos de pKa, esta ligação seria responsável pelo decréscimo de 0,6 unidades no pKa do resíduo catalítico Glu32, este fato pode justificar em parte a atividade da lisozima digestiva atuar em um pH mais ácido.

7. Referências bibliográficas

- Adler, A.J.; Greenfield, N.J. and Fasman, G.D. (1973). *Methods Enzymol*, **27**: 675-735
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). Basic local alignment search tool. *J Mol Biol.*, **215**: 403-410.
- Ausubel, F.M.; Brent, R.; Kingstone, R.E.; Moore, D.D.; Seidmen, J.G.; Smith, J.A. and Struhl, K. (eds) (1995): Current Protocols in Molecular Biology. John Wiley & Sons, Inc.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1997). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**: 535-542.
- Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C. and Sarma, V.R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **206**:757–761.
- Bohm, G.; Murh, R.; Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5** 191-195
- Brunger, A.T., Adams, P.D., Clore, G. M., Delano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J. –S., Kuszewsk J., Nilges M., Pannu N.S., Read R. J., Rice L.M., Simonson T., Warren G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst. D Biol Crystallogr.*, **54**:905-21.
- Cançado F.C., Valério, A.A., Marana S. R., Barbosa, J.A.R.G. (2007). The crystal structure of a lysozyme *c* from housefly *Musca domestica*, the first structure of a digestive lysozyme. No prelo: *Journal of Structural Biology*.
- Cheetham, J.C., Artymiuk, P.J. and Phillips, D.C. (1992). Refinement of an enzyme complex with inhibitor bound at partial occupancy: Hen egg-white lysozyme and tri-*N*-acetylchitotriose at 1.75 Å resolution. *J. Mol. Biol.* **224**: 613-628.
- Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D* **50**: 760-763.
- Coutinho R.S.R., De Lima J.E.O., Miranda V.S., Carlos E.F. (1996). Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Dilobopterus costalimai*, *Acrogonia terminalis*, *Oncometopia facialis* em citros. *Fitopatol. Bras.*, **21**:517-518.

- Coutinho, P. M. and Henrissat, B. (1999). Carbohydrate active enzymes server. <http://afmb.cnrs-mrs.fr/CAZY/>.
- da Silva, F.R., Vettore, A.L., Kemper, E.L., Leite, A. & Arruda, P. (2001). Fastidian gum: the *Xylella fastidiosa* exopolysaccharide possibly involved in bacterial pathogenicity. *FEMS Microbiology Letters* **203**: 165-171.
- Dow, J.M., Daniels M.J. (2000). *Xylella* genomics and bacterial pathogenicity to plants. *Yeast* **17**: 263-271.
- Davies, G. and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**: 853-859.
- Davies G.J., Wilson K.S., Henrissat B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem.J.* **321**:557-9.
- DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA.
- Deléage, G. & Geourjon, C. (1993). An interactive graphic program for calculating the secondary structures content of proteins from circular dichroism spectrum., *Comp. Appl.Biosc.*, **9**:197-199 .
- Dobson, D.E., Prager, E.M. and Wilson, A.C. (1984) Stomach lysozymes of ruminant. 1 . distribution and catalytic properties. *J. Biol. Chem.*, **18**: 11607.11616.
- Emsley, P. and Cowtan, K. (2004) *Coot*: model-building tools for molecular graphics. *Acta Cryst.*, **D60**: 2126-2132.
- Gilkes N.R., Henrissat B., Kilburn D.G., Miller R.C.Jr, Warren, R.A. (1991). Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiol Rev.*, **55(2)**:303-15.
- Harata, K., Abe, Y. and Muraki, M. (1998). Full-matrix least-squares refinement of lysozymes and analysis of anisotropic thermal motion. *Proteins* **30**: 232-243.
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarity. *Biochem. J.*, **280**: 309-316.
- 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.
- Henrissat, B. and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.*, **316**: 695-696.

- Henrissat B. and Davies G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr.Opin.Struct.Biol.*,**7**: 637-644.
- Hilge M., Gloor S.M., Rypniewski W., Sauer O., Heightman T.D., Zimmermann W., Winterhalter K., Piontek K. (1998). High-resolution native and complex structures of thermostable b-mannanase from *Theromonospora fusca* – substrate specificity in glycosyl hydrolase family 5. *Structure*, **6**: 1433-1444.
- Iversen, O. & Grov, A. (1973). Studies on lysostaphin. Separation and characterization of three enzymes. *Eur. J. Biochem.*, **38**: 293-300.
- Ito, Y., Nakamura, M., Hotani, T. and Imoto, T. (1995). Insect lysozyme from house fly (*Musca domestica*) larvae: possible digestive function base don sequence and enzymatic properties. *J. Biochem. Tokyo* **118**: 546 – 551.
- Jain, D., Nair, D.T., Swaminathan, G.J., Abraham, E.G., Nagaraju, J. and Salunke, D.M. (2001). Structure of the induced antibacterial protein from tasar silkworm, *Antheraea mylitta*. *J. Biol. Chem.* **44**: 41377 – 41382.
- Jollès, P. and Jollès, J. (1984). What's new in lysozyme research? *Molec. Cell. Biochem.*, **63**: 165-189.
- 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.
- Karlsen, S., Eliassen, B.E., Hansen, L.K., Larsen, R.L., Riise, B.W., Smalas, A.O., Hough, E. and Grinde, B. (1995). Refined Crystal-structure of lysozyme from the raibow-trout (*Oncorhynchus-mykiss*). *Acta Cryst. D* **51**: 354-367.
- Kelly, S.M. and Price, N.C. (1997). *Biochimica et Biophysica Acta.* **1338**, 161-185
- Kleywegt, G.J. and Jones, T.A. (1998). Databases in proteincrystallography. *Acta Cryst* **D54**: 1119-1131.
- Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* **28**: 416-436.
- Kraulis per J. (1991). "Molscrip: a program to produce both detailed and schematic plots of protein structures". *Journal of Appl. Cryst.* **24**: 946-950.

- Krissinel, E. and Henrick, K. 2005. *Detection of Protein Assemblies in Crystals*, in: Berthold et al.(Ed.), M.R.CompLife 2005, LNBI 3695. Springer-Verlag Berlin Heidelberg, pp. 163-174.
- Lamzin, V.S. & Wilson, K.S. (1997) Automated refinement for protein crystallography. In *Meth. Enz.* (Carter, C. & Sweet, B. eds.) **277**: 269-305.
- 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. Cryst.*, **26**: 283-291.
- Le Gouill, C. et al., (1994). *Anal. Biochem.* **219**: 164.
- Lemos F. J. A. & Terra W. R. (1991). Digestion of bacteria and the role of midgut lysosome in some insect larvae. *Comp. Biochem.Physiol.B*, **100**: 265-268.
- Li, H., Robertson, A.D. and Jensen, J.H. (2005)."Very Fast Empirical Prediction and Interpretation of Protein pKa Values". *Proteins* **61**: 704-721.
- Luthy R, Bowie JU, Eisenberg D. (1992). Assessment of protein models with three-dimensional profiles. *Nature.* **5;356 (6364)**:83-5.
- Machado, M.A. *Xylella fastidiosa* and the citrus variegated chlorosis (CVC), a new destructive citrus disease in Brazil. Agronomic Institute of Campinas. Citrus Research Center (<http://aeg.lbi.ic.unicamp.br/xf/home/mmachado.html>).
- Marana S.R., Cançado F.C., Valério A.A., Ferreira C., Terra W.R., Barbosa J.A.R.G. (2006). Crystallization data collection and phasing of two digestive lysosomes from *Musca domestica*. *Acta Crystallogr. F*, **62**: 750-752.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol Biol.*, **33**:491-497.
- Matsuura, A., Yao, M., Aizawa, T., Koganesawa, N., Masaki, K., Miyazawa, M., Demura, M., Tanaka, I., Kawano, K. and Nitta, K. (2002). Structural analysis of an insect lysozyme exhibiting catalytic efficiency at low temperatures. *Biochemistry* **41**: 12086 – 12092.
- McCarter, J.D. & Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**: 885-892.
- McKenzie H. A., White F. H., Jr., (1991). Lysozyme and α -lactoalbumin: Structure function and interrelationship. *Adv. Protein chem.*, **41**: 173-315.

- Mellors, A. O-Sialoglycoprotein endopeptidase. In *Handbook of Proteolytic Enzymes* (Barrett,A.J., Rawlings,N.D. & Woessner,J.F. eds), p.1471-1473, Academic Press, London, 1998.
- Mytelka D.S. & Chamberlin M.J. (1996). Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. *Nucl.Acid Res.* **24 (14)**: 2744-2781.
- Murshudov, G.N., Vagin A. A., Dodson E. J., (1997) .Refinement of Macromolecular Structures by the Maximum-Likelihood Method" *Acta Cryst.* **D53**: 240-255.
- Otwinowski, Z. and Minor, W. (1997). Macromolecular Crystallography, Part A,edited by C. W. Carter Jr. and R. M. Sweet, pp. 307-326. New York: Academic Press.*Methods in Enzymology*: 276.
- Phillips, D.C. (1967).The hem egg white lysozyme molecule. *Proc. Nat. Acad. Sci. U.S.A.* **57**: 484-495.
- Prager, E.M. (1996). Adaptative evolution of lysozymes: Changes in amino acid sequence, regulation of expression and gene number. *In: Lysozymes: Model in Biochemistry and Biology* pp. 323-345. Edited by Jollès, P., Birkhäuser Verlag Basell Press.
- Rees, W.A., Yager, T.D., Korte J., Hippel, P.H. (1993). Betaine can eliminate the base pair composition dependence of DNA Melting. *Biochemistry* **32**: 137-144.
- Roman A., Laskowski, Malcolm W., MacArthur, Moss D.S. and M Thornton J.M. (1993). PROCHEK: a program to check the stereochemical quality of protein structures. *J. App. Cryst.* **26** 283-291.
- Rossetti, V., Garnier, M., Bové, J.M., Beretta, M.-J.-G., Teixeira, A.R.R., Quaggiio, J.A. & Negri, J.D. (1990). Occurrence of xylem-restricted bacteria in sweet orange trees affected by chlorotic variegation, a new citrus disease in Brazil. *C. R. Acad. Sci. Série III*, **310**: 345-349.
- Šali, A. and Blundell T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**: 779-815.
- Sakon, J.; Adney, W.S.; Himmel, M.E.; Thomas, S.R. & Karplus, P.A. (1996). Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose.*Biochemistry*, **35**: 10648-10660.

- Sambrook, J.; Fritsch E.F.; Maniatis, T. (1989). *Molecular cloning: a laboratory manual* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LM, Araya JE, Baia GS, Baptista CS, Barros MH, Bonaccorsi ED, Bordin S, Bove JM, Briones MR, Bueno MR, Camargo AA, Camargo LE, Carraro DM, Carrer H, Colauto NB, Colombo C, Costa FF, Costa MC, Costa-Neto CM, Coutinho LL, Cristofani M, Dias-Neto E, Docena C, El-Dorry H, Facincani AP, Ferreira AJ, Ferreira VC, Ferro JA, Fraga JS, Franca SC, Franco MC, Frohme M, Furlan LR, Garnier M, Goldman GH, Goldman MH, Gomes SL, Gruber A, Ho PL, Hoheisel JD, Junqueira ML, Kemper EL, Kitajima JP, Krieger JE, Kuramae EE, Laigret F, Lambais MR, Leite LC, Lemos EG, Lemos MV, Lopes SA, Lopes CR, Machado JA, Machado MA, Madeira AM, Madeira HM, Marino CL, Marques MV, Martins EA, Martins EM, Matsukuma AY, Menck CF, Miracca EC, Miyaki CY, Monteriro-Vitorello CB, Moon DH, Nagai MA, Nascimento AL, Netto LE, Nhani A Jr, Nobrega FG, Nunes LR, Oliveira MA, de Oliveira MC, de Oliveira RC, Palmieri DA, Paris A, Peixoto BR, Pereira GA, Pereira HA Jr, Pesquero JB, Quaggio RB, Roberto PG, Rodrigues V, de M Rosa AJ, de Rosa VE Jr, de Sa RG, Santelli RV, Sawasaki HE, da Silva AC, da Silva AM, da Silva FR, da Silva WA Jr, da Silveira JF, Silvestri ML, Siqueira WJ, de Souza AA, de Souza AP, Terenzi MF, Truffi D, Tsai SM, Tshako MH, Vallada H, Van Sluys MA, Verjovski-Almeida S, Vettore AL, Zago MA, Zatz M, Meidanis J & Setubal JC.(2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, **406**: 151-159.
- Sinnott, M.L. (1990). Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **90**: 1171-1202.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**: 4876-4882.
- Tull, D., Withers, S.G., Gilkes N.R., Kilburn D.G., Warren R.A., Aebersold R. (1991). Glutamic acid 274 is the nucleophile in the active site of a "retaining" exoglucanase from *Cellulomonas fimi*. *J.Biol.Chem.*, **266(24)**: 15621-5.
- Varrot A., Schulein M., Davies G. J. (2000). Insights into ligand-induced conformational change in Cel5A *Bacillus agaradhaerens* revealed by a catalytically active crystal form. *J. Mol.Biol.* **297**, 819-828.

- Vagin, A.A. and Teplyakov, A. (1997). *MOLREP*: an Automated Program for Molecular Replacement. *J. Appl. Cryst.*, **30**:1022-1025.
- Vocadlo, D.J., Davies, G.J., Laine, R. and Withers, S.G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**: 835-838.
- Vollan, V.B., Hough E., Karlsen S. (1999). Structural studies on the binding of 4-methylumbelliferone glycosides of chitin to rainbow trout lysozyme. *Acta Crystallogr D Biol Crystallogr*, **55**: 60-6.
- Von Itzsen M., Colman P., (1996). Design and synthesis of carbohydrate-based inhibitors of protein-carbohydrate interactions. *Curr. Opin. Struct. Biology*, **6**: 703-709.
- Yang, Y. and Hamaguchi, K. (1980). Hydrolysis of 4-Methylumbelliferyl *N*-Acetyl-Chitotrioside catalyzed by hen and turkey lysozymes. *J. Biochem.* **87**: 1003-1014.
- Wallace, A.C., Laskowski, R. A. and Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot. Eng.* **8**: 127-134.
- Wang, Q., Tull, D., Meinke, A., Gilkes, N.R., Warren, R.A.J., Aebersold, R. & Withers, S.G. (1993). Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-beta-1,4-glucanase. *J. Biol. Chem.* **268**: 14096-14102.
- Woody, R.W. Circular dichroism of peptides and proteins, in: K. Nakanishi, N. Berova, R.W. Woody (Eds.), *Circular Dichroism: Principles and Applications*, VCH Publishers, New York, 1994, pp. 473-488.

8. Anexos

8.1. Clorose variegada de citros (CVC) e *Xylella fastidiosa*

O estado de São Paulo (Brasil) e o estado da Flórida (USA) são as principais áreas de citricultura do mundo. São Paulo tem 164 milhões de laranjeiras e uma produção anual de 374 milhões de caixas de laranja, correspondendo a 87% da produção nacional e 30% da produção mundial. Em torno de 70% das laranjas produzidas no estado paulista é destinada a companhias de suco concentrado de laranja (Machado <http://aeg.lbi.ic.unicamp.br/xf/home/mmachado.html>). Apesar do grande potencial produtivo, que pode chegar a mais de 40 toneladas por hectare (Donadio & Moreira, 1997), a cultura de citros é um alvo constante de inúmeras pragas e doenças que podem causar danos irreversíveis e que resultam na queda da produtividade e na qualidade dos frutos. Dentre as principais doenças com grande impacto econômico destacam-se a doença clorose variegada de citros (CVC) que foi constatada pela primeira vez no estado de São Paulo em 1987 (Rosetti et al., 1990).

A CVC é considerada a doença mais devastadora da citricultura (Donadio & Moreira, 1998) afetando todas as variedades comerciais de laranja doce. Os sintomas que a doença apresenta são: redução de tamanho das folhas, frutas pequenas com casca muito rígida, maturação prematura e variação das folhas antigas (figura 42). A transmissão da doença é feita através de vetores (insetos). As possibilidades de controle da doença incluem a remoção de galhos infectados, controle químico do vetor de transmissão através do uso de inseticidas e a total

queima da planta infectada (Machado

<http://aeg.lbi.ic.unicamp.br/xf/home/mmachado.html>).

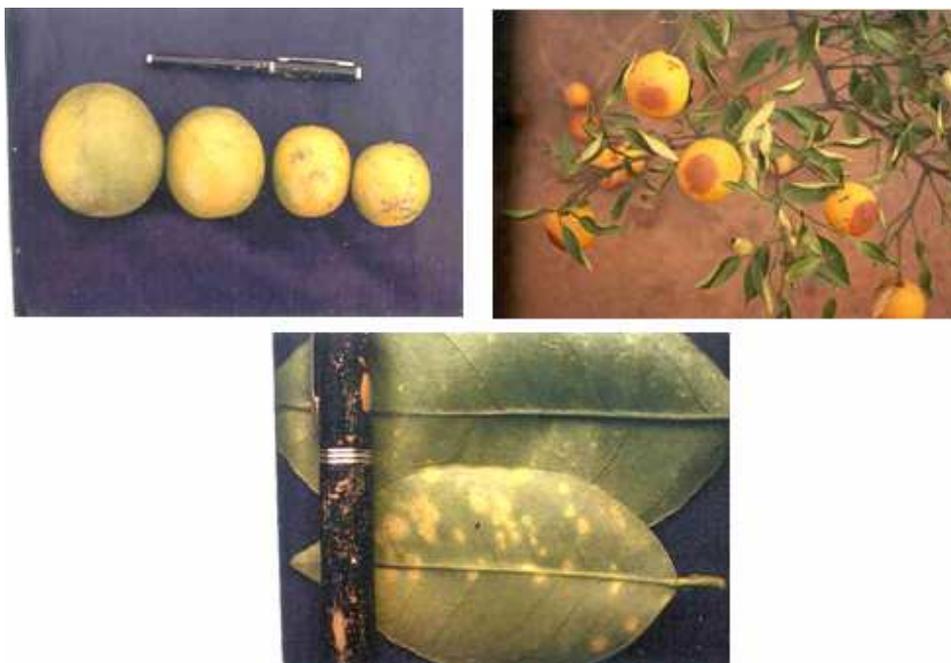


Figura 42: Sintomas da doença clorose variegada de citros (CVC). Machado <http://aeg.lbi.ic.unicamp.br/xf/home/mmachado.html>

O agente causador da CVC é a *Xylella fastidiosa* (Rossetti *et al.*, 1990), uma bactéria Gram-negativa em forma de bastão com parede celular rugosa (Chang *et al.*, 1993). Sua transmissão é feita por vetores, insetos homópteros que se alimentam do líquido contido no xilema (Coutinho *et al.*, 1996). Estes insetos utilizam-se de seu aparelho bucal, onde estão as bactérias, para perfurar os vasos da planta e sugar a água e nutrientes do xilema. Portanto, as bactérias têm acesso direto ao xilema durante a alimentação dos insetos. *X. fastidiosa* é também o patógeno causador de outras doenças em plantas economicamente importantes

como a videira, alfafa, pessegueiro, amendoeira e cafezeiro (Purcell & Hopkins, 1996).

Não existe um consenso acerca do mecanismo de patogenicidade da doença, porém há evidências indicando que a agregação das bactérias no xilema leva a uma oclusão vascular, gerando um estresse hídrico (figura 43). Os agregados de bactérias aparecem fixados nas paredes dos vasos por fitas extracelulares produzidas pela bactéria. Estas fitas seriam formadas por fibras de exopolissacarídeos produzidos pelas bactérias. Além desta função de fixação, especula-se que os exopolissacarídeos constituem uma rede que auxilia na concentração de enzimas e nutrientes (da Silva *et al.*, 2001).

Após a infecção de um vaso de xilema, a *X. fastidiosa* deve ser capaz de colonizar outros vasos. O mecanismo de espalhamento da infecção proposto envolve a quebra parcial das membranas pontuadas. Estas membranas contêm pontuações, regiões que não possuem a parede celular secundária, somente a primária, portanto são menos espessas. Especula-se que as celulasas são utilizadas para romper as paredes da membrana primária das pontuações, permitindo a migração das bactérias de um vaso a outro. Os sintomas que aparecem nas plantas dependem da extensão da colonização. (Dow & Daniels, 2000).

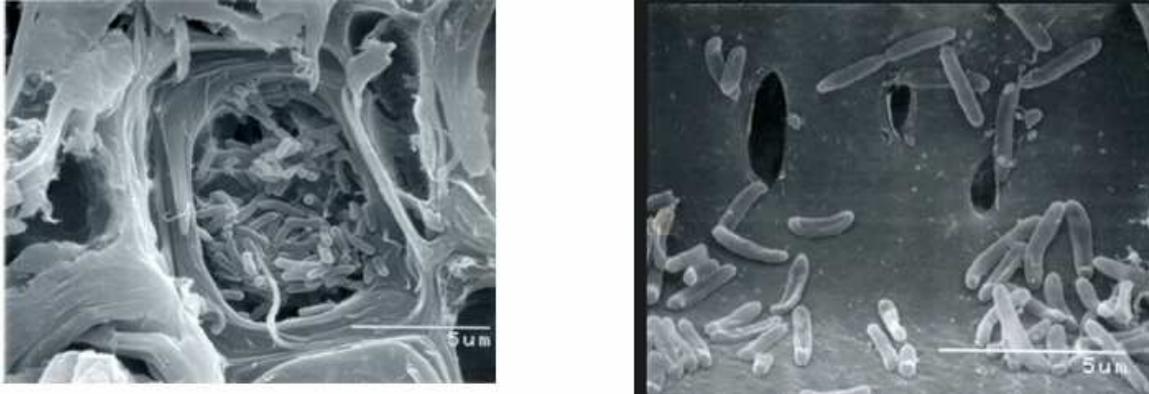


Figura 43: Microscopia eletrônica de *Xylella fastidiosa* em vaso do xilema. Machado <http://aeg.lbi.ic.unicamp.br/xf/home/mmachado.html>

No ano de 2000 o genoma da *X. fastidiosa* foi seqüenciado (Simpson *et al.*, 2000), sendo este o primeiro genoma de um fitopatógeno, representando destaque na ciência brasileira. Para este estudo foram selecionados genes da *X. fastidiosa* que codificam hidrolases (celulases, glicosidases e peptidases) (tabela 10), esta seleção incluiu a predição de regiões transmembrana a partir da seqüência primária da proteína usando o servidor DAS (<http://www.sbc.su.se/~miklos/DAS/maindas.html>, Cserzo *et al.*, 1997), visando a eliminação de proteínas de membrana. As hidrolases estão presentes no genoma da *X. fastidiosa* e atuam na degradação de polímeros de açúcares e aminoácidos da planta hospedeira, afetando, desta forma o desenvolvimento das plantas infectadas. Além disso, as celulases são importantes para a *X. fastidiosa*, pois estão associadas a obtenção de glicose e colonização do sistema vascular da planta.

Tabela 10 - Lista dos genes selecionados de *X. fastidiosa* com características das proteínas correspondentes: # aa – número de aminoácidos, MM – massa molecular, pI – ponto isoelétrico calculado.

Hidrolase	Família	Código	# aa	MM (Da)	pI
Glicosilases	Glicosil-hidrolase 3	XF 2355	335	35324.3	5.3
	Celulase 5	XF 0810	565	61620.7	6.2
		XF 0818	592	59967.4	5.7
		XF 2708	356	39273.5	5.4
Peptidases	Metalopeptidase M22	XF 0435	348	36798.4	5.8
	Metalopeptidase M37	XF 0167	417	46554.2	9.6
		XF 0805	319	34505.0	9.5
		XF 0862	275	29633.6	6.9
	XF 0987	285	30520.7	9.9	

8.2 Artigos:

1. **Artigo publicado:** Marana, S.R., Cançado F.C., **Valério A.A.**, Ferreira C., Terra W.R. and Barbosa J.A.R.G. Crystallization, data collection and phasing of two digestive lysozymes from *Musca domestica*. (2006) Acta Crystallographica Section F. Structural Biology and Crystallization Communications, 62, 750-752.
2. **Artigo aceito:** Cançado F.C., **Valério A.A.**, Ferreira C., Terra W.R., Marana S.R., Barbosa J.A.R.G. The crystal structure of a digestive lysozyme from housefly *Musca domestica*, the first digestive lysozyme from insects. Journal of Structural Biology.
3. **Artigo em preparação:** **Valério A.A.**, Barbosa J.A.R.G. XF0810 is not an exocellulase.

Crystallization, data collection and phasing of two digestive lysozymes from *Musca domestica*

S. R. Marana,^{a,‡} F. C. Cançado,^{a,‡}
A. A. Valério,^{b,c} C. Ferreira,^a
W. R. Terra^a and
J. A. R. G. Barbosa^{b,c,d*}

^aDepartamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, ^bCentro de Biologia Molecular e Estrutural (CeBiME), Laboratório Nacional de Luz Síncrotron (LNLS), CP 6192, Campinas, SP 13084-971, Brazil, ^cDepartamento de Bioquímica, Universidade Estadual de Campinas, Campinas, Brazil, and ^dDepartamento de Genética e Evolução, Universidade Estadual de Campinas, Campinas, Brazil

‡ These authors contributed equally to this work.

Correspondence e-mail: joao@lnls.br

Received 4 April 2006
Accepted 23 June 2006



© 2006 International Union of Crystallography
All rights reserved

Lysozymes are mostly known for their defensive role against bacteria, but in several animals lysozymes have a digestive function. Here, the initial crystallographic characterization of two digestive lysozymes from *Musca domestica* are presented. The proteins were crystallized using the sitting-drop vapour-diffusion method in the presence of ammonium sulfate or PEG/2-propanol as the precipitant. X-ray diffraction data were collected to a maximum resolution of 1.9 Å using synchrotron radiation. The lysozyme 1 and 2 crystals belong to the monoclinic space group $P2_1$ (unit-cell parameters $a = 36.52$, $b = 79.44$, $c = 45.20$ Å, $\beta = 102.97^\circ$) and the orthorhombic space group $P2_12_12$ (unit-cell parameters $a = 73.90$, $b = 96.40$, $c = 33.27$ Å), respectively. The crystal structures were solved by molecular replacement and structure refinement is in progress.

1. Introduction

Lysozymes (EC 3.2.1.17) catalyze the hydrolysis of 1,4- β -linkages between *N*-acetylmuramic acid and *N*-acetylglucosamine residues in the peptidoglycan of the bacteria cell wall. Based on sequence and structure similarities, these enzymes are classified into families 22, 23, 24, 25 and 73 of the glycoside hydrolases (Coutinho & Henrissat, 1999).

Family 22 contains 219 type C lysozymes from vertebrates (amphibians, fishes, birds and mammals), invertebrates (crustaceans, arachnids and insects) and fungi (Coutinho & Henrissat, 1999). Hen egg-white lysozyme (HEWL) is probably the best characterized member of this family.

Lysozymes from family 22 are usually part of the defence system against bacteria (Jollès & Jollès, 1984). However, some of these lysozymes are also involved in the digestion of bacteria in vertebrates that have a fermentation chamber in their foregut (for example, ruminants such as *Bos taurus* and primates such as *Presbytis entellus*) and in insects that feed on decomposing organic material (for example, the housefly *Musca domestica*; Lemos & Terra, 1991; Prager, 1996). These lysozymes present several adaptations to digestive function: a high expression level in the gut, a resistance to proteinase hydrolysis and bacteriolytic activity with an acidic pH optimum (Dobson *et al.*, 1984). Sequence alignments have been used to propose the molecular basis of these adaptations (Prager, 1996; Regal *et al.*, 1998), but these hypotheses still remain to be tested.

The tertiary structure of digestive lysozymes is not known, although 15 different lysozymes from family 22 have already been crystallized, corresponding to 452 structures in the PDB. Despite the fact that all lysozymes from family 22 share the same fold, slight differences in their structures may be correlated to adaptations of lysozyme to the digestive function.

Two digestive lysozymes are found in the *M. domestica* midgut (lysozyme 1, AAQ20048; lysozyme 2, AAQ20047). Lysozyme 1 (122 amino acids, 13 816 Da) has previously been submitted to biochemical characterization, showing that the lytic activity of this lysozyme has a pH optimum of 4.5 and its affinity for bacterial cell walls decreases as the ionic strength of the medium becomes higher (Lemos *et al.*, 1993; Ito *et al.*, 1995). Lysozyme 2 (122 amino acids; 13 890 Da) is very similar to lysozyme 1 (81% similarity; 70% identity), but still remains to be characterized.

In this paper, the crystallization of lysozymes 1 and 2 is reported. The tertiary structure of these digestive lysozymes may contribute to the comprehension of the molecular basis of the adaptation of these lysozymes to digestive function.

2. Methods

2.1. Crystallization

Both lysozymes were expressed as recombinant protein in *Pichia pastoris*. An initial step of purification was ammonium sulfate precipitation followed by ion-exchange chromatography (Cançado *et al.*, 2006). Samples of purified lysozyme 1 (9.6 mg ml⁻¹) and lysozyme 2 (9.8 mg ml⁻¹) from *M. domestica* were submitted to crystallization trials using the sitting-drop vapour-diffusion method. Both samples were in 10 mM 2-morpholinoethanesulfonic acid (MES) pH 5.5 buffer. The experiment was set up by mixing equal volumes (1 µl) of protein and reservoir solution in Cryschem plates. Initial screening was performed using the commercial kits Crystal Screen and Crystal Screen II from Hampton Research according to the manufacturer's instructions. Conditions showing crystalline structures were refined by varying the pH of the buffers, the concentrations of precipitants and the drop volumes to yield suitable crystals for X-ray data collection. In the case of lysozyme 2, the condition was further refined with additive solutions in two ways: (i) by using the Additive Screen from Hampton Research according to the manufacturer's instructions and (ii) by mixing the solutions of the Crystal Screen from Hampton Research with the crystallization drop (Birtley & Curry, 2005). In the latter case, 0.2 µl of the Crystal Screen was mixed with 1.5 µl of protein solution and 1.3 µl of the reservoir solution, leading to a total drop volume of 3 µl. All experiments were carried out at 291 K.

2.2. Data collection, processing and phasing

Crystallographic data were collected at the protein crystallography beamline D03B-MX1 at Laboratório Nacional de Luz Síncrotron (LNLS), Campinas, Brazil. This beamline is equipped with a MAR CCD detector with a circular X-ray-sensitive surface of 165 mm in diameter combined with a MAR DTB goniostat. Crystals were

scooped straight from the drop and cooled directly in a nitrogen-gas stream to 100 K in order to minimize radiation damage to the crystals. Data collection was carried out using the oscillation method with a 1.0° oscillation per frame and radiation of wavelength around 1.43 Å. D03B-MX1 is a monochromatic beamline and operates at this wavelength as a compromise between flux and absorption. The solution in which the crystals were grown provided partial protection against ice formation for lysozyme 1. A solution containing 20% glycerol, 9% polyethyleneglycol (PEG) 4000, 12% 2-propanol and 0.05 M sodium citrate pH 4.2 was used as a cryoprotectant in the case of lysozyme 2. The crystals were soaked for 5–10 s in this solution prior to flash-cooling in nitrogen gas. The data set was processed using *HKL-2000* (Otwinowski & Minor, 1997) and the *CCP4* package (Collaborative Computational Project, Number 4, 1994). Molecular replacement was performed with the *MOLREP* program (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 1997) using one monomer of the lysozyme structure from hen egg white deposited in the PDB with code 1hew (Cheetham *et al.*, 1992) depleted of its waters and ligand as a search model. The identity of the 1hew sequence with lysozymes 1 and 2 is 38.5 and 36.2%, while the similarity is 55.4 and 53.8%, respectively. In the case of lysozyme 2, molecular replacement was performed with the partially refined lysozyme 1 as a search model.

3. Results and discussion

Initial crystallization trials with lysozyme 1 and lysozyme 2 resulted in microcrystals in some conditions. In order to improve the quality of these crystals, refinement of these conditions was performed. Crystals grew in one to two weeks, leading to crystals of between 50 and 1000 µm in the longest dimension. Crystals usually formed clusters of plates, especially for lysozyme 2. To minimize this clustering in the crystallization of lysozyme 2, additives were tested.

The best condition found for lysozyme 1 was 1.4 M ammonium sulfate, 0.1 M 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid sodium salt (sodium HEPES) pH 7.5 and 1% PEG 400 (Fig. 1). The best condition for crystals of lysozyme 2 was the drop mixture described in §2, where the reservoir solution consisted of 28%

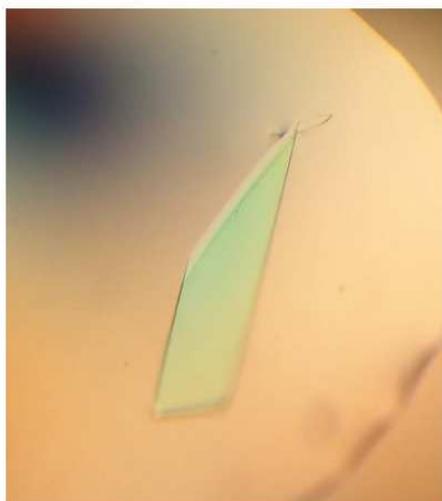


Figure 1
(a) Crystals of lysozyme 1 measuring approximately 1000 µm in the longest dimension.

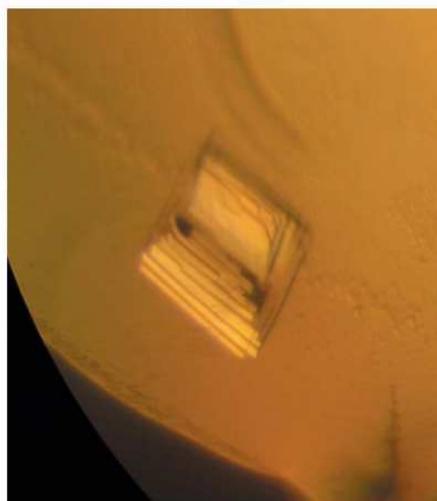


Figure 2
(a) Crystals of lysozyme 2 measuring approximately 200 µm in the longest dimension.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

	Lysozyme 1	Lysozyme 2
Space group	$P2_1$	$P2_12_12$
Unit-cell parameters (Å, °)	$a = 36.52, b = 79.44,$ $c = 45.20,$ $\alpha = \gamma = 90.00,$ $\beta = 102.97$	$a = 73.90, b = 96.40,$ $c = 33.27,$ $\alpha = \beta = \gamma = 90.00$
Mosaicity (°)	0.7	0.8
Temperature (K)	100	100
Wavelength (Å)	1.431	1.427
Oscillation (°)	1.0	1.0
Crystal-to-detector distance (mm)	80.0	80.0
No. of frames	171	127
Resolution limits (Å)	40.00–1.9 (1.97–1.90)	30.00–1.9 (1.97–1.90)
$I/\sigma(I)$ after merging	23.8 (5.6)	14.4 (2.1)
Completeness (%)	100.0 (99.8)	90.7 (74.1)
Multiplicity	3.5 (3.4)	3.9 (1.8)
R_{sym}	0.052 (0.222)	0.089 (0.304)
No. of reflections	70004	69179
No. of unique reflections	19857 (1953)	17698 (1423)
B factor (Wilson plot) (Å ²)	20.0	23.3

2-propanol, 0.115 M sodium citrate pH 4.2 and 21% PEG 4000 and the additive solution was No. 26 of Hampton Research Crystal Screen (0.2 M ammonium acetate, 0.1 M sodium citrate pH 5.6 and 30% 2-methyl-2,4-pentanediol) (Fig. 2). The cluster shown in Fig. 2 is not a single crystal. In order to avoid multiple diffraction patterns, we needed to break the cluster by touching its vertex, thus separating the plates, which could be used as single crystals.

The crystals of both lysozymes 1 and 2 yielded diffraction data that were processed to a maximum resolution of 1.9 Å. The rotation method was used for data collection and no prediction of the best strategy (starting angle) was made for the data collection. Space group $P2_1$ was assigned for lysozyme 1, while the crystal of lysozyme 2 showed the symmetry and systematic absences of the orthorhombic space group $P2_12_12$. Table 1 summarizes the data-collection statistics.

The Matthews coefficient (Matthews, 1968) for two molecules of lysozyme 1 in the asymmetric unit was calculated to be 2.3 Å³ Da⁻¹, giving a solvent content of 45.7%. For two molecules of lysozyme 2 in the asymmetric unit, the coefficient was 2.1 Å³ Da⁻¹, with a solvent content of 40.4%. Clear peaks in the rotation and translation func-

tions were observed for the molecular-replacement solutions of lysozyme 1. The initial electron-density maps of the proteins clearly showed the expected differences in the side chains between 1hew and lysozyme 1. After a few cycles of refinement, the model of lysozyme 1 was used as a search model for lysozyme 2. Once again, clear peaks in the rotation and translation functions were found and the electron-density maps showed the differences between lysozyme 1 and 2. The crystallographic models of these proteins are being built and refined.

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grants 04/02017-9, 04/02225-0, 02/10634-2, 01/08198-7, 01/07531-4 and 00/10266-8; SMolBNet), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Associação Brasileira de Tecnologia de Luz Síncrotron (ABTLuS). We also thankfully acknowledge the support from Dr Rogério Meneghini and the technical assistance of Andréia N. Meza.

References

- Birtley, J. R. & Curry, S. (2005). *Acta Cryst.* **D61**, 646–650.
 Cançado, F. C., Effio, P. C., Terra, W. R. & Marana, S. R. (2006). Submitted.
 Cheetham, J. C., Artymiuk, P. J. & Phillips, D. C. (1992). *J. Mol. Biol.* **224**, 613–628.
 Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
 Coutinho, P. M. & Henrissat, B. (1999). *Carbohydrate-Active Enzymes Server*. <http://afmb.cnrs-mrs.fr/CAZY/>.
 Dobson, D. E., Prager, E. M. & Wilson, A. C. (1984). *J. Biol. Chem.* **18**, 11607–11616.
 Ito, Y., Nakamura, M., Hotani, T. & Imoto, T. (1995). *J. Biochem. (Tokyo)*, **118**, 546–551.
 Jollès, P. & Jollès, J. (1984). *Mol. Cell. Biochem.* **63**, 165–189.
 Lemos, F. J. A., Ribeiro, A. F. & Terra, W. R. (1993). *Insect Biochem. Mol. Biol.* **23**, 533–541.
 Lemos, F. J. A. & Terra, W. R. (1991). *Comput. Biochem. Physiol.* **B**, **100**, 265–268.
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
 Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
 Prager, E. M. (1996). *Lysozymes: Model Enzymes in Biochemistry and Biology*, edited by P. Jollès, pp. 323–345. Basel: Birkhäuser Verlag.
 Regel, R., Matioli, S. R. & Terra, W. R. (1998). *Insect Biochem. Mol. Biol.* **28**, 309–319.
 Vagin, A. A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Title: The crystal structure of a lysozyme *c* from housefly *Musca domestica*, the first structure of a digestive lysozyme

Authors: Fabiane Chaves Cançado^{1,‡}, Amanda Abdalla Valério^{2,3,‡}, Clélia Ferreira¹, Walter Ribeiro Terra¹, Sandro Roberto Marana^{1,*}, João Alexandre Ribeiro Gonçalves Barbosa^{2,3,4,*}

Addresses: ¹ Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, CP 26077, São Paulo, 05513-970, São Paulo, Brasil. ² Center for Structural Molecular Biology (CeBiMe), Brazilian Synchrotron Light Laboratory (LNLS), CP 6192, Campinas, SP, 13084-971, Brazil. ³ Departamento de Bioquímica, Universidade Estadual de Campinas, Brasil. ⁴ Departamento de Genética e Evolução, Universidade Estadual de Campinas, Brasil.

[‡] These authors contributed equally to this work.

* Corresponding authors: joao@lnls.br and srmarana@iq.usp.br

phone 55-19-3512-1107 fax 55-19-3512-1004

Running title: First atomic structure of a digestive lysozyme.

Total number of:

-manuscript pages: 31

-supplementary material pages: 0

-tables: 3

-figures: 2

Keywords: Digestive lysozyme 1; *Musca domestica*; X-ray crystal structure; tetra-*N*-acetylchitotetraose

Abstract

Lysozymes (EC 3.2.1.17) catalyze the hydrolysis of 1,4- β -linkages between *N*-acetylmuramic acid and *N*-acetylglucosamine residues in peptidoglycan of the bacterial wall. Lysozymes from family 22 are usually part of the defense system against bacteria. However in ruminant artiodactyls and saprophagous insects, lysozymes are involved in the digestion of bacteria. Here, we report the first atomic structure of a digestive lysozyme in its native and complexed forms, the crystallographic structure of lysozyme 1 from *Musca domestica* larvae midgut (MdL1). The structural core is similar, but a careful analysis of a structural alignment generated with lysozymes c from four different classes of animals reveals that significant differences occur in coil regions. The loop from MdL1 defined by residues 98-100 has one deletion previous to residue Gln100, which leads to a less exposed conformation and might justify the resistance to proteolysis observed for MdL1 and other digestive lysozymes from insects. Furthermore, Gln100 is directly involved in a few hydrogen bonds to the ligand in a yet unobserved substrate binding mode. Another unique feature of MdL1 is a hydrogen bond between Thr107 O γ and Glu32 carboxylate group (catalytic acid/base). In other lysozymes Thr107 is replaced by alanine or glycine. Theoretical calculations suggest that the presence of the hydrogen bond between Thr107 and Glu32 is responsible for a decrease of 0.6 units in the pK_a of the catalytic acid/base Glu32. Thus, this decrease might explain the

acidic pH optimum presented by MdL1 and other insect digestive lysozymes.

Introduction

Lysozymes (EC 3.2.1.17) catalyze the hydrolysis of 1,4- β -linkages between N-acetylmuramic acid and N-acetylglucosamine (NAG) residues in the peptidoglycan of the bacterial cell wall. Based on sequence and structure similarity, these enzymes are classified in the glycoside hydrolase families 22, 23, 24, 25 and 73 (Coutinho and Henrissat, 1999). Lysozymes *c* are found in the family 22, for which the hen egg white lysozyme (HEWL) is the best characterized member.

Family 22 lysozymes are usually part of a defense system against bacteria (Jollès and Jollès, 1984), although in some animals these enzymes are also involved in digestion. For instance, ruminant artiodactyls use lysozymes *c* to cleave the walls of bacteria growing in the fermentation chamber of their foregut (Dobson et al., 1984). Besides those vertebrates, insects that feed on bacteria growing in decomposing material (for instance housefly *Musca domestica* and fruitfly *Drosophila melanogaster*) also employ lysozymes *c* to disrupt the bacteria present in their diet (Lemos and Terra, 1991; Hultmark, 1996; Prager, 1996). Digestive lysozymes *c* share some specific properties, such as high expression in the gut, resistance to proteinase hydrolysis and bacteriolytic activity exhibiting an acid pH optimum (Dobson et al., 1984). Sequence alignments have been used to propose the molecular basis of these adaptations (Prager, 1996) but these hypotheses still remain to be tested.

HEWL was the first enzyme structure to be solved by X-ray methods (Blake et al., 1965). The crystallographic structure of the cocrystallised complex with tri-N-acetylchitotriose (NAG)₃ led immediately to proposals for the catalytic mechanism with the indication of the two catalytic residues and of the six substrate binding subsites labeled -4 to +2 (originally named A to F), one per monosaccharide (Philips, 1967; Blake et al., 1967; Davies et al., 1997). Besides that, it was proposed that the Glu35 acted as a general acid protonating the glycosidic oxygen while the other residue, Asp52, acted as an electrostatic stabilizing group to promote and stabilize the carbonium ion intermediate (Phillips, 1967). Recently, this mechanism has been revisited and catalysis via a covalent intermediate formed with the catalytic aspartate residue has been proposed (Vocadlo et al., 2001).

The tertiary structures of two lysozymes *c* from the insect immune system are known. For instance, the core structure of the antibacterial protein from Tasar silkworm *Antheraea mylitta* (AmL) (Jain et al., 2001) and the lysozyme from the silkworm *Bombix mori* (BmL) (Ito et al., 1995) are similar to that of HEWL, although differences in the amino acid residues forming the active site and in some external loops had been observed. Nevertheless, the tertiary structure of digestive lysozymes from insects has not been previously reported.

Lysozyme 1 (MdL1; GenBank accession no. **AY344589**) is found in the *M. domestica* larvae midgut and has 122 amino acids (13816 Da). Its sequence shares 55.4% similarity and 38.5% identity with the HEWL. It exhibits a pH optimum of 4.5 upon *Micrococcus lysodeikticus* at an ionic strength of 0.1, is resistant to the cathepsin D (an acidic proteinase) and its affinity for bacterial

walls is affected by the ionic strength of the reaction medium (Lemos et al., 1993; Ito et al., 1995).

The tertiary structure of *M. domestica* lysozyme 1 (MdL1) reported here is the first obtained for a digestive lysozyme. The comparison of this structure with digestive and non-digestive lysozymes was used in the attempt to elucidate the molecular basis of the properties of insects' digestive lysozymes.

Results and discussion

Overall structure

Both crystallographic structures of the MdL1, with and without (NAG)₃, present a dimer in the asymmetric unit. These structures have been refined, validated and deposited in the PDB, where they received the codes 2H5Z and 2FBD, respectively. The final $R_{\text{cryst}}/R_{\text{free}}$ were 0.175/0.242 and 0.153/0.198, respectively, and are presented on table 1 together with other refinement statistic data. The dimer was scrutinized about its possible existence in solution and biological significance in two different ways. The first approach was *in silico* via the program PISA (Krissinel and Henrick, 2005) at http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html. The calculated dimer interface area is 490.9 Å² (less than 8% of the monomer surface), presenting seven hydrogen bonds and a complexation significance score of zero, indicating that the interface of the dimer does not occur naturally and is probably a result of crystal packing. We also measured the dynamic light scattering of the sample which had a monodisperse pattern with particles of 13 kDa of estimated mass. Both methods used are in accordance and imply that the dimer present in the crystal is not the biological form of the enzyme.

The monomers respect the lysozyme *c* fold, consisting of a central core of three long and one short α -helices (Thr4-Leu15, Glu21-Ser34, Ile86-Gly101 and Thr107-Ser112), an antiparallel β -sheet with two β -strands (Asp50-Tyr51 and Ile56-Asn57) and four 3_{10} -helices (fig. 1). The eight cysteines form four conserved disulfide bonds. The two stranded β -sheet differs from that of other lysozymes where it usually has three β -strands. A careful inspection of the region corresponding to the absent first β -strand, in the region of residue Val40, shows an extended structure that barely misses the Kabsch and Sander parameters (Kabsch and Sander, 1983) used to define a β -strand.

Three sulfate ions, two polyethyleneglycol (PEG) molecules and 311 waters are the final components of the 2FBD structure. One of the sulfates is in the interface of the two molecules in the asymmetric unit and makes no contacts with the protein, only with solvent molecules. The other two sulfates are making hydrogen bonds to Thr43B and Thr4B and solvent molecules. None of the sulfates are close to the active site. Each PEG molecule was modeled as a diethyleneglycol and was located in the active site of each lysozyme monomer in the asymmetric unit. In 2H5Z, there are 173 waters and the soaked ligand (NAG)₄ modeled as a trisaccharide in subsites -4 to -2. The interactions of these molecules will be discussed in a following section.

Comparison with other lysozymes c

Although lysozymes *c* are one of the most abundant molecules in the PDB, there are only four classes of animals represented: mammals, birds, fishes and insects. Furthermore, all of these are representatives of lysozymes associated with the immune function, so far there are no digestive lysozymes deposited in the PDB. We

have selected one example of each of these classes, 1JSF from *Homo sapiens* placenta (Harata et al., 1998), 1LMN from *Oncorhynchus mykiss* kidney (Karlsen et al., 1995), 1HEW from *Gallus gallus* egg white (Cheetham et al., 1992), 1GD6 from *Bombyx mori* (Matsuura et al., 2002) and 1IIZ from *Antheraea mylitta* (Jain et al., 2001), to carry out a comparison to the structure presented in this study in order to find possible reasons for the characteristics found in the digestive lysozymes. The two structures available from insects (1GD6 and 1IIZ) were included since they are closer related to the one presented here and thus may provide more information.

The RMSD between the superposed structures (tab. 2) shows that MdL1 is just slightly closer to the other insect structures (mean RMSD of 1.09 Å) than to the lysozymes representing birds, mammals and fishes (mean RMSD of 1.19 Å). This contrasts with the greater similarity between the structures of the other insects (RMSD of 0.66 Å) and indicates that divergence towards a digestive function required more adaptations of the structure. Another interesting feature is that the mean RMSD between birds, mammals and fishes (0.56 Å) is smaller than that between silkworms mentioned above.

A structural alignment shows nine regions of insertions or deletions, with five of them at the C-terminal region of the molecule containing residue 110 (fig. 1). The deletion of three amino acids in the region of residue Ser20, where the residues adopt a short 3_{10} -helix configuration, seems to be a trait of the insect lysozymes irrespective of its defensive or digestive function. The structures in a second region containing gaps in the alignment, between Val40 and the catalytic Asp50, show different ways to accommodate and maintain residues Asn44, Asn/Asp46 and Ser48, which together with Asn57, are responsible for the

placement of the Asp50 side chain (fig. 1 and 2a). The biggest unaligned region, around residue 70, includes a unique feature of MdL1: the increment of the conserved hydrophobic cluster containing Tyr60 and Trp61 by residues Tyr59, Phe69 and Tyr71 (fig. 2A). The deletion next to residue Gln100 is another characteristic of insect lysozymes and renders the helix 5 with three extra residues when compared with other lysozymes. Another feature of the insect structures is the addition of a Trp/Phe102, usually an isoleucine or methionine, to the aromatic cluster formed by Trp25, Phe54, Trp108 and Trp105. The differences in the main chain in the region of residues 40-50, around 70 and 100, have been noticed before even in the closely related structures, such as turkey and hen egg lysozymes, which are only seven residues different (Harata, 1993).

Ligand binding

The electron density found for the ligand showed a triose corresponding to (NAG)₃ in sites -4 (A) to -2 (C) of asymmetric unit's molecule B in a similar configuration as that of the 1HEW structure (Cheetham et al., 1992) (fig. 2B). Sites -2 and -3 present good densities, while site -4 is poor either due to low occupancy or disorder. The low occupancy of site -4 has been associated with the hydrolysis of the ligand to shorter tri-, di- and monosaccharides (Cheetham et al., 1992). Disorder may also contribute to the weak density in site -4 since the sugar unit in this site has no interactions with the protein or water molecules. The same reasoning of low occupancy due to hydrolysis and disorder can be used to explain the fact that we could only find a trisaccharide instead of the tetrasaccharide that was soaked, (NAG)₄.

Molecule A of the asymmetric unit is not occupied by the sugar, even though its active site is exposed to the same region of solvent as molecule B's occupied active site. One possible reason is that if the trisaccharide was present in molecule A in the same configuration as in molecule B, the saccharide in site -4 of molecule A would be in steric clash with Gln100 of molecule B. The only other difference between the ligand sites of molecules A and B is a different rotamer for the side chain of Gln100. The unbound molecule A rotamer is similar to the rotamers found in both molecules of the unsoaked structure.

Contrasting to the absence of interactions with the saccharide in site -4, there are six hydrogen bonds between the ligand and sites -3 and -2 of MdL1 (fig. 2B-C). Comparing these interactions with those that occur in HEWL, it becomes clear that the conservation is higher in site -2 than in -3, correlating with the proximity of the reaction site. In fact, the hydrogen bonds between the saccharide atoms in site -2 and the protein residues or waters (O1 with W55 and W171; N2 with Ala104; O3 with Trp61; O7 with Asn57) are the same in MdL1 and HEWL (where the waters are numbered 74 and 76, respectively) (fig. 2D). The only exception is a hydrogen bond between O3 and Gln100, which is unique to MdL1. Gln100 shares two other hydrogen bonds with the ligands' O4 and O6 atoms of the saccharides in positions -2 and -3, respectively. In HEWL, these interactions are replaced by hydrogen bonds between residues Asp101 and Asn103 (HEWL numbering) and the O6 of saccharides in positions -3 and -4, respectively. In other insect structures, this region is similar to MdL1, where there is a deletion and the main chain is accommodated. But neither AmL nor BmL has a Gln at position 100, this is occupied by Thr or His, respectively. Both amino acids are unable to form the same

interactions with a substrate as the Gln from MdL1. Thus it is expected that for the lysozymes of these insects, other residues in the neighbourhood, possibly Arg99, can compensate for the lack of the Gln100 interactions.

The ligand makes six hydrogen bonds to water molecules (fig. 2C). Two of these waters, W55 and W171, together with waters W172 and W18 form a network in the region of saccharide in position -1, as seen in the 1H6M structure (Vocadlo et al., 2001). This network is present in HEWL, except for W18 (fig. 2D). These structural waters clearly show the space reserved for the saccharide in position -1 and help maintain the side chains of the catalytic residues in position. In the native structure of MdL1 there is a PEG molecule in the active site, partially occupying the position -1 at O6 and -2 at the N-acetyl. This molecule changes the disposition of the waters and prevents a comparison of the water positioning prior and after (NAG)₃ binding.

Stability in the digestive tract environment

Digestive lysozymes are active in an environment that combines acidic pH and high proteolytic activity. Resistance to the inactivation by acidic proteinases was described for cow lysozyme, which is approximately six times more resistant to pepsin than HEWL (Dobson et al., 1984). Similarly, MdL1 activity is totally recovered after incubation with an extract containing cathepsin D, a digestive proteinase from the midgut of that insect (Espinoza-Fuentes and Terra, 1987).

This MdL1 property may be related to a reduction in the number of sites recognized by cathepsin D, however this hypothesis

can not be further analyzed since the substrate specificity of insect cathepsin D is unknown. A low exposition to the solvent may also be related to the resistance to proteinases. That might be the case of the region between helix 5 and 6 of MdL1, specifically residues Q99 through G101, which forms a loop that is shorter and less exposed than in HEWL, making this region more protected from an attack by proteinases.

Properties related to stability in acidic medium shared by digestive lysozymes (Prager, 1996) are also present in MdL1, such as a low Arg content (5 in total) and fewer acid-labile Asp, Asn and Gln bonds (20 in total), specially the lack of the most acid-hydrolysis labile Asp99-Pro99A bond (101 and 102 in Prager numbering scheme).

pH optimum

A distinguishing feature of MdL1 and other digestive lysozymes is an acidic pH optimum, usually lower than those from non-digestive lysozymes (Lemos et al., 1993; Prager, 1996). This feature of the MdL1 catalytic activity depends on the carboxylates of Glu32 and Asp50, whose functionalities in catalysis were defined based on comparison with HEWL. Glu32 is the catalytic acid/base while Asp50 acts as a nucleophile and forms a covalent intermediate (Vocadlo et al., 2001). In this mechanism, Glu32 starts uncharged and Asp50 is negatively charged. Theoretical pK_a s for these two catalytic groups were estimated based on the MdL1 structure (tab. 3). Considering that Lys1 activity depends on these two ionizable groups, its pH optimum is expected to be an arithmetical mean between their pK_a s (Segel, 1993). Indeed, MdL1

theoretical pH optimum (4.4) agrees with that experimentally obtained (Lemos et al., 1993).

Based on this pH optimum, considering that pHs at each third of the *M. domestica* midgut are 6.1, 3.1 and 6.8 (Espinoza-Fuentes and Terra, 1987) and that bacterial wall is negatively charged at these pHs (Price and Pethig, 1986), MdL1, whose pI is 7.6, should be positively charged and would be more active at the second third of the *M. domestica* midgut. In this region, MdL1 could contribute to the lysis and digestion of bacteria (Espinoza-Fuentes and Terra, 1987). Besides that, MdL1 acidic pH optimum would prevent damage to the natural flora in the last part of midgut, where MdL1 activity is lower.

A reduction in the pK_a s of the catalytic groups is expected in order to attain a more acidic pH optimum. In agreement with that, theoretical pK_a for Glu32 (5.1) is lower than that observed for HEWL catalytic acid/base (6.5), whereas the catalytic nucleophiles have similar pK_a s (3.5) (Yang and Hamaguchi, 1980). Coulombic interactions, desolvation effects and hydrogen bonds will affect the pK_a , with the latest being the most important (Warshel, 1981; Mehler and Guarnieri, 1999; Porter et al., 2006; Davies et al., 2006). The structure of the MdL1 shows a unique hydrogen bond between Thr107 O γ and Glu32 carboxylate group (fig. 2D) when compared to the other structures in the alignment of figure 1B. This threonine is replaced by alanine or glycine in the other lysozyme sequences, which prevents such an interaction from occurring. Further analysis of 98 sequences deposited in the databank shows that this threonine is conserved in the digestive lysozymes from *M. domestica* (AY344588 and AY344589) and *Drosophila melanogaster* (LysD; AE003807), while being always alanine or

glycine in all other lysozymes (data not shown). Theoretical calculations of the effect produced by the change from Thr107 to glycine or alanine suggest that the presence of the hydrogen bond between Thr107 and Glu32 is responsible for a decrease of 0.6 units in the pK_a of the catalytic acid/base (tab. 3). Indeed, an analysis of the complex between HEWL and either (NAM-NAG-NAM) (Strynadka and James, 1991) or (NAG)₃ (Cheetham et al., 1992) revealed that the residue A110, which corresponds to Thr107, is part of a hydrophobic pocket that could contribute to increase the pK_a of the catalytic acid/base.

Other residues in the neighboring area of Glu32 and Asp50 (Asn44, Gln55, Asn57 and Ser106) were also scrutinized for specific changes that could lead to a pK_a change. However no other alteration was detected. The fact that only Glu32 is being influenced is not unexpected since Asp50 has a much lower pK_a , which is similar to that of a typical carboxylate, whereas Glu32 presents an unusually higher pK_a .

Materials and methods

Protein production, purification and crystallization

A single colony of *Pichia pastoris* transformed with pPic9kf1 coding for MdL1 was grown in YPD medium (160 rpm, 28°C, 18 h). Following that, the cells were harvested by centrifugation (2,500 x g, 4°C, 5 min), washed three times with autoclaved water, resuspended in BMM medium containing 1% methanol (v/v) and cultivated at 160 rpm and 28°C for 72 h. Aliquots of methanol (final concentration 1%) were periodically (each 24h) added to the

culture. Finally, the cells were discarded and supernatant containing recombinant MdL1 was stored at -20°C .

Ammonium sulfate (0.652 g/mL) was added to the culture supernatant containing MdL1 and incubated at 25°C for 18 h. Following that, the mixture was centrifuged ($10,000 \times g$, 1 h, 4°C) and the pellet was resuspended in 20 mM sodium acetate pH 5.5. This sample was dialyzed against the same buffer for 18 h at 4°C . After that, this sample was loaded onto a Mono S HR5/5 column (GE) previously equilibrated with 20 mM sodium acetate pH 5.5 and proteins were eluted with a linear NaCl gradient (0 to 1 M; 20 mL; flow rate: 1 mL/min). Eluted fractions (1.0 mL) containing lysozyme activity were pooled and dialyzed against 10 mM 2-morpholinoethanesulfonic acid (MES) pH 5.5. The purification of MdL1 was checked by SDS-PAGE.

The purified MdL1 at a concentration of 9.6 mg ml^{-1} and in 10 mM 2-morpholinoethanesulfonic acid (MES) pH 5.5 buffer, was crystallized by vapour-diffusion sitting-drop, using a solution of 1.4 M ammonium sulfate, 0.1 M 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid sodium salt (sodium HEPES) pH 7.5 and 1% polyethyleneglycol (PEG) 400 at 18°C (Marana et al., 2006). The crystals of the complex between lysozyme 1 and tetra-N-acetylchitotetraose ((NAG)₄) were obtained by soaking the sugar into native crystals. (NAG)₄ was dissolved in water to prepare a 1 mM solution and 1 μl of this solution was added to the solution containing a crystal of MdL1, where the protein concentration was 0.7 mM prior to crystallization. After this procedure, the crystals were stored for one week at 18°C .

Data collection and structure solution

Crystallographic data were collected at the D03B-MX1 beam line of the Brazilian Synchrotron Light Laboratory (LNLS). This beamline is equipped with a MAR CCD detector combined with a MAR DTB goniostat. Crystals were cooled directly in a nitrogen-gas stream to 100 K. Data collection was carried out using the oscillation method with 1.0° oscillation per frame and radiation wavelength 1,431 Å for the native dataset and 1.427 for the complex one. Data processing was performed with the HKL2000 package (Otwinowski and Minor, 1997) and CCP4 package (Collaborative Computational Project, Number 4, 1994). Initial phases for the native structure were obtained by molecular replacement using the HEWL structure deposited in the PDB with code lhew (Cheetham et al., 1992) as a search model (Marana et al., 2006). For the complex, one monomer of the MdL1 native structure deposited in PDB with code 2FBD without waters was used as a search model. In both cases, the MOLREP program (Vagin and Teplyakov, 1997) was used.

Crystallographic refinement and validation

In the first cycle of structure refinement, a simulated annealing procedure was used to reduce any bias towards the template structures. This was accomplished with the program CNS (Brunger et al., 1998). All further cycles were carried out with the REFMAC5 program (Murshudov et al., 1997) from the CCP4 suite (Collaborative Computational Project, Number 4, 1994). Visualization and model building was performed with the O (Jones et al., 1991) and COOT programs (Emsley and Cowtan, 2004), where $3F_o - 2F_c$ and $F_o - F_c$ electron density maps were used to guide the positioning of the atoms. Water molecules were added with the

automatic procedure of ARP/WARP (Lamzin and Wilson, 1997) and by visual inspection. The tri-*N*-acetylchitotriose (CTO or (NAG)₃) coordinates and topology parameters were obtained at the Hic-Up data base (Kleywegt and Jones, 1998). The refined structure was validated with the PROCHECK (Laskowski et al., 1993) program and the tools available in COOT.

Structure analysis

Besides the graphics programs O and COOT used above, the program PYMOL (DeLano, 2002) was employed to analyze the structure of the molecule. Another five crystallographic structures and their sequences were used in this analysis: 1JSF from *Homo sapiens* placenta (Harata et al., 1998), 1LMN from *Oncorhynchus mykiss* kidney (Karlsen et al., 1995), 1HEW from *Gallus gallus* egg white (Cheetham et al., 1992), 1GD6 from *Bombyx mori* (Matsuura et al., 2002) and 1IIZ from *Antheraea mylitta* (Jain et al., 2001). All the structures were superimposed and structurally aligned using the program COOT. Position of insertions or deletions were checked visually.

DLS measurement

Dynamic light scattering (DLS) measurements of the purified lysozyme were performed in a DynaPro™ DLS system from Wyatt Technology Corp. The sample, at 1.0 mg/ml concentration in 10 mM MES pH 5.5, was centrifuged to 7,000 x g for 7 minutes and the supernatant was submitted to 100 measurements.

Acknowledgments

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grants 02/10634-2; SMolBNet), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Associação Brasileira de Tecnologia de Luz Síncrotron (ABTLuS).

References

Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C. and Sarma, V.R. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **206**:757-761.

Blake, C.C., Johnson, L.N., Mair, G.A., North, A.C., Phillips, D.C. and Sarma V.R. 1967. Crystallographic studies of the activity of hen egg-white lysozyme. *Proc R Soc Lond B Biol Sci.* **167**:378-388.

Brunger, A.T., Adams, P.D., Clore, G. M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J. -S., Kuszewsk, J., Nilges, M., Pannu, N.S., Read, R. J., Rice, L.M., Simonson, T. and Warren, G.L. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst. D* **54**: 905-21.

Cheetham, J.C., Artymiuk, P.J. and Phillips, D.C. 1992. Refinement of an enzyme complex with inhibitor bound at partial

occupancy : Hen egg-white lysozyme and tri-*N*-acetylchitotriose at 1.75 Å resolution. *J. Mol. Biol.* **224**: 613-628

Collaborative Computational Project, Number 4. 1994. The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D* **50**: 760-763.

Coutinho, P.M. and Henrissat, B. 1999. Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering*. (eds. H.J. Gilbert, G. Davies, B. Henrissat and B. Svensson), pp. 3-12. The Royal Society of Chemistry, Cambridge.

Davies, G.J., Wilson, K.S. and Henrissat, B. 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**:557-559.

Davies, M.N, Toseland, C.P., Moss, D.S. and Flower, D.R. 2006. Benchmarking pK(a) prediction. *BMC Biochem.* **7**: 18.

DeLano, W.L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

Dobson, D.E., Prager, E.M. and Wilson, A.C. 1984. Stomach lysozymes of ruminant. 1 - Distribution and catalytic properties. *J. Biol. Chem.* **259**: 11607-11616.

Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Cryst. D* **60**: 2126-2132.

Espinoza-Fuentes, F.P. and Terra, W.R. 1987. Physiological adaptations for digesting bacteria. Water fluxes and distribution of digestive enzymes in *Musca domestica* larval midgut. *Insect Biochem.* **17**: 809-817.

Harata, K. 1993. X-ray Structure of Monoclinic Turkey Egg Lysozyme at 1.3 Å Resolution. *Acta Cryst. D* **49**: 497-504.

Harata, K., Abe, Y. and Muraki, M. 1998. Full-matrix least-squares refinement of lysozymes and analysis of anisotropic thermal motion. *Proteins* **30**: 232-243.

Hultmark D. 1996. Insect lysozymes. In: *Lysozymes: Model Enzymes in Biochemistry and Biology.* (eds. P. Jollès), pp 87 - 102 Birkhauser Verlag, Basel.

Ito, Y., Nakamura, M., Hotani, T. and Imoto, T. 1995. Insect lysozyme from house fly (*Musca domestica*) larvae: possible digestive function base don sequence and enzymatic properties. *J. Biochem. Tokyo* **118**: 546 - 551.

Jain, D., Nair, D.T., Swaminathan, G.J., Abraham, E.G., Nagaraju, J. and Salunke, D.M. 2001. Structure of the induced antibacterial protein from tasar silkworm, *Antheraea mylitta*. *J. Biol. Chem.* **44**: 41377 - 41382.

Jollès P. and Jollès J. 1984. What's new in lysozyme research? *Molec. Cell. Biochem.* **63**: 165 - 189.

Jones, T.A., Zou, J-Y., Cowan, S.W. and Kjeldgaard, M. 1991. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models. *Acta Cryst. A* **47**: 110-119.

Karlsen, S., Eliassen, B.E., Hansen, L.K., Larsen, R.L., Riise, B.W., Smalas, A.O., Hough, E. and Grinde, B. 1995. Refined Crystal-structure of lysozyme from the raibow-trout (*Oncorhynchus-mykiss*). *Acta Cryst. D* **51**: 354-367.

Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577-2637.

Kleywegt, G.J. and Jones, T.A. 1998. Databases in protein crystallography. *Acta Cryst. D* **54**: 1119-1131.

Krissinel, E. and Henrick, K. 2005. *Detection of Protein Assemblies in Crystals*. In: M.R. (eds. Berthold et al.), pp. 163-174. *CompLife 2005*, LNBI 3695. Springer-Verlag Berlin Heidelberg.

Lakowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283-291.

Lamzin, V.S., and Wilson, K.S. 1997. Automated refinement for protein crystallography. In *Meth. Enz.:277*.(eds. C. Carter and B. Sweet) pp. 269-305.

Lemos, F.J.A., Ribeiro, A.F. and Terra, W.R. 1993. A bacteria-digesting midgut lysozyme from *Musca domestica* (Diptera) larvae. Purification, properties and secretory mechanism. *Insect Biochem. Mol. Biol.* **23**: 533 - 541

Lemos, F.J.A. and Terra, W.R. 1991. Digestion of bacteria and the role of midgut lysozyme in some insect larvae. *Comp. Biochem. Physiol.* **100B**: 265 - 268.

Li, H., Robertson, A.D. and Jensen, J.H. 2005. "Very Fast Empirical Prediction and Interpretation of Protein pKa Values". *Proteins* **61**: 704-721.

Marana, S.R., Cançado, F.C., Valério, A.A., Ferreira, C., Terra, W.R. and Barbosa, J.A.R.G. 2006. Crystallization data collection and phasing of two digestive lysosymes from *Musca domestica*. *Acta Cryst. F* **62**: 750-752.

Matsuura, A., Yao, M., Aizawa, T., Koganesawa, N., Masaki, K., Miyazawa, M., Demura, M., Tanaka, I., Kawano, K. and Nitta, K. 2002. Structural analysis of an insect lysozyme exhibiting catalytic efficiency at low temperatures. *Biochemistry* **41**: 12086 - 12092.

Mehler, E.L. and Guarnieri, F. 1999. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J.* **77**: 3-22.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D* **53**: 240-255.

Otwinowski, Z. and Minor, W. 1997. Macromolecular Crystallography, Part A, *Methods in Enzymology*: 276. (eds. C.W. Carter Jr. and R.M. Sweet), pp. 307-326. New York: Academic Press.

Phillips, D.C. 1967. The hen egg white lysozyme molecule. *Proc. Nat. Acad. Sci. U.S.A.* **57**: 484-495.

Porter, M.A., Hall, J.R., Locke, J.C., Jensen, J.H. and Molina, P.A. 2006. Hydrogen bonding is the prime determinant of carboxyl pKa values at the N-termini of alpha-helices. *Proteins* **63**: 621-635.

Prager, E.M. 1996. Adaptive evolution of lysozyme: Changes in amino acid sequence, regulation of expression and gene number. In: *Lysozymes: Model Enzymes in Biochemistry and Biology*. (eds. P. Jollès) pp 323-345. Birkhauser Verlag, Basel.

Price, J.A. and Pethig, R. 1986. Surface charge measurements on *Micrococcus lysodeikticus* and the catalytic implications for lysozyme. *Biochim. Biophys. Acta* **889**: 128-135.

Segel, I.H. 1993. *Enzyme kinetics. Behavior and analysis of rapid equilibrium and steady-state enzyme systems*. Edited by Wiley and Sons. NY-USA.

Strynadka, N.C.J. and James, M.N.G. 1991. Lysozyme revisited: Crystallographic evidence for distortion of an N-acetylmuramic acid residue bound in site D. *J Mol Biol* **220**:401-424.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**: 4876-4882.

Vagin, A.A. and Teplyakov, A. 1997. *MOLREP*: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* **30**: 1022-1025.

Vocadlo, D.J., Davies, G.J., Laine, R. and Withers, S.G. 2001. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**: 835-838.

Yang, Y. and Hamaguchi, K. 1980. Hydrolysis of 4-Methylumbelliferyl *N*-Acetyl-Chitotrioside catalyzed by hen and turkey lysozymes. *J. Biochem.* **87**: 1003-1014.

Wallace, A.C., Laskowski, R. A. and Thornton, J. M. 1995. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot. Eng.* **8**: 127-134.

Warshel, A. 1981. Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry* **20**: 757-761.

Table 1 - Data collection and refinement statistics.

Data collection	Native (2FBD)	Complex (2H5Z)
Space group	P2 ₁	P2 ₁
Unit cell (Å; °)	a=36.52, b=79.44, c=45.20; • = • = 90.00, • = 102.97	a=35.95, b=78.81, c=44.79; • = • = 90.00, • = 102.16
Mosaicity (°)	0.7	1.0
Resolution range (Å)	30.00-1.90 (1.97-1.90)	30.00-1.92 (1.99-1.92)
I/•(I) after merging	23.8 (5.6)	16.2 (2.4)
Completeness (%)	100.0 (99.8)	88.1 (82.0)
Multiplicity	3.5 (3.4)	3.8 (2.8)
R _{sym}	0.052 (0.222)	0.078 (0.307)
No. of reflections	70004	62924
No. of unique reflections	19857 (1953)	16455 (1521)
Refinement		
R _{cryst} ^a	0.153	0.175
R _{free} ^a	0.198	0.242
r.m.s.d. bond lengths (Å)	0.012	0.018
r.m.s.d. bond angles (°)	1.365	1.660
Average B-factors (Å ²)	28.33	36.12
Ramachandran plot statistics by region (%)		
Most favored	93.1	91.2
Allowed	6.9	8.8
Generously allowed	0.0	0.0
Disallowed	0.0	0.0

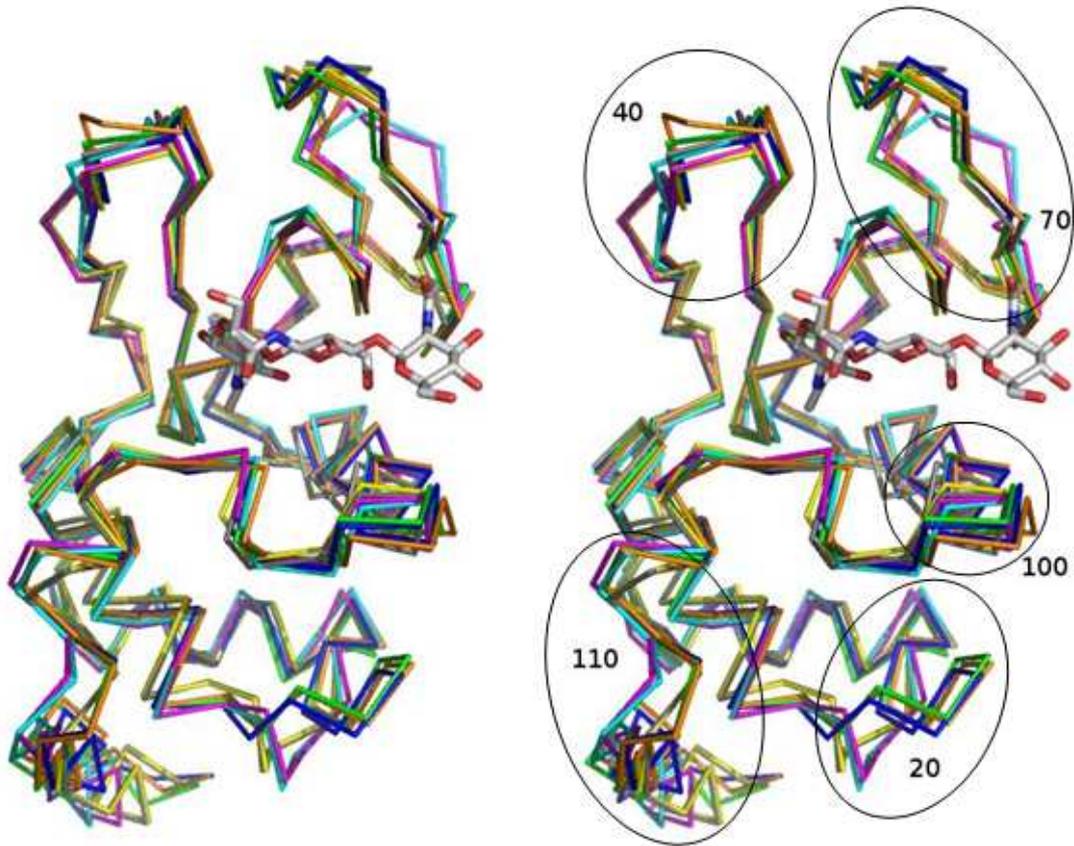
^a $R_{cryst} = \frac{\sum |F_o(h) - F_c(h)|}{\sum F_o(h)}$, where $F_o(h)$ and $F_c(h)$ are observed and calculated amplitudes for reflection h . R_{free} is calculated by the same equation using 5 % of the data, chosen randomly and omitted from the refinement.

Table 2 - Superposition of lysozyme *c* representatives from insects: 2FBD_A, 2FBD_B, 2H5Z_B (*Musca domestica*), 1GD6 (*Bombyx mori*) and IIZ (*Anthera mylitta*); birds: 1HEW (*Gallus gallus*); mammals: 1JSF (*Homo sapiens*) and fishes: 1LMN (*Oncorhynchus mykiss*). The _A and _B at the end of the PDB codes denote the monomer. The numbers in the last column represent the mean values of the numbers shaded in the respective gray scale.

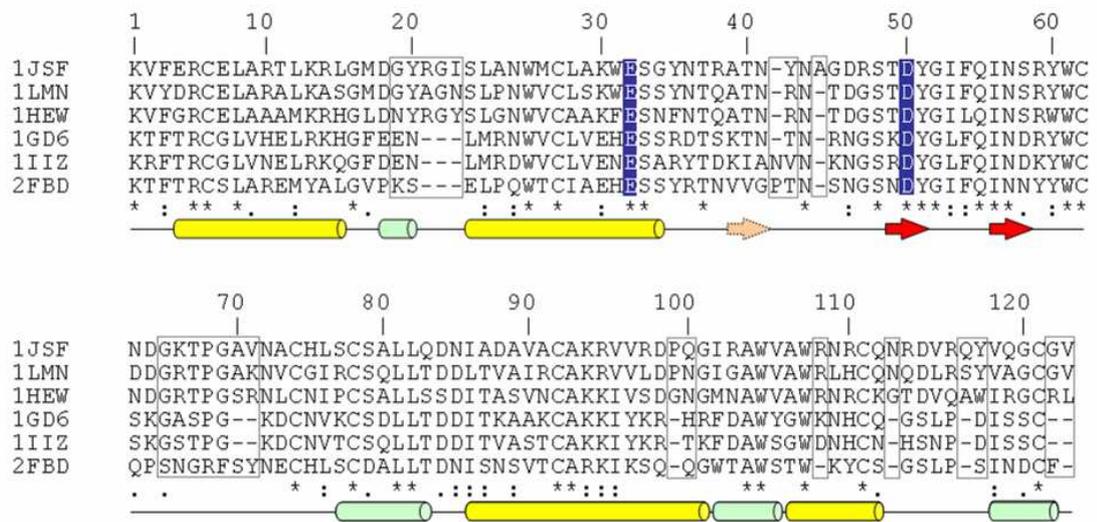
	Resolution		RMSD (Å)							
	(Å)	(%)	2FBD_B	2FBD_A	2H5Z_B	1GD6	1IIZ	1HEW	1JSF	mean
2FBD_A	1.90	100	0.24							0.27
2H5Z_B	1.92	100	0.34	0.24						1.09
1GD6	2.50	46	1.12	1.08	1.14					0.66
1IIZ	2.40	43	1.11	1.05	1.06	0.66				1.19
1HEW	1.75	43	1.21	1.21	1.22	1.04	1.22			1.02
1JSF	1.15	46	1.17	1.20	1.19	0.98	1.20	0.66		1.20
1LMN	1.80	42	1.10	1.23	1.18	1.03	1.17	0.56	0.47	0.56

Table 3 - Theoretical pK_a of the MdL1 catalytic residues calculated with the program PROPKA (Li et al., 2005). The numbers separated by slashes correspond to the values of pK_a for monomer A and B, respectively.

	Glu32	Asp50
Native (2FBD)	4.96/5.18	3.94/3.66
Complex (2H5Z)	4.99/5.15	3.83/4.00
Complex T107A	5.63/5.79	3.81/3.98
Complex T107G	5.56/5.73	3.80/3.97



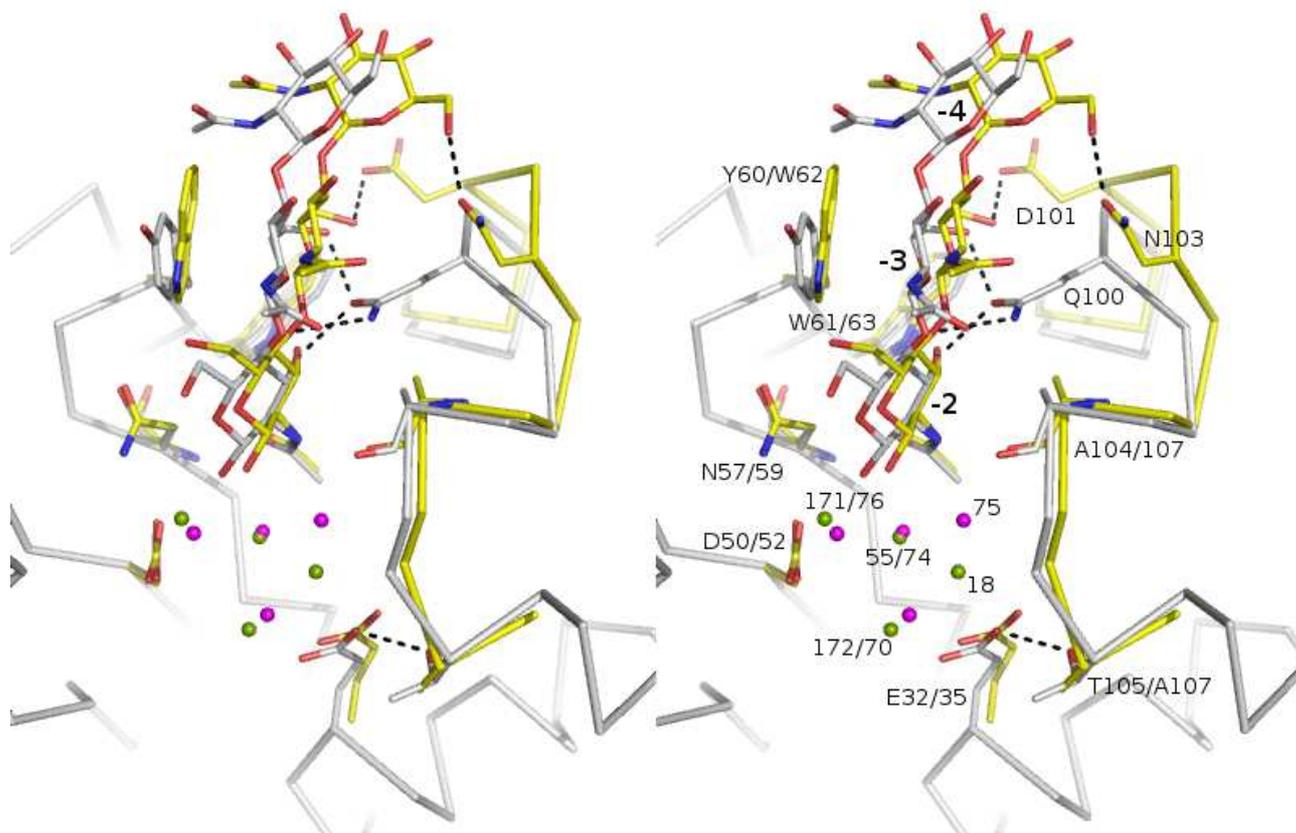
(A)



(B)

Legenda na próxima página.

Figure 1 - (A) Stereoview of the overall fold of lysozyme c shown by the superimposed C_α-traces of all animal classes represented in the PDB: 2H5Z-*Musca domestica* (housefly)-yellow, 2FBD-*Musca domestica* (housefly)-gray, 1IIZ-*Antheraea mylitta* (tasar silkworm)-magenta, 1GD6-*Bombyx mori* (silkworm)-cyan, 1HEW-*Gallus gallus* (hen) egg white-blue, 1JSF-*Homo sapiens* (human) placenta-orange and 1LMN-*Oncorhynchus mykiss* (rainbow trout) kidney-green. The circles illustrate regions presenting differences between the structures, where the numbers indicate an amino acid residue within the region. (B) Multiple alignment of representative structures of lysozymes C: 1JSF, 1LMN, 1HEW, 1GD6, 1IIZ and 2FBD. Residues involved directly in catalysis are marked in color and residues which are not structurally aligned are in boxes. The numbering refers to the 2FBD structure. This is a modified version of an alignment produced with ClustalX (Thompson et al., 1997). Asterisks (*) represent strictly conserved residues, while colon (:), and dot (.) represent conserved residues in decreasing degrees of conservation, respectively. The bottom line shows the secondary structure of 2FBD with arrows for α-strands and yellow cylinders for α-helices and green ones for 3₁₀-helices.



(D)

Figure 2 - (A) Stereoview of the stabilization of Asp50 side chain (left) and a unique aromatic cluster in MdL1 (right). Black dashed lines are representing hydrogen bonds. All the structures are represented by the C.-trace from their equivalents of residues Val40 to Cys74. The color scheme is the same as figure 1. The (NAG)₃ molecule is represented only by its sugar rings. (B) Stereoview of the (NAG)₃ binding to the sites -4 to -2 of MdL1 showing a simulated annealing omit map contoured at 1.0 sigma around the (NAG)₃ and the interactions with the protein (yellow) and solvent molecules. (C) Schematic plot of the ligand interactions with their respective distances and waters in cyan. Produced with LIGPLOT (Wallace et al., 1995). (D) Stereoview of the superposition of MdL1 (gray and waters in green) and HEWL (yellow and waters in magenta) showing the differences in ligand binding. The hydrogen bonds are represented by dashed lines.

JAP 02.1

Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança da ABTLuS – Associação Brasileira de Tecnologia de Luz Síncrotron

Título do projeto: Estudos estruturais e funcionais de hidrolases de *Xylella fastidiosa*

Pesquisador responsável: João Alexandre Ribeiro Gonçalves Barbosa

Experimentador: Amanda Abdalla Valério

Nível do treinamento do experimentador: []-Iniciação científica, []-mestrado, [x]-doutorado, []-doutorado direto, []-pós-doutorado, []-nível técnico, []-outro, especifique: _____

Resumo do projeto:

A clorose variegada de citros é uma doença que atinge plantações de laranja comprometendo seriamente a produção agrícola. O agente causador da clorose variegada de citros é a bactéria gram-negativa *Xylella fastidiosa*. Esta bactéria se encontra no xilema das plantas ou no aparelho bucal do inseto transmissor da doença. Recentemente o genoma completo da *X. fastidiosa* foi seqüenciado.

As hidrolases são enzimas que catalisam reações de hidrólise, e portanto englobam as celulases, hexosaminidases e peptidases. Estas enzimas estão presentes no genoma de *X. fastidiosa* e atuam na degradação de polímeros de açúcares e aminoácidos da planta hospedeira. Desta forma, estas hidrolases afetam o desenvolvimento das plantas infectadas. Além disso, as celulases são importantes para a *X. fastidiosa* pois estão associadas à obtenção de glicose e colonização do sistema vascular da planta.

Neste projeto pretendemos estudar os mecanismos de ação de algumas das hidrolases de *X. fastidiosa* através de técnicas de biologia molecular estrutural. Um total de nove genes de hidrolases foi selecionado do genoma e estão distribuídos em quatro famílias distintas. Duas destas famílias não possuem estrutura atômica conhecida. Experimentos desde a clonagem à resolução da estrutura serão executados visando contribuir para o preenchimento do espaço estrutural, o conhecimento dos mecanismos da clorose variegada de citros e quiçá para o controle desta doença.

A CIBio analisou este projeto em reunião realizada no dia: 16.4.2007.

Parecer final: -projeto aprovado, []-projeto recusado, []-projeto com deficiências, favor comentários abaixo:

Jörg Kobarg
 Presidente da CIBio - ABTLuS
 Prof. Dr. Jörg Kobarg

Albino Benedetti
 Membro da CIBio - ABTLuS
 Prof. Dr. Celso Eduardo Benedetti

Nilson Ivo Tonin Zanchin
 Membro da CIBio - ABTLuS
 Prof. Dr. Nilson Ivo Tonin Zanchin