

**UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**“TÉCNICAS DE BIOINFORMÁTICA APLICADAS AO  
ESTUDO DE POLIGALACTURONASES DE FUNGOS”**

**Adriana Miranda dos Santos\***

Dissertação apresentada como parte dos requisitos para obtenção do título de MESTRE EM QUÍMICA, área de concentração: QUÍMICA ORGÂNICA.

**Orientadora: Prof<sup>a</sup> Dr<sup>a</sup> Dulce Helena Ferreira de Souza**

**\* bolsista**

**São Carlos - SP  
2012**

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

S237tb Santos, Adriana Miranda dos.  
Técnicas de bioinformática aplicadas ao estudo de poligalacturonases de fungos / Adriana Miranda dos Santos.  
-- São Carlos : UFSCar, 2012.  
110 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2012.

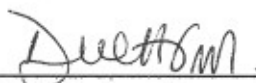
1. Química orgânica. 2. Poligalacturonase. 3. Fungos. 4. Alinhamento de sequências. 5. Sequência de aminoácidos. 6. Relações filogenéticas. I. Título.

CDD: 547 (20<sup>a</sup>)

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**  
*Centro de Ciências Exatas e de Tecnologia*  
*Departamento de Química*  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**  
**Curso de Mestrado Acadêmico**

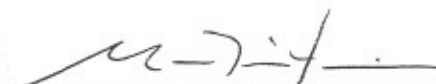
---

*Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de dissertação de Mestrado Acadêmico da candidata Adriana Miranda dos Santos, realizada em 18 de abril de 2012:*



---

Profª. Dra. Dulce Helena Ferreira de Souza



---

Prof. Dr. Moacir Rossi Forim



---

Dr. Humberto D'Muniz Pereira

## DEDICATÓRIA

Ao meu pai, João B. dos Santos. À minha mãe, Eliza M. dos Santos. Aos meus irmãos Wilian e Eduardo. Ao meu namorado John.

## **AGRADECIMENTOS**

À Profª Drª Dulce Helena Ferreira de Souza pela excelente orientação, paciência, ensinamentos e amizade;

À Profª Drª Quezia B. Cass pelos conselhos e principalmente por ter me apresentado à minha orientadora;

Aos meus queridos amigos de Laboratório pelo clima agradável e descontraído de trabalho;

Ao Thiago Cass pela ajuda na confecção do abstract;

Ao Wilian e John pelas discussões construtivas e apoio incondicional;

À CAPES pela ajuda financeira.

## LISTA DE TABELAS

TABELA 1.4 - Número de sequências de PG de fungos depositadas no NCBI e número de sequências únicas por gênero encontradas neste trabalho. ....	38
TABELA 2.4 - Maiores escores obtidos para os alinhamentos múltiplos entre as sequências de PG sem classificação (Sequência A) e as sequências de Endo-PG ou Exo-PG (sequência B). ....	43
TABELA 3.4 - Motivos sequenciais presentes nas 63 sequências de Endo-PG e 38 sequências de Exo-PG, encontrados pelo programa MEME (Bailey & Elkan, 1994). ....	49
TABELA 4.4 - Representação das sequências dos motivos presentes nas 63 sequências de Endo-PG e 38 sequências de Exo-PG .....	49
TABELA 5.4 – Melhores aminoácidos correspondentes possíveis e o grau de similaridade entre os motivos. ....	49

## LISTA DE FIGURAS

FIGURA 1.1 - Modelo da parede celular de plantas, mostrando os principais polímeros estruturais e seu arranjo nas células vegetais. ( <a href="http://www.sobiologia.com.br/conteudos/Morfofisiologia_vegetal/morfovegetal15.php">http://www.sobiologia.com.br/conteudos/Morfofisiologia_vegetal/morfovegetal15.php</a> ).....	2
FIGURA 2.1 - Estrutura esquemática da pectina (HARHOLT et al., 2010). Kdo, 3-Deoxy-d-manno-2-octulosonic acid; DHA, 3-deoxy-d-lyxo-2-heptulosaric acid. ....	3
FIGURA 3.1 - Estrutura de homogalacturonana. Resíduos de D-ácido galacturônico unidos através de ligação $\alpha$ - (1 $\rightarrow$ 4) e grupos metil e acetil ésteres (RIDLEY et al., 2001). ....	4
FIGURA 4.1 - Estrutura de ramnogalacturonana-I ( <a href="http://www.ccr.uga.edu/~mao/rg1/rg1.htm">http://www.ccr.uga.edu/~mao/rg1/rg1.htm</a> ). ....	4
FIGURA 5.1 - Sequência glicosil de RG II (VIDAL et al., 2000). A cadeia principal de GalA é substituída com quatro diferentes cadeias laterais de oligossacarídeos (cadeias A-D). A localização das cadeias laterais é arbitrária uma vez que a distribuição das mesmas na cadeia principal não é conhecida. ....	5
FIGURA 6.1 - Modo de ação das pectinases sobre a molécula de pectina e pectato (UENOJO & PASTORE, 2007). PMGL: polimetilgalacturonato liase. PMG: polimetilgalacturonase. PMGE: polimetilgalacturonato esterase (pectina esterase). PGL: poligalacturonato liase (pectato liase). PG: poligalacturonase. ....	7
FIGURA 7.1 - Comparação entre os mecanismos de clivagem de ligação glicosídica pela ação de polissacarídeo liases, PL, via $\beta$ -eliminação e pela ação de glicosídeo hidrolases, GH, via hidrólise (LOMBARD et al., 2010). ....	8
FIGURA 8.3 - Exemplo de alinhamento entre duas sequências de aminoácidos de Endo-poligalacturanase (Endo-PG) de <i>Athelia rolfsii</i> . O alinhamento foi obtido com o uso do programa <i>Clustal-W</i> (THOMPSON et al., 1994). ....	18
FIGURA 9.3 - Exemplo de alinhamento múltiplo de sequência. Foram alinhadas 6 sequências de Endo-PG de <i>A. tenuissima</i> pelo <i>ClustalW</i> (THOMPSON et al., 1994). ....	19
FIGURA 10.3 - Alinhamento global e local entre duas sequências. ....	20
FIGURA 11.3 - Matriz de valores de alinhamentos PAM 250 ( <a href="http://research.nmsu.edu/molbio/bioinfo/tutorials/comp-bio/distance.htm">http://research.nmsu.edu/molbio/bioinfo/tutorials/comp-bio/distance.htm</a> ). ....	21
FIGURA 12.3 - Matriz de valores de alinhamentos BLOSUM 62 ( <a href="http://en.wikipedia.org/wiki/File:BLOSUM62.gif">http://en.wikipedia.org/wiki/File:BLOSUM62.gif</a> ). ....	21
FIGURA 13.3 - Exemplo de uma sequência no formato FASTA. Sequência de uma poligalacturonase de <i>Aspergillus niger</i> . A primeira linha é referente à identificação da sequência e as demais linhas se referem à sequência. ....	23
FIGURA 14.3 - Representação de árvore filogenética com raiz e suas principais terminologias. ....	26
FIGURA 15.3 - Representação de árvore filogenética sem raiz. ....	26
FIGURA 16.3 - Representação dos ângulos torsionais. ( <a href="http://www2.bioqmed.ufrj.br/prosdocimi/courses/.../ProteinStructure.ppt">www2.bioqmed.ufrj.br/prosdocimi/courses/.../ProteinStructure.ppt</a> ).....	33
FIGURA 17.3 - Mapa de Ramachandran inicial. ( <a href="http://kpwu.wordpress.com/2007/02/15/making-ramachandran-plot-by-your-self/">http://kpwu.wordpress.com/2007/02/15/making-ramachandran-plot-by-your-self/</a> ). ....	33
FIGURA 18.4 - Gráfico representando a classificação das PG conforme depositadas no banco de dados do NCBI.....	39
FIGURA 19.4 - Alinhamento das sequências de PG (S24156) e Endo-PG (ABF50893.1; ABF50868.1) de <i>E. nidulans</i> e escores obtidos para os alinhamentos. Os alinhamentos foram realizados com o programa <i>ClustalW</i> (THOMPSON et al., 1994). ....	40
FIGURA 20.4 - Alinhamento das sequências de PG (S24156) e Exo-PG de <i>E. nidulans</i> e escores obtidos para o alinhamento. Os alinhamentos foram realizados com o programa <i>ClustalW</i> (Thompson et al., 1994). ....	41
FIGURA 21.4 - Gráfico representando o número total de sequências de Endo-PG e Exo-PG após classificação por alinhamentos múltiplos (68 sequências) e literatura (3 sequências). ....	44
FIGURA 22.4 - Árvore filogenética construída a partir de 38 sequências Exo-PG e 63 sequências Endo-PG de fungos. ....	47
FIGURA 23.4 - Árvore filogenética de sequências de 63 Endo-PG, 38 Exo-PG e 24 PG de fungos sem classificação. ....	51
FIGURA 24.4 - Número total de sequências de Endo-PG e Exo-PG de fungos após a classificação realizada neste trabalho. ....	52
FIGURA 25.4 - Ilustração dos dois mecanismos propostos para hidrólise de ligação glicosídica, com retenção (a) e inversão (b) da configuração do açúcar (DAVIES & HENRISSAT, 1995).....	56

- FIGURA 26.4 - Predição da estrutura secundária da PG feita com o programa Psipred (<http://bioinf.cs.ucl.ac.uk/psipred/>) Os cilindros rosas representam  $\alpha$ -hélices, as setas amarelas representam folhas- $\beta$  e os traços em preto representam loops e voltas.....57
- FIGURA 27.4 - Alinhamento e tabela de escore dados pelo programa *ClustalW* (THOMPSON et al., 1994) entre as sequências de PG do *L. gongylophorus* (ADV30326.1) e de Endo-PG do *C. purpureum* (1K5C).....60
- FIGURA 28.4 - Sequência de aminoácidos da PG de *Leucoagaricus gongylophorus* deduzida da sequência de nucleotídeo. Os resíduos em vermelho são os conservados em PG de fungos (MARKOVIC & JANECEK, 2001).....61
- FIGURA 29.4 - Representação esquemática da estrutura  $\beta$ -hélice paralela destra do modelo estrutural da PG do *L. gongylophorus* construído com o programa SWISS-MODEL. A molécula é visualizada pelo lado N-terminal.....62
- FIGURA 30.4 - Representação esquemática do modelo estrutural da PG de *L. gongylophorus* visualizada do lado C-terminal focalizando o sítio catalítico (grande fenda aberta formada pela folha- $\beta$  PB1 e dos *loops* que se encontram dos dois lados da folha- $\beta$  PB1). Os resíduos de ácido aspártico envolvidos provavelmente na catálise (Asp<sup>179</sup>, Asp<sup>199</sup> e Asp<sup>200</sup>) estão representados na forma de linhas pretas.....62
- FIGURA 31.4 - Representação 3D de algumas Endo-PG de fungos e do modelo da PG de *L. gongylophorus* (D). Sendo o N-terminal à esquerda e o C-terminal à direita. (A) Endo-PG do *A. niger* (código PDB 1CFZ). (B) Endo-PG do *C. lupini* (código PDB 2IQ7). (C) Endo-PG do *C. purpureum* (código PDB 1K5C).....63
- FIGURA 32.4 - Representação esquemática das folhas- $\beta$  PB1, PB2a, PB2b e PB3 mostradas em azul escuro, lilás, azul claro e marrom, respectivamente. Juntas elas formam a estrutura  $\beta$ -hélice da PG do *L. gongylophorus*. Os 8 resíduos (Asn<sup>177</sup>, Asp<sup>179</sup>, Asp<sup>199</sup>, Asp<sup>200</sup>, His<sup>221</sup>, Gly<sup>222</sup>, Arg<sup>252</sup> e Lys<sup>254</sup>) extremamente conservados tanto em Endo- quanto em Exo-PG são representados em forma de linhas pretas.....64
- FIGURA 33.4 - Mapa de Ramachandran do modelo sem qualquer refinamento da PG do *L. gongylophorus*. Os ângulos Phi e Psi estão em graus. Os triângulos representam os resíduos de glicina.....65
- FIGURA 34.4 - Modelo 3D da PG de *L. gongylophorus* destacando os aminoácidos localizados nas regiões generosas e proibidas do mapa de Ramachandran.....66
- FIGURA 35.4 Estatísticas de alguns parâmetros estereoquímicos da cadeia principal (representado pelo quadrado em preto) do modelo 3D da PG de *L. gongylophorus* quando comparados a estruturas refinadas a mesma resolução: a. qualidade do mapa de Ramachandran, b.planaridade da ligação peptídica (ângulo  $\omega$ ), c. medidas ruins de interações entre não-ligantes, d. distorção tetraédrica, e. energia de ligação de hidrogênio e f.  $G_{factor}$ .....67
- FIGURA 36.4 - Mapa de Ramachandran do modelo refinado pelo "coot" (EMSLEY & COWTAN, 2004) da PG do *L. gongylophorus*. Os ângulos Phi e Psi estão em graus. Os triângulos representam os resíduos de glicina.....69
- FIGURA 37.4 - Modelo 3D refinado da PG de *L. gongylophorus* destacando os aminoácidos (Ala36, Met125, Ser145, Pro149, Asn266, Arg321 e Phe353) localizados nas regiões generosas do mapa de Ramachandran.....69
- FIGURA 38.4 Estatísticas de alguns parâmetros estereoquímicos da cadeia principal (representado pelo quadrado em preto) do modelo 3D refinado da PG de *L. gongylophorus* quando comparados a estruturas refinadas a mesma resolução: a. qualidade do mapa de Ramachandran, b.planaridade da ligação peptídica (ângulo  $\omega$ ), c. medidas ruins de interações entre não-ligantes, d. distorção tetraédrica, e. energia de ligação de hidrogênio e f.  $G_{factor}$ ..... 71
- FIGURA 39.4 - Gráfico representando os resultados obtidos pelo programa VERIFY 3D (LUTHY et al, 1992) para o modelo após refinamento. O menor valor médio 3D-1D foi 0,12 (Val293) e o maior 0,78 (Asn314). ..... 71
- FIGURA 40.4 - Sobreposição da estrutura tridimensional de Endo-PG de *C.purpureum* (azul) e do modelo 3D da PG de *L.gongylophorus* (vermelho). O valor de rms é de 0,029Å. ....72
- FIGURA 41.4 – Alinhamento entre Endo-PG de fungos de estrutura 3D conhecida. 1IA5 (*A. aculeatus*), 1NHC (*Aspergillus niger*), 1CZF (*A. niger*), 2IQ7 (*Colletotrichum lupini*), 1HG8 (*Fusarium moniliforme*), ADV30326.1 (*L. gongylophorus*) e 1K5C (*C. purpureum*). As sequências grifadas em azul corresponde ao motivo 1, a amarela ao motivo 2, a vermelha ao motivo 3, a verde ao motivo 4 e a rosa ao motivo 5. Os resíduos em azul correspondem aos 8 resíduos conservados em PG de fungos (Kester et al, 1996). Os resíduos em vermelho quando em posição *cis*-peptídica em conjunto com a K (grifada em cinza) são responsáveis pelo



reconhecimento do substrato (Shimizu et al, 2002).....	75
FIGURA 42.4 - Modelo tridimensional da PG de <i>L. gongylophorus</i> destacando os motivos 2 (em amarelo), 4 (em verde) e 5 (em rosa). ....	76
FIGURA 43.4 - Modelo tridimensional da PG de <i>L. gongylophorus</i> , destacando a região do sítio catalítico com os três resíduos de ácido aspárticos envolvidos na catálise (em azul escuro) e a molécula de açúcar (ácido galacturônico, em azul claro).....	76
FIGURA 44.4 - Região do sítio catalítico da PG de <i>L. gongylophorus</i> com os três resíduos de ácido aspárticos envolvidos na catálise (em azul escuro) e a molécula de açúcar (ácido galacturônico, em azul claro). ....	77
FIGURA 45.4 - Desenho esquemático propondo o possível mecanismo de ação para PG de <i>L. gongylophorus</i> com base no modelo 3D obtido no presente trabalho. O conservado resíduo Asp200 atua como um doador de próton, enquanto os resíduos Asp179 e Asp199 ativam a molécula de água. Arg252 e Lys254 interagem com o Asp199 e com o açúcar. ....	78

## RESUMO

A parede celular de células de plantas é composta de aproximadamente 65% de rede de celulose e rede de pectato. Essa última é proteoliticamente degradada por enzimas chamadas pectinases ou enzimas pectinolíticas. As pectinases podem ser despolimerizantes ou desesterificantes e são produzidas por plantas, fungos filamentosos, bactérias e leveduras. Impulsionado pelas aplicações comerciais que pectinases representam, é crescente o estudo para o melhor entendimento sobre as atividades dessas enzimas. Ferramentas de bioinformática foram usadas neste trabalho para estudar poligalacturonases (PG) de fungos, sob dois aspectos. Na primeira parte do trabalho, todas as sequências de PG de fungos depositadas em bancos de dados (2093) foram analisadas e avaliadas quanto à possibilidade de classificação das enzimas em Endo- ou Exo-PG através de busca de motivos sequenciais e estudos filogenéticos. Após exclusão daquelas com menos de 70 aminoácidos, das que correspondiam a 'proteína hipotética' e ainda das que não eram PG ou mesmo não eram PG de fungos restaram 957 sequências. Essas sequências foram separadas por gênero e espécie e para cada grupo foram realizados alinhamentos múltiplos de sequências usando o programa *ClustalW*. Os alinhamentos foram analisados e as sequências com 100% de identidade foram retiradas, resultando em um conjunto de dados com 417 sequências únicas de PG de fungos. Através dos alinhamentos, do estudo de motivos estruturais e da construção de árvores filogenéticas foi possível classificar todas as sequências de acordo com seu modo de ação em Endo- ou Exo-PG. Na segunda parte do trabalho foi usado o método de modelagem por homologia na construção do modelo tridimensional de uma poligalacturonase do fungo *L. gongylophorus*, simbiote de formigas cortadeiras. O modelo foi validado com os programas PROCHECK, VERIFY 3D e WHAT IF. Através da análise do modelo 3D da PG do *L. gongylophorus* foi possível propor um mecanismo de ação da enzima, que deve ocorrer com inversão da configuração do carbono anomérico do açúcar (substrato).

## ABSTRACT

Plant cell walls are composed of approximately 65% cellulose microfibrils and pectin. The latter is proteolytically degraded by the so-called pectinases enzymes (also known as pectinolytic enzymes). Pectinases may be either depolymerizing. They are produced by plants, filamentous fungi, bacteria, and yeast. Due to the wide commercial use of pectinase, one has testified a growing number of studies devoted to understand the activities of such enzymes. Bioinformatics tools were employed throughout this work in order to study fungal polygalacturonase (PG) under two aspects. Firstly, all stored fungal PG sequences in databases (2093) were analyzed in order to evaluate – through searching for sequential motifs and phylogenetic studies – the possibility to classifying the enzymes as either Endo- or Exo-PG. After excluding those with less than 70 amino acids, those that corresponded to the ‘hypothetical protein’, and those that were neither PG nor fungal PG, there were 957 sequences left. Those sequences were separated according to genus and species. For each group, multiple sequence alignments were made by using *ClustalW* software. The alignments were analyzed and the sequences displaying 100% identity were then expunged, thus resulting in a database of unique sequences of fungal PG. By means of the alignment, the study of structural motifs, and the construction of phylogenetic trees, one was able to classify all the sequences according to their mode of action in either Endo- or Ex-PG. Throughout the second part of our research, protein homology-modeling methods were employed while constructing a three-dimensional model of a *L. gongylophorus* fungal polygalacturonase, symbiont of leaf-cutting ant. The model was validated by PROCHECK, VERIFY 3D, and WHAT IF software. By analyzing the 3D model of the *L. gongylophorus* PG, a catalytic mechanism of the enzyme was outlined, which may take place by inverting the configuration of sugar anomeric carbon (substrate).

## SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1 - Parede celular de plantas .....	1
1.2 - Enzimas que degradam parede celular de plantas .....	6
1.2.1- Enzimas desesterificantes ou pectinas esterases .....	7
1.2.2 – Enzimas despolimerizantes .....	7
1.3 – Aplicações (bio)tecnológicas de enzimas pectinolíticas.....	9
1.4 – Bioinformática.....	11
CAPÍTULO 2 - OBJETIVOS.....	13
CAPÍTULO 3 - FERRAMENTAS DE BIOINFORMÁTICA .....	15
3.1 – Bancos de dados.....	15
3.2 – Alinhamentos de sequências .....	16
3.2.1 – Alinhamento global e local.....	19
3.2.2 – Matrizes de substituição.....	20
3.2.3 – Programa <i>ClustalW</i> .....	22
3.3 – Filogenia Molecular.....	24
3.3.1 – Árvores Filogenéticas .....	25
3.3.2 - Reconstrução filogenética: método da <i>máxima verossimilhança</i> .....	27
3.4 – Busca por motivos sequenciais – Programa MEME.....	28
3.5 – Modelagem molecular .....	29
3.5.1 - Modelagem por homologia.....	30
3.5.2 - O programa SWISS-MODEL .....	31
3.5.3 - Validação de modelos estruturais .....	32
CAPÍTULO 4 - RESULTADOS E DISCUSSÕES.....	37
4.1 - Estudo teórico de poligalacturonases (PG) de fungos.....	37
4.1.1- Formação do banco de dados com PG de fungos .....	37
4.1.2- Classificação das sequências de PG por alinhamento.....	38
4.1.3- Classificação das sequências de PG por árvore filogenética .....	44
4.1.4- Construção de uma árvore filogenética teste .....	45
4.1.5 - Construção da árvore filogenética para classificação de PG de fungos .....	49
4.2 - Estudo estrutural teórico de uma PG do fungo <i>L. gongylophorus</i> através de modelagem por homologia .....	52
4.2.1 - Predição da estrutura secundária da PG do <i>L. gongylophorus</i> .....	56
4.2.2 - Construção do modelo tridimensional da PG do <i>L. gongylophorus</i> .....	57
4.2.3 - Descrição do modelo estrutural da PG de <i>L. gongylophorus</i> .....	60

4.2.4 - Validação do modelo tridimensional da PG do <i>L. gongylophorus</i> .....	64
4.2.5 - O modelo estrutural e o sítio ativo da enzima.....	72
CAPÍTULO 5 - CONCLUSÕES .....	79
CAPÍTULO 6 - REFERÊNCIAS BIBLIOGRÁFICAS.....	83
Apêndice A – Tabela das 63 sequências de Endo-PG de fungos e 38 de Exo-PG com seus respectivos motivos conservados.....	91
Apêndice B – Alinhamento representativo das PG de fungos (63 Endo-PG e 38 Exo-PG). ...	92
Apêndice C – Resultado da avaliação feito pelo programa WHAT IF para o modelo 3D do <i>L. gongylophorus</i> . .....	106
Apêndice D – Escore resultante do alinhamento entre Endo-PG de fungos de estrutura 3D conhecida e a sequência de PG do <i>L. gongylophorus</i> .....	110

## CAPÍTULO 1 - INTRODUÇÃO

### 1.1 - Parede celular de plantas

As células de plantas, assim como de diversos seres vivos, são constituídas por uma estrutura semi-rígida chamada parede celular, que separa o meio intracelular do extracelular. Geralmente a parede celular de células de plantas é composta de aproximadamente 30% de celulose, 30% de hemicelulose, 35% de pectinas e 1 a 5% de proteínas estruturais (COSGROVE, 1997). As primeiras camadas formam a parede primária, que mantém a sua elasticidade permitindo que a célula possa crescer (COSGROVE, 1997). Na parede primária, as microfibrilas não apresentam uma direção definida e encontram-se ligadas por ligações de hidrogênio, o que torna a estrutura mais flexível. Novas camadas de celulose depositadas dentro da parede primária geram espessamento da parede, inclusive com impregnação de lignina e formam a parede secundária - com a qual as células não podem mais crescer (COSGROVE, 1997). A parede celular é uma estrutura muito complexa que possui uma grande diversidade de funções durante a vida da planta: proporcionam às células robustez mecânica, mantêm a sua morfologia, controlam a expansão celular e o transporte intercelular, protegem a célula contra a maioria dos organismos potencialmente patogênicos e predadores, participam na comunicação intercelular e contribui em alguns casos como reserva alimentar (NORTHCOTE, 1972).

Pectina é, na verdade, um termo genérico para um grupo extremamente complexo de polissacarídeos de alto peso molecular cuja composição, estrutura e quantidade diferem em diferentes espécies, idade e partes da planta (RIDLEY et al. 2001). É o principal componente da lamela média, uma fina camada de material adesivo extracelular entre as paredes primárias de células de vegetais superiores. Em conjunto com celulose e hemicelulose exerce a função de cimentação intercelular: as pectinas formam um gel no qual a celulose e a hemicelulose estão embebidas (Figura 1.1) e devem agir como um gel hidrofílico que evita a agregação e colapso da rede de celulose (COSGROVE, 1997). As ramificações na molécula servem para aprisionar moléculas de água fato que dá à pectina a característica de gel (UENOJO & PASTORE, 2007).

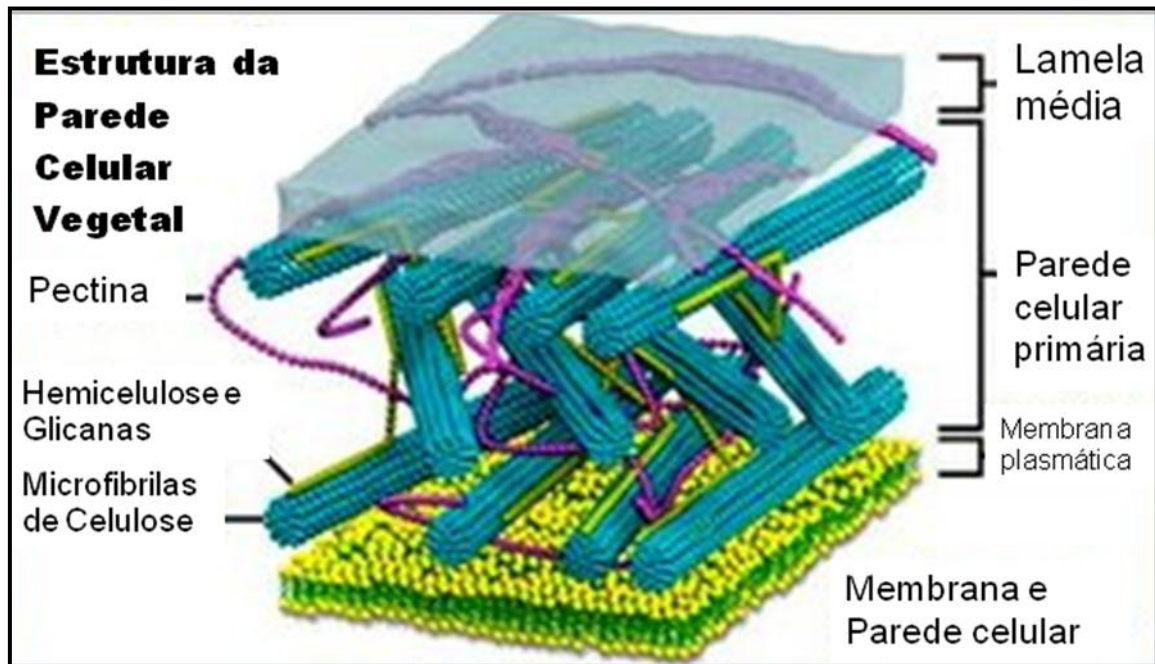


FIGURA 1.1 - Modelo da parede celular de plantas, mostrando os principais polímeros estruturais e seu arranjo nas células vegetais. ([http://www.sobiologia.com.br/conteudos/Morfofisiologia\\_vegetal/morfovegetal15.php](http://www.sobiologia.com.br/conteudos/Morfofisiologia_vegetal/morfovegetal15.php)).

Quimicamente, as pectinas são um complexo coloidal de polissacarídeos ácidos, composto de resíduos de  $\alpha$ -D-ácido galacturônico (GalA) unidos por ligações  $\alpha$ -1,4 (UNOJO & PASTORE, 2007). Os três principais polissacarídeos pécnicos melhor estudados são homogalacturonana (HG), ramnogalacturonana I (RG-I) e ramnogalacturonana II (RG-II) (RIDLEY et al., 2001), todos contendo resíduos de ácido galacturônico em maior ou menor grau. O xilogalacturonana, um dos polissacarídeos pécnicos que constituem a pectina, tem sido menos estudado. Os polissacarídeos mais abundantes na pectina são o homogalacturonana (HG) (aproximadamente 65% da pectina) e o ramnogalacturonana I (RG-I) (25 a 35%) (HARHOLT et al., 2010). A Figura 2.1 ilustra a estrutura da pectina e seus constituintes, proposta por Harholt e colaboradores (HARHOLT et al., 2010).

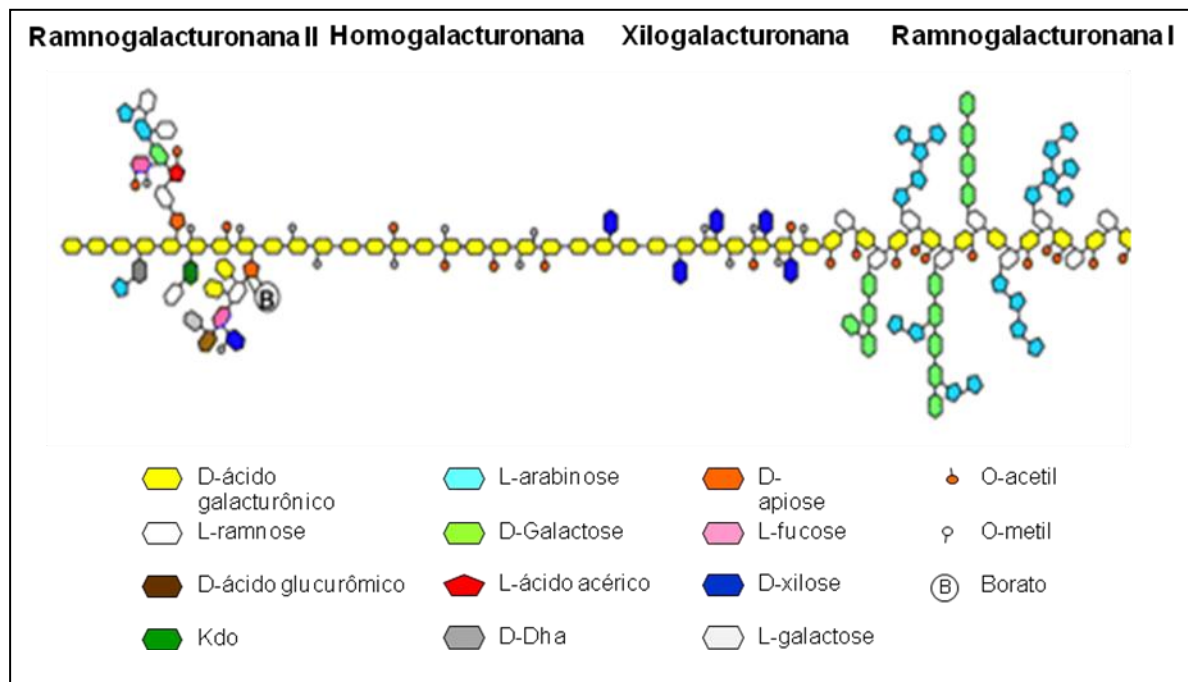


FIGURA 2.1 - Estrutura esquemática da pectina (HARHOLT et al., 2010). Kdo, 3-Deoxy-d-manno-2-octulosonic acid; DHA, 3-deoxy-d-lyxo-2-heptulosaric acid.

A homogalacturonana (HG) é um polissacarídeo linear cuja cadeia principal é formada por resíduos de  $\alpha$ -D-ácido galacturônico (GalA) que pode encontrar-se acetilada no C-2 ou C-3 e/ou metil-esterificada no C-6 (Figuras 2.1 e 3.1). Os resíduos de GalA estão unidos entre si por ligações  $\alpha$  (1 $\rightarrow$ 4) e alguns grupos carboxilas são em geral parcialmente esterificados com metanol, razão pela qual são resistentes à ação de enzimas Endo-poligalacturonases (MCNEIL et al., 1984). Quanto ao grau de esterificação, o HG pode apresentar diferentes graus: totalmente metil-esterificado (pectina), parcialmente metil-esterificado (ácido pectínico) ou isenta de metilação (ácido péctico ou poligalacturônico). O termo pectina também é frequentemente utilizado para designar qualquer um dos três tipos de substâncias pécticas (ROMBOUTS & PILNIK, 1980; NAIDU & PANDA, 1998; LANG & DÖRNENBURG, 2000).



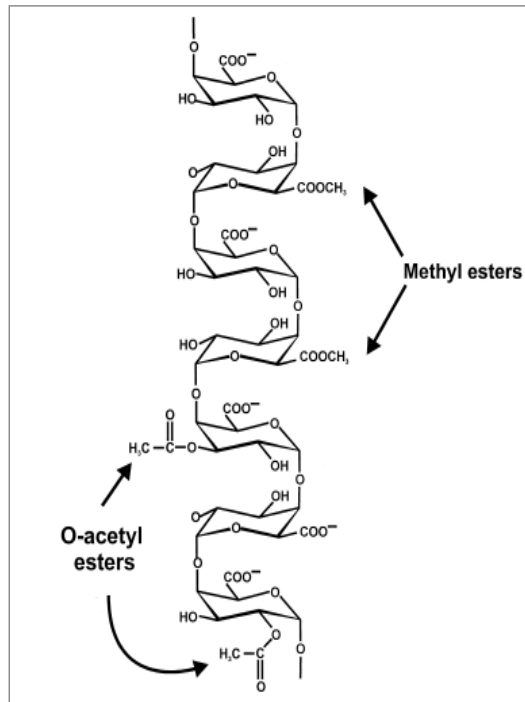


FIGURA 3.1 - Estrutura de homogalacturonana. Resíduos de D-ácido galacturônico unidos através de ligação  $\alpha$ - (1 $\rightarrow$ 4) e grupos metil e acetil ésteres (RIDLEY et al., 2001).

Ramnogalacturonana I (RG-I) é um polissacarídeo péctico cuja cadeia principal consiste em resíduos alternados de GalA com resíduos de ramnose [ $\alpha$ - (1 $\rightarrow$ 4)- D-ácido galacturônico,  $\alpha$ - (1 $\rightarrow$ 2)- L-ramnose]. Cerca de 20 a 80% dos resíduos de ramnose encontram-se ligados covalentemente a resíduos de açúcares neutros, tais como arabinose e galactose (MCNEIL et al., 1984; RIDLEY et al., 2001). Algumas das cadeias laterais podem ser terminadas com resíduos de  $\alpha$ -L-fucose,  $\beta$ -D-GlcpA, e 4-O-metil  $\beta$ -D-GlcpA (Figura 4.1).

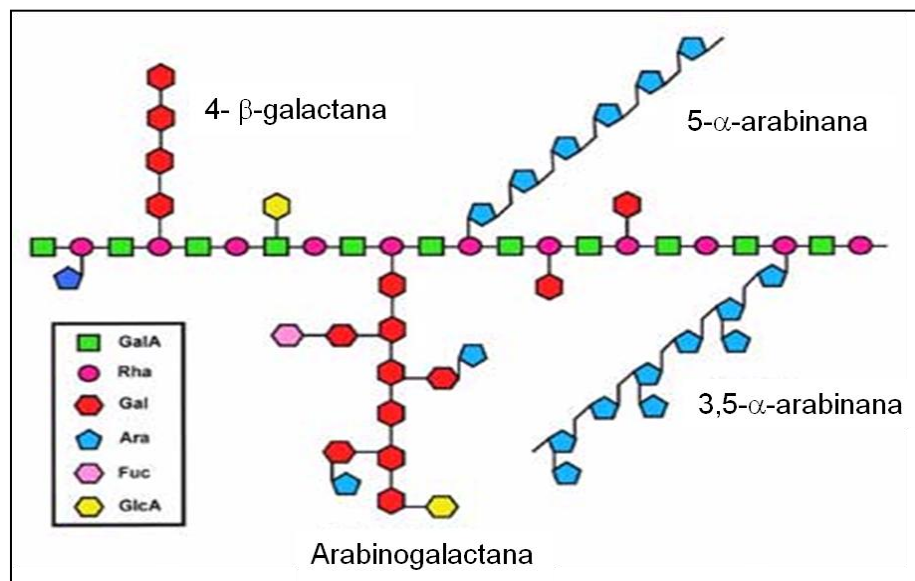


FIGURA 4.1 - Estrutura de ramnogalacturonana-I (<http://www.ccruc.uga.edu/~mao/rg1/rg1.htm>).

Ramnogalacturonana II (RG-II), diferente do que o nome propõe, é composta por uma cadeia principal de HG (somente resíduos de GalA). Os resíduos de ramnose estão presentes nas cadeias laterais da RG-II e em quantidade menor que na RG-I. Na RG-II observa-se a ocorrência de quatro diferentes cadeias laterais como apresentado na Figura 5.1 (VIDAL et al., 2000). As cadeias laterais contêm 12 diferentes resíduos, alguns dos quais sendo monossacarídeos raros tais como 2-O-metil-L-fucose, 2-O-metil-D-xilose, apiose, ácido acérico, Kdo e Dha. As RGs II estão presentes na parede primária de plantas como dímero covalentemente ligado a um éster de borato, que parece ter como função o auxílio na formação da rede péctica macromolecular da parede celular (KOBAYASHI et al., 1999).

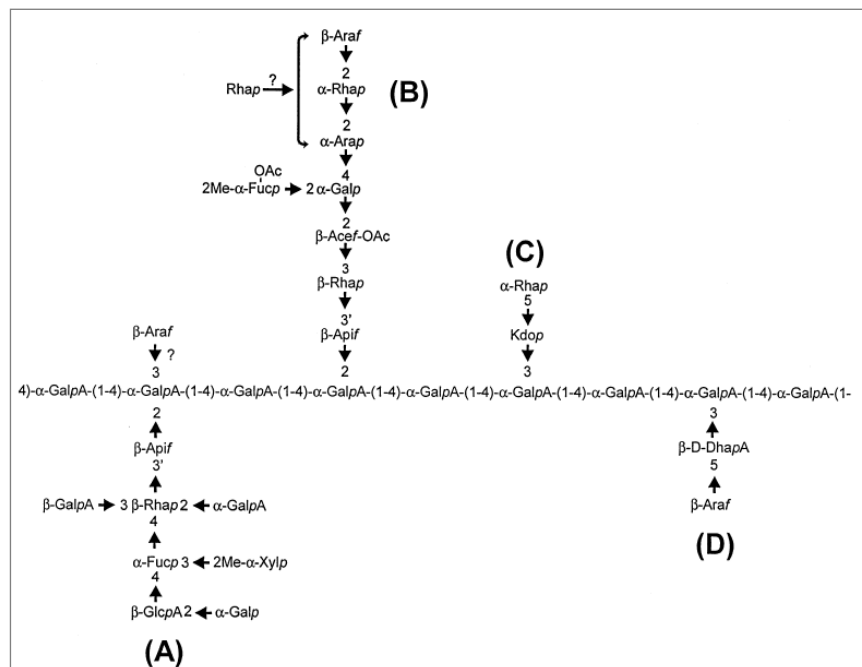


FIGURA 5.1 - Sequência glicosil de RG II (VIDAL et al., 2000). A cadeia principal de GalA é substituída com quatro diferentes cadeias laterais de oligossacarídeos (cadeias A-D). A localização das cadeias laterais é arbitrária uma vez que a distribuição das mesmas na cadeia principal não é conhecida.

Diversas funções têm sido atribuídas à matriz péctica durante a deposição e montagem da parede celular e também durante a expansão da célula. A modulação da pectina dentro das paredes das células pode refletir mudanças progressivas durante o desenvolvimento celular (WILLATS et al., 2001). A rede de pectinas influencia o pH e a força iônica da matriz e, devido a sua capacidade de formar gel, está também intimamente envolvida nas propriedades mecânicas e de porosidade da parede celular, que por sua vez interferem na hidratação e no movimento e acesso de proteínas que modificam a parede celular através da matriz

(WILLATS et al., 2001). Além de suas diferentes funções na fisiologia e crescimento celular, a pectina está na linha de frente de qualquer ataque contra células vegetais por patógenos que utilizam de enzimas para degradar a rede pectica de plantas.

## **1.2 - Enzimas que degradam parede celular de plantas**

A parede celular de plantas é composta basicamente de dois tipos de polissacarídeos: a rede de pectato e a rede de celulose. As enzimas que degradam pectina podem ser fatores que determinam a virulência do patógeno (ROGERS et al., 2000) e são chamadas de pectinases ou enzimas pectinolíticas. Estas enzimas degradam substâncias pecticas hidrolisando ligações glicosídicas ao longo da cadeia carbônica podendo ser classificadas em desesterificantes (pectina esterase) ou despolimerizantes (liases e hidrolases) e são produzidas por plantas, fungos filamentosos, bactérias e leveduras (UENOJO & PASTORE, 2007). Os diferentes tipos de pectinases serão brevemente descritos a seguir e seus modos de ação são ilustrados na Figura 6.1.

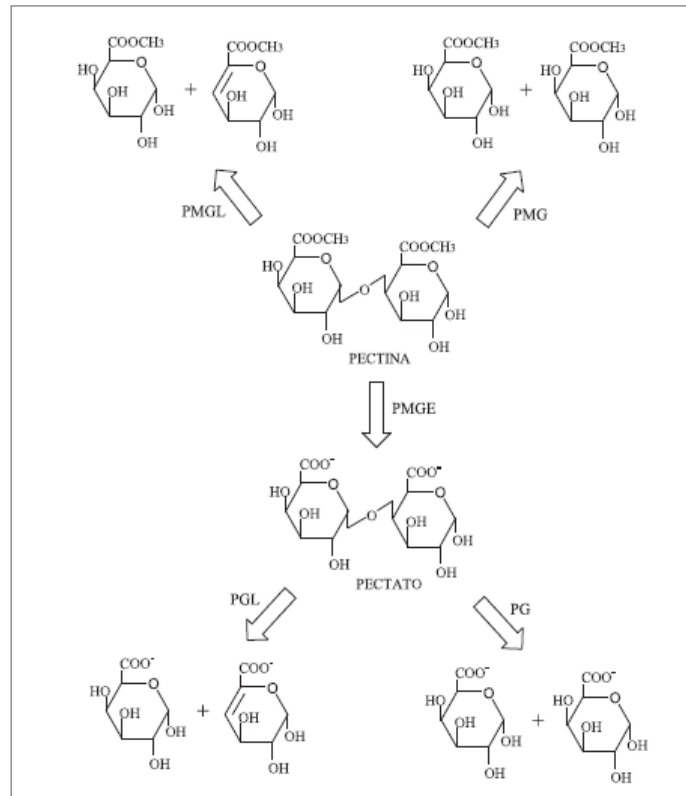


FIGURA 6.1 - Modo de ação das pectinases sobre a molécula de pectina e pectato (UENOJO & PASTORE, 2007). PMGL: polimetilgalacturonato liase. PMG: polimetilgalacturonase. PMGE: polimetilgalacturonato esterase (pectina esterase). PGL: poligalacturonato liase (pectato liase). PG: poligalacturonase.

### 1.2.1- Enzimas desesterificantes ou pectinas esterases

A pectina esterase (polimetilgalacturonato esterase, PMGE) remove grupos metil da pectina expondo grupos carboxil formando pectina ácida (pectato) que na presença de cátions bi ou multivalentes, como o cálcio, forma sais insolúveis (UNDERKOFLEER et al, 1958). A enzima atua preferencialmente em grupo metil-éster da unidade de galacturonato próxima a uma unidade de galacturonato não esterificada (KASHYAP et al., 2001).

### 1.2.2 – Enzimas despolimerizantes

As despolimerizantes são enzimas que clivam ligações  $\alpha$ -1,4 entre unidades de GalA das substâncias pécnicas e são classificadas em duas classes diferentes, de acordo com seus mecanismos de ação: mecanismo via  $\beta$ -eliminação (pela ação de polissacarídeo liases, PL) ou via hidrólise (pela ação de glicosídeo

hidrolases, GH). Uma ilustração das diferenças nos dois mecanismos possíveis para ação de enzimas despolimerizantes é mostrada na Figura 7.1, onde o mesmo substrato é clivado através de mecanismo de  $\beta$ -eliminação (pela enzima PL), resultando na formação de um resíduo glicosídeo insaturado pela clivagem eliminativa da ligação  $O_4-C_4$ , e através da ação de uma GH via hidrólise, que resulta na formação de dois resíduos saturados pela clivagem da ligação  $C_1'-O_4$  (LOMBARD et al., 2010).

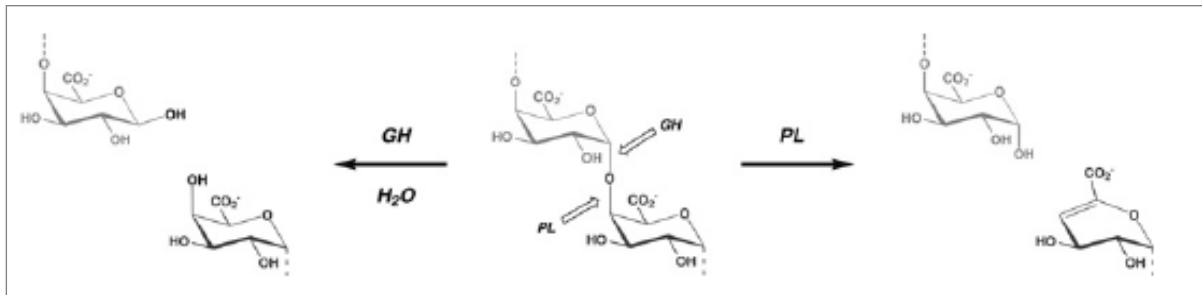


FIGURA 7.1 - Comparação entre os mecanismos de clivagem de ligação glicosídica pela ação de polissacarídeo liases, PL, via  $\beta$ -eliminação e pela ação de glicosídeo hidrolases, GH, via hidrólise (LOMBARD et al., 2010).

As polissacarídeos liases (EC 4.2.2.-), também chamadas transeliminases, incluem as pectina liases e as pectato liases. Como ilustrado na Figura 6.1, as pectina liases (polimetilgalacturonato liase, PMGL) catalisam a  $\beta$ -eliminação entre dois resíduos de GalA da pectina. Podem atuar pelo modo Endo- (catalisam o substrato randômicamente) ou Exo- (catalisam o substrato a partir do final não redutor do substrato).

As pectato liases (poligalacturonato liase, PGL) catalisam a clivagem de ligações  $\alpha$ -1,4 entre unidades de GalA por transeliminação e assim como a PMGL, podem atuar pelo modo Endo- ou Exo-. Diferente da PMGL, atuam preferencialmente sobre o pectato.

As glicosídeo hidrolases (EC 3.2.1.-) são classificadas em polimetilgalacturonases (PMG) e poligalacturonases (PG). Ambas podem ser classificadas em Endo- e Exo- de acordo com o modo de atuação sobre o substrato. PMG catalisa a clivagem de ligações  $\alpha$ -1,4 entre unidades de GalA por hidrólise liberando 6-metil-galacturonato. As Endo-PMG clivam randômicamente ligações  $\alpha$ -1,4 da pectina, preferencialmente aquela com alto grau de esterificação. Já as Exo-PMG clivam sequencialmente ligações  $\alpha$ -1,4 do final não reduzido da pectina (KASHYAP et al., 2001). A PG age sobre o ácido péctico liberando D-galacturonato. As Endo-PG clivam randômicamente ligações  $\alpha$ -1,4 do pectato ou

ácido péctico, enquanto que as Exo-PG clivam ligações  $\alpha$ - 1,4 de extremidades não redutoras do pectato.

Além das enzimas descritas, envolvidas na clivagem de ligações glicosídicas entre resíduos de GalA, outras enzimas são necessárias para a degradação completa da pectina. Estas incluem enzimas responsáveis pela catálise de reações de clivagem envolvendo resíduos de ramnose tanto na cadeia principal do polissacarídeo como clivagem das porções substituídas das cadeias laterais do polímero (MUTTER et al., 1994).

### **1.3 – Aplicações (bio)tecnológicas de enzimas pectinolíticas**

Embora o uso de enzimas para realização de algumas reações como fermentação de pães e fabricação de cerveja e vinho datam muitos séculos atrás em uma época em que não havia qualquer conhecimento sobre elas, somente na década de 1930 a primeira aplicação industrial foi observada. Dr. Jokichi Takamine (1854-1922) foi a primeira pessoa a perceber a possibilidade técnica de enzimas cultivadas e a apresentá-las para a indústria. Ele estava preocupado principalmente com enzimas fúngicas, enquanto BOIDIN & EFFRONT (1917), na França, foram os pioneiros na produção de enzimas bacterianas cerca de 20 anos mais tarde (UNDERKOFLEER et al, 1958). A partir dos anos 60 a possibilidade de uma melhor visualização dos tecidos vegetais somada a um maior conhecimento da natureza química dos mesmos o uso das enzimas tornou-se mais eficiente e diversificado (KASHYAP et al, 2001; LEISOLA et al, 2002). Atualmente, as enzimas são parte integrada da indústria alimentícia e têxtil além de apresentarem outras aplicações biotecnológicas (LANG & DÖRNENBURG, 2000; TURNER et al., 2007; FERNANDES, 2010; MALLER et al 2010; KUHAD et al. 2011). Em 1995 a estimativa de vendas de enzimas foi de 1 bilhão de dólares, sendo 75 milhões só de pectinases (KASHYAP et al, 2001). Nos EUA a demanda de enzimas em 2009 chegou a 2,2 bilhões de dólares e espera-se um aumento na demanda mundial de 6,3% ao ano até 2013 (FREEDONIA GROUP, 2010). O crescimento observado nos últimos anos está relacionado aos avanços em enzimas específicas que são aplicadas em pesquisas, no ramo farmacêutico e biotecnológico, bem como no mercado de biocombustíveis emergentes (FREEDONIA GROUP, 2010).

Algumas das aplicações das pectinases nas indústrias alimentícias

incluem amadurecimento de frutas, clarificação e redução de viscosidade em sucos de frutas, tratamento preliminar do suco de uva para indústrias vinícolas, extração de polpa de tomate, fermentação de chá e café, tratamento de resíduos vegetais, nutrição animal, enriquecimento proteico de alimentos infantis e extração de óleos. Além de serem aplicadas na degomagem de fibras nas indústrias têxteis e de papel, (UENOJO & PASTORE, 2007).

As substâncias pécticas são responsáveis pela consistência, turbidez e aparência dos sucos de frutas. Durante o processamento do suco é possível observar duas fases distintas: a solúvel ou líquida (suco) que contém pectinas solúveis responsáveis pela turbidez do suco, e a insolúvel chamada de polpa, onde estão presentes as substâncias pécticas insolúveis que, por sua vez, dificultam a extração do suco por prensagem ou por outros métodos mecânicos. Além disso, tais substâncias causam aumento na viscosidade que acaba por atrapalhar o processo de filtração e conseqüentemente o de concentração do suco. A adição de enzimas pécticas à polpa causa degradação da rede de pectina, tornando a extração do suco mais fácil e conseqüentemente aumenta o rendimento do mesmo. Já no suco, a adição de pectinases é realizada para remover substâncias pécticas suspensas promovendo sua clarificação.

No caso de processamentos de alimentos, as preparações enzimáticas são quase exclusivamente derivadas de *Aspergillus sp.* e são tradicionalmente misturas de PG, PL e PME (LANG & DÖRNENBURG, 2000). No caso do uso de pectinases como suplemento de alimentos infantis a fim de melhorar a nutrição e o desenvolvimento físico infantil, a indústria prefere utilizar preparados enzimáticos contendo atividade de PG, pois este tratamento mantém as células intactas conservando as vitaminas, cor e aroma do alimento. Isto é conseguido, através da purificação da PG do *Aspergillus* ou usando cepas que sintetizam apenas PG (LANG & DÖRNENBURG, 2000).

Pectinases também desempenham um papel importante na fermentação de chás e café. Na fermentação do café são utilizados microrganismos pectinolíticos para a remoção da camada de mucilagem presente nos grãos. Celulases e hemicelulases também ajudam nessa tarefa. Na fermentação do chá as pectinases aceleram tal processo (KASHYAP et al., 2001; UENOJO & PASTORE, 2007).

Na fabricação de papel as pectinases são usadas para despolimerizar

substâncias pécticas e, conseqüentemente, diminuir a demanda catiônica das soluções pécticas e do filtrado resultantes do branqueamento com peróxido (THORNTON, 1994; THORNTON et al, 1996; RICARD et al, 2005).

Na indústria têxtil a aplicação de pectinases resulta na degradação da camada de pectina que recobre as fibras de celulose e hemicelulose, liberando-as para posterior processamento (MARTINS, 2006). No processo de degomagem de fibras de rami, juta, linho, cânhamo, kenaff (*Hibiscus sativa*), e fibra de casca de coco seu uso também se faz presente (CHESSON, 1980; BRUHLMANN et al., 1994).

## 1.4 – Bioinformática

Em 1869 Friedrich Miescher descobriu um tipo de ácido fraco, abundante nos núcleos das células brancas do sangue, uma substância química que hoje chamamos de DNA (ácido desoxirribonucleico) (HARTL & JONES, 2004). Durante muitos anos a função biológica do DNA ficou desconhecida e após muitas pesquisas envolvendo cientistas de diferentes áreas chegou-se a conclusão que o DNA era a molécula que armazenava a informação genética. Em 1953 James Watson e Francis Crick, na Universidade de Cambridge, propuseram a primeira estrutura tridimensional essencialmente correta da molécula de DNA. Posteriormente, F. Sanger e, independentemente, A. Maxam e W. Gilbert, desenvolveram métodos para o sequenciamento de DNA (LESK, 2008).

Em 1990 foi lançado o Projeto Internacional Genoma Humano (Projeto Genoma) tendo como principal objetivo o uso das informações obtidas para estudo da biologia e da medicina com uma previsão de 15 anos para ser completado. Na mesma década começaram a surgir sequenciadores automáticos de DNA aumentando consideravelmente a quantidade de sequências genéticas a serem armazenadas exigindo conseqüentemente cada vez mais recursos computacionais (PROSDOCIMI et al, 2002). A enorme quantidade de informações gerada todos os dias, principalmente pelos projetos genomas, precisava ser cuidadosamente decifrada, organizada e analisada para poder trazer informações interessantes a respeito dos mecanismos biológicos. A computação contribuiu muito neste aspecto, não só com os meios de armazenamentos rápidos e de alta capacidade, como também com o desenvolvimento de programas simples ou mesmo sofisticados, para



análise dessas informações. Essa união entre a biologia e ciência da computação criou um novo campo da ciência chamado bioinformática (LESK, 2008).

O foco da bioinformática é o desenvolvimento e aplicação de técnicas intensivas em computação, por exemplo, reconhecimento de padrões, mineração de dados -*data mining*-, algoritmos e visualização. Algumas pesquisas neste ramo da ciência incluem alinhamento de sequências, busca de genes, montagem de genoma, desenho de fármacos, alinhamento de estrutura proteica, predição de estruturas de proteínas, previsão da expressão gênica e interação proteína-proteína. Portanto, a bioinformática agrega conhecimento em diferentes áreas tais como biologia, computação, matemática, física e medicina (BAYAT, 2002).

## CAPÍTULO 2 - OBJETIVOS

A proposta deste trabalho foi utilizar a bioinformática para o estudo de poligalacturonases de fungos sob dois aspectos:

1- Realizar o estudo de poligalacturonases de fungos que atuam via mecanismo hidrolítico (as glicosídeo hidrolases - EC 3.2.1.-, PG e PMG) utilizando programas computacionais apropriados, bancos de dados de sequência de proteína e de estruturas proteicas, além de construção de árvores filogenética a fim de estudar a enzima no que diz respeito a sua classificação, busca de padrões sequenciais filogenéticos e análise das estruturas tridimensionais determinadas no que diz respeito à busca por padrões estruturais conservados.

2- Construir o modelo estrutural tridimensional de uma poligalacturonase (PG) do fungo *Leucoagaricus gongylophorus*, simbiote de formigas cortadeiras.



## CAPÍTULO 3 - FERRAMENTAS DE BIOINFORMÁTICA

### 3.1 – Bancos de dados

Um banco de dados é um conjunto de dados que relacionam entre si e são criados a fim de organizar e estruturar as informações de modo a facilitar consultas, atualizações e deleções de dados. Na área da bioinformática os bancos de dados permitem integrar e consultar dados de sequência de ácidos nucleicos (DNA) ou de aminoácidos (proteína), estruturas de proteínas, padrões de expressão de genes, genomas completos, vias metabólicas e publicações.

Os bancos de dados envolvendo sequência de ácidos nucleicos, de aminoácidos ou estruturas de proteínas podem ser classificados em bancos de sequências primários ou secundários. Os bancos primários são derivados diretamente dos dados obtidos a partir do sequenciamento de ácidos nucleicos, aminoácidos ou de estruturas proteicas, sem qualquer processamento ou análise dos mesmos. Os principais bancos primários são o *GenBank*, o EBI (*European Bioinformatics Institute*), o DDBJ (*DNA Data Bank of Japan*) e o PDB (*Protein Data Bank*). Vale ressaltar que os três primeiros bancos de dados trocam informações diariamente entre si, e dessa forma todos eles possuem informações atualizadas de todas as sequências depositadas mundialmente. Normalmente, as sequências identificadas pelos laboratórios são submetidas aos bancos de dados primários antes da publicação do artigo científico. Uma vez submetidas, as sequências ganham um número de acesso que deve ser citado no manuscrito do artigo e que servirá como um sinal de liberação dessa informação para consulta pública (PROSDOCIMI et al, 2002). Os bancos de dados secundários, como GDB (banco de dados oficial do projeto genoma humano), PIR (*Protein Information Resource*) ou SWISS-PROT, PROSITE são derivados dos bancos primários e apresentam dados tratados, como por exemplo, contém dados não redundantes.

Os principais *websites* que contém banco de dados são:

- *National Center for Biotechnology Information* ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) - mantém ferramentas de bioinformática e bancos de dados.

- *National Center for Genome Resources* ([www.ncgr.org/](http://www.ncgr.org/)) - conecta cientistas a soluções bioinformáticas por colaborações, dados e desenvolvimento de softwares.
- *Genbank* ([www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)) - armazena sequências de DNA de projetos genoma de pequenos e grandes laboratórios.
- *Ensembl* ([www.ensembl.org](http://www.ensembl.org)) – apresenta anotação automática de dados de genoma.
- *Unigene* ([www.ncbi.nlm.nih.gov/UniGene](http://www.ncbi.nlm.nih.gov/UniGene)) – apresenta sequência de genes contendo dados sobre a localização de genes nos cromossomas.
- *European Bioinformatic Institute* ([www.ebi.ac.uk](http://www.ebi.ac.uk)) - centro de pesquisa e serviços em bioinformática; gerenciamento de bancos de dados de dados biológicos.
- *SWISS-PROT* ([www.expasy.org/sprot](http://www.expasy.org/sprot)) - Importante banco de dados de proteína com dados de sequência de todos os organismos, com alto nível de anotação (inclui função, estrutura) e é minimamente redundante (poucos dados duplicados).
- *ExPASy (Expert Protein Analysis System)* servidor de proteômica do *Swiss Institute of Bioinformatics* (SIB) - Dedicado a análise de sequências e estruturas de proteínas e gel bidimensional (2-D PAGE).

O estudo realizado nesta dissertação de mestrado, sobre as PG de fungos foi iniciado com a busca em bancos de dados de sequência de aminoácidos de todas as PG de fungos depositadas no *site* do **National Center for Biotechnology Information** (NCBI). A palavra de busca para sequências de PG de fungos no NCBI foi *polygalacturonase fungi*.

### 3.2 – Alinhamentos de sequências

Alinhamento de sequências é um método de comparação que tem por finalidade determinar o grau de similaridade entre sequências. Vale lembrar que similaridade e homologia são conceitos diferentes. Similaridade é a observação ou mensura de semelhança e diferença, independente da origem da semelhança. A similaridade de sequências é visível nos dados coletados e não envolve hipóteses históricas, enquanto que homologia significa que as sequências e os organismos

nos quais ocorrem descendem de um ancestral comum, com a implicação de que as similaridades são características ancestrais compartilhadas (LESK, 2008). A homologia envolve eventos históricos os quais na sua maioria não são visíveis. Assim, duas sequências serão homólogas somente se derivarem de um ancestral comum.

Na bioinformática o alinhamento de sequências é a ferramenta básica, pois permite estabelecer identidades entre sequências, deduzir função proteica, definir domínios proteicos conservados, além de possibilitar estudo da evolução de proteínas. Assim, alinhar uma sequência compreende a identificação de correspondências entre pares de resíduos de sequências primárias de DNA ou proteínas (Figura 8.3) a fim de identificar regiões similares que possam ser consequência de relações funcionais ou mesmo estruturais entre elas. Lacunas (*gaps*) podem ser introduzidas com a finalidade de obter o melhor alinhamento possível.

Na Figura 8.3 é possível observar sinais como (\*), (:), e (.) logo abaixo das sequências alinhadas. O (\*) sob a sequência indica uma posição conservada, ou seja, o resíduo é idêntico em ambas as sequências. Já os símbolos (:) ou (.) indicam posições em que as sequências contêm resíduos de características físico-químicas muito similares (:), ou com pouca similaridade (.). Quando não há símbolo algum, significa troca de resíduo com características totalmente diferentes.

A Figura 8.3 ilustra um alinhamento chamado 'par a par' e trata-se de um alinhamento entre apenas duas sequências. Quando ocorre um alinhamento simultâneo de mais de duas sequências o processo é chamado alinhamento múltiplo de sequência, como exemplificado para o alinhamento de seis sequências de Endo-PG de *Alternaria tenuissima* (Figura 9.3).

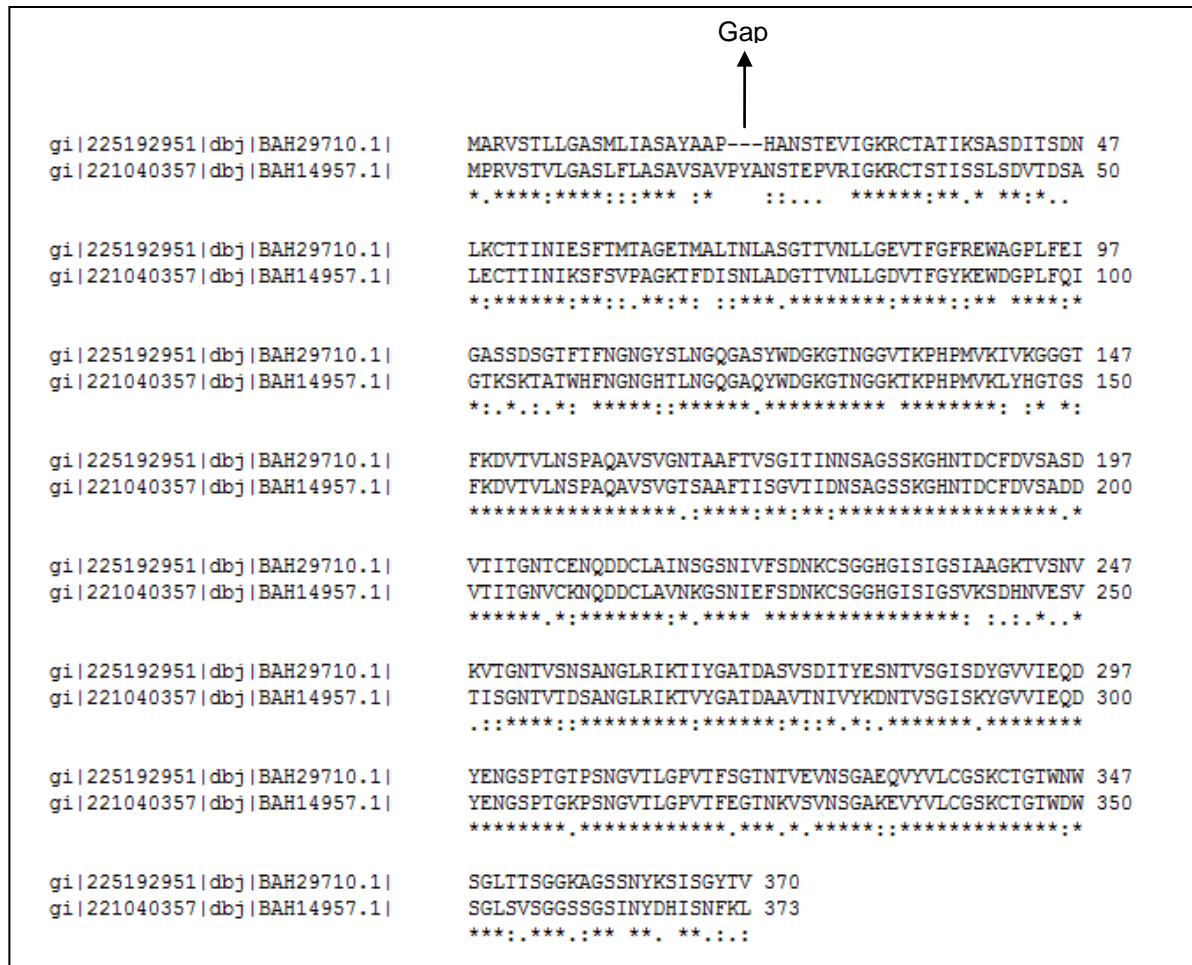


FIGURA 8.3 - Exemplo de alinhamento entre duas sequências de aminoácidos de Endo-polygalacturanase (Endo-PG) de *Athelia rolfsii*. O alinhamento foi obtido com o uso do programa *Clustal-W* (THOMPSON et al., 1994).

```

gi|154433925|gb|ABS82361.1|      SDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVK 50
gi|154433895|gb|ABS82346.1|      SDVAVPSGTTLDLSSLADGTTVVFEFEGTTTWGYSEWKGPLLDIQGKKITVK 50
gi|154433983|gb|ABS82390.1|      SDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVK 50
gi|154433819|gb|ABS82308.1|      SDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVK 50
gi|154433795|gb|ABS82296.1|      SDVAVPSGTTLDLSSLADGTTVIFEGTTTWGYSEWKGPLLDIQGKKITVK 50
gi|154433755|gb|ABS82276.1|      SDVAVPSGTTLDLSSLADDTTVIFEGTTTWGYSEWKGPLLDIRGKKITVK 50
*****.***:*****:*****:*****

gi|154433925|gb|ABS82361.1|      GAEGSVLNGDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPV 100
gi|154433895|gb|ABS82346.1|      GAEGSVLNGDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPV 100
gi|154433983|gb|ABS82390.1|      GAEGSVLNGDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPV 100
gi|154433819|gb|ABS82308.1|      GAEGSVLNGDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPV 100
gi|154433795|gb|ABS82296.1|      GAEGSVLNGDGARWWDGKGGNGGKTKPKFFSAHKLTDSTITGITIKNPPV 100
gi|154433755|gb|ABS82276.1|      GAEGSVLNGDGARWWDGKGGNGGKTKPKFFSAHKLTDSSITGITIKNPPV 100
*****:*****:*****:*****

gi|154433925|gb|ABS82361.1|      QVVSINGCDGLTITDMTIDASDGDKEEQGHNTDGFDIGSSNNVIIDGAKV 150
gi|154433895|gb|ABS82346.1|      QVVSINGCDGLTITDMTIDASDGDKEEQGHNTDGFDIGSSNNVIIDGAKV 150
gi|154433983|gb|ABS82390.1|      QVVSINGCDGLTITDMTIDASDGDKEEQGHNTDGFDIGSSNNVIIDGAKV 150
gi|154433819|gb|ABS82308.1|      QVVSINGCDGLTITDMTIDASDGDKEEQGHNTDGFDIGSSNNVIIDGAKV 150
gi|154433795|gb|ABS82296.1|      QVVSINGCDGLTITDMTIDASDGDKEEQGHNTDGFDIGSSNNVIIDGAKV 150
gi|154433755|gb|ABS82276.1|      QVVSINGCDGLTITDMTIDASDGDKEEQGHNTDGFDIGSSNNVIIDGAKV 150
*****:*****:*****:*****

gi|154433925|gb|ABS82361.1|      YN 152
gi|154433895|gb|ABS82346.1|      YN 152
gi|154433983|gb|ABS82390.1|      YN 152
gi|154433819|gb|ABS82308.1|      YN 152
gi|154433795|gb|ABS82296.1|      YN 152
gi|154433755|gb|ABS82276.1|      YN 152
**

```

FIGURA 9.3 - Exemplo de alinhamento múltiplo de sequência. Foram alinhadas 6 sequências de Endo-PG de *A. tenuissima* pelo *ClustalW* (THOMPSON et al., 1994).

### 3.2.1 – Alinhamento global e local

Os alinhamentos podem ser classificados basicamente em dois tipos: o global e o local. No alinhamento global todos os resíduos de uma sequência são alinhados com todos os resíduos da outra sequência. Esse tipo de alinhamento é mais comumente usado para comparar sequências cujas semelhanças sejam esperadas em toda a extensão da sequência. Programas computacionais que usam esse tipo de alinhamento são *ClustalW* (THOMPSON et al., 1994) e *Multalin* (CORPET, 1988).

Já o alinhamento local procura encontrar uma região, geralmente conservada, em uma determinada sequência que se alinhe a uma região de outra sequência. Esse alinhamento torna-se útil para realizar pesquisa em bancos de dados ou mesmo em situações onde não há conhecimento sobre a similaridade entre as sequências que se almeja comparar. Um dos programas mais conhecidos e usados para busca de sequências similares, o *BLAST* (**B**asic **L**ocal **A**lignment



*Search Tool*) é um exemplo de programa que utiliza esse tipo de alinhamento. A Figura 10.3 ilustra exemplos de alinhamento global e local

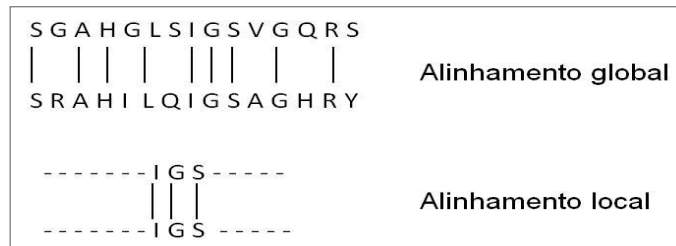


FIGURA 10.3 - Alinhamento global e local entre duas sequências.

### 3.2.2 – Matrizes de substituição

Ao fazer um estudo de similaridade entre sequências, além do alinhamento é preciso definir medidas quantitativas de similaridade e de diferenças entre as sequências. Esta medida é importante e leva em consideração que certas alterações têm maior probabilidade de ocorrer que outras, como na substituição de aminoácidos que tendem a ser conservativos: a substituição de um aminoácido por outro de tamanho ou propriedades físico-químicas similares é mais provável de ter ocorrido do que a substituição por outro com propriedades muito diferentes. Logo, um sistema de pontuação (de score) que leve em conta não só as substituições de resíduos como também as inserções e deleções se faz necessário.

As matrizes de substituição são normalmente utilizadas no alinhamento de sequências proteicas e analisam com qual frequência as diferentes substituições de aminoácidos ocorrem. Com bases nestas frequências, juntamente com a quantidade relativa de cada aminoácido na proteína, é possível atribuir um score que reflete a probabilidade daquela mutação ocorrer, sendo as substituições mais prováveis expressa por score positivo.

Os dois tipos mais utilizados de matrizes de substituição são a PAM 250 (*Point Accepted Mutation*) (DAYHOFF et al., 1978) e a Blossum 62 (*Blocks Substitution Matrix*) (HENIKOFF & HENIKOFF, 1992), ilustradas nas figuras 11.3 e 12.3, respectivamente.



Em 1992 Steven e Jorja Henikoff desenvolveram a família de matrizes BLOSUM com o objetivo de substituir a matriz de Dayhoff por outra com um melhor desempenho na identificação de relações distantes entre sequências. A matriz BLOSUM foi obtida a partir de um enorme volume de dados pertencentes ao banco de dados *BLOCKS*, e por isso o nome **BLO**cks **SU**bstitution **M**atrix (matriz de substituição baseada em *BLOCK*). Para evitar a presença de sequências muito semelhantes na comparação, as sequências eram agrupadas em blocos de acordo com o grau de identidade, além disso, era atribuído aos blocos o mesmo peso independente da quantidade de sequências que o compunha. Na matriz BLOSUM 62, por exemplo, a pontuação de substituição foi obtida a partir dos alinhamentos de sequências que não tiveram a identidade maior que 62%. Essa matriz é oferecida por todos os programas como uma opção, sendo a escolha-padrão na maioria, por exemplo, no programa *ClustalW* (THOMPSON et al., 1994).

### 3.2.3 – Programa *ClustalW*

Alinhamento múltiplos de sequências é uma ferramenta essencial na biologia molecular, sendo usada para diversos estudos como a análise evolutiva molecular como também na busca por padrões de diagnóstico para caracterizar as famílias de proteínas; detectar ou demonstrar homologia entre novas sequências e famílias de sequências existentes; ajudar a prever as estruturas secundárias e terciárias de novas sequências e sugerir *primers* para PCR- Reação em Cadeia da Polimerase (THOMPSON et al., 1994).

O *Clustal* é um programa amplamente usado para alinhamentos múltiplos de sequências, sendo que o site EBI/Clustal recebe por ano milhões de trabalhos de alinhamentos múltiplos (LARKI et al., 2007). Além de possuir a vantagem de ser rápido, fácil de utilizar e resultar em alinhamentos de qualidade suficiente sem a necessidade de ajustes frequentes, o programa está livremente disponível para diversos sistemas operacionais em duas variações principais: o *ClustalW* e o *ClustalX*. No presente trabalho utilizou-se a primeira variação por sua disponibilização *on-line*.

O programa *ClustalW* (THOMPSON et al., 1994) aceita diversos formatos de arquivos de entrada como NBRF/PIR, FASTA, EMBL/Swiss-Prot, Clustal, GCC/MSF, GCG9 RSF, e GDE. Quanto aos formatos dos arquivos de saída

podem ser Clustal, NBRF/PIR, GCG/MSF, PHYLIP, GDE, ou NEXUS. O formato FASTA (Figura 13.3), adotado no presente trabalho, é iniciado pelo sinal > seguido pelo título ou identificação da sequência (linha de descrição) e as linhas subsequentes compõem a sequência, um caractere por resíduo, ou seja, utiliza-se o código de uma letra para nucleotídeos ou aminoácidos especificados pela União Internacional de Bioquímica e Biologia Molecular (IUBMB) e pela União Internacional de Química Pura e Aplicada (IUPAC). Para os alinhamentos resultantes o formato Clustal foi preferido como ilustrado nas Figuras 8.3 e 9.3.

```
>gi|2383|emb|CAA41693.1| polygalacturonase [Aspergillus niger]
MHSYQLLGLAAVGLVSAAPAPSRVSEFAKKASTCTFTSASEASESISCSDEVVLSIEVPAGETLDLSD
AADGSTITFEGTTSFGYKEWKGPLIRFGGKDLVTMADGAVIDGDGSRWWDSKGTNGGKTKPKFMYIHDV
EDSTFKGINIKNTPVQAISVQATNVHLNDFITDINSDDGDDNGGHNTDGFDISESTGVYISGATVKNQDDCI
AINSGESISFTGGTCSGGHGLSIGSVGGRDDNTVKNVTISDSTVNSANGVRIKTIYKETGDVSEITYSN
IQLSGITDYGIVIEQDYENGSPSTGIPITDVTVDGVTGTLEDDATQVYILCGDGS DWTWSGVDL
SGGKTSKDCENVPSGASC
```

FIGURA 13.3 - Exemplo de uma sequência no formato FASTA. Sequência de uma poligalacturonase de *Aspergillus niger*. A primeira linha é referente à identificação da sequência e as demais linhas se referem à sequência.

Basicamente, o programa *ClustalW* é composto por três etapas principais:

- Alinhamento em pares de todas as sequências com a finalidade de determinar a similaridade entre elas. Nessa etapa foi utilizada a matriz de substituição BLOSUM 62. Além da escolha da matriz de substituição é possível escolher duas penalidades para *gap* (uma para introdução de lacunas e outra para extensão de lacunas); as penalidades usadas foram as dos padrões fornecidos pelo próprio programa (*default*).
- Construção de árvore filogenética, a fim de refletir as semelhanças entre as sequências, pelo método *Neighbour-Joining* (NJ). Este método produz árvores não enraizadas com comprimentos de ramos proporcionais à divergência estimada ao longo de cada ramo. O método começa com uma árvore em forma de estrela e o próximo passo é separar o par de sequência mais próxima das demais a partir dos valores de uma matriz de distância ou, no caso, de substituição. Este procedimento é repetido várias

vezes até que todos os ramos sejam encontrados.

- Uso da árvore filogenética para a realização do alinhamento múltiplo. Nesta etapa a ordem de alinhamento das sequências será a mesma da árvore. Assim os alinhamentos serão combinados começando pelos grupos que apresentam maiores semelhanças para os mais dessemelhantes.

### **3.3 – Filogenia Molecular**

A publicação do livro "A Origem das Espécies" por Charles Darwin (1859), despertou e acirrou o interesse dos cientistas em entender a respeito da evolução dos organismos (ramo da ciência chamado de filogenética) e expressá-lo de forma explicativa. Uma das maneiras de fazer isso é usar as informações contidas em fósseis, mas na maioria das vezes estas se encontram fragmentadas e incompletas. Outra maneira é realizar comparações de características morfológicas e fisiológicas. Mas esses métodos, pertencentes às técnicas evolucionárias clássicas, apresentam limitações já que por si só as mudanças morfológicas e fisiológicas são complexas e não retratam com clareza a história evolucionária (NEI & KUMAR, 2000).

Todavia, avanços na biologia molecular acarretaram em mudanças no campo da filogenia. Um desses avanços foi a descoberta do DNA, que por conter instruções genéticas capazes de coordenar o desenvolvimento de todos os seres vivo e de alguns vírus (enquanto outros possuem apenas RNA – ácido ribonucléico), logo passaram a ser utilizados em estudos filogenéticos apresentando algumas vantagens sobre as abordagens clássicas como a possibilidade de comparar DNAs de diversos organismos (bactérias, fungos, plantas e animais), uma vez que a composição do DNA é a mesma para todos os organismos, sendo a diferença advinda das sequências de nucleotídeos. Além disso, as mudanças evolucionárias do DNA seguem mais ou menos um padrão, já que não sofrem a ação do ambiente, como no caso das características morfológicas, e seus mecanismos de modificação são bem conhecidos pela genética molecular (replicação do DNA, mutações e mecanismos de reparo). Dessa forma é possível desenvolver e usar modelos matemáticos para formular mudanças e comparar DNAs de organismos com

relações distantes.

Segundo LESK (2008) um evento crucial para o aceite dos métodos moleculares ocorreu em 1967 quando V. M. Sarich e A. C. Wilson através de dados imunológicos dataram o tempo de divergência entre humanos e chimpanzés em 5 milhões de anos atrás, enquanto que naquela época os paleontólogos datavam 15 milhões de anos atrás e estavam relutantes em aceitar o método molecular. Passado algum tempo, o registro fóssil foi reinterpretado levando à aceitação da data e também do método.

Atualmente não somente as sequências de nucleotídeos fornecem a melhor medida de similaridade entre espécies para análises filogenéticas como também as sequências de aminoácidos.

### 3.3.1 – Árvores Filogenéticas

Geralmente as relações filogenéticas são representadas em forma de árvore (árvore filogenética). Estas são confeccionadas a partir de uma matriz contendo os dados disponíveis (morfológicos, químicos, genéticos ou proteicos) sobre os táxons estudados. Os dados são comparados, e os táxons são agrupados em clados ou ramos de acordo com as semelhanças e diferenças entre si. As árvores podem ser representadas na forma enraizada (Figura 14.3) ou não enraizada (Figura 15.3). Ambas demonstram a topologia da relação, porém na primeira o ancestral comum é observável, enquanto que na segunda não.

Na figura 14.3 é mostrado um exemplo de árvore filogenética. As folhas são representadas pelas letras (A-E) também conhecidas como táxons ou Unidade Taxonômica Operacional - OTUs (do inglês *Operational Taxonomic Units*). Os nós externos correspondem a entidades biológicas (espécies, populações, indivíduos ou genes), enquanto que os nós internos correspondem aos possíveis ancestrais. Os ramos expressam as relações evolutivas.

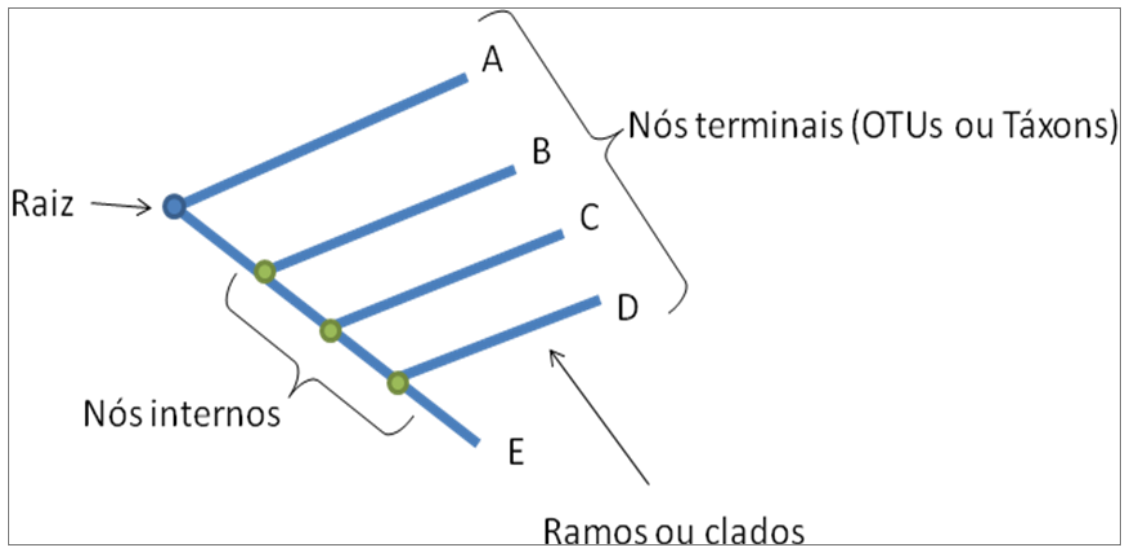


FIGURA 14.3 - Representação de árvore filogenética com raiz e suas principais terminologias.

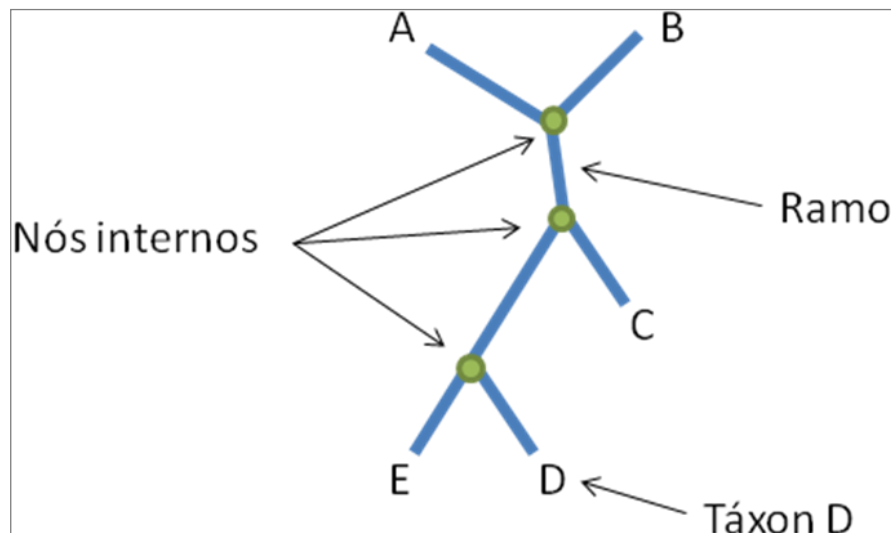


FIGURA 15.3 - Representação de árvore filogenética sem raiz.

As árvores filogenéticas podem ser de 3 tipos:

- Cladograma – representa apenas a topologia da árvore, ou seja, como os táxons estão relacionados entre si.
- Filograma – representa a árvore onde o comprimento dos ramos é proporcional ao número de mudanças aferidas.
- Dendograma - representa a árvore onde o comprimento dos ramos reflete o tempo transcorrido.

De forma geral, há três métodos para a derivação de árvores filogenéticas: os fenéticos, os cladísticos e os probabilísticos. Os métodos fenéticos são aqueles que não fazem referência a qualquer modelo histórico das relações, ou

seja, não são baseados em modelos evolutivos. As árvores são construídas de acordo com a similaridade entre as sequências, baseando-se nos métodos de distância, onde para as diferenças existentes entre duas sequências são atribuídos valores. São exemplos desse método o UPGMA (*Unweighted Pair Group Method using arithmetic Averages*) e o *Neighbor-Joining* (junção de vizinhos).

Os métodos cladísticos são aqueles que se baseiam em modelos evolutivos, ou seja, em modelos matemáticos que tentam representar os processos de substituição de bases nucleotídicas ou de aminoácidos entre diferentes espécies (LIÒ & GOLDMAN, 1998). O método cladístico mais popular é o da máxima parsimonia que adota como árvore filogenética ótima aquela que pressupõe o menor número de mutações.

Os métodos probabilísticos também são baseados em modelos evolutivos, no entanto eles usam cálculos probabilísticos para encontrar a árvore que melhor explique as variações observadas num determinado conjunto de sequência. A *máxima verossimilhança* e a inferência *Bayesiana* são exemplos desses métodos. Como no presente trabalho foi utilizado apenas o método da *máxima verossimilhança*, este será mais detalhado a seguir.

### **3.3.2 - Reconstrução filogenética: método da *máxima verossimilhança***

O modelo evolutivo conhecido como *máxima verossimilhança* (ML – do inglês *Maximum Likelihood*) foi primeiramente descrito pelo estatístico e também biólogo evolutivo R. A. Fisher (FISHER, 1922). Mais tarde, CAVALLI-SFORZA e EDWARDS (1967) tiveram a ideia de utilizá-lo para inferência filogenética. No entanto, o método exigia grandes recursos computacionais que na época não estavam disponíveis.

Com o desenvolvimento de computadores mais potentes somadas a uma formulação matemática desenvolvida por FELSENSTEIN (1981), que reduziu bastante o número de operações algébricas necessárias, o método passou a ser extensivamente utilizado a partir dos anos 80. Porém, a sua complexidade computacional ainda impedia a análise para um grande número de sequências (FELSENSTEIN, 1993; STRIMMER & HAESELER, 1996), uma vez que o número de árvores possíveis cresce exponencialmente de acordo com o número de



sequências. Dessa forma, meios de otimização para acelerar os cálculos ou o uso de técnicas aproximadas passaram a ser utilizados (ADACHI & HASEGAWA 1994; OLSEN et al 1994; ANISIMOVA & GASCUEL, 2006). Atualmente há um grande número de programas computacionais práticos para reconstrução de árvores filogenéticas que utilizam o método ML como o PAML (YANG, 1997), TREE-PUZZLE (SCHMIDT et al, 2002), PHYML (GUINDON et al, 2005) e Phylogeny.fr (DEREEPER et al., 2008). Sumários de ferramentas filogenéticas, incluindo listas úteis de sites na *web* e uma listagem geral de programas para filogenia pode ser consultada em: (<http://evolution.genetics.washington.edu/phylip/software.html>).

As principais vantagens de utilização do método ML são sua capacidade de usar todos os dados disponíveis (LIÒ & GOLDMAN, 1998) e ser um método que trabalha diretamente com os caracteres biológicos observados e não com uma estimativa de distância entre eles como no método NJ- *Neighbor-Joining* (SILVA, 2007).

O principal objetivo do algoritmo ML é gerar uma árvore filogenética com comprimentos de ramos e topologia que correspondem aos dados fornecidos. Essa tarefa, segundo STRIMMER & HAESLER (1996), pode ser dividida em duas partes: determinar a topologia da árvore e em seguida atribuir os comprimentos dos ramos a essa topologia.

### **3.4 – Busca por motivos sequenciais – Programa MEME**

Um motivo de sequência de proteína é uma sequência padrão de aminoácidos que ocorre repetidamente em um grupo de proteínas relacionadas e está quase sempre relacionada a uma função bioquímica específica (BORK & KOONIN, 1996). A ocorrência de motivos sequenciais geralmente implica em uma origem comum (homologia) para um dado grupo de proteínas. O fato dos motivos sequenciais serem como 'assinaturas' de famílias de proteínas, frequentemente eles são usados como ferramentas para predição de função de proteínas.

Diversos programas computacionais têm sido desenvolvidos com o intuito de identificar motivos sequenciais (principalmente nos novos genomas sequenciados) e um desses programas é o MEME (*Multiple EM For Motif Elicitation*) (BAILEY & ELKAN, 1994). O arquivo de entrada para o programa, uma sequência de aminoácido ou DNA, pode ter diferentes formatos tais como IG/Stanford,

GenBank/GB, NBRF, EMBL e outros. No modo padrão (*default*) o MEME procura por até três motivos na sequência analisada sendo que o comprimento do motivo procurado fica entre 6 e 50 caracteres no modo *default*, mas o usuário pode mudar este assim como vários outros aspectos da busca de motivos (BAILEY et al, 2006).

### 3.5 – Modelagem molecular

O conhecimento da estrutura terciária de uma proteína traz informação valiosa a respeito da sua função, por permitir a identificação de domínios conhecidos, como por exemplo de sítios catalíticos. Além disso, o conhecimento da estrutura possibilita a realização de pesquisas mais direcionadas no sentido de encontrar inibidores, ativadores enzimáticos ou outros ligantes.

A resolução de estruturas tridimensionais pode ser feita através de técnicas experimentais como a cristalografia de raios X e a espectroscopia de Ressonância Magnética Nuclear (RMN). No entanto, tais técnicas apresentam limitações e dificuldades, como no caso da cristalografia, a necessidade de obtenção de proteínas puras e em quantidade suficiente para os ensaios de cristalização, os cristais obtidos nem sempre têm a qualidade necessária para o trabalho experimental e algumas proteínas, como as de membrana, não cristalizam por completo (MAGGIO & RAMNARAYAN, 2001). As limitações da técnica de RMN se referem principalmente no tamanho das proteínas a ter sua estrutura determinada.

As estruturas tridimensionais de proteínas podem também ser estudadas sob o ponto de vista teórico, com a predição da estrutura tridimensional sem o uso de dados experimentais. Diferentes métodos têm sido desenvolvidos para a construção de modelos 3D de proteínas. Os métodos *ab initio*, por exemplo, são utilizados para prever estruturas proteicas a partir da sequência de aminoácidos da mesma (estrutura primária) sem qualquer comparação com uma estrutura tridimensional conhecida, apenas fragmentos da sequência são comparados com estruturas conhecidas obtidas pelo *Protein Data Bank* (PDB) (FLOUDAS et al, 2006). No entanto esses métodos apresentam algumas dificuldades que ainda precisam ser contornadas, já que o cálculo das energias e entropias de diferentes conformações ainda não é tão preciso e os computadores não são tão potentes ao ponto de permitir a análise de um conjunto grande das possíveis conformações. Os métodos denominados comparativos, como a modelagem por homologia, é um método

alternativo, não experimental, que permite a predição de estruturas tridimensionais de proteínas a partir das estruturas terciárias já resolvidas. Esses métodos se baseiam na observação de que similaridade sequencial implica em similaridade estrutural, logo, é possível construir o modelo 3D da proteína de estudo se houver o conhecimento da estrutura 3D de uma proteína de razoável similaridade com a proteína de interesse. A precisão e a qualidade do modelo 3D gerado dependem fundamentalmente da similaridade entre as sequências. Se essa similaridade for acima de 50% os modelos gerados são de alta qualidade e tem se mostrado tão precisos quanto às estruturas obtidas por raios X de baixa resolução. Se a similaridade entre as sequências for entre 30-50% espera-se que mais de 80% dos átomos  $C\alpha$  estejam dentro de suas verdadeiras posições a 3,5 Å (KOPP & SCHWEDE, 2004). Neste contexto, o método de modelagem por homologia é bastante confiável para a construção de estruturas terciárias, limitando-se pela dependência da existência de uma proteína semelhante à de estudo e de estrutura conhecida.

### 3.5.1 - Modelagem por homologia

O método de modelagem molecular por homologia para construção de um modelo 3D de uma proteína (molécula-alvo) de sequência conhecida, parte da ideia de que se esta estiver relacionada a uma ou mais proteínas de estrutura tridimensional conhecida (molécula molde), então grande parte da estrutura da proteína-alvo será parecida com a da proteína de estrutura conhecida, que servirá de molde para a construção de um modelo estrutural para proteína-alvo. A construção do modelo estrutural por modelagem por homologia pode ser realizada através de *softwares* como o MODELLER (SALI & BLUNDELL, 1993) e SWISS-MODEL (SCHWEDE et al. 2003). Um estudo realizado por NAYEEM e colaboradores (2006), com objetivo de comparar *softwares* disponíveis para modelagem por homologia de alta precisão (incluindo MODELLER e SWISS-MODEL), mostrou que quando a identidade sequencial entre a proteína alvo e o molde é maior que 40%, os modelos obtidos são igualmente satisfatórios para todos os *softwares* comparados no estudo. Neste trabalho foi utilizado o programa SWISS-MODEL para a construção do modelo estrutural de uma PG do *L. gongylophorus*.

### 3.5.2 - O programa SWISS-MODEL

O SWISS-MODEL (SCHWEDE et al. 2003) é um programa computacional disponível livremente ([www.expasy.org/swissmod/SWISS-MODEL.html](http://www.expasy.org/swissmod/SWISS-MODEL.html)) na *web*. Ele fornece ao usuário três principais formas de atuação:

1. O modo mais simples é "Modo de primeira abordagem" e é totalmente automático. Ele exige apenas a sequência de aminoácidos da molécula-alvo como dado de entrada. Em seguida o servidor irá buscar moléculas que tenham similaridade com a molécula-alvo e que tenham estruturas conhecidas. Essa busca é feita através de alinhamento sequencial, da molécula-alvo com todas as sequências de proteínas depositadas nos bancos de dados. Esse modo também permite ao usuário especificar até cinco estruturas-molde obtidas pela biblioteca ExpDB ou por *upload* de arquivos de coordenadas (por exemplo .pdb). No entanto a modelagem automática só será iniciada se pelo menos um molde estiver disponível e se ele possuir uma identidade acima de 25% com a sequência alvo submetida.
2. No "modo de alinhamento" o arquivo de entrada para o programa é o alinhamento entre as sequências da molécula-alvo com a(s) molécula(s)-molde(s). Nesse modo o servidor irá construir o modelo com base no alinhamento realizado pelo usuário.
3. O "modo de projeto" permite ao usuário enviar uma solicitação de modelagem manualmente otimizada para o servidor SWISS-MODEL através de um arquivo de projeto *DeepView*. Este último é um programa de visualização e neste arquivo deverá conter as estruturas (alvo e molde) sobrepostas e o alinhamento entre os mesmos. Este modo dá ao usuário controle sobre uma ampla gama de parâmetros, por exemplo, seleção de molde ou adição de lacuna no alinhamento. Além disso, ele pode ser utilizado para melhorar os resultados obtidos no "modo de primeira abordagem"

De maneira geral, todos os métodos de modelagem por homologia consistem das seguintes etapas: seleção da molécula-molde; alinhamento entre as sequências da molécula-alvo e molécula—molde, sendo esta fase de primordial importância; construção do modelo e validação do mesmo, sendo que cada uma

dessas etapas envolve métodos, programas e servidores específicos. Além disso, essas etapas podem ser repetidas várias vezes até que uma estrutura-modelo satisfatória seja alcançada.

Neste trabalho, como um forma de comparação, foram utilizados os dois métodos, modo de primeira abordagem e o de alinhamento, para a construção do modelo tridimensional da PG.

### 3.5.3 - Validação de modelos estruturais

#### 3.5.3.1 - PROCHECK

A qualidade do modelo tridimensional construído depende muito da qualidade da estrutura da proteína escolhida como molde e do alinhamento entre a sequência alvo e do molde. Dessa forma, é importante que a estrutura da molécula-molde tenha sido obtida em boa resolução. Após a construção do modelo estrutural da molécula alvo, este deve ser avaliado no que diz respeito a parâmetros estereoquímicos. Isso é feito através de programas computacionais como, por exemplo, o PROCHECK (LASKOWSKI et al., 1993), que avalia os comprimentos de ligação; ângulos planos; planaridade dos anéis de cadeias laterais, das ligações peptídicas; quiralidade; conformações das cadeias laterais; ângulos torcionais da cadeia principal e das cadeias laterais; impedimento estérico entre pares de átomos não-ligados e qualidade do mapa de Ramachandran (RAMACHANDRAM et al., 1963).

O mapa de Ramachandran é particularmente útil porque define os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis e orienta a avaliação da qualidade de modelos teóricos ou experimentais de proteínas (RAMACHANDRAM et al., 1963). O mapa mostra os ângulos torsionais  $\phi$  e  $\psi$  para todos os resíduos da molécula. Um fragmento da cadeia polipeptídica linear, comum a todas as estruturas de proteínas é mostrado na Figura 16.3. Por esta figura é possível observar os ângulos torsionais que definem a conformação da cadeia principal polipeptídica e que rotações são permitidas em tornos das ligações simples N-C $\alpha$  e C $\alpha$ -C. Partindo do princípio de que dois átomos não podem ocupar o mesmo lugar no espaço, os valores para os ângulos torsionais acabam por serem limitados. Dessa forma, as faixas permitidas para  $\phi$  e  $\psi$

situam-se em regiões definidas pelo mapa de Ramachandran. As regiões mais escuras (em vermelho) do mapa (Figura 17.3) delimitam as regiões onde os valores de  $\phi$  e  $\psi$  são energeticamente preferidas; as áreas em branco são regiões estericamente proibidas. As áreas mais claras (amarela) são regiões onde os resíduos assumem estados energéticos menos favoráveis, porém aceitáveis.

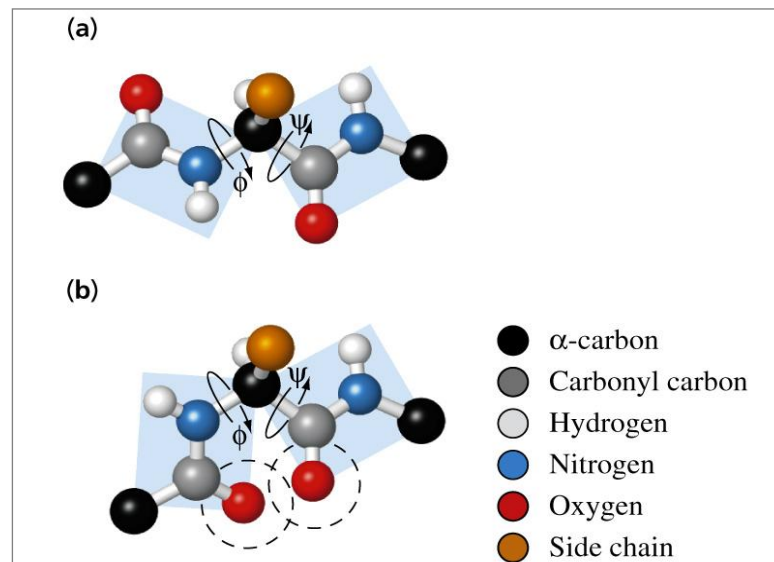


FIGURA 16.3 - Representação dos ângulos torsionais. (www2.bioqmed.ufrj.br/prosdocimi/courses/.../ProteinStructure.ppt).

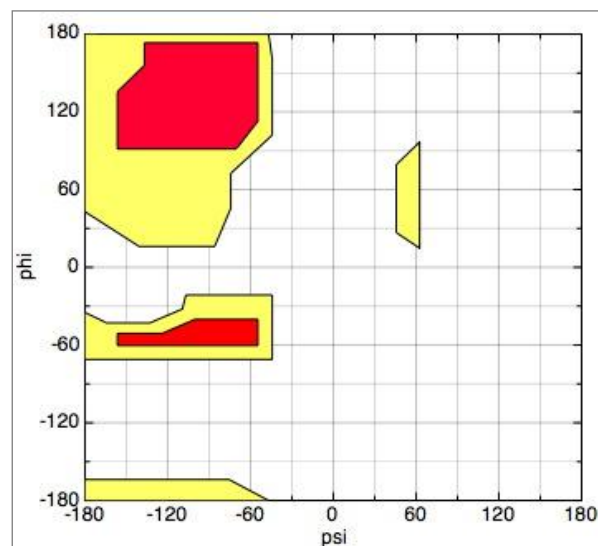


FIGURA 17.3 - Mapa de Ramachandran inicial. (<http://kpwu.wordpress.com/2007/02/15/making-ramachandran-plot-by-your-self/>).

### 3.5.3.2 - VERIFY 3D

O programa VERIFY 3D (LUTHY et al, 1992) consiste em medir a compatibilidade entre a estrutura tridimensional de uma proteína com a sua própria sequência de aminoácidos (1D). A cada resíduo da proteína é atribuída uma classe

estrutural com base em sua localização (estruturas secundárias alfa, beta, loop) e ambiente (polar, apolar). Em seguida, é construída uma matriz de comparação entre os possíveis ambientes e o número total de aminoácidos da estrutura a ser validada. O cálculo do ambiente é realizado para cada um dos aminoácidos de acordo com a probabilidade de encontrar cada um dos 20 aminoácidos em cada uma das classes estruturais através do bando de dados PDB. Com base nesta matriz, um alinhamento é realizado entre a sequência primária e o seu perfil tridimensional e o resultado é a medida da compatibilidade da sequência com sua estrutura 3D. Grande parte dessas medidas deve ficar acima de 0,2 e nunca abaixo de zero.

### 3.5.3.3 – WHAT IF

O WHAT IF (VRIEND, 1990) é um programa que oferece diversas ferramentas relacionadas à modelagem de estruturas. Neste trabalho foi utilizada a ferramenta OLDQUA para avaliar a normalidade estrutural do modelo. Isso é realizado através dos parâmetros de enovelamentos que envolvem proximidades atômicas.

Um conjunto de átomos, denominados grupos rígidos, representante de um determinado resíduo é usado para a superposição em todos os resíduos do mesmo tipo no arquivo PDB, após a superposição os átomos vizinhos de cada resíduo será mostrado. Em seguida, com as coordenadas dos átomos vizinhos, uma função é calculada para atribuir valores maiores aos locais onde um número maior de átomos é encontrado, levando em consideração o número de resíduos e suas exposições ao solvente. Por fim, esses valores obtidos são comparados com os valores padrões. O resultado para o modelo como um todo pode ser lido da seguinte forma:

- Maior do que -0,5: meio perfeito;
- -0,5: modelo bom, na média;
- -1,0 até -0,5: modelo ainda de boa ou muito boa qualidade;
- -1,5: provavelmente ainda OK, mas com muitos erros pequenos, modelo normal.
- -2,0: um modelo projetado *ab initio* ou um modelo verdadeiramente muito

pobre;

- -3,0: modelo errado ou um mau modelo.

Com relação ao desvio médio para cada resíduo, quando este apresenta um valor de -5,0, ou pior, é recomendado checá-lo, pois pode vir a significar que:

- Está envolvido em simetria de contatos.
- Está ligado a um co-fator, ligante, ou íons.
- Está no sítio ativo.
- Está errado.





## CAPÍTULO 4 - RESULTADOS E DISCUSSÕES

### 4.1 - Estudo teórico de poligalacturonases (PG) de fungos

Neste sub capítulo serão mostrados os resultados obtidos no estudo de PG de fungos, cujas sequências estão depositadas em banco de dados.

#### 4.1.1- Formação do banco de dados com PG de fungos

A busca no *site* NCBI por sequências de PG de fungos (*polygalacturonase fungi*) resultou em 2093 sequências. Estas sequências foram analisadas e após exclusão daquelas com menos de 70 resíduos de aminoácidos, das que correspondiam a 'proteína hipotética' e ainda das que não eram PG ou mesmo não eram PG de fungos restaram 957 sequências de aminoácidos. Essas sequências foram separadas por gênero e espécie e para cada grupo foram realizados alinhamentos múltiplos de sequências usando o programa *ClustalW* (THOMPSON et al., 1994). Os alinhamentos das sequências de PG de cada espécie foram analisados e as sequências com 100% de identidade foram retiradas. Assim, o conjunto de dados ficou com 417 sequências únicas de PG de fungos (Tabela 1.4).

TABELA 1.4 - Número de seqüências de PG de fungos depositadas no NCBI e número de seqüências únicas por gênero encontradas neste trabalho.

	Gênero	Filo	Número de seqüências de poligalacturonase de fungos depositadas no NCBI	Número de seqüências únicas de poligalacturonases de fungos
1	<i>Alternaria</i>	Ascomycota	176	61
2	<i>Aspergillus</i>	Ascomycota	91	53
3	<i>Athelia</i>	Basidiomycota	2	2
4	<i>Bispora</i>	Ascomycota	1	1
5	<i>Botryotinia</i>	Ascomycota	251	33
6	<i>Botrytis</i>	Ascomycota	9	5
7	<i>Chondrostereum</i>	Basidiomycota	11	9
8	<i>Chrysosporthe</i>	Ascomycota	1	1
9	<i>Claviceps</i>	Ascomycota	2	2
10	<i>Cochliobolus</i>	Ascomycota	1	1
11	<i>Colletotrichum</i>	Ascomycota	2	2
12	<i>Cryphonectria</i>	Ascomycota	1	1
13	<i>Cryptococcus</i>	Basidiomycota	2	2
14	<i>Diaporthe</i>	Ascomycota	1	1
15	<i>Emericella</i>	Ascomycota	6	6
16	<i>Fusarium</i>	Ascomycota	278	135
17	<i>Galactomyces</i>	Ascomycota	5	4
18	<i>Gibberella</i>	Ascomycota	23	15
19	<i>Glomerella</i>	Ascomycota	2	2
20	<i>Hypocrea</i>	Ascomycota	1	1
21	<i>Kluyveromyces</i>	Ascomycota	3	3
22	<i>Leptosphaeria</i>	Ascomycota	1	1
23	<i>Leucoagaricus</i>	Basidiomycota	1	1
24	<i>Metarhizium</i>	Ascomycota	2	2
25	<i>Neosartorya</i>	Ascomycota	12	12
26	<i>Neurospora</i>	Ascomycota	1	1
27	<i>Ophiostoma</i>	Ascomycota	3	2
28	<i>Penicillium</i>	Ascomycota	14	13
29	<i>Phomopsis</i>	Ascomycota	1	1
30	<i>Postia</i>	Basidiomycota	2	2
31	<i>Pyrenophora</i>	Ascomycota	5	5
32	<i>Rhizopus</i>	Fungi incertae sedis	2	1
33	<i>Saccharomyces</i>	Ascomycota	4	2
34	<i>Sclerotinia</i>	Ascomycota	15	11
35	<i>Talaromyces</i>	Ascomycota	6	5
36	<i>Thanatephorus</i>	Basidiomycota	3	3
37	<i>Venturia</i>	Ascomycota	5	4
38	<i>Verticillium</i>	Ascomycota	11	11
	<b>Total de seqüências</b>		<b>957</b>	<b>417</b>

#### 4.1.2- Classificação das seqüências de PG por alinhamento

Como discutido no capítulo 1, as PG são classificadas em Endo- e Exo- de acordo com o seu modo de atuação sobre o substrato. No entanto, nem todas as seqüências depositadas no banco de dados do NCBI estão classificadas como Endo- ou Exo-PG. Das 417 seqüências únicas encontradas para os fungos, Tabela 1.4, 95 estavam sem classificação, ou seja, foram depositadas como PG sem uma classificação mais precisa (Figura 18.4).

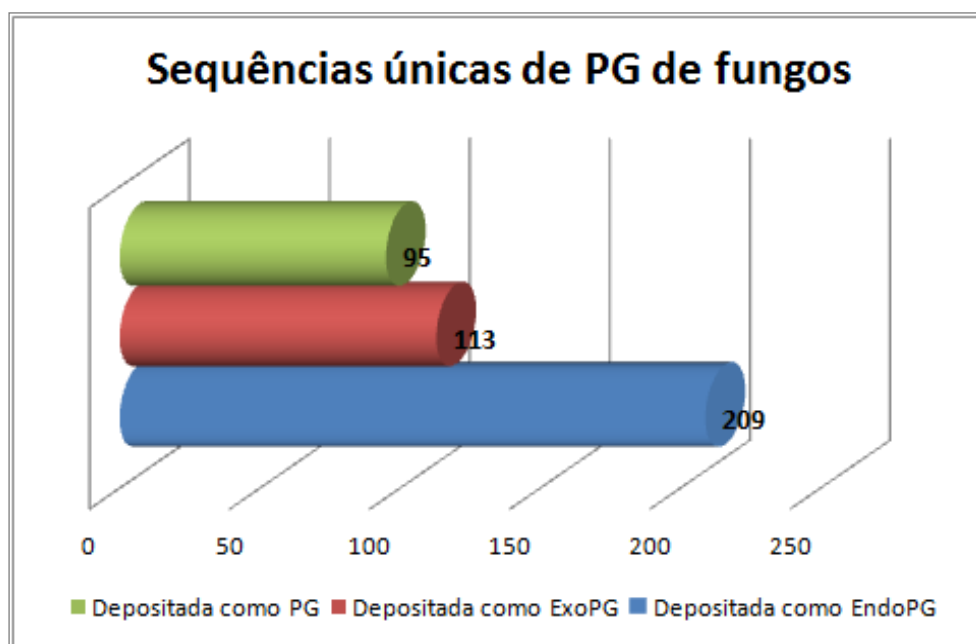


FIGURA 18.4 - Gráfico representando a classificação das PG conforme depositadas no banco de dados do NCBI.

Como uma tentativa de classificar as 95 sequências em Endo- ou Exo-PG, alinhamentos múltiplos de sequências por espécie foram realizados (e em alguns casos por gênero). Para a classificação de uma determinada sequência esta foi alinhada com sequências já classificadas em Endo- ou Exo- PG e o escore obtido do alinhamento foi avaliado. Como exemplo do que foi feito neste trabalho, com todas as sequências não classificadas, passa-se a descrever a classificação de uma PG de *Emericella nidulans*.

*E. nidulans* (Tabela 1.4), apresenta uma sequência como PG (S24156), 2 sequências depositadas como Endo-PG (ABF50893.1; ABF50868.1), 3 como Exo-PG (ABF50897.1; ABF50895.1; AAO61898.1). Primeiramente, foi feito o alinhamento múltiplo da sequência não classificada (S24156) com as 2 sequências de Endo-PG (Figura 19.4). O escore obtido para os alinhamentos foram de 51 e 55 e é interessante observar que o escore do alinhamento entre as sequências das duas Endo-PG (escore 52) não é tão diferente da obtida entre a ABF50868.1 e S24156 (escore 51).

O alinhamento da sequência não classificada (S24156) com as 3 sequências de Exo-PG (Figura 20.4) mostrou escores de 19 a 22, bem abaixo dos obtidos para os alinhamentos com as Endo-PG. Assim, infere-se que a sequência S24156 é uma possível Endo-PG.

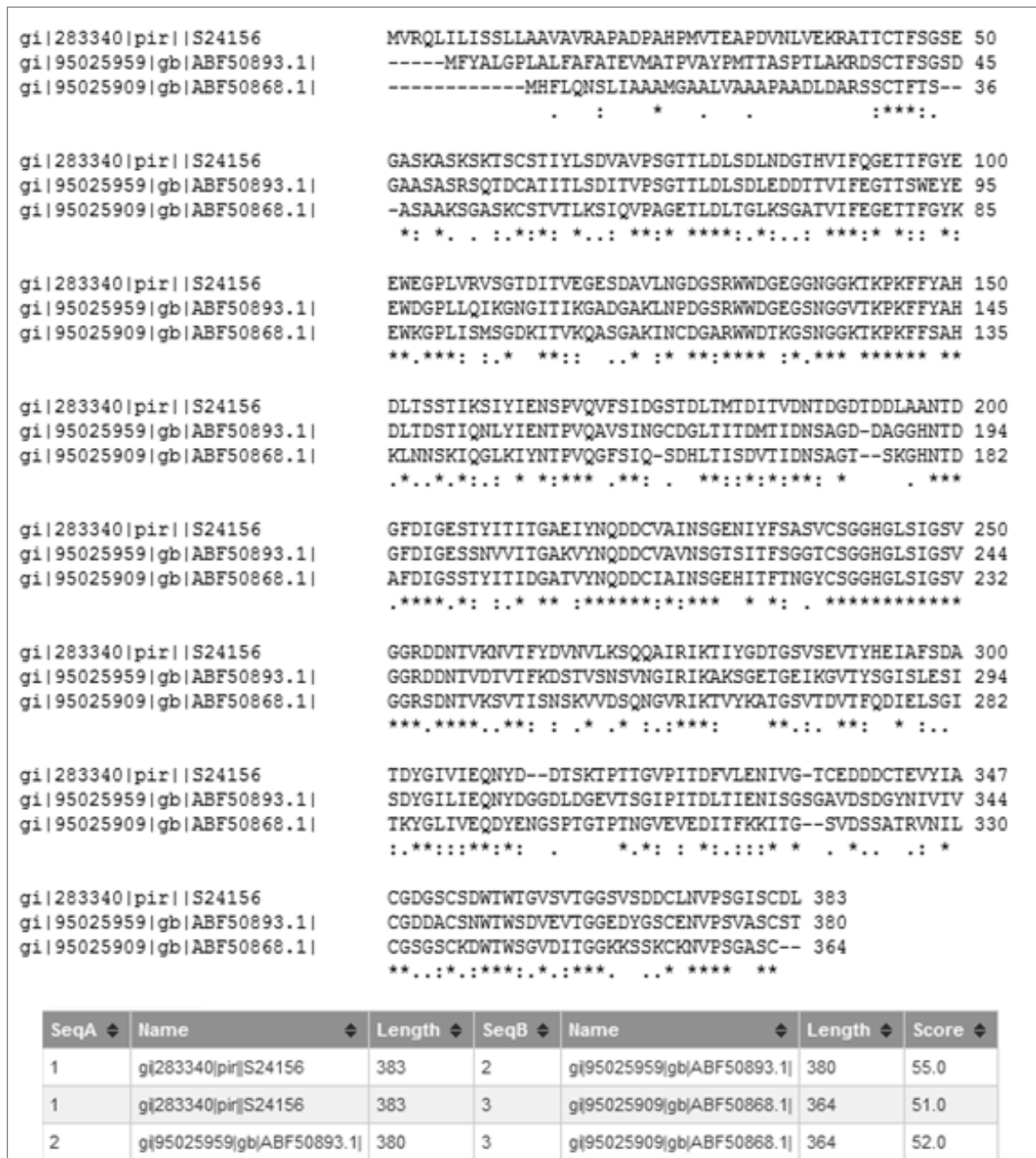


FIGURA 19.4 - Alinhamento das seqüências de PG (S24156) e Endo-PG (ABF50893.1; ABF50868.1) de *E. nidulans* e escores obtidos para os alinhamentos. Os alinhamentos foram realizados com o programa *ClustalW* (THOMPSON et al., 1994).

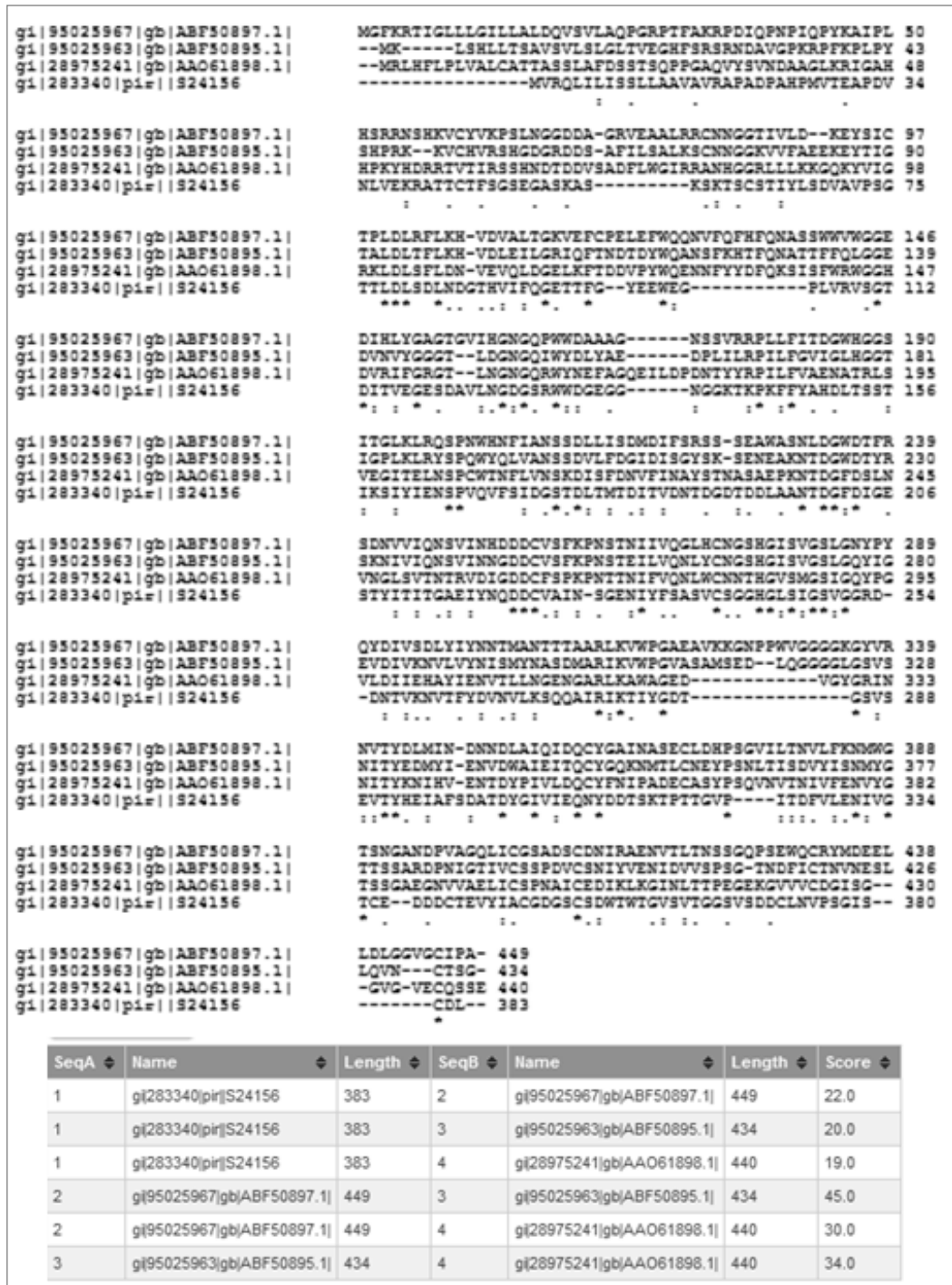


FIGURA 20.4 - Alinhamento das sequências de PG (S24156) e Exo-PG de *E. nidulans* e escores obtidos para o alinhamento. Os alinhamentos foram realizados com o programa *ClustalW* (Thompson et al., 1994).

Esse procedimento foi realizado na classificação de sequências olhando-se primeiramente para sequências Endo e Exo-PG dentro da mesma espécie e não sendo possível, passando para sequências do mesmo gênero. Desta maneira, 68 sequências que haviam sido depositadas como PG de fungos puderam ser classificadas em Endo- ou Exo-PG, como mostrado de forma resumida na Tabela 2.4. Nesta tabela estão representados os maiores escores obtidos entre as

sequências que não possuem classificação (sequência A) e as sequências já depositadas no banco de dados do NCBI com classificação Endo ou Exo-PG (sequência B). De acordo com o valor do score é possível inferir se as sequências A são possíveis sequências de Endo-PG ou Exo-PG. Por exemplo, a sequência CAE46195.1, na primeira linha da tabela, (PG de *Aspergillus aculeatus*) possui uma maior similaridade com a sequência CAA74744.1 (Endo-PG de *Aspergillus niger*), logo a sequência CAE46195.1 é uma Endo-PG. O mesmo princípio foi realizado para todas as sequências A presentes na Tabela 2.4.

TABELA 2.4 - Maiores escores obtidos para os alinhamentos múltiplos entre as sequências de PG sem classificação (Sequência A) e as sequências de Endo-PG ou Exo-PG

Sequência A	Sequência B	Escore
<i>Aspergillus aculeatus</i>	<i>Aspergillus niger</i>	
CAE46195.1	CAA74744.1 EndoPG	71
CAE46194.1	CAK47020.1 EndoPG	71
CAE46193.1	CAA74744.1 EndoPG	73
<i>Aspergillus clavatus</i>	<i>Aspergillus niger</i>	
EAW10813.1	CAK47020.1 EndoPG	67
<i>Aspergillus flavus</i>	<i>Aspergillus flavus</i>	
AAA85280.1	EED53217.1 EndoPG	56
AAA85279.1	EED53217.1 EndoPG	99
EED53441.1	EED53217.1 EndoPG	58
<i>Aspergillus fumigatus</i>	<i>Aspergillus fumigatus</i>	
EDP56939.1	EDP49013.1 EndoPG	55
EDP50743.1	EDP49013.1 EndoPG	57
<i>Aspergillus kawachii</i>	<i>Aspergillus niger</i>	
BAE94655.1	CAK47667.1 EndoPG	97
BAC10597.1	CAK47667.1 EndoPG	62
BAC10596.1	CAK47020.1 EndoPG	98
<i>Aspergillus niger</i>	<i>Aspergillus niger</i>	
CAA41693.1	CAK47020.1 EndoPG	69
CAA38085.1	CAK47667.1 EndoPG	63
CAA41694.1	CAK47667.1 EndoPG	62
ACS44814.1	CAK47020.1 EndoPG	66
CAA45707.1	CAA74744.1 EndoPG	74
CAK44164.1	CAA74744.1 EndoPG	98
<i>Aspergillus oryzae</i>		
BAA03244.1	EED53217.1 EndoPG A. <i>flavus</i>	98
BAA34782.1	CAK47020.1 EndoPG A. <i>niger</i>	62
<i>Aspergillus terreus</i>	<i>Aspergillus niger</i>	
EAU34060.1	CAK47020.1 EndoPG	67
EAU32010.1	CAA74744.1 EndoPG	59
<i>Aspergillus tubingensis</i>	<i>Aspergillus niger</i>	
CAA41695.1	CAK47667.1 EndoPG	63
<i>Botryotinia allii</i>	<i>Botryotinia fuckeliana</i>	
ABR20949.1	CAM91338.1 EndoPG	61
<i>Botryotinia calthae</i>	<i>Botryotinia fuckeliana</i>	
ABP88779.1	CAM91322.1 EndoPG	96
ABR20984.1	CAM91235.2 EndoPG	96
<i>Botryotinia fuckeliana</i>	<i>Botryotinia fuckeliana</i>	
ABR20983.1	CAM91235.2 EndoPG	99
ABR20977.1	CAM91235.2 EndoPG	99
ABR20947.1	CAM33189.2 EndoPG	61
ABR20942.1	CAM33189.2 EndoPG	61
ABR20939.1	CAM91338.1 EndoPG	62
ABR20928.1	CAM91338.1 EndoPG	63
ABR20927.1	CAM91338.1 EndoPG	62
ABR20925.1	CAM91338.1 EndoPG	62
ABR20921.1	CAM91338.1 EndoPG	62
ABR20922.1	CAM91338.1 EndoPG	63
ABP88778.1	CAM91322.1 EndoPG	96
ABP88777.1	CAM91322.1 EndoPG	98
ABP88776.1	CAM91322.1 EndoPG	97
ABP88772.1	CAM91322.1 EndoPG	99
ABP88768.1	CAM91322.1 EndoPG	98
ABP88766.1	CAM91322.1 EndoPG	98
ABP88748.1	CAM91338.1 EndoPG	99
ABP88747.1	CAM91338.1 EndoPG	98
ABP88746.1	CAM91338.1 EndoPG	98
<i>Emmericella nidulans</i>	<i>Emmericella nidulans</i>	
S24156	ABF50893.1 EndoPG	56
<i>Fusarium oxysporum f. sp. Lycopersici</i>	<i>Fusarium oxysporum f. sp. Lycopersici</i>	
BAE45849.1	BAE97122.1 EndoPG	45
BAE45848.1	BAE97122.1 EndoPG	46
BAE45846.1	BAE97122.1 EndoPG	46
<i>Galactomyces geotrichum</i>	<i>Galactomyces geotrichum</i>	
BAC67212.1	BAB97386.1 EndoPG	81
BAC67213.1	BAB97386.1 EndoPG	82
<i>Kluyveromyces wickerhamii</i>	<i>Kluyveromyces marxianus</i>	
BAB69585.1	AAR84199.1 EndoPG	68
<i>Neosartorya fischeri</i>	<i>Neosartorya fischeri</i>	
EAW24760.1	EAW19942.1 EndoPG	55
EAW23158.1	EAW20012.1 ExoPG	23
EAW22138.1	EAW19942.1 EndoPG	51
EAW21675.1	EAW20325.1 ExoPG	24
EAW20375.1	EAW20012.1 ExoPG	34
EAW20306.1	EAW19942.1 EndoPG	45
EAW19696.1	EAW20325.1 ExoPG	26
EAW15592.1	EAW20325.1 ExoPG	33
EAW17738.1	EAW20325.1 ExoPG	23
<i>Sclerotinia sclerotiorum</i>	<i>Sclerotinia sclerotiorum</i>	
AAR87657.1	CAF05670.1 EndoPG	68
<i>Verticillium albo-atrum</i>	<i>Verticillium albo-atrum</i>	
EEY22323.1	EEY19557.1 ExoPG	55
EEY21535.1	EEY19557.1 ExoPG	20
EEY20545.1	EEY19557.1 ExoPG	25
EEY15660.1	EEY19557.1 ExoPG	51
EEY15055.1	EEY20199.1 ExoPG	52
EEY22322.1	EEY19557.1 ExoPG	25



Três sequências de PG, embora não estivessem classificadas no NCBI, estavam classificadas na literatura:

- As sequências CAA71247.2 e CAA71246.2 de *Claviceps purpurea* são Endo-PG segundo TENBERGE e colaboradores (1996).
- A sequência ACA48699.1 de *Rhizopus oryzae* é uma Exo-PG segundo YOSHIDA e colaboradores (2004).

Portanto, das 95 sequências de PG de fungos não classificadas inicialmente, 71 sequências foram classificadas em Endo- ou Exo-PG (68 através de análise dos alinhamentos e 3 segundo a literatura), restando 24 sequências sem classificação (Figura 21.4). Essas sequências foram, então, classificadas através do estudo de filogenia.

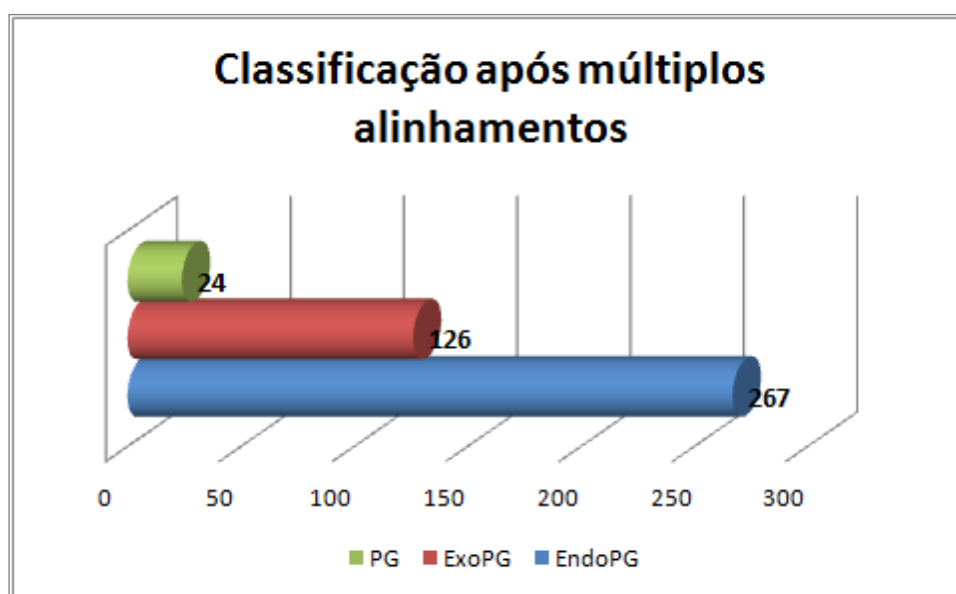


FIGURA 21.4 - Gráfico representando o número total de sequências de Endo-PG e Exo-PG após classificação por alinhamentos múltiplos (68 sequências) e literatura (3 sequências).

#### 4.1.3- Classificação das sequências de PG por árvore filogenética

A classificação de sequências de PG através da análise de árvore filogenética não é novidade. YOSHIDA e colaboradores (2004) construíram uma árvore filogenética contendo seis sequências de Exo-PG de fungos e onze Endo-PG de fungos, além de duas sequências de Endo-PG de bactérias a fim de inferir classificação para uma sequência de PG do fungo *Rhizopus oryzae*. Como resultado, a análise do filograma demonstrou que as sequências de Endo-PG e Exo-

PG de fungos foram visivelmente separadas das sequências de EndoPG de bactérias. Além disso, as sequências de Endo-PG de fungos formaram um grupo a parte das Exo-PG de fungos, sendo a PG de *R. oryzae* pertencente ao grupo das Exo-PG.

Estudos de PG através da análise de árvore filogenética, e outras ferramentas de bioinformática, como os de MARKOVIC & JANECEK (2001) e YADAV e colaboradores (2009), foram realizados. Neles as PG de fungos foram estudadas em conjunto com as PG de plantas, bactérias e insetos, além de outras enzimas pertencentes à família 28 das glicosil hidrolases a fim de demonstrar as relações filogenéticas entre os diferentes organismos, bem como das enzimas. Porém esses estudos apresentaram o número de sequências de PG de fungos analisada pequeno, quarenta e oito no trabalho de MARKOVIC & JANECEK (2001) e sete no do YADAV e colaboradores (2009).

O programa computacional escolhido para construção da árvore filogenética foi o PhyML através da plataforma Phylogeny.fr ([http://www.phylogeny.fr/version2\\_cgi/index.cgi](http://www.phylogeny.fr/version2_cgi/index.cgi)). Este programa não aceita número maior que 200 sequências proteicas como arquivo de entrada. Logo, o banco de dados formado pelas 417 sequências únicas de PG de fungos não poderia ser utilizado. Um novo banco de dados - a partir das 417 sequências - com um número menor de sequência era necessário.

Para formar esse novo banco de dados alguns critérios foram estabelecidos: foi selecionada uma sequência de cada espécie sendo que as sequências deveriam possuir classificação, ou seja, serem designadas (depositadas) como Endo-PG ou Exo-PG. Aplicando esses critérios, o banco de dados formado passou a ser constituído por 63 sequências de Endo-PG e 38 sequências de Exo-PG (Apêndice A).

#### **4.1.4- Construção de uma árvore filogenética teste**

Formado o novo banco de dados com apenas sequências de Endo-PG e Exo-PG, uma árvore filogenética foi construída com estes dados com o objetivo de observar como as sequências se comportavam frente aos parâmetros computacionais escolhidos, além de observar se realmente as sequências de Endo-PG e Exo-PG formariam grupos a parte, ou seja, se elas seriam separadas de

acordo com a classificação.

A Figura 22.4 mostra a árvore filogenética resultante. Por ela é possível inferir que as sequências foram separadas em basicamente dois grandes grupos ou clados: Endo-PG e Exo-PG. Além disso, ocorreu separação entre filos distintos. As sequências pertencentes aos fungos 'Basidiomycota' formaram um pequeno grupo a parte no meio das sequências pertencentes aos fungos 'Ascomycota' enquanto a única sequência do fungo *Rhizopus oryzae*, não pertencente a nenhum filo definido, não se agrupou a nenhuma outra sequência.

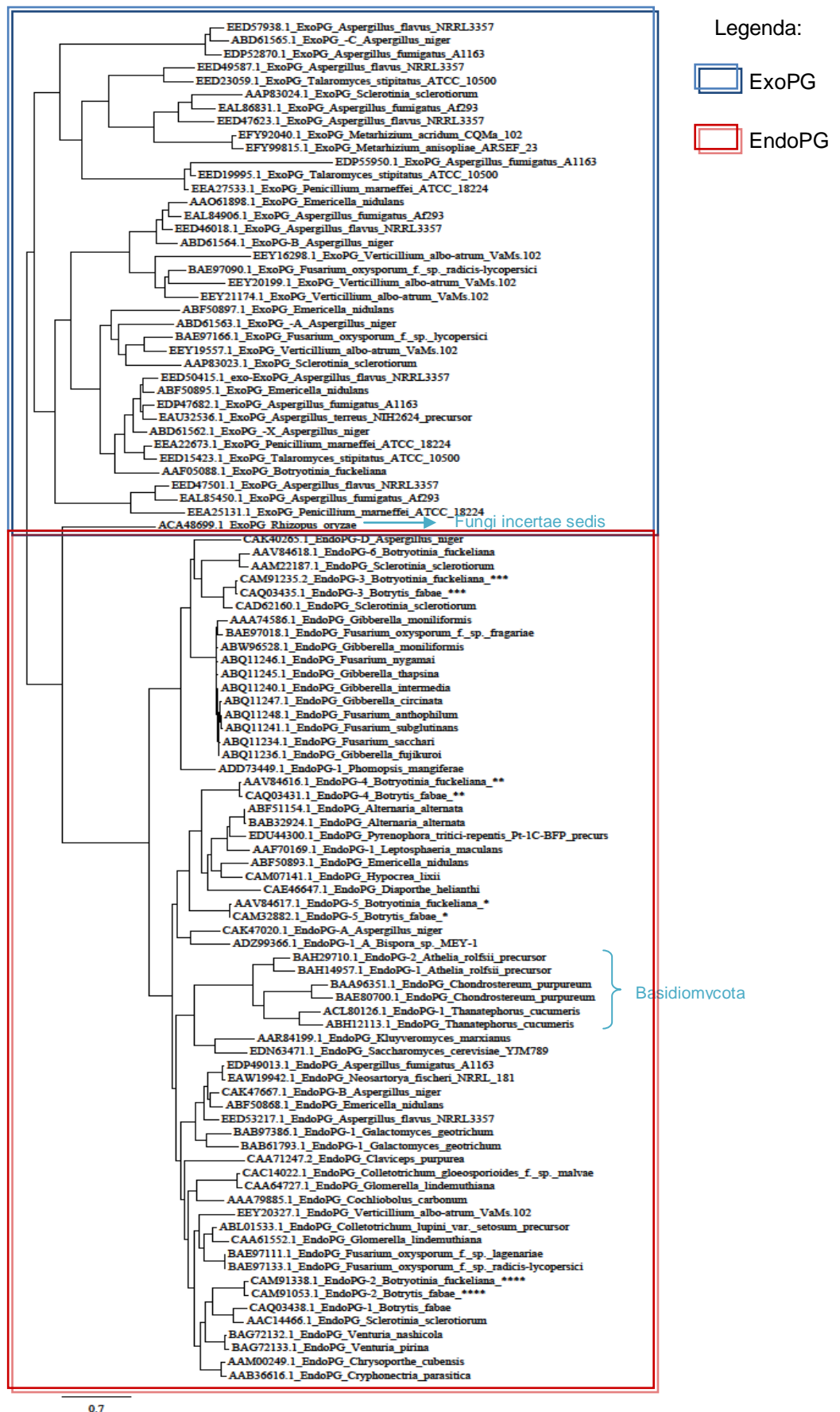


FIGURA 22.4 - Árvore filogenética construída a partir de 38 seqüências Exo-PG e 63 seqüências Endo-PG de fungos.

Após a análise da árvore filogenética (FIGURA 22.4), onde foi possível notar dois grandes clados - um com predominância de Exo-PG (parte superior da árvore filogenética) e outro com predominância de Endo-PG (parte inferior da árvore filogenética), o próximo passo seria mostrar o porquê destes clados se diferirem. Neste contexto, as mesmas sequências utilizadas na construção da árvore filogenética foram utilizadas como arquivo de entrada no programa MEME com o objetivo de encontrar motivos conservados, e se possível encontrar algum motivo que seja específico para Endo-PG ou Exo-PG. Cinco motivos sequenciais foram encontrados pelo programa MEME nas sequências analisadas (Apêndice A), sendo os resultados obtidos resumidos nas tabelas abaixo. A Tabela 3.4 mostra que das 63 sequências de Endo-PG 57 tem o motivo 1, 62 tem o motivo 2, 57 tem o motivo 3, 61 tem o motivo 4 e 61 sequências tem o motivo 5. Estes resultados mostraram que a grande maioria das sequências de Endo-PG apresenta os 5 motivos sequenciais. Já a análise das 38 sequências de Exo-PG de fungos mostrou que em 36 delas encontra-se o motivo 4 e em 32 delas encontra-se o motivo 5. Os motivos 1, 2 e 3 não são encontrados em Exo-PG. As sequências correspondentes aos motivos estão representadas nas Tabelas 4.4 e 5.4. Na primeira tabela alguns aminoácidos encontram-se entre colchetes, indicando que um destes aminoácidos dentro dos colchetes pode ser encontrado nesta posição. Os aminoácidos fora dos colchetes são aminoácidos bastante conservados, ou seja, quase não variam entre as sequências analisadas. Já na segunda tabela (5.4) os colchetes são excluídos pelo programa MAST, pois ele próprio calcula o "best possible match", ou seja, qual o melhor aminoácido a ocupar aquele determinado sítio.

Com a finalidade de inferir função biológica aos 5 motivos encontrado pelo MEME, os motivos foram analisados pelo BLAST (<http://blast.ncbi.nlm.nih.gov/>). Este programa encontra regiões de similaridade local entre sequências, comparando a(s) sequência(s) a ser analisada com as sequências depositadas em seu banco de dados. Trata-se de uma ferramenta computacional muito utilizada para inferir função e relações evolucionárias. Como resultado desta análise os motivos 1, 2 e 5 estão relacionados ao domínio Glyco\_hydro\_28 enquanto os motivos 3 e 4 não estão relacionados a nenhuma função.

Esses resultados sugerem que, apesar das Endo- e Exo-PG atuarem com o mesmo mecanismo de ação - hidrólise da cadeia do substrato por inversão da configuração do carbono anomérico - existem regiões que são bastante conservadas

em uma das classes e não em outra, o que pode ser entendido como regiões com função de especificidade da enzima.

TABELA 3.4 - Motivos sequenciais presentes nas 63 seqüências de Endo-PG e 38 seqüências de Exo-PG, encontrados pelo programa MEME (BAILEY & ELKAN, 1994).

	Número de seqüências	motivo 1	motivo 2	motivo 3	motivo 4	motivo 5
Endo	63	57	62	57	61	61
Exo	38	0	0	0	36	32
Total	101	57	62	57	97	93

TABELA 4.4 - Representação das seqüências dos motivos presentes nas 63 seqüências de Endo-PG e 38 seqüências de Exo-PG

Motivo	Tamanho	Seqüência	Ocorrência nas seqüências de PGs
1	37	G[TS][VI][SN][NG][VI][YF][QS][ND][IL][ST][NG][ST]KYG[IV][VD][VI][EQ]QDY[EL]NG[SG]PTG[TK]PT	57
2	21	NQDDC[VL]A[IV][NT]SG[TS]NIT[FV][TS][NG]GYC	62
3	28	[AQ]VP[AS]G[TK]TLDL[TS][GK]L[KN][DA]GTTV[IT]F[EK]G[TK]TTF[GA]	57
4	15	SG[GS]HG[L]S[IV]GS[VL]G[GQ][RY]S	97
5	21	GA[SG]G[HG]V[IL][DN]GNG[QA][RA][WY][WY]DGKG[SG]N	93

Fonte: MEME (Bailey & Elkan, 1994)

TABELA 5.4 – Melhores aminoácidos correspondentes possíveis e o grau de similaridade entre os motivos.

Motivo	Tamanho	Melhor correspondente possível	Similaridade entre os motivos				
			1	2	3	4	5
1	37	GTVSNVTYQNITLSNISKYGIVVEQDYENGSPGKPT	-	0.29	0.21	0.22	0.22
2	21	NQDDCVAINSGTNITFTNMYC	0.29	-	0.36	0.24	0.13
3	28	QVPAGETLDTLTKLDGTTVIFEGETTFG	0.21	0.36	-	0.30	0.27
4	15	SGGHLSIGSVGQRS	0.22	0.24	0.30	-	0.19
5	21	GASGHVIDGNGQRWWDGKGSN	0.22	0.13	0.27	0.19	-

Fonte: MAST (Bailey & Elkan, 1998)

#### 4.1.5 - Construção da árvore filogenética para classificação de PG de fungos

Utilizando os mesmos parâmetros usados para obtenção da árvore filogenética da Figura 22.4, foi construída uma árvore filogenética (Figura 23.4) com as 63 seqüências de Endo-PG, as 38 seqüências de Exo-PG e as 24 seqüências de PG não classificadas.

Analisando a árvore filogenética (Figura 23.4) é possível inferir que novamente dois grandes grupos são formados: um com predominância de Exo-PG

(parte superior da árvore filogenética) e outro com predominância de Endo-PG (parte inferior da árvore filogenética). Além disso, as 24 sequências de PG sem classificação definida foram homoganeamente distribuídas na árvore. Destas 24 sequências, apenas 3 sequências de PG estão visivelmente relacionadas às Exo-PG, enquanto as 21 sequências restantes estão distribuídas no grupo pertencente às Endo-PG mostrando, conseqüentemente, melhores relações filogenéticas com as sequências de Endo-PG.

A Figura 24.4 mostra número total de sequências de Endo-PG e Exo-PG após o estudo de classificações.

Além da possibilidade de classificar as PG quanto ao seu modo de ação sobre o substrato (Endo- ou Exo-), a árvore filogenética demonstra as relações de semelhança e diferenças entre as sequências analisadas filogeneticamente. Assim, sequências muito semelhantes são distribuídas ao longo da árvore de maneira a ficarem próximas uma das outras enquanto as dissimilares formam clados que são dispostos afastadamente. Isso explica o porquê das sequências de fungos basideomicetos formarem grupos a parte dos fungos ascomicetos (Figura 23.4).

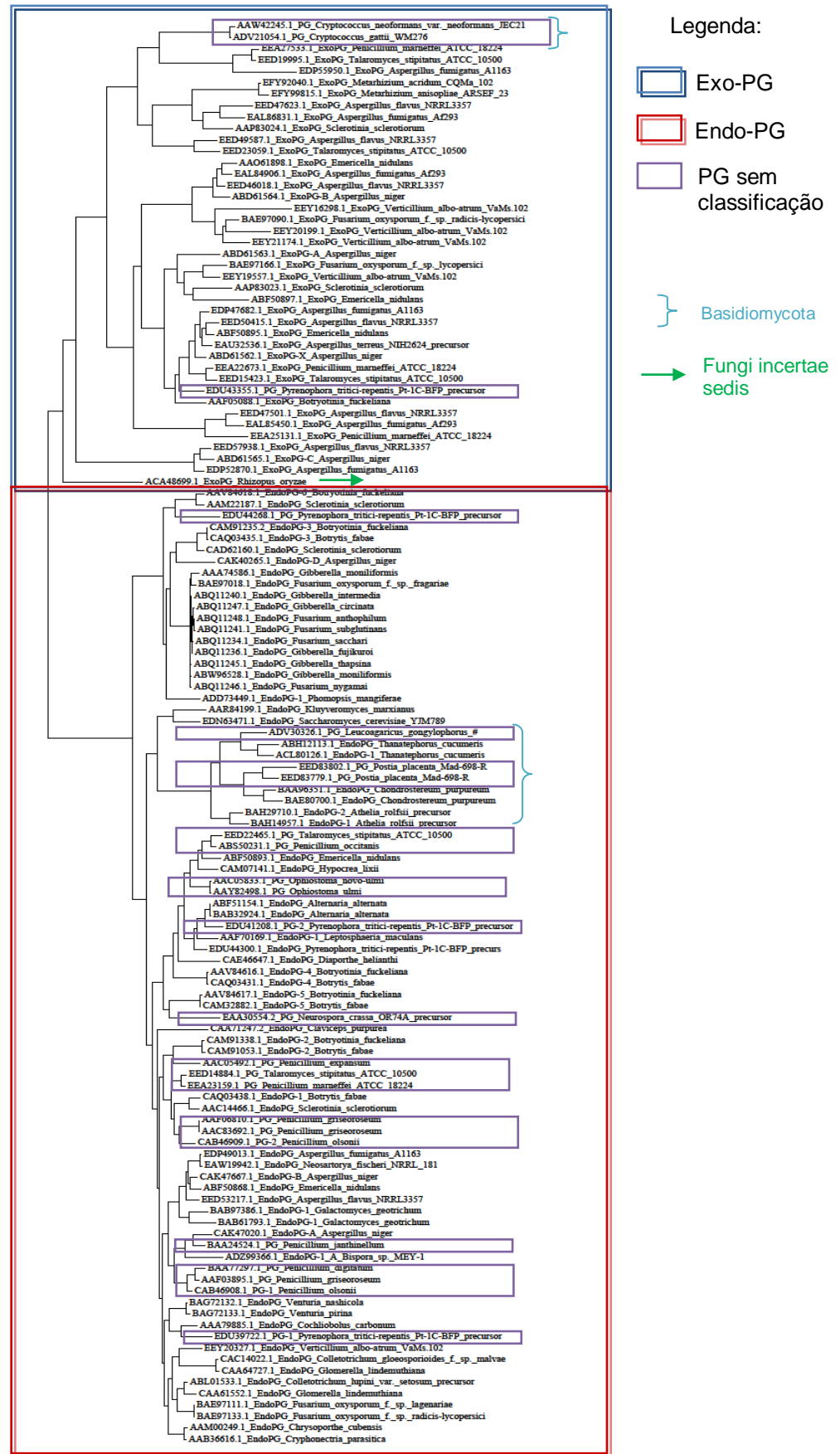


FIGURA 23.4 - Árvore filogenética de seqüências de 63 Endo-PG, 38 Exo-PG e 24 PG de fungos sem classificação.



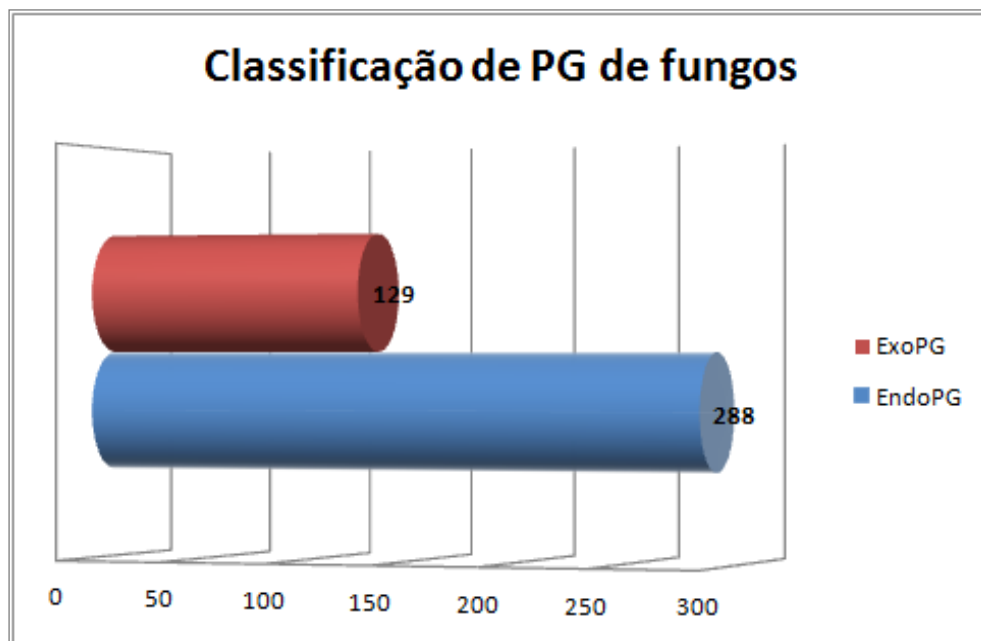


FIGURA 24.4 - Número total de sequências de Endo-PG e Exo-PG de fungos após a classificação realizada neste trabalho.

#### 4.2 - Estudo estrutural teórico de uma PG do fungo *L. gongylophorus* através de modelagem por homologia

As formigas são animais pertencentes à família Formicidae, o grupo mais numeroso dentre os insetos, com 12644 espécies descritas até 01 de março de 2012 ([http://osuc.biosci.ohio-state.edu/hymenoptera/tsa.sppcount?the\\_taxon=Formicidae](http://osuc.biosci.ohio-state.edu/hymenoptera/tsa.sppcount?the_taxon=Formicidae)). Esses insetos são encontrados em todas as regiões do planeta, exceto nas regiões polares. Dentre os diferentes tipos de formigas são consideradas formigas cortadeiras todas as espécies do gênero *Atta*, popularmente conhecidas como saúvas; do gênero *Acromyrmex*, conhecidas como quenquéns e também algumas dos gêneros *Trachymyrmex*, *Sericomyrmex* e *Apterostigma*. Com exceção das três últimas (*Trachymyrmex*, *Sericomyrmex* e *Apterostigma*), que têm colônias muito pequenas (Justi-Junior et al., 1996), as formigas cortadeiras são conhecidas pelo poder de destruição de um grande número de espécies vegetais e pelo prejuízo econômico causado à agricultura (BERTI FILHO et al., 1992).

De uma forma bastante curiosa e eficiente as formigas cortadeiras não se alimentam diretamente das folhas que cortam e sim indiretamente, através da simbiose com um fungo, o basídeoomiceto *Leucoagaricus gongylophorus*, que habita os ninhos das formigas (WEBER, 1955). Dentre as diferentes castas no formigueiro, com funções específicas na manutenção da colônia (operárias, soldados, operárias

do jardim), existem aquelas que cortam e/ou carregam folhas, flores e ramos, outras cuidam da limpeza e da defesa da colônia, e outras ainda do cultivo do fungo e do cuidado com os filhotes, chamados larvas. As formigas da casta das "jardineiras" cortam as folhas e, ao fazê-lo, aproveitam para se alimentarem da seiva exsudada. Estas folhas são carregadas para o interior do formigueiro, onde formigas de outra casta se encarregarão de triturá-las para o cultivo do fungo, base da sua alimentação. O fungo *L. gongylophorus* supre as necessidades alimentares de todas as formigas que vivem exclusivamente dentro do formigueiro, como as larvas, e da rainha.

Diversos estudos têm sido realizados para melhor entender a evolução da relação entre as formigas cortadeiras e o fungo simbiote (SINGER, 1986, FISHER et al. 1994). O cultivo de fungo dentro do ninho como alimento surgiu há cerca de 45-65 milhões de anos atrás em um ancestral de formigas criadoras de fungos (Formicidae, tribo Attini) representando uma transição evolutiva da vida da formiga caçadora-coletora de vegetais e outros alimentos para a vida da formiga subsistindo de fungos cultivados (MUELLER et al., 2001). Estudos sugerem que o fungo, ao selecionar um determinado substrato, tem um certo controle sobre o comportamento de forrageamento de formigas por meio de um mecanismo de *feedback* químico (NORTH et al., 1997). Portanto, a associação mutualística entre as formigas cortadeiras e o fungo *L. gongylophorus* é muito forte e existe uma total dependência entre ambos. Por esse motivo, eles têm sido alvos de diversas pesquisas com a finalidade de desenvolvimento de metodologias de controle e assim garantir a produtividade das plantações.

As pectinas presentes nos substratos vegetais desempenham funções de agente hidratante e de material cimentador para a rede de fibras de celulose (SAKAI et al., 1993; THARKUR et al., 1997). Para promover a degradação da pectina no tecido vegetal é necessário um complexo enzimático, que tem sido detectado no fungo simbiote *L. gongylophorus* (SIQUEIRA et al., 1998; SILVA, 1999; SILVA et al., 2006). Essas enzimas também foram encontradas no líquido fecal de *A. colombica tonsipes* (MARTIN et al., 1975) e *A. sexdens rufopilosa* (SIQUEIRA, 1997; SILVA, 1999). Em um trabalho recente (SCHIØTT et al., 2010) foi demonstrado, através de estudos de proteoma, que as enzimas pectinolíticas produzidas no gongilídio do fungo simbiote são ingeridas mas não digeridas pela cortadeira *Acromyrmex* de modo que elas acabam no líquido fecal e se misturam

com o substrato novamente no ninho.

Dentre as enzimas produzidas pelo fungo *L. gongylophorus* que degradam a matéria vegetal gerando carboidratos simples que são assimilados pelas formigas, as amilases e pectinases parecem ser as mais importantes para a nutrição das formigas (SILVA et al., 2006).

Uma abordagem bastante interessante de controle de formigas cortadeiras é explorar a forte dependência da sobrevivência das formigas com o fungo simbiote *L. gongylophorus*: o controle do fungo levaria ao controle das formigas. Neste contexto, o melhor conhecimento das enzimas pectinolíticas produzidas pelo fungo se faz necessário.

Em um trabalho recente Schiøtt e colaboradores (SCHIØTT et al., 2010) identificaram, dentre as enzimas pectinolíticas, uma PG do *L. gongylophorus* e sua sequência foi depositada no banco de dados do NCBI sendo até o momento (fevereiro de 2011) a única sequência de PG do *L. gongylophorus* (ADV30326.1) depositada. Esta sequência foi então colhida e passou a fazer parte do banco de dados composto pelas 417 sequências únicas de PG de fungos deste trabalho.

No banco de dados do NCBI a sequência ADV30326.1 encontra-se depositada sem classificação (Endo- ou Exo-), além disso, o trabalho de referência da mesma (SCHIØTT et al., 2010) não menciona em nenhum momento uma possível classificação, já que a enzima não foi ainda obtida na forma isolada. Por se tratar de uma sequência única, classificá-la por meio de alinhamentos múltiplos tornou-se inviável, assim recorreu-se a classificação através do estudo de filogenia. Como pode ser constatado pela árvore filogenética representada na Figura 23.4, a sequência da PG do *L. gongylophorus* apresentou melhores relações filogenéticas com as sequências de Endo-PG especialmente com as sequências de Endo-PG do *Thanatephorus cucumeris* (ABH12113.1 e ACL80126.1).

A fim de obter um maior conhecimento a respeito da função desta PG do *L. gongylophorus*, sua sequência foi submetida ao banco de dados Pfam (<http://pfam.sanger.ac.uk/>) onde a mesma seria classificada quanto à família a qual pertence. O programa Pfam realiza um alinhamento múltiplo entre a sequência alvo, no caso a ADV30326.1, e as outras sequências que fazem parte do seu banco de dados. Este banco de dados é composto por sequências de proteínas que por sua vez estão devidamente distribuídas ou agrupadas em famílias com suas respectivas funções biológicas.

Como resultado a sequência ADV30326.1 foi incluída na família Glyco\_hydro\_28 (família 28 das glicosil hidrolases) e observou-se a presença do clã CL0268 (Pec\_lyase). Clã é um agrupamento de alto nível de famílias aparentadas realizado pelo Pfam. O clã CL0268 é uma superfamília onde todos os membros possuem  $\beta$ -hélice destra (sentido horário) semelhante à encontrada pela primeira vez em pectato liase (JENKINS et al. 1998).

De acordo com o site CAZY (<http://www.cazy.org>) os membros da família 28 das glicosil hidrolases - polygalacturonase (EC 3.2.1.15) incluindo endo- / exo-polygalacturonase (EC 3.2.1.67); exo-polygalacturonosidase (EC 3.2.1.82); rhamnogalacturonase (EC 3.2.1.171); endo-xylogalacturonan hydrolase (EC 3.2.1) e rhamnogalacturonan  $\alpha$ -L-rhamnopyranohydrolase (EC 3.2.1.40) - compartilham do mesmo mecanismo de ação com inversão de configuração do açúcar e a estrutura predominante é a ( $\beta$ ) - hélice.

Quanto ao mecanismo de ação, sabe-se que a hidrólise enzimática de ligações glicosídicas ocorre via catalise ácida na presença de dois resíduos de aminoácidos ácidos: um doador de próton e um nucleófilo/base. A hidrólise pode ocorrer via dois mecanismos diferentes que levam a retenção ou inversão da configuração anomérica do açúcar (DAVIES & HENRISSAT, 1995). Os mecanismos de retenção e de inversão são mostrados na figura 25.4, onde se pode observar que em ambos os casos o resíduo doador de próton (representado como ---A-H na Figura 25.4) encontra-se próximo do oxigênio glicosídico. No entanto, o resíduo nucleófilo/base (representado como ---B- na Figura 25.4) está mais distante do oxigênio do açúcar no caso da enzima de inversão (Figura 25.4 b), pois existe a necessidade de acomodar uma molécula de água entre o açúcar e o nucleófilo/base. Em enzimas que agem via retenção da configuração essa molécula de água, que ataca o carbono anomérico do açúcar em ambos os casos, encontra-se mais próxima do grupo doador de prótons (Figura 25.4 a). A diferente posição da molécula de água nos dois mecanismos resulta na distância média entre os dois resíduos catalíticos de aproximadamente 5,5Å nas enzimas de retenção enquanto que esta distância é de aproximadamente 10Å nas enzimas de inversão (MCCARTER & WITHERS, 1994).

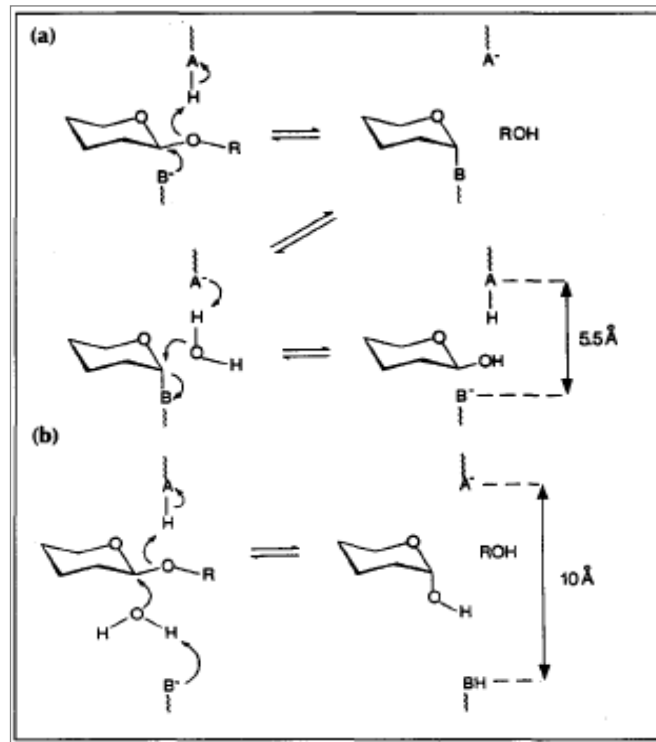


FIGURA 25.4 - Ilustração dos dois mecanismos propostos para hidrólise de ligação glicosídica, com retenção (a) e inversão (b) da configuração do açúcar (DAVIES & HENRISSAT, 1995).

#### 4.2.1 - Predição da estrutura secundária da PG do *L. gongylophorus*

O estudo estrutural teve início com a previsão de estrutura secundária da PG a fim de se obter uma primeira informação estrutural dada pela comparação da sequência da PG do *L. gongylophorus* com um banco de dados de estruturas secundárias estabelecidas experimentalmente. A predição foi feita com o programa PSIPRED v3.0 (<http://bioinf.cs.ucl.ac.uk/psipred>), e mostrou que a molécula é constituída por 3,5 % de alpha-hélices, 45,6 % de folhas-beta e 50,9 % de conexões (Figura 26.4).

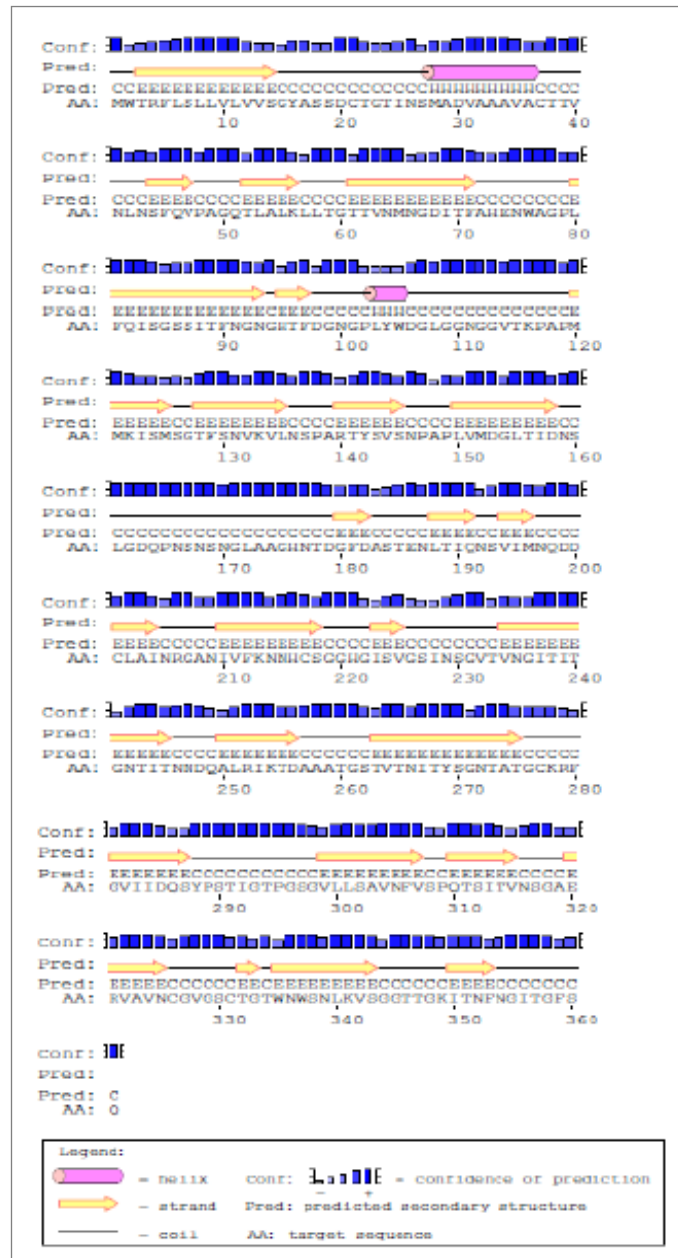


FIGURA 26.4 - Predição da estrutura secundária da PG feita com o programa Psipred (<http://bioinf.cs.ucl.ac.uk/psipred/>) Os cilindros rosas representam  $\alpha$ -hélices, as setas amarelas representam folhas- $\beta$  e os traços em preto representam loops e voltas.

#### 4.2.2 - Construção do modelo tridimensional da PG do *L. gongylophorus*

Após a conclusão do sequenciamento do genoma humano era natural que os esforços científicos fossem voltados para o estudo dos produtos gênicos (as proteínas). A era pós-genômica tem sido um período onde essas moléculas têm sido estudadas com o objetivo de uma maior compreensão de suas atividades em vias

metabólicas e doenças e tem sido responsável pelo desenvolvimento de novas técnicas de análise. Uma das mais importantes maneiras de se estudar uma proteína e obter informações a nível molecular é determinar a estrutura da molécula e esse conhecimento tem sido fundamental no delineamento experimental racional de diferentes áreas.

O crescente número de estruturas 3D resolvidas de proteínas gerou uma grande quantidade de informação estrutural das mesmas. Com o objetivo de organizar tais informações foi criada uma base de dados, que armazena as coordenadas atômicas de macromoléculas biológicas, chamada **Protein Data Bank** (PDB), disponível no site [www.rcsb.org/pdb](http://www.rcsb.org/pdb). Atualmente estão depositadas no banco de dados PDB mais de 79.000 estruturas 3D de proteínas. Destas, até fev/2012, treze são de PG, sendo quatro de bactérias e nove de eucariotos (todos de fungos). Das nove estruturas de fungos conhecidas, três delas são, na verdade, estruturas das proteínas complexadas. Portanto, existem somente seis estruturas 3D de PG de fungos descritas na literatura. Não há qualquer conhecimento a respeito de estrutura 3D da PG do *L. gongylophorus* na literatura e nenhuma estrutura da mesma depositada no banco de dados PDB. Assim, construir um modelo 3D da mesma por meio de predição estrutural pode vir a fornecer informações interessantes a respeito da enzima no que diz respeito à função e mecanismo de ação.

O modelo tridimensional da PG do *L. gongylophorus* foi construído por “modelagem por homologia” que, como descrito no capítulo 3.5, é uma técnica baseada na constatação de que moléculas com bom grau de semelhança sequencial (identidade) apresentam enovelamento (estrutura 3D) semelhantes. A utilização deste método somente seria possível se houvesse uma molécula de identidade razoável com a PG de interesse (molécula alvo) e que tivesse sua estrutura determinada (molécula molde). A procura da 'molécula molde' sucedeu-se de duas maneiras distintas:

- analisando as sequências na árvore filogenética, construída neste trabalho e mostrada no capítulo 4.3, e procurando por estruturas de proteínas próximas à PG de *L. gongylophorus*;
- fazendo busca automática pelo programa SWISS-MODEL.

Através do uso de ambas as técnicas chegaram-se à estrutura 3D da

Endo-PG de *Chondrostereum purpureum* (antes conhecido como *Stereum purpureum*) cujo código de entrada no PDB é 1K5C. O *C. purpureum* é um fungo patogênico de macieiras que causa a doença conhecida como *silver-leaf disease* e é conhecido por secretar grandes quantidades de diversas Endo-PG, que têm sido associadas aos sintomas causados pela doença (MIYAIRI et al., 1985). Existem outras duas entradas de coordenadas para a mesma Endo-PG de *C. purpureum* no PDB, que se referem a dois complexos. Um complexo binário (PDB 1KCC) contendo uma molécula simples de ácido monogalacturônico como ligante, que foi obtido pela co-cristalização da enzima com ácido trigalacturônico e, portanto, o ligante deve ser o produto da catálise. O outro complexo, ternário, (PDB 1KCD) foi preparado através de *soaking* com duas moléculas de ácido monogalacturônico, sendo que uma delas ocupa a mesma posição do ácido galacturônico do complexo binário (SHIMIZU et al., 2002).

A estrutura tridimensional da Endo-PG de *C. purpureum* 1K5C foi usada na construção do modelo 3D da PG do *L. gongylophorus*. No entanto, como algumas mudanças conformacionais haviam sido identificadas no sítio ativo das três estruturas (SHIMIZU et al., 2002), a estrutura do complexo ternário foi usada para gerar um modelo da PG para uma melhor análise do sítio ativo.

A fim de se obter o grau de identidade entre a sequência do molde (1K5C) e da PG de *L. gongylophorus* realizou-se um alinhamento entre estas sequências (Figura 27.4) com o programa *ClustalW* (THOMPSON et al., 1994) mostrando uma identidade de aproximadamente 50% entre as sequências. Trata-se de uma identidade considerável uma vez que segundo KOPP e colaboradores (2004) uma identidade acima de 50% o modelo gerado é de alta qualidade e tem se mostrado tão preciso quanto às estruturas obtidas por raios-X de baixa resolução.

O modelo 3D da PG do fungo foi construído com o pacote de programas SWISS-MODEL (ARNOLD et al, 2006), como mencionado no capítulo 3.5 deste trabalho, partindo das coordenadas da molécula molde (1K5C) e do alinhamento entre as sequências de PG do *Leucoagaricus* e Endo-PG do *Chondrostereum*.



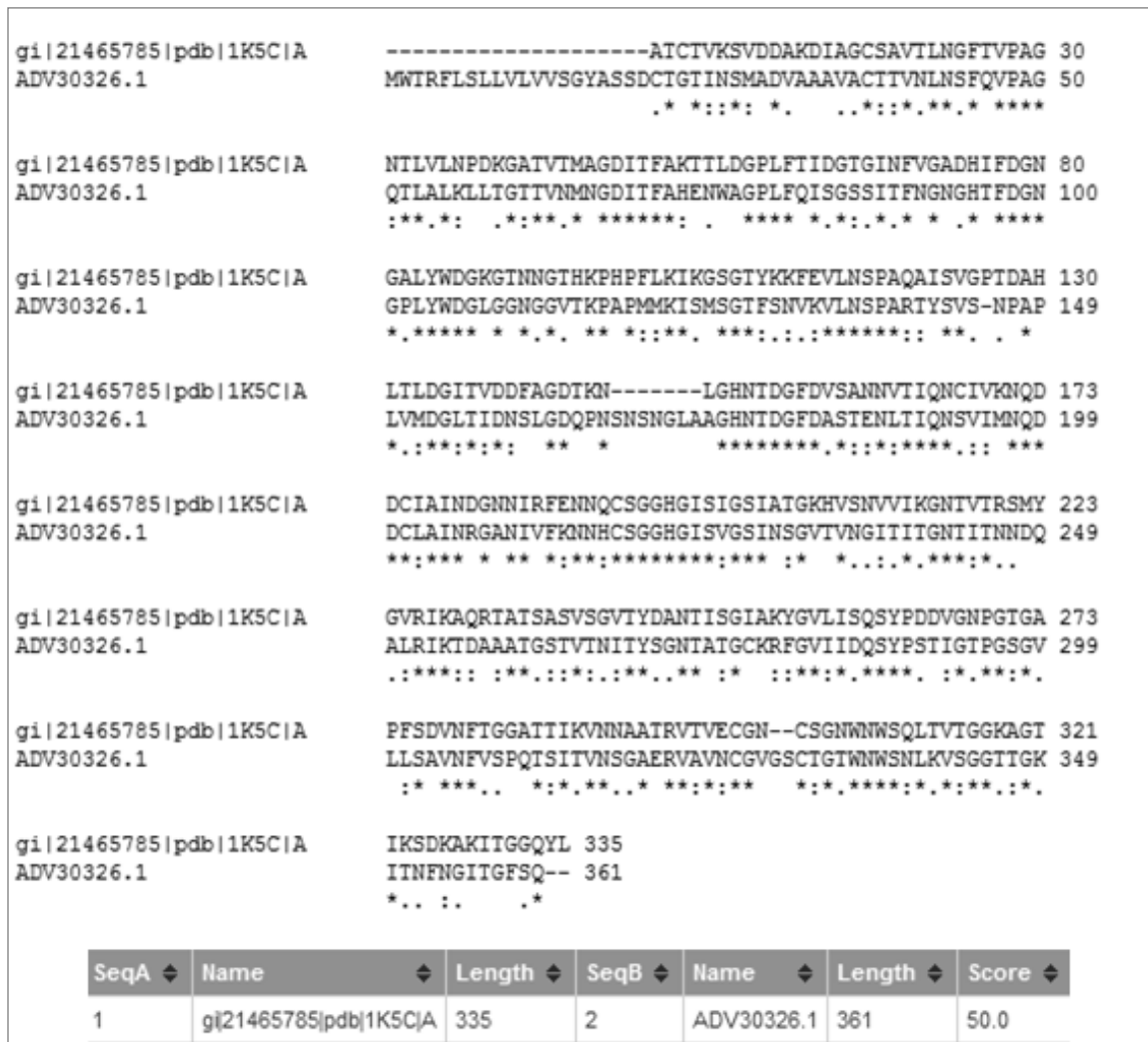


FIGURA 27.4 - Alinhamento e tabela de escore dados pelo programa *ClustalW* (THOMPSON et al., 1994) entre as seqüências de PG do *L. gonylophorus* (ADV30326.1) e de Endo-PG do *C. purpureum* (1K5C).

### 4.2.3 - Descrição do modelo estrutural da PG de *L. gonylophorus*

A PG de *L. gonylophorus* é uma molécula constituída por 361 resíduos de aminoácidos, de massa molecular aproximada de 37 kDa e, de acordo com o descrito na literatura para outras PG de fungos, deve ser monomérica. A seqüência de aminoácidos, deduzida da seqüência de nucleotídeos, está ilustrada na Figura 28.4 onde estão em destaque os segmentos que têm sido identificados como conservados em PG de fungos (MARKOVIC & JANECEK, 2001). Um alinhamento representativo de PG de fungos composto pelas 63 seqüências de Endo-PG e 38 seqüências de Exo-PG de fungos (item 4.3.1 do capítulo 4) foi realizado (Apêndice B) comprovando o quão conservado são os resíduos em

destaque na Figura 28.4. A importância de alguns desses resíduos serão detalhados a seguir.

10	20	30	40	50	60
MWTRFLSLLV	LVVSGYASSD	CTGTINSMAD	VAAAVACTIV	NLNSFQVPAG	QTLALKLLTG
70	80	90	100	110	120
TTVNMNGDIT	FAHENWAGPL	FQISGSSITF	NGNGHTFDCN	GPLYWDGLGG	NGGVTKPAPM
130	140	150	160	170	180
MKISMSGTFS	NVKVLNSPAR	TYSVSNPAPL	VMDGLTIDNS	LGDPNSNSM	GLAAGHNTDG
190	200	210	220	230	240
FDASTENLII	QNSVIMNQDD	CLAINRGANI	VFKNNHCSGG	HGISVGSINS	GVTVNGITIT
250	260	270	280	290	300
GNTITNNDQA	LRIKTDAAT	GSTVTNITYS	GNTATGCKRF	GVIIDQSYPS	TIGTPGSGVL
310	320	330	340	350	360
LSAVNFVSPQ	TSITVNSGAE	RVAVNCGVGS	CTGTWNWSML	KVSGGTTGKI	TNENGITGFS
Q					

FIGURA 28.4 - Sequência de aminoácidos da PG de *Leucoagaricus gongylophorus* deduzida da sequência de nucleotídeo. Os resíduos em vermelho são os conservados em PG de fungos (MARKOVIC & JANECEK, 2001).

A estrutura tridimensional da PG do *Leucoagaricus gongylophorus* é um monômero com um enovelamento de beta-hélice de mão direita (Figura 29.4) e apresenta uma fenda onde se localiza o sítio catalítico, mostrado na parte superior da Figura 30.4. Nesta figura estão representados os três resíduos de ácido aspártico conservados nas sequências de PG de fungos e que provavelmente estão envolvidos na catálise (Asp<sup>179</sup>, Asp<sup>199</sup> e Asp<sup>200</sup>). Esta estrutura  $\beta$ -helicoidal destra é comum em EndoPG de *Aspergillus niger* (VAN SANTEN et al, 1999), *Chondrostereum purpureum* (SHIMIZU et al, 2000) e *Colletotrichum lupini* (BONIVENTO et al, 2008) (Figura 31.4), assim como nas demais pectinases (SHIMIZU et al., 2001).

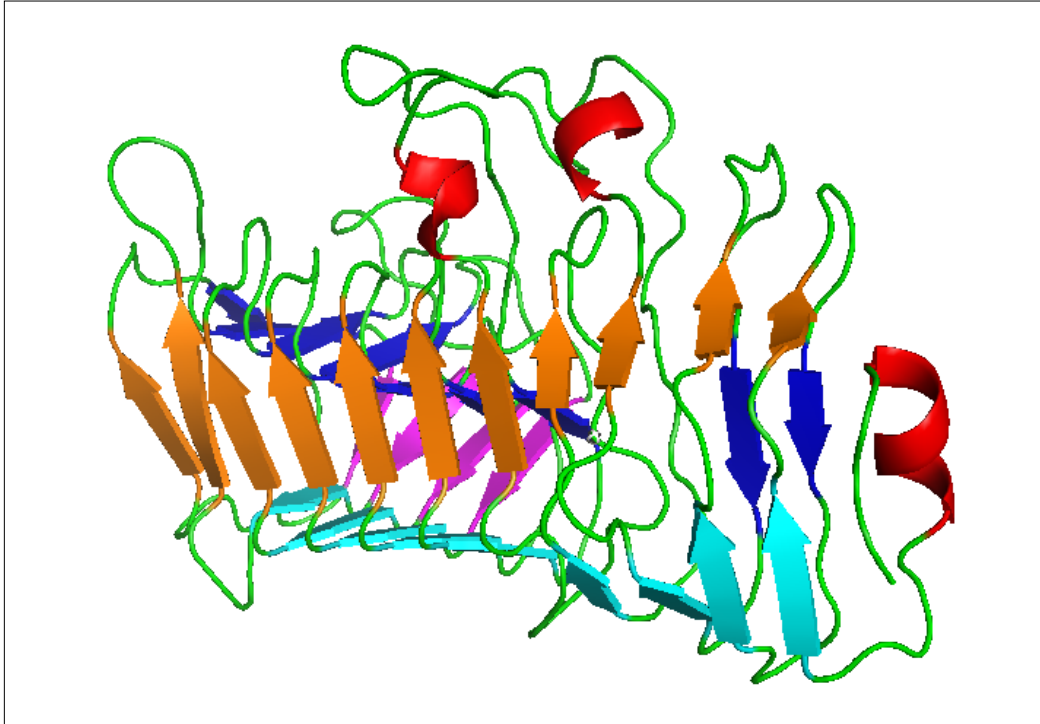


FIGURA 29.4 - Representação esquemática da estrutura  $\beta$ -hélice paralela destra do modelo estrutural da PG do *L. gongylophorus* construído com o programa SWISS-MODEL. A molécula é visualizada pelo lado N-terminal.

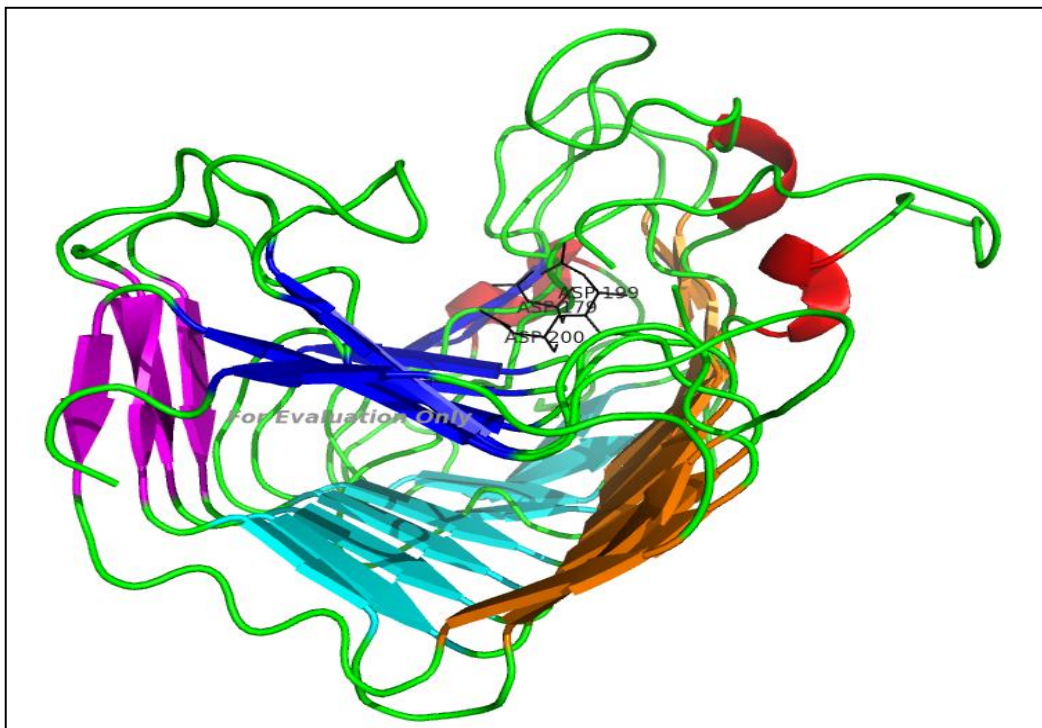


FIGURA 30.4 - Representação esquemática do modelo estrutural da PG de *L. gongylophorus* visualizada do lado C-terminal focalizando o sítio catalítico (grande fenda aberta formada pela folha- $\beta$  PB1 e dos *loops* que se encontram dos dois lados da folha- $\beta$  PB1). Os resíduos de ácido aspártico envolvidos provavelmente na catálise (Asp<sup>179</sup>, Asp<sup>199</sup> e Asp<sup>200</sup>) estão representados na forma de linhas pretas.

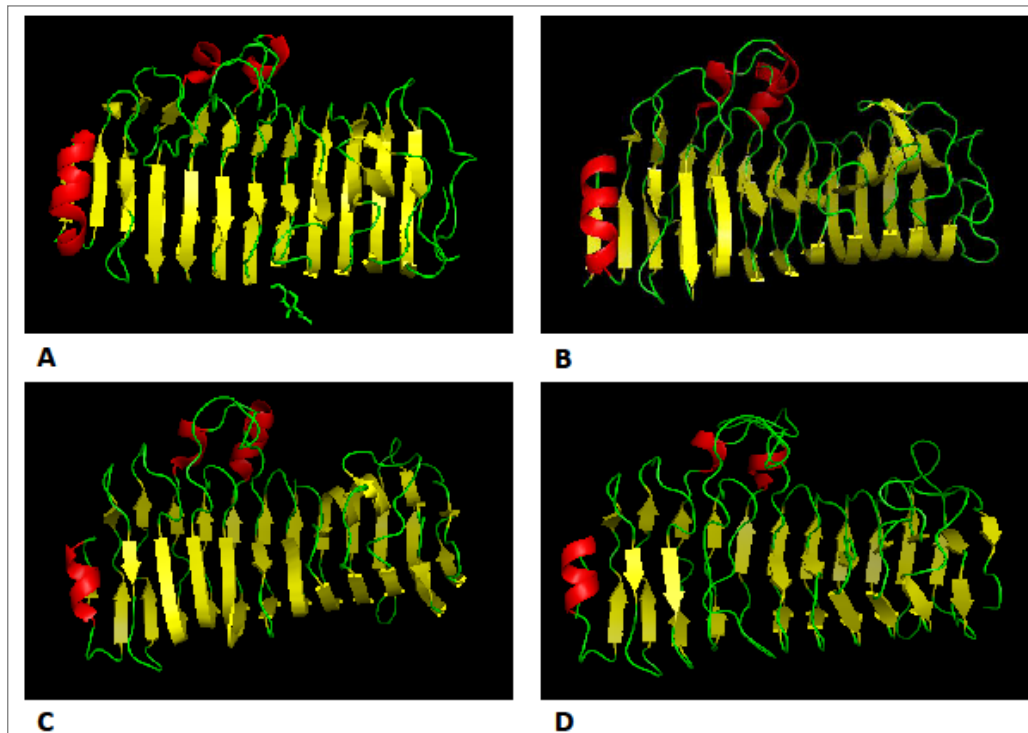


FIGURA 31.4 - Representação 3D de algumas Endo-PG de fungos e do modelo da PG de *L. gongylophorus* (D). Sendo o N-terminal à esquerda e o C-terminal à direita. (A) Endo-PG do *A. niger* (código PDB 1CFZ). (B) Endo-PG do *C. lupini* (código PDB 2IQ7). (C) Endo-PG do *C. purpureum* (código PDB 1K5C).

A estrutura  $\beta$ -hélice do *L. gongylophorus* é formada por 4 folhas- $\beta$  designadas PB1, PB2a, PB2b e PB3 compreendendo 10 voltas completas da fita 1 na PB2b até a fita 32 na PB3 (Figura 32.4).

As regiões de *loop* formam uma grande fenda aberta na parte exterior da  $\beta$ -hélice juntamente com a folha- $\beta$  PB1 (Figura 30.4). É nesta região que se localiza o sítio catalítico da molécula. Nele estão localizados os 8 resíduos (Asn<sup>177</sup>, Asp<sup>179</sup>, Asp<sup>199</sup>, Asp<sup>200</sup>, His<sup>221</sup>, Gly<sup>222</sup>, Arg<sup>252</sup> e Lys<sup>254</sup> – numeração de acordo com o modelo) extremamente conservados (Figura 32.4) tanto em Endo- como em Exo-PG (KESTER et al, 1996). A saber, os três resíduos de ácido aspártico estão diretamente relacionados com a catálise enquanto os resíduos de arginina e lisina podem desempenhar um papel importante na ligação com o substrato (VAN SANTEN et al, 1999). Os demais resíduos de aminoácidos são aqueles ressaltados em vermelho na Figura 28.4)

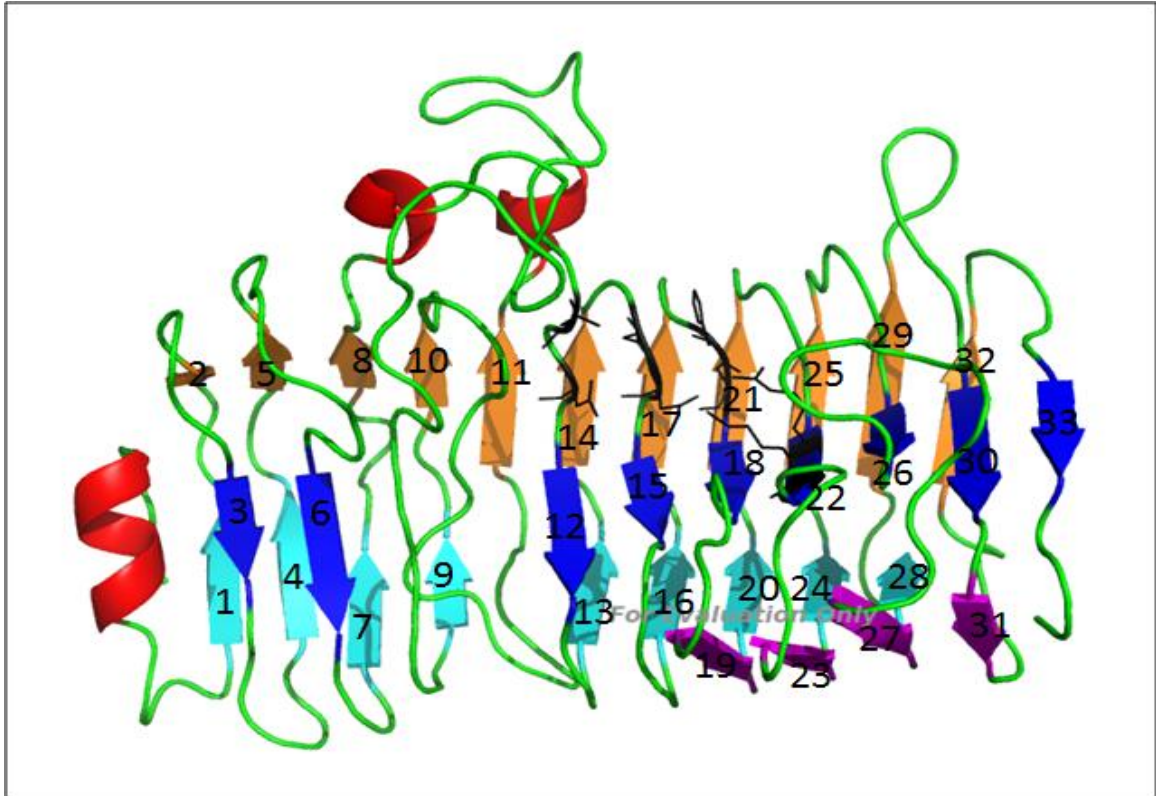


FIGURA 32.4 - Representação esquemática das folhas- $\beta$  PB1, PB2a, PB2b e PB3 mostradas em azul escuro, lilás, azul claro e marrom, respectivamente. Juntas elas formam a estrutura  $\beta$ -hélice da PG do *L. gongylophorus*. Os 8 resíduos (Asn<sup>177</sup>, Asp<sup>179</sup>, Asp<sup>199</sup>, Asp<sup>200</sup>, His<sup>221</sup>, Gly<sup>222</sup>, Arg<sup>252</sup> e Lys<sup>254</sup>) extremamente conservados tanto em Endo- quanto em Exo-PG são representados em forma de linhas pretas.

#### 4.2.4 - Validação do modelo tridimensional da PG do *L. gongylophorus*

O modelo tridimensional da PG de *L. gongylophorus* foi submetido à avaliação de sua qualidade em relação aos fatores estereoquímicos com o programa computacional PROCHECK (LASKOWSKI *et al.*, 1993). A Figura 33.4 representa o mapa de Ramachandran (RAMACHANDRAM *et al.*, 1963) resultante do modelo obtido através da molécula molde 1K5C. Quanto à posição dos resíduos analisados no mapa (333 resíduos no total), o modelo apresentou 79,4% dos resíduos de aminoácidos em áreas consideradas muito favoráveis, 17,7% em regiões favoráveis, 2,2% em regiões generosamente permitidas e, finalmente, 0,7% de resíduos em regiões não permitidas. Quanto aos resíduos de glicina (44 no total), estes são representados por triângulos no mapa de Ramachandran pelo fato deles não possuem quiralidade no C $\alpha$ , já que eles possuem como cadeia lateral um átomo de hidrogênio.

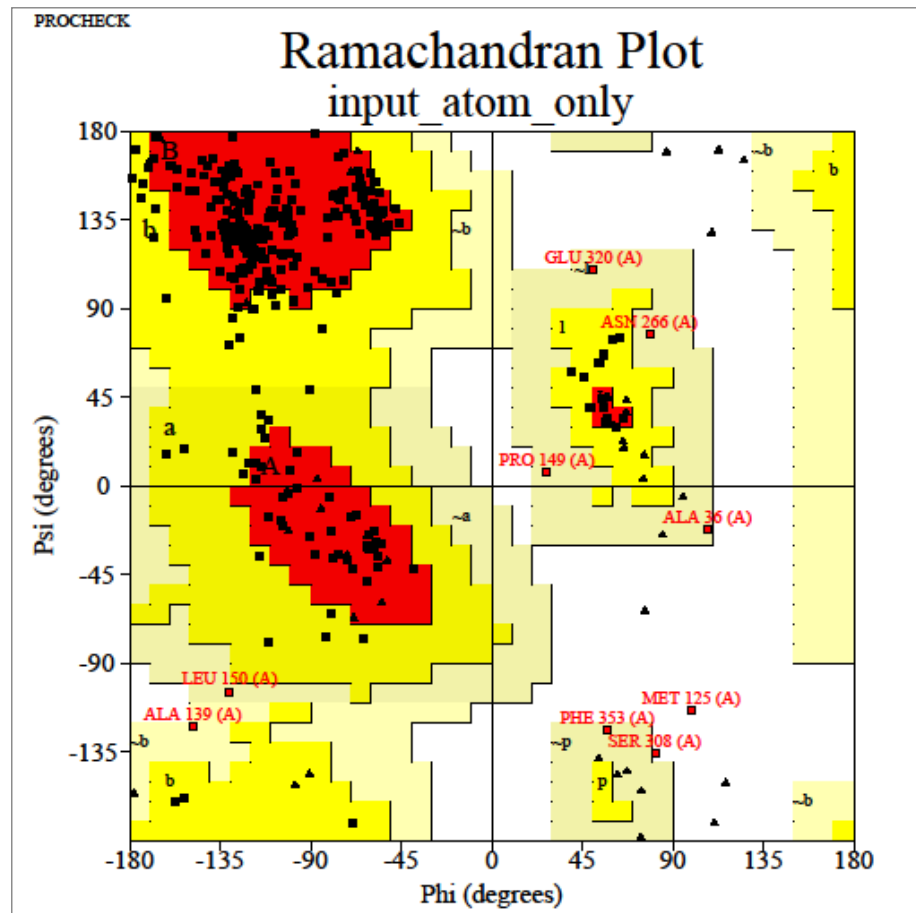


FIGURA 33.4 - Mapa de Ramachandran do modelo sem qualquer refinamento da PG do *L. gongylophorus*. Os ângulos Phi e Psi estão em graus. Os triângulos representam os resíduos de glicina.

Os aminoácidos em destaques no mapa de Ramachandran (Figura 34.4) são Ala<sup>36</sup>, Met<sup>125</sup>, Ala<sup>139</sup>, Pro<sup>149</sup>, Leu<sup>150</sup>, Asn<sup>266</sup>, Ser<sup>308</sup>, Glu<sup>320</sup> e Phe<sup>353</sup>. Estes aminoácidos encontram-se longe do sítio ativo da proteína com exceção da Ala<sup>139</sup> e Glu<sup>320</sup>. Uma localização mais precisa destes aminoácidos pode ser observada na Figura 34.4.



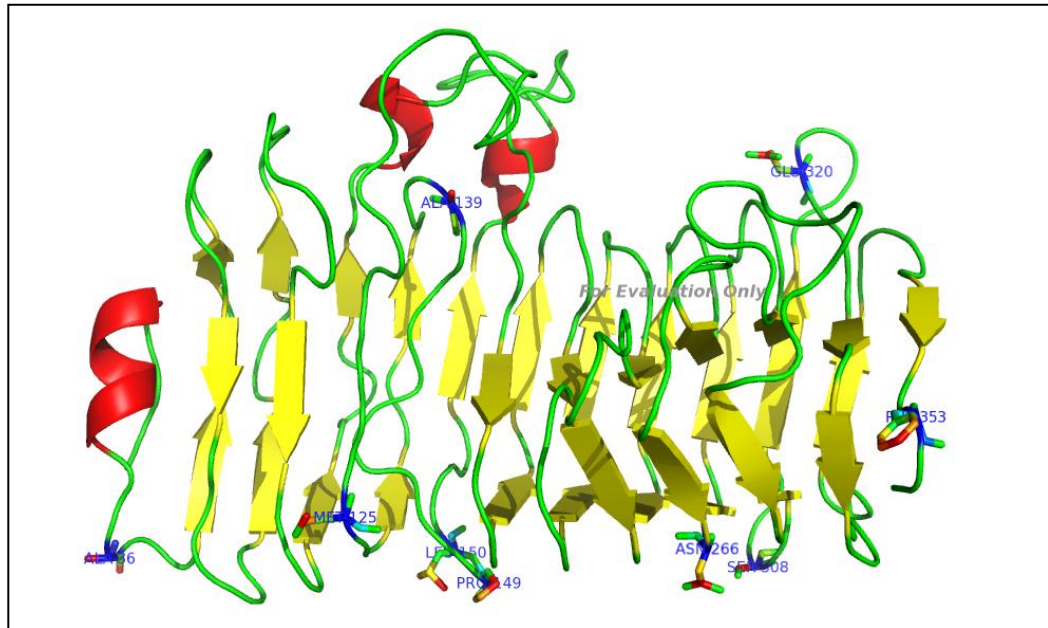


FIGURA 34.4 - Modelo 3D da PG de *L. gongylophorus* destacando os aminoácidos localizados nas regiões generosas e proibidas do mapa de Ramachandran.

Na Figura 35.4 estão mostrados os parâmetros da cadeia principal obtidos pelo PROCHECK através da comparação estrutural do modelo com outras estruturas refinadas na resolução de 1,0Å (por modelagem molecular a resolução de 2,0Å é comumente usada). As regiões destacadas na cor roxa dos gráficos representam as faixas de resultados obtidos por essas estruturas. A linha central corresponde ao ajuste mínimo quadrático para a inclinação média em função da resolução enquanto que a largura da banda corresponde à variação de um desvio padrão em torno da média. A Figura 35.4 **a** ilustra a qualidade do mapa de Ramachandran com base no número de resíduos presentes nas regiões ditas mais favorecidas. A Figura 35.4 **b** ilustra a planaridade da ligação peptídica, ou seja, o desvio padrão para o ângulo torsional ômega ( $\omega$ ) da estrutura. Quanto menor for este desvio, mais próximo de 180° estará o ângulo, o que implica em uma ligação peptídica perfeitamente planar. A Figura 35.4 **c** representa as más interações de não-ligantes medida pelo número de maus contatos por 100 resíduos. A distorção tetraédrica do C alfa (Figura 35.4 **d**) é medida por meio do cálculo do desvio padrão do ângulo de torção zeta, que é um ângulo de torção fictício na medida em que não está definida sobre qualquer ligação efetiva na estrutura. Sendo definido pelos seguintes átomos: C alfa, N, C, e C beta. A energia de ligação de hidrogênio da cadeia principal (Figura 35.4 **e**) é dada pelo desvio padrão das energias de ligação de hidrogênio para ligações da cadeia principal de hidrogênio. E finalmente, o

parâmetro  $G_{\text{factor}}$  (Figura 35.4 f), uma medida da normalidade total da estrutura. O valor global é obtido a partir de uma média de todos os diferentes  $G_{\text{factores}}$  para cada resíduo na estrutura.

Observando os gráficos verifica-se que o modelo apresenta alguns fatores estereoquímicos (a, c e f) fora dos padrões determinados para estruturas resolvidas a 1Å.

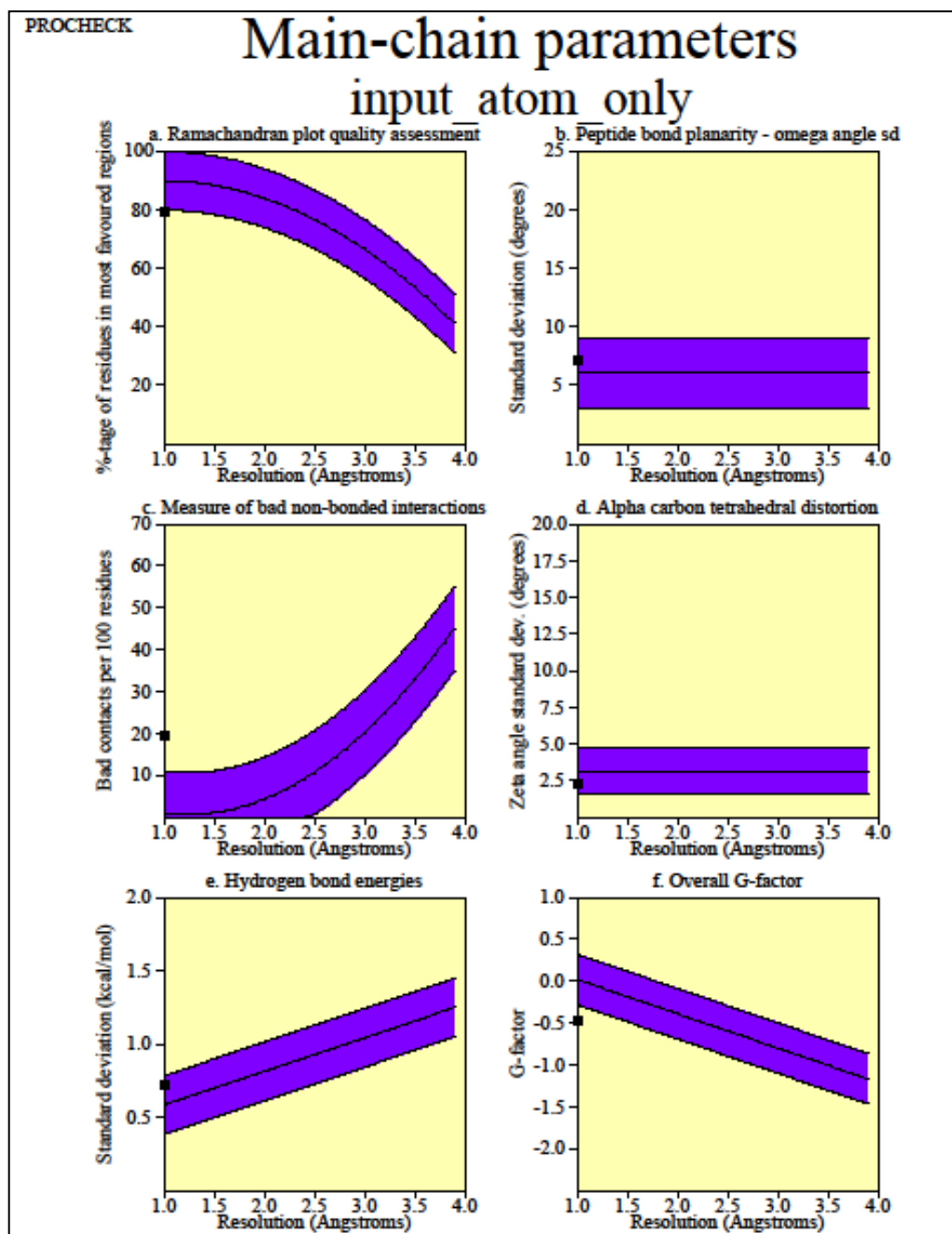


FIGURA 35.4 Estatísticas de alguns parâmetros estereoquímicos da cadeia principal (representado pelo quadrado em preto) do modelo 3D da PG de *L. gongylophorus* quando comparados a estruturas refinadas a mesma resolução: **a.** qualidade do mapa de Ramachandran, **b.** planaridade da ligação peptídica (ângulo ômega), **c.** medidas ruins de interações entre não-ligantes, **d.** distorção tetraédrica, **e.** energia de ligação de hidrogênio e **f.**  $G_{\text{factor}}$ .



Como o modelo apresentou 0,7% de seus resíduos em regiões não permitidas, técnicas de refinamento pelo programa Coot (EMSLEY & COWTAN, 2004) foram aplicadas.

O modelo refinado foi analisado com o programa PROCHECK (LASKOWSKI *et al.*, 1993), VERIFY 3D (LUTHY *et al.*, 1992) e o WHAT IF (VRIEND, 1990). Os resultados são mostrados a seguir. A Figura 36.4 representa o mapa de Ramachandran (RAMACHANDRAM *et al.*, 1963) resultante do modelo obtido através da molécula molde 1K5C, porém refinado. Quanto à posição dos resíduos analisados no mapa (333 resíduos no total), o modelo apresentou 84,5% dos resíduos de aminoácidos em áreas consideradas muito favoráveis, 13,4% em regiões favoráveis, 2,2% em regiões generosamente permitidas e finalmente 0% de resíduos em regiões não permitidas. Quanto aos resíduos de glicina (44 no total), estes são representados por triângulos no mapa de Ramachandran. Esses resultados mostram significativa melhora no modelo uma vez que 2 resíduos (Met<sup>125</sup> e Ser<sup>308</sup>), que se encontravam em regiões proibidas do mapa de Ramachandran, após o refinamento passaram a ocupar posições mais favoráveis, ou seja, passaram a apresentar fatores estereoquímicos melhores. Além disso, a porcentagem de resíduos em áreas consideradas muito favoráveis aumentou, passando de 79,4% para 84,5%. Comparando com o mapa de Ramachandran do molde (1K5C) (não demonstrado), onde 85,9% dos resíduos estão em áreas consideradas muito favoráveis e 0,4% dos resíduos em áreas proibidas, o modelo após refinamento mostra-se muito satisfatório principalmente por não apresentar resíduos em áreas proibidas do mapa.

Quanto aos aminoácidos (Ala<sup>36</sup>, Met<sup>125</sup>, Ser<sup>145</sup>, Pro<sup>149</sup>, Asn<sup>266</sup>, Arg<sup>321</sup> e Phe<sup>353</sup>) em destaques no mapa de Ramachandran e representados na Figura 38.4, todos eles encontram-se longe do sítio ativo da proteína. Uma localização mais precisa destes aminoácidos pode ser vista na Figura 37.4.

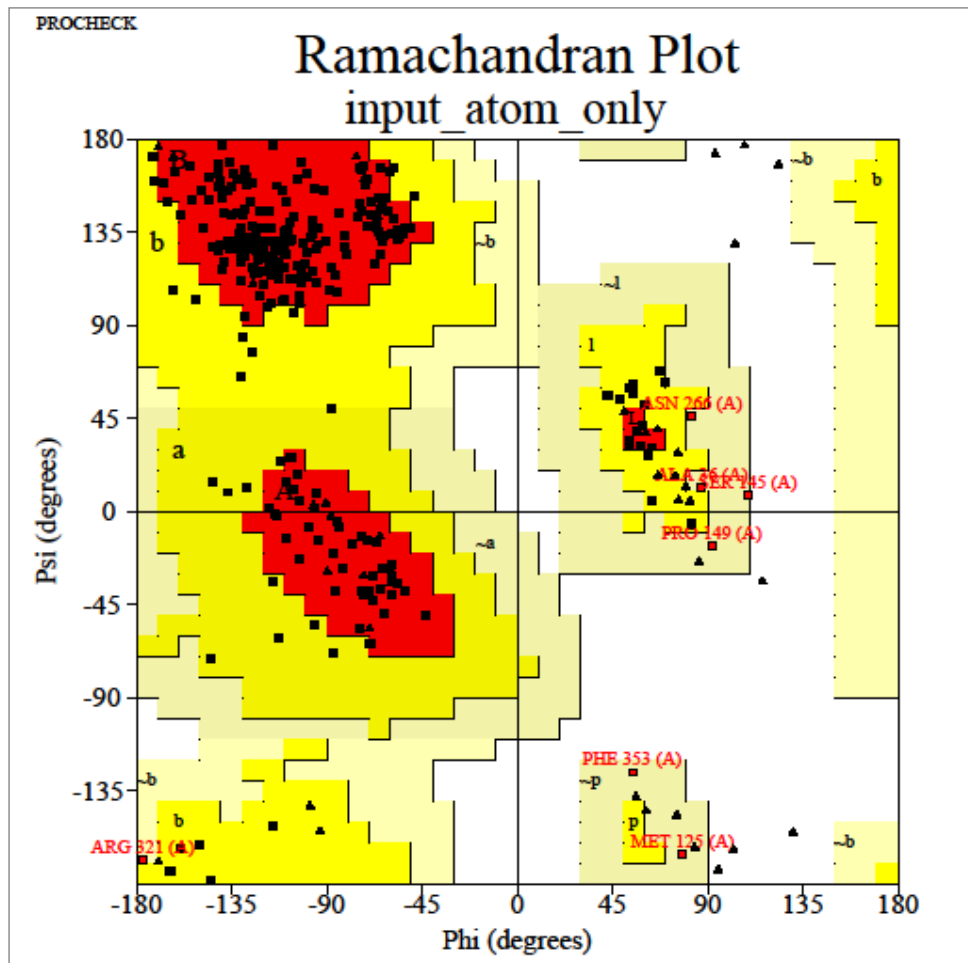


FIGURA 36.4 - Mapa de Ramachandran do modelo refinado pelo “coot” (EMSLEY & COWTAN, 2004) da PG do *L. gongylophorus*. Os ângulos Phi e Psi estão em graus. Os triângulos representam os resíduos de glicina.

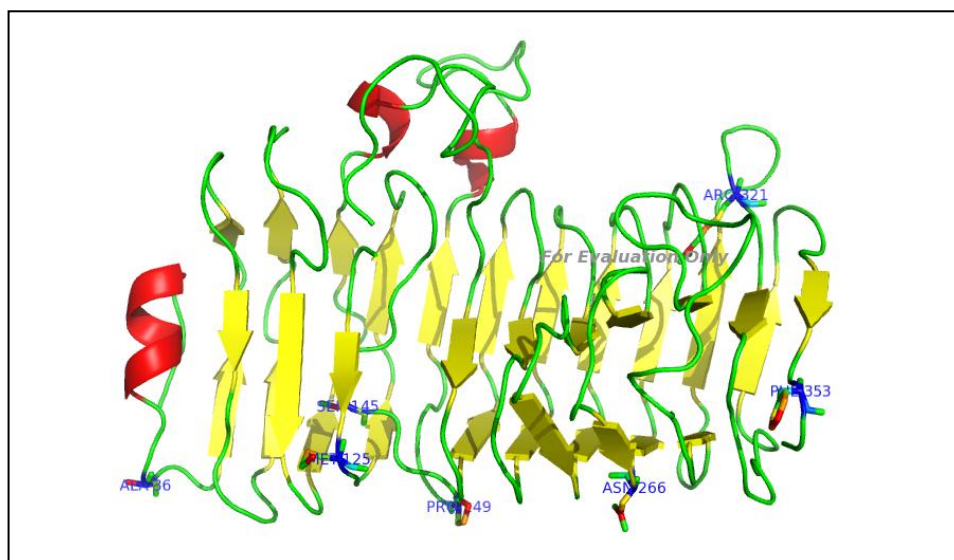


FIGURA 37.4 - Modelo 3D refinado da PG de *L. gongylophorus* destacando os aminoácidos (Ala36, Met125, Ser145, Pro149, Asn266, Arg321 e Phe353) localizados nas regiões generosas do mapa de Ramachandran.

Com relação aos parâmetros da cadeia principal obtidos pelo PROCHECK através da comparação estrutural do modelo refinado com outras estruturas refinadas na resolução de 1,0Å (Figura 38.4), os parâmetros **b**, **c** e **f** encontram-se fora dos padrões estabelecidos. Porém, quando esses resultados são comparados com os do modelo não refinado verifica-se melhora na qualidade do mapa de Ramachandran.

A análise do modelo pelo programa VERIFY 3D ([http://nihserver.mbi.ucla.edu/Verify\\_3D/](http://nihserver.mbi.ucla.edu/Verify_3D/)) mostrou-se bastante satisfatória, pois nenhum resíduo se encontrou abaixo da linha zero, o que significa que todos os aminoácidos da PG do *L. gongylophorus* apresentam uma conformação favorável (Figura 39.4).

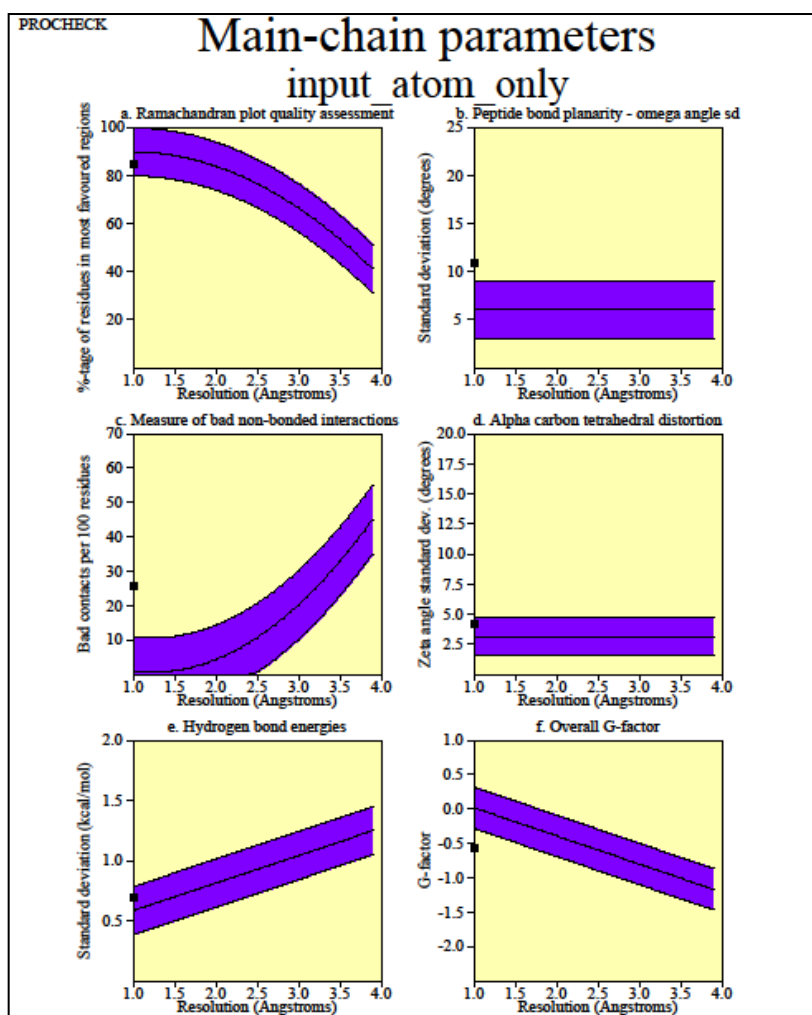


FIGURA 38.4 Estatísticas de alguns parâmetros estereoquímicos da cadeia principal (representado pelo quadrado em preto) do modelo 3D refinado da PG de *L. gongylophorus* quando comparados a estruturas refinadas a mesma resolução: **a.** qualidade do mapa de Ramachandran, **b.** planaridade da ligação peptídica (ângulo ômega), **c.** medidas ruins de interações entre não-ligantes, **d.** distorção tetraédrica, **e.** energia de ligação de hidrogênio e **f.**  $G_{\text{factor}}$ .

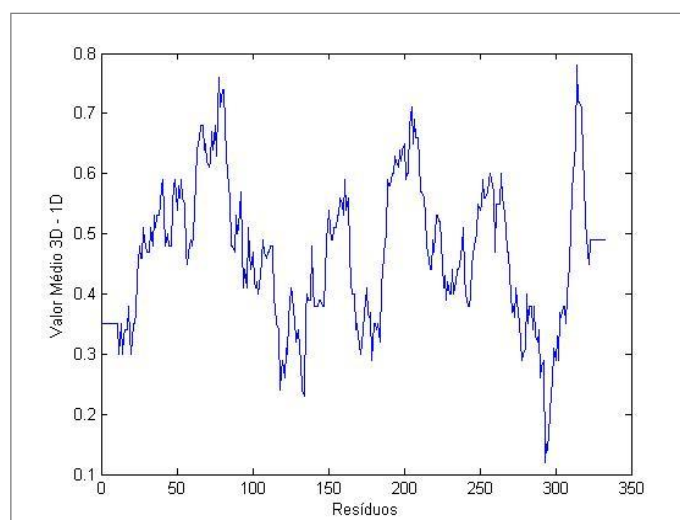


FIGURA 39.4 - Gráfico representando os resultados obtidos pelo programa VERIFY 3D (LUTHY et al, 1992) para o modelo após refinamento. O menor valor médio 3D-1D foi 0.12 (Val293) e o maior 0.78

Por fim, o modelo refinado foi avaliado pela ferramenta OLDQUA pertencente ao programa QUALTY que está inserido no pacote de programas WHAT IF (VRIED, 1990). Esta ferramenta avaliou as vizinhanças atômicas do modelo e como resultado observou-se que dos 333 resíduos avaliados pelo programa, apenas 5 resíduos adquiriram valores inferior a -5 (Apêndice C). Embora os resíduos que estão abaixo de -5 necessitem de maneira geral ser checados, isso não foi necessário uma vez que não é incomum alguns resíduos sofrerem pequenos desvios, além disso, a pontuação média referente aos 328 resíduos (os 5 resíduos que apresentaram valores abaixo do padrão não foram computados) foi de -0,388, o que significa que o modelo é de muito boa qualidade de acordo com os valores de referências do programa.

#### 4.2.5 - O modelo estrutural e o sítio ativo da enzima

A sobreposição (Figura 40.4) da cadeia principal da estrutura tridimensional de Endo-PG de *C. purpureum* (em azul) e do modelo 3D da PG de *L. gongylophorus* (em vermelho) mostra uma grande homologia estrutural entre as duas enzimas (Figura 40.4), sendo que 318 átomos estão em posições estruturalmente equivalentes com rms de 0,029 Å. Como é possível observar na Figura 40.4 há alguns desvios de sobreposição em regiões de conexões, como nos três *loops* destacados. Dois desses *loops* são maiores na molécula de PG do *L. gongylophorus* e um deles é maior na molécula de Endo-PG do *C. purpureum*, o que leva a pior superposição das moléculas nessas regiões.

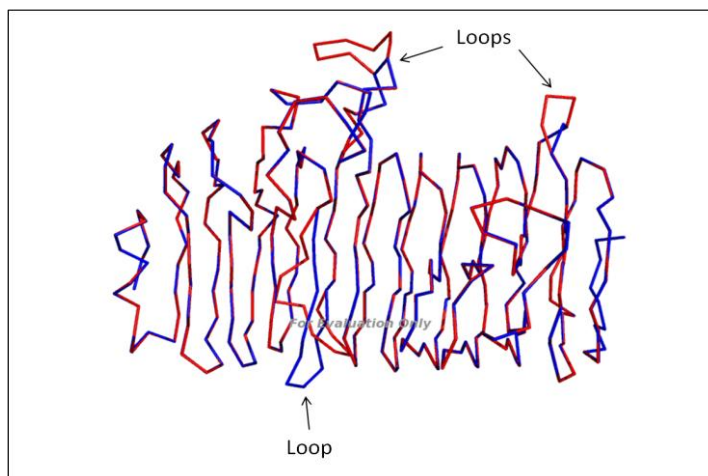


FIGURA 40.4 - Sobreposição da estrutura tridimensional de Endo-PG de *C.purpureum* (azul) e do modelo 3D da PG de *L.gongylophorus* (vermelho). O valor de rms é de 0,029Å.

Membros da família 28 das glicosil hidrolases (na qual a PG em estudo foi classificada) têm sido descritos por compartilharem do mesmo mecanismo de ação, com inversão de configuração do açúcar. O mecanismo proposto para essas enzimas sugere que os dois aminoácidos aspárticos envolvidos na catálise estão a aproximadamente 9 Å um do outro (MCCARTER & WITHERS, 1994; DAVIES & HENRISSAT, 1995). Para que a enzima aja por meio do mecanismo de inversão seria necessário haver uma grande separação entre os dois ácidos carboxílicos essenciais presentes em faces opostas da cavidade catalítica a fim de que a água juntamente com o substrato se ajustem entre esses dois grupos carboxílicos. Já para o mecanismo de retenção a distância média entre os ácidos carboxílicos seria de 4,8 a 5,3 Å (conforme Figura 25.4). Estruturas 3D de PG resolvidas mostram que estas enzimas têm três aminoácidos aspárticos (e não dois) em posições que poderiam favorecer suas atuações na catálise e em todos esses casos a distância entre os aspárticos é de aproximadamente 5 Å e não 9 Å (VAN SANTEN et al, 1999; ARMAND et al., 2000; FEDERICI et al, 2001; CHO et al, 2001; SHIMIZU et al, 2002).

A estrutura da PG de *C. purpureum*, usada como molécula molde neste trabalho, também mostra que os três resíduos aspárticos conservados presentes no sítio catalítico (e envolvidos no mecanismo catalítico da enzima) encontram-se a uma distância de 4,8 a 5,3 Å, muito menor do que a esperada, de aproximadamente 9 Å (SHIMIZU et al, 2002). Portanto, o mecanismo de inversão da configuração do carbono anomérico do substrato proposto para as enzimas da família 28 das glicosil hidrolases não pode ser generalizado e algumas enzimas, embora atuem com esse mecanismo, apresentam o sítio ativo diferente do geralmente proposto (MCCARTER & WITHERS, 1994; DAVIES & HENRISSAT, 1995). O nosso entendimento é que, a princípio, a molécula de água não afetaria a distância entre os aspárticos catalíticos.

Na estrutura da PG de *C. purpureum* e também no modelo construído para *L. gongylophorus*, existem três ácidos aspárticos situados no sítio catalítico, que são conservados em todas as sequências analisadas de PG de fungos (Asp<sup>179</sup>, Asp<sup>199</sup> e Asp<sup>200</sup>, numeração da PG de *L. gongylophorus*) (Figura 41.4). Nesta figura foi realizado o alinhamento de todas as proteínas de fungos com estrutura 3D conhecida e nela estão representados os 5 motivos conservados encontrados para o grupo formado pelas 63 sequências de Endo-PG e 38 de Exo-PG na primeira parte deste trabalho. Ao analisar estas 63 sequências de Endo-PG (Apêndice A) observou-se a presença das sequências BAA96351.1 do fungo *C. purpureum* e da sequência

ABL01533.1 do fungo *Colletotrichum lupini*. Ambas as sequências possuem estrutura 3D conhecida cujo PDB são respectivamente 1K5C e 2IQ7. Quando estas sequências foram analisadas pelo MEME, como resultado a sequência ABL01533.1 apresentou todos os 5 motivos enquanto a BAA96351.1 apresentou apenas os motivos 2, 4 e 5 (Apêndice A). Com base nestes resultados em conjunto com o alinhamento (Figura 41.4) e o escore obtido (Apêndice D) foi possível inferir que os possíveis motivos 2, 4 e 5 estão presentes na sequência do *L. gongylophorus*. Vale ressaltar a enorme semelhança dos motivos encontrados para *C. purpureum* com os inferidos para sequência do *L. gongylophorus* o que de certa forma era esperado, uma vez que o maior escore obtido (50%) entre a sequência *L. gongylophorus* e as sequências de estrutura conhecida foi com a sequência do *C. purpureum* (1K5C). A Figura 42.4 mostra os motivos 2, 4 e 5 destacados na estrutura 3D do *L. gongylophorus*. Nela é possível observar que os motivos 2 e 4 estão localizados no sítio catalítico da molécula, sendo que apenas dois dos três resíduos aspárticos conservados estão presentes no motivo 2 e os resíduos conservados Gly226 e Ser227 estão presentes no motivo 4. Estes dois últimos resíduos, além de conservados em todas as sequências de EndoPG de estrutura conhecida (Figura 41.4) apresentam entre si ligação *cis*-peptídica no modelo 3D do *L. gongylophorus* que em conjunto com a Lys254 (também conservadas nas demais Endo-PG de estrutura conhecida) são consideradas importantes no mecanismo de reconhecimento do substrato. Tal fato é também observado para estrutura 1K5C (SHIMIZU et al, 2002).

A Figura 43.4 mostra a região do sítio catalítico com destaque (em azul escuro) para os três aspárticos e a molécula de açúcar modelada (em azul claro, ácido galacturônico) e a Figura 44.4 apresenta esta mesma região mais próxima do observador. As distâncias entre os 3 resíduos aspárticos conservados no sítio catalítico do modelo são de 4,5 Å (Asp<sup>179</sup> – Asp<sup>199</sup>); 5,3 Å (Asp<sup>179</sup> – Asp<sup>200</sup>) e 4,8 Å (Asp<sup>199</sup> – Asp<sup>200</sup>).

```

1IA5|A -----ATTCTFSGSNGASSASKSKTSCSTIVLSNVAVPSGTT 37
1NHC|A -----STCTFTS---ASEASES ISSCSVDVLLSSIEVPAGET 33
1CZF|A MHSFASLLAYGLVAGATFASASPIEARDSCFTTAAAKAGKAKCSTITLNNIEVPAGT 60
2IQ7|A -----ASCTFTDAAAAIKGKASCTSIILNGI VVPAGT 33
1HG8|A -----DPCSVTEYSGLATAVSSCKNIVLNGFQVPTGKQ 33
ADV30326.1 -----MWTRFLLSLVLVVSQYASSDCTGTINSMADVAAVACTTVNLSNFQVPAGQT 52
1K5C|A -----ATCTVKSVDDAKDIAGCSAVTLNLFVTPAGNT 32
                                     * . : *... **:*

1IA5|A LDLTKLNDGTHVIFSGETTFGYKEWGGP-LISVSGSDLTITGASGHSINGDGRWWDGEG 96
1NHC|A LDLSDAADGSTITFEGTTSFGYKEWKGP-LIRFGGKDLTVMADGAVIDDGRWWDGSKG 92
1CZF|A LDLTGLTSGTKVIFEGTTTFQYEEWAGP-LISMSGEHITVTVGASGHLINCDGARWWDGKG 119
2IQ7|A LDMPGLKSGTIVTFQKRTTFGYKEWEGP-LISFSGTNININGASGHSIDCQGRWWDGSKG 92
1HG8|A LDLSLQNDSTVTFKGTTFATTADNDFNPIVISGSNITITGASGHVIDGNQAYWDGKG 93
ADV30326.1 LALKLLT-GTTVMNGDITFAHENWAGP-LFQISGSSITFNGNG-HTFDGNGPLYWDGLG 109
1K5C|A LVLNPKD-GATVTMAGDITFAKTTLDGP-LFTIDGTGINFVGAD-HIFDGNALYWDGKG 89
* . : . : : * : * . : ..* :.. . : : * : ** . *

1IA5|A GNGG---KTKPKFFAAHSLTNSVISGLKIVNSPVQVFSVAGSDYLTLDK-ITIDNSDGDD 152
1NHC|A TNGG---KTKPKFMYIHDVEDSTFKGINIKNTPVQAI SVQATN-VHLND-FTIDNSDGDD 147
1CZF|A TSG----KTKPKFFYAHGLDSSITGLNINKTPLMAF SVQAN-DITFTD-VTINNADGDT 173
2IQ7|A SNGG---KTKPKFFYAHSLKSSNIKGLNVLNTPVQAFSINSATTLGVYD-VIIDNSAGDS 148
1HG8|A SNSNSNQKPDHFIVVQKTGNSKIPNLNIQNWVHCFDITGSSQLTISG-LILDNRAGDK 152
ADV30326.1 GNGG---VTKPAPMMKISMSGTFSNVKVLSNPARTYSVS-NPAPLVMDGLTIDNSLDGQ 164
1K5C|A TNNG---THKPHPFLLIKSGSYKKEVLNSPAQAI SVGPTDAHLTLDGITVDDFAGDT 145
. . : . . . : : * * . : . . : : * *

1IA5|A NGG-----HNTDAFDIGTSTYVTISGATVYNQDDCVAVNSGENIYFSGGYCSGGHGL 204
1NHC|A NGG-----HNTDGFDI SESTGVYISGATVKNQDDCIAINSGESISFTGGTCSGGHGL 199
1CZF|A QGG-----HNTDAFDVGNVGVNI IKPWVHNQDDCCLAVNSGENIWF TGGT CIGGHGL 225
2IQ7|A AGG-----HNTDAFDVGSSTGVYISGANVKQDDCCLAINSGTNIFTGGTCSGG I 200
1HG8|A PNAKSGSLPAAHNTDGFDISSSDHVTLDNNHVYNQDDCVAVTSGTNIVVSNMYCSGGHGL 212
ADV30326.1 PNSNSNGLAAGHNTDGFDASTEN-LTIQNSVIMNQDDCCLAINRGANIVFKNNHCSSGGHI 223
1K5C|A KN-----LGHNTDGFVDSANN-VTIQNCIVKQDDCCLAINDGNNIRFENQCSGG I 197
.          ****.*. . : : : *****:*. * . * . * ****:

1IA5|A SIGSVGGRSDNTVKNVTFVDSTIINSDNGVRIKTNIDTTG-SVSDVYTKDITLTSIAKYG 263
1NHC|A SIGSVGGRDDNTVKNVTISDSTVSNANGVRIKTIYKETG-SVSEITYSNIQLSGITDYG 258
1CZF|A SIGSVGDRSNVVKVNTIEHSTVSNSENAVRIKTISGATG-SVSEITYSNIVMSGISDYG 284
2IQ7|A SIGSVGGRSDNTVKTVTISNSKIVNSDNGVRIKTVSGATG-SVSGVTYSGITLSNIAKYG 259
1HG8|A SIGSVGGKSDNVVDGVQFLSSQVNSQNGCRIKSNSGATG-TINNVTYQNIALTNISTYG 271
ADV30326.1 SVGSINSG--VTVNGITITGNTITNNDQALRIKTDAAATGSTVTNITYSGNTATGCKRFG 281
1K5C|A SIGVIATG--KHVSNVVIKGNVTRSMYVRIKAQRTATSASVSGVTYDANTISGIKYG 255
*:**:* . * . : . . . * ** : * . : : ** . : . : *

1IA5|A IVVQQNYG--DTSSTPTGVPITDFVLDN--VHGSVVSSTGNILISCGSGSCS-DWTWTD 318
1NHC|A IVIEQDYENGSPGTGPTGIPITDVTDG--VTGTLEDDATQVYIILCGDGS-CS-DWTWSG 315
1CZF|A VVIQQDYEDGKPTGKPTNGVTIQDVKLES--VTGSDVSGATEIYLLCGSGSCS-DWTWDD 341
2IQ7|A IVIEQDYENGSPGTGPTNGVPIITGLTLSK--ITGSVASSGTNVIYILCASGACS-NWKWSG 316
1HG8|A VDVQQDYLNGGPTGKPTNGVKISNIKFIK--VTGTVASSAQDWFILCGDGS-CS-GFTTFSG 328
ADV30326.1 VIIDQSYG--STIGTPGSGVLLSAVNFVSPQTSITVNSGAERVAVNCVGSCTGTWNWSN 339
1K5C|A VLISQSYG--DDVGNPGTGAPFSDVNFTGGATTIKVNNAATRVTVECGN--CSGNWNWSQ 311
: : *. * . . * . * : . . . : . . . : * . * : : :

1IA5|A VSVSGGKTSKCTNVPSGASC--- 339
1NHC|A VDLGGKTSKCNVPSGASC--- 336
1CZF|A VKVTGGKSTACKNFPSVASC--- 362
2IQ7|A VSVTGGKSTKCSNIPSGGAAC- 339
1HG8|A NAITGGKTSNCPYPTNTCPN--- 349
ADV30326.1 LKVSOGTTGKITNFNGITGFSQ-- 361
1K5C|A LTVTGGKAGTIKSDKAKITGGQYL 335
: : **

```

FIGURA 41.4 – Alinhamento entre Endo-PG de fungos de estrutura 3D conhecida. 1IA5 (*A. aculeatus*), 1NHC (*Aspergillus niger*), 1CZF (*A. niger*), 2IQ7 (*Colletotrichum lupini*), 1HG8 (*Fusarium moniliforme*), ADV30326.1 (*L. gongylophorus*) e 1K5C (*C. purpureum*). As seqüências grifadas em azul corresponde ao motivo 1, a amarela ao motivo 2, a vermelha ao motivo 3, a verde ao motivo 4 e a rosa ao motivo 5. Os resíduos em azul correspondem aos 8 resíduos conservados em PG de fungos (Kester et al, 1996). Os resíduos em vermelho quando em posição *cis*-peptídica em conjunto com a K (grifada em cinza) são responsáveis pelo reconhecimento do substrato (Shimizu et al, 2002).





FIGURA 42.4 - Modelo tridimensional da PG de *L. gongylophorus* destacando os motivos 2 (em amarelo), 4 (em verde) e 5 (em rosa).



FIGURA 43.4 - Modelo tridimensional da PG de *L. gongylophorus*, destacando a região do sítio catalítico com os três resíduos de ácido aspártico envolvidos na catálise (em azul escuro) e a molécula de açúcar (ácido galacturônico, em azul claro).

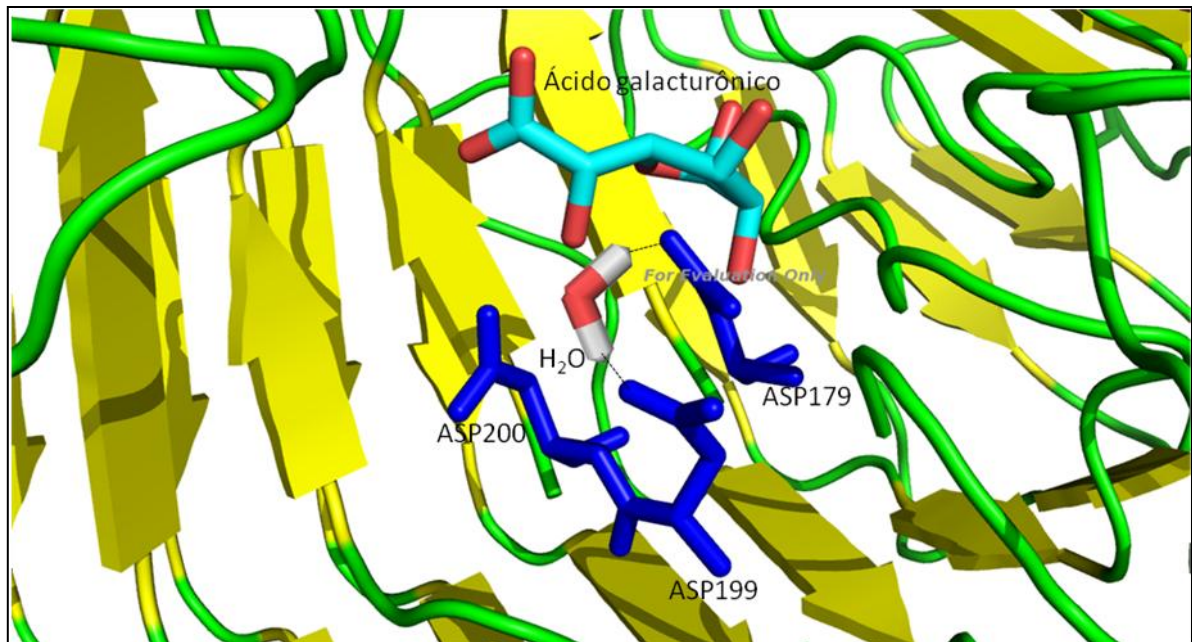


FIGURA 44.4 - Região do sítio catalítico da PG de *L. gongylophorus* com os três resíduos de ácido aspárticos envolvidos na catálise (em azul escuro) e a molécula de açúcar (ácido galacturônico, em azul claro).

A função exata dos três ácidos aspárticos no mecanismo catalítico da PG de *C. purpureum* não foi totalmente esclarecida (SHIMIZU *et al.*, 2002) assim como também em outras estruturas (VAN SANTEN *et al.*, 1999; ARMAND *et al.*, 2000; FEDERICI *et al.*, 2001; CHO *et al.*, 2001; SHIMIZU *et al.*, 2002), mas com base no modelo construído neste trabalho e nas proposições da literatura podemos sugerir que dois resíduos aspárticos (Asp<sup>179</sup> e Asp<sup>199</sup>) interajam com a molécula de água, que ataca o carbono do açúcar substrato como esquematizado na Figura 45.4 e o conservado resíduo Asp<sup>200</sup> atue como um doador de próton.

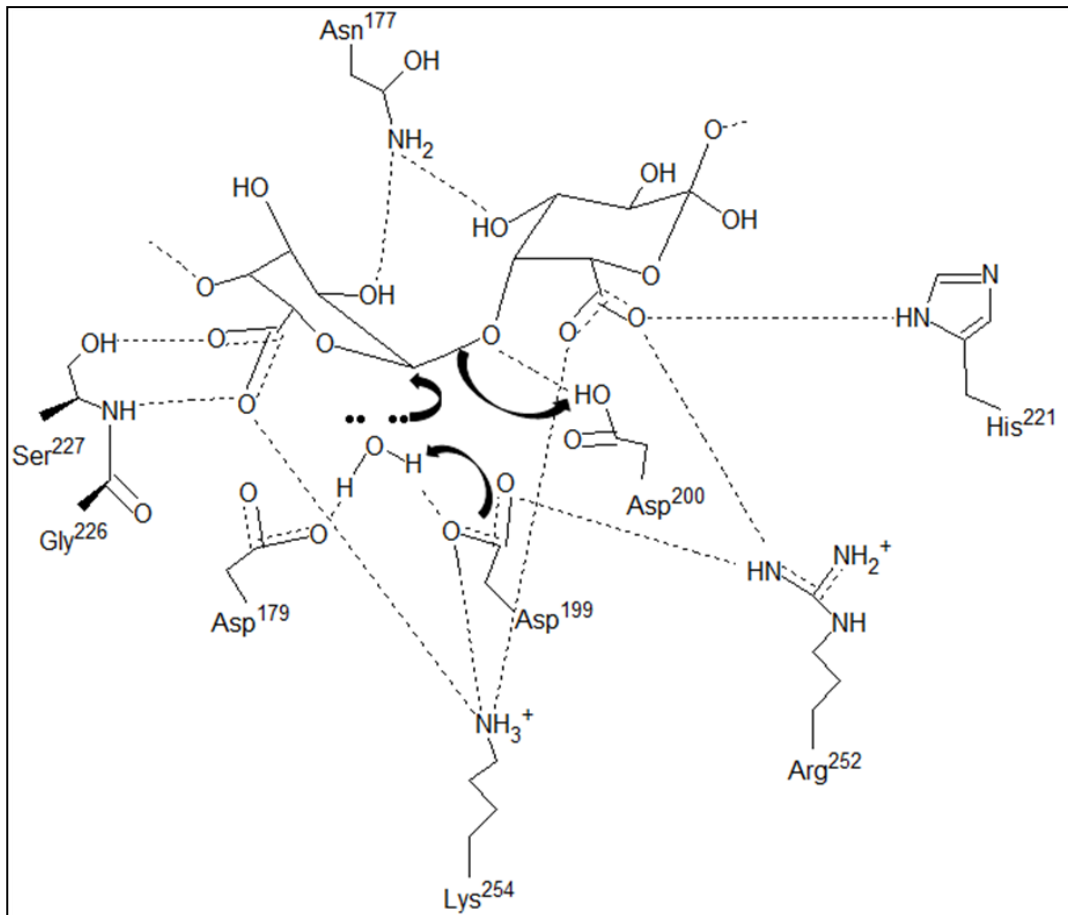


FIGURA 45.4 - Desenho esquemático propondo o possível mecanismo de ação para PG de *L. gongylophorus* com base no modelo 3D obtido no presente trabalho. O conservado resíduo Asp200 atua como um doador de próton, enquanto os resíduos Asp179 e Asp199 ativam a molécula de água. Arg252 e Lys254 interagem com o Asp199 e com o açúcar.

Neste trabalho foi possível construir o modelo tridimensional de uma poligalacturonase de *L. gongylophorus*, um fungo simbiote de formiga cortadeira e fazer uma análise do possível mecanismo de atuação da enzima. Poucas informações são disponíveis a respeito de estruturas de proteínas da família 28 das glicosil hidrolases e o mecanismo proposto para essa família de proteínas começou a ser questionado a partir da análise das estruturas determinadas. Aliar o conhecimento das estruturas com técnicas de biologia molecular, como por exemplo mutagênese sítio dirigida, pode auxiliar a melhor esclarecer o mecanismo de atuação dessa classe de proteínas (DAVIES & HENRISSAT, 1995).

## CAPÍTULO 5 - CONCLUSÕES

Neste trabalho a bioinformática foi utilizada com duas abordagens para o estudo do mesmo tema: poligalacturonases (PG) de fungos. Na primeira parte do trabalho, foi realizado um estudo de motivos estruturais e filogenia com todas as sequências de PG de fungos depositadas em bancos de dados. Para isso, todas as sequências depositadas no banco de dados NCBI como '*polygalacturonase fungi*', 2093 sequências, foram utilizadas no início do trabalho. Como muitas das sequências depositadas podem ser relativas a peptídeos e/ou estão depositadas como 'proteína hipotética' foi feita uma redução do número de sequências do banco, retirando-se sequências com menos de 70 aminoácidos, as que correspondiam a 'proteína hipotética' e ainda das que não eram PG ou mesmo não eram PG de fungos. Este novo banco de dados continha 957 sequências de PG de fungos. Após um estudo de alinhamento das 957 sequências, as que tinham 100% de identidade foram retiradas e assim o banco de dados do trabalho ficou com 417 sequências únicas de PG de fungos. Após ordenação das sequências por gêneros, foi observado que 95 sequências não haviam sido classificadas em Endo- ou Exo-PG, e através de alinhamentos entre as sequências foi possível classificar 71 das sequências. As 24 sequências que permaneceram sem classificação foram submetidas à análise, através de árvores filogenéticas, na tentativa de classificação em Endo e ExoPG. Os resultados permitiram inferir a respeito da classificação das sequências de PG de fungos que ainda não haviam sido classificadas e dessa maneira todas as sequências únicas de PG de fungos foram classificadas.

O estudo de presença de motivos sequenciais nas PG de fungos mostrou a ocorrência de cinco motivos, sendo que das sequências classificadas como sendo de Endo-PG e usadas no estudo (63) a maioria apresenta os cinco motivos, mas das sequências avaliadas como Exo-PG e usadas no estudo (38) nenhuma apresenta três dos motivos. Esses resultados sugerem regiões nas moléculas responsáveis por especificidade das enzimas. Ao analisar os 5 motivos pelo BLAST os motivos 1, 2 e 5 mostraram relacionamento com o domínio Glyco\_hydro\_28, enquanto que nenhuma relação funcional foi estabelecida para os motivos 4 e 3. A exata função dos 5 motivos e sua influência na atividade catalítica nas PG devem ser melhor investigadas.

A segunda parte do trabalho teve como objetivo o estudo da estrutura

tridimensional de uma PG de *L. gongylophorus*, fungo simbiote de formigas cortadeiras. A sequência depositada no NCBI relacionada a uma PG deste fungo (ADV30326.1) foi submetida à análise com o programa *Pfam*, que, através de alinhamentos desta sequência e as depositas no seu banco de dados, classificou a proteína como um membro da família Glyco\_hydro\_28 (família 28 das glicosil hidrolases). Na primeira parte do trabalho esta sequência havia sido classificada, na árvore filogenética, em uma Endo-PG, o que confirma o resultado obtido pelo *Pfam*.

O método utilizado para a construção do modelo 3D da PG de *L. gongylophorus* foi o de modelagem por homologia, que requer o conhecimento da estrutura 3D de uma proteína com identidade ao menos razoável com a proteína de estudo. A busca por estruturas conhecidas de proteínas homólogas à PG de *L. gongylophorus* foi feita através da análise das sequências na árvore filogenética construída neste trabalho e procurando por estruturas de proteínas próximas à PG de *L. gongylophorus* e também fazendo busca automática pelo programa SWISS-MODEL. Existe somente uma estrutura depositada no banco de dados de estruturas, de uma Endo-PG do fungo *C. purpureum* que mostra um bom grau de identidade (50%) com a PG de *L. gongylophorus* e esta estrutura foi utilizada para a construção do modelo estrutural através do programa SWISS-MODEL. O modelo gerado foi validado por diferentes programas e conclui-se que o modelo 3D da PG do *L. gongylophorus* apresenta bons parâmetros esteroquímicos. Apesar de alguns resultados negativos terem sido apresentados, de maneira geral a modelagem por homologia trouxe resultados muitos bons, já que os demais parâmetros e servidores (WHAT IF e VERIFY 3D) de validação mostraram-se satisfatórios.

O modelo apresenta um enovelamento conhecido como beta-hélice de mão direita, comumente encontrado em estruturas de PG e pectato liases, e apresenta quatro folhas-beta paralelas compreendendo dez voltas completas. A região do sítio catalítico pode ser descrita como uma fenda (ladeada por *loops*) compreendendo os resíduos conservados nas PG e relatados como envolvidos na catálise: Asn<sup>177</sup>, Asp<sup>179</sup>, Asp<sup>199</sup>, Asp<sup>200</sup>, His<sup>221</sup>, Gly<sup>222</sup>, Arg<sup>252</sup> e Lys<sup>254</sup>.

Os resíduos Asp<sup>179</sup>, Asp<sup>199</sup>, Asp<sup>200</sup> são propostos atuarem diretamente na catálise, como bases e nucleófilo e os resíduos Arg<sup>252</sup> e Lys<sup>254</sup> podem estar envolvidos em interações com o substrato. Um mecanismo, ocorrendo por inversão da configuração do carbono anomérico, foi proposto e discutido.

A determinação da estrutura de uma molécula de proteína é

fundamental para o conhecimento de sua forma de atuação. As técnicas disponíveis para determinar estrutura de proteínas requerem quantidades e grande grau de pureza da amostra, o que dificulta a determinação experimental da estrutura. Assim, técnicas de modelagem molecular têm sido uma excelente ferramenta no estudo de moléculas que ainda não puderam ser estudadas experimentalmente.



## CAPÍTULO 6 - REFERÊNCIAS BIBLIOGRÁFICAS

ADACHI, J. & M. HASEGAWA. MOLPHY: programs for molecular phylogenetics. Version 2.2. Institute of Statistical Mathematics, Tokyo, 1994.

ANISIMOVA, M. & GASCUEL, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 2006; 55 (4): 539-552.

ARMAND, S.; WAGEMAKER, M. J.; SANCHEZ-TORRES, P.; KESTER, H. C.; VAN SANTEN, Y.; DIJKSTRA, B. W.; VISSER J. & BENEN, J. A. The active site topology of *Aspergillus niger* endopolygalacturonase II as studied by site directed mutagenesis. *J. Biol. Chem.* 2000; **275**: 691-696.

ARNOLD, K.; BORDOLI, L.; KOPP, J. & SCHWEDE, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics.* 2006; 22 (2): 195–201.

BAILEY T.L. & ELKAN, C. Fitting a mixture model by expectation-maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1994; **2**: 28-36.

BAILEY, T.L.; WILLIAMS, N.; MISLEH, C. & LI, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucl. Acids Res.* July 2006; 34 (suppl 2): W369-W373.

BAYAT, A. Science, medicine, and the future bioinformatics. *BMJ.* 2002; **324**: 1018-1022.

BERTI FILHO, E.; MARCHINI, L. C & NAKANO, O. “Curso de Entomologia Aplicada à Agricultura”. IN: *Formigas Cortadeiras e Cupins*. Piracicaba, FEALQ, 1992. p. 631-671.

BONIVENTO, D.; PONTIGGIA, D.; DI MATTEO, A.; RECIO, SALVI, G.; TSERNOGLOU, D.; CERVONE, F.; DE LORENZO, G. & FEDERICI, L. Crystal structure of the endopolygalacturonase from the phytopathogenic fungus *Colletotrichum lupini* and its interaction with polygalacturonase-inhibiting proteins. *Proteins.* 2008; 70 (1): 294-299.

BORK, P. & KOONINT, E. V. Protein sequence motifs. *Current Opinion in Structural Biology.* 1996, **6**: 366-376.

BRUHLMANN, F.; Kim, K. S.; Zimmerman, W. & Fiechter, A. Pectinolytic enzymes from actinomycetes for the degumming of ramie bast fibers. *Appl. Environ. Microbiol.* 1994; 60 (6): 2107-2112.

CAVALLI-SFORZA, L.L. & EDWARDS, A.W.F. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics.* 1967; **19**: 233-257.

CHESSON, A. Maceration in relation to the post handling and processing of plant material. *J. Appl. Biotechnol.* 1980; **48**: 1–45.



- CHO, S.W.; LEE, S. & SHIN, W. The X-ray structure of *Aspergillus aculeatus* polygalacturonase and a modeled structure of the poligalacturonase octagalacturonate complex. *J. Mol. Biol.* 2001; **314**: 863-878.
- CORPET, F. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* 1988; 16 (22): 10881-10890.
- COSGROVE, D.J. Assembly and enlargement of the primary cell wall in plants. *Annu. Rev. Cell. Dev. Biol.* 1997; **13**: 171–201.
- DAVIES, G & HENRISSAT, B. Structures and mechanisms of glycosyl hydrolases. *Structure* 1995, **3**: 853-859.
- DAYHOFF, M. O., SCHWARTZ, R. M. & ORCUTT, B. C. A model for evolutionary change in proteins. *Atlas of Protein Sequence and Structure.* 1978; **5**: 345-352.
- DEREEPER, A.; GUIGNON, V.; BLANC, G.; AUDIC, S.; BUFFET, S.; CHEVENET, J.; DUFAYARD, F.; GUINDON, S.; LEFORT, V.; LESCOT, M.; CLAVERIE, J. M. & GASCUEL O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research.* 2008; **36**: W465-W469.
- EMSLEY, P. & COWTAN, K. Coot: model-building tools for molecular graphics. *Acta Crystallografica Section D-Biological Crystallography.* 2004; **60**: 2126-2132.
- FEDERICI, L.; CAPRARI, C.; MATTEI, B.; SAVINO, C.; DI MATTEO, A.; DE LORENZO, G.; CERVONE, F. & TSERNOGLOU, D. Structural requirements of endopolygalacturonase for the interaction with PGIP (poligalacturonase inhibiting protein). *PNAS.* 2001; 98 (23): 13425–13430.
- FELSENSTEIN, J. PHYLIP: phylogenetic inference package. Version.3.5. Department of Genetics, University of Washington, Seattle, 1993.
- FELSENSTEIN, J. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biological Journal of the Linnean Society.* 1981; 1 (16): 183-196.
- FERNANDES, P. Enzymes in food processing: a condensed overview on strategies for better biocatalysts. *Enzyme Research.* 2010.
- FILHO, O. A. S. & ALENCASTRO, R. B. de. Modelagem de proteínas por homologia. *Quim. Nova.* 2003; 26 (2): 253-259.
- FISHER, P. J.; STRADLING, D. J. & PEGLER, D. N. Leaf-cutting ants, their fungus gardens and the formation of basidiomata of *Leucoagaricu gongylophorus*. *Mycologist.* 1994; **8**: 128-132.
- FISHER, R.A. "On the mathematical foundations of theoretical statistics". *Philosophical Transactions of the Royal Society of London.* 1922; **222**: 309-368.
- FLOUDAS, C.A; FUNG, H.K.; MCALLISTER, S.R.; MÖNNIGMANN, M. & RAJGARIA, R. Advances in protein structure prediction and de novo protein

design: A review. *Chemical Engineering Science*. 2006; 61 (3): 966–988.

FREEDONIA GROUP INC. WORLD ENZYMES—Industry Study with Forecasts for 2014 & 2019: Study #2670, Setembro 2010, <http://www.freedoniagroup.com/FractionalDetails.aspx?DocumentId=523442>

GUINDON, S.; LETHIEC, F.; DUROUX, P. & GASCUEL, O. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*. 2005; **33**: W557–W559.

HARHOLT, J.; SUTTANGKAKUL, A. & SCHELLER, H. V. Biosynthesis of Pectin. *Plant Physiology*. 2010; 153 (2): 384-395.

HARTL, D. L. & JONES, E. W. *Introduction to Molecular Genetics and Genomics*. In: Genetics: analysis of genes and genomes. 6<sup>a</sup> ed. Jones & Bartlett Publishers, 2004. 854 p.

HENIKOFF, S. & HENIKOFF, J. G. Amino acid substitution matrices from protein blocks. *Biochemistry*. 1992; **89**: 10915-10919.

JENKINS J.; MAYANS, O & PICKERSGILL, R. Structure and evolution of parallel beta-helix proteins. *J. Struct. Biol*. 1998; 122(1-2): 236-246.

JUSTI-JUNIOR, J.; IMENES, S. DE LAMONICA; BERGAMANN, E.C., CAMPOS-FARINHA, A.E. DE. & ZORZENON, F.J. Formigas Cortadeiras. *Boletim Técnico*. Instituto Biológico, São Paulo. 1996; **4**: 5-31.

KASHYAP, D.R.; VOHRA, P.K.; CHOPRA, S. & TEWARI, R. Applications of pectinases in the commercial sector: a review *Bioresource Technology*. 2001; 77(3): 215-227.

KESTER, H. C.; KUSTERS-VAN SOMEREN, M. A.; MULLER, Y. & VISSER, J. *Eur. J. Biochem*. 1996; **240**, 738-746

KOBAYASHI, M.; NAKAGAWA, H.; ASAKA, T. & MATOH, T., Borate rhamnogalacturonan II bonding reinforced by Ca retains pectic polysaccharides in higher-plant cell walls. 1999; *Plant Physiol*. **119**: 199–203.

KOPP, J. & SCHWEDE, T. Automated protein structure homology modeling: a progress report. *Pharmacogenomics Journal*. 2004; 5 (4): 405–416.

KUHAD, R. C; GUPTA, R. & SINGH, A. Microbial cellulases and their industrial applications. *Enzyme Research*. 2011.

LANG, C. & DÖRNENBURG, H. Perspectives in the biological function and the technological application of polygalacturonases. *Appl Microbiol Biotechnol*. 2000; **53**: 366-375.

LARKI, M. A.; BLACKSHIELDS, G.; BROWN, N.P.; CHENNA, R.; MCGETTIGAN, P.A.; MCWILLIAM, H.; VALENTIN, F.; WALLACE, I.M.; WILM, A.; LOPEZ, R.; THOMPSON, J. D.; GIBSON, T. J. & HIGGINS, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23 (21): 2947–2948.

- LASKOWSKI, R.A.; MACARTHUR, M. W.; MOSS, D. S. & THORNTON, J. M. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 1993. **26**: 283-291.
- LEISOLA, M.; JOKELA, J.; PASTINEN, O.; TURUNEN, O. & H. SCHOEMAKER, H. Industrial use of enzymes. In: *Encyclopedia of Life Support Systems (EOLSS)*, O. O. P. Hänninen and M. Atalay, Eds., pp. 1–25, EOLSS, Oxford, UK, 2002.
- LESK, A. M. Introdução à bioinformática. 2 ed – Porto Alegre: Artmed, 2008. 384 p.
- LIÒ, P. & GOLDMAN, N. Models of Molecular Evolution and Phylogeny. *Genome Research. Cold Spring Harbor Laboratory Press.* 1998; **8**: 1233-1244.
- LOMBARD, V; BERNARD, T; RANCUREL, C; BRUMER, H; COUTINHO, P.M. & HENRISSAT, B. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* 2010; **432**: 437-444.
- LUTHY, R.; BOWIE, J. U. & EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature.* 1992; **356**: 83-85.
- MAGGIO, E. T. & RAMNARAYAN, K. Recent developments in computational proteomics. *Trends Biotechnol.* 2001; 19 (7): 266-272
- MALLER, A.; DAMÁSIO, A.R.L.; DA SILVA, T. M.; JORGE, J.A.; TERENCE, H.F. & M. DE L.T. DE M. POLIZELI. Biotechnological potential of agro-industrial wastes as a carbon source to thermostable polygalacturonase production in *Aspergillus niveus*. *Enzyme Research.* 2010; **2011**:289206.
- MARKOVIC, O. & JANECEK, S. Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. *Protein Engineering.* 2001; 14 (9): 615-631.
- MARTIN, M. M.; BOYD, M. J.; GIESELMANN, M. J. & SILVER, R. G. "Activity of faecal fluid of a leaf-cutting ant toward plant cell wall polysaccharides". *J. Insect Phys.* 1975; **21**: 1887.
- MARTINS, E. da S. Purificação e Caracterização Bioquímica de Poligalacturonases Termoestáveis Produzidas pelo Fungo *Thermoascus aurantiacus* Através de Fermentação Submersa e Fermentação em Estado Sólido. Rio Claro. Programa de Pós-Graduação em Biociências, 2006. Tese de doutorado, 108 p.
- MCCARTER, J.D. & WITHERS, G.S. Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology.* 1994; 4 (6): 885–892.
- MCNEIL, M.; DARVILL, A. G.; FRY, S. C. & ALBERSHEIM, P. Structure and function of the primary cell walls of plants. *Ann. Rev. Biochem.* 1984; **53**: 625-663.
- MIYAIRI, K.; OKUNO, T. & SAWAI, K. Purification and properties of endopolygalacturonase I from *Stereum purpureum*, a factor inducing silver-leaf symptoms on apple trees. *Agricultural and Biological Chemistry.* 1985; 49 (4): 1111-1118.

MUELLER, U.G.; SCHULTZ, T.R.; CURRIE, C.R.; ADAMS, R.M. & MALLOCH, D. The origin of the attine ant-fungus mutualism. *Q Rev Biol.* 2001; 76 (2):169-197.

MUTTER, M.; BELDMAN, G.; SCHOLS, H. A. & VORAGEN, A. G. J. Rhamnogalacturonan  $\alpha$ -L-Rhamnopyranohydrolase: a novel enzyme specific for the terminal nonreducing rhamnosyl unit in rhamnogalacturonan regions of pectin. *Plant Physiology.* 1994; **106**: 241-250.

NAIDU, G. & PANDA, T. Production of pectolytic enzymes, a review. *Bioproc. Eng.* 1998; **19**: 355-361.

NAYEEM, A.; SITKOFF, D. & KRISTEK JR. S. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Science.* 2006; 15 (4): 808–824.

NEI, M. & KUMAR, S. *Molecular basis of evolution*. In: Molecular evolution and phylogenetics. Páginas exibidas com permissão de Oxford University Press. Direitos autorais. Ano 2000

NORTH, R.D.; JACKSON, C.W. & HOWSE, P.E. Evolutionary aspects of ant fungus interactions in leaf-cutting ants. *Trends in Ecology and Evolution.* 1997; 12 (10): 386-389.

NORTHCOTE, D.H. Chemistry of the plant cell wall. *Ann.Rev.Plant Physiol.* 1972; **23**: 113-132.

*Number of species recorded in Formicidae* -[http://osuc.biosci.ohio.state.edu/hymenoptera/tsa.sppcount?the\\_taxon=Formicidae](http://osuc.biosci.ohio.state.edu/hymenoptera/tsa.sppcount?the_taxon=Formicidae). Página visitada em 1 de março de 2012.

OLSEN, G. J.; MATSUDA, H.; HAGSTROM, R. & OVERBEEK, R. FastDNAML: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci.* 1994; 10 (1): 41-48.

PANDEY & MANN. Proteomics to study genes and genomes. 2000; *Nature.* **405**: 837-846

PROSDOCIMI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A. F.; REIS, A. N. DOS; JUNQUEIRA, A. C. M.; SANTOS, A. C. F. DOS; JÚNIOR, A. N.; WUST, C. I.; FILHO, F. C.; KESSEDJIAN, J. L.; PETRETSKI, J. H.; CAMARGO, L. P.; FERREIRA, R. DE G. M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. DE S. & FLATSCHART, A. V. F-. Manual do usuário. *Biotecnologia Ciência & Desenvolvimento.* 2002; **29**: 12-25.

RAMACHANDRAN, G.N.; RAMAKRISHNAN, C. & SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology.* 1963; **7**: 95–9.

RICARD, M.; ORCCOTOMA, J.-A.; LING, J. & WATSON, R. Pectinase reduces cationic chemical costs in peroxide-bleached mechanical grades. *Pulp & Paper Canada.* 2005; 106 (12): T264-268.

RIDLEY, B.L.; O'NEILL, M. A. & MOHNEN, D. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*. 2001; 57 (6): 929-967.

ROGERS, L.M.; KIM, Y.-K.; GUO, W.; GONZÁLEZ-CANDELAS, L.; LI, D. & KOLATTUKUDY, P.E. Requirement for either a host- or pectin-induced pectate lyase for infection of *Pisum sativum* by *Nectria hematococca*. *Proc. Natl. Acad. Sci. USA*. 2000; **97**: 9813– 9818.

ROMBOUTS, F. M. & PILNIK, W. Pectic enzymes. In: *Microbial Enzymes and Bioconversions*. v. 5, ed. A. H. Rose. Academic Press, London, 1980, p. 227-202.

SAKAI, T.; SAKAMOTO, T. & HALLAERT, J. Pectin, pectinase and protopectinase: production, properties and applications. *Advances in Applied Microbiology*. 1993; **39**: 213.

SALI, A. & BLUNDELL, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993; **234**: 779–815.

SCHIØTT, M.; ROGOWSKA-WRZESINSKA, A.; ROEPSTORFF, P. & BOOMSMA, J.J. Leaf-cutting ant fungi produce cell wall degrading pectinase complexes reminiscent of phytopathogenic fungi. *BMC Biology*. 2010; **8**:156

SCHMIDT, H. A.; STRIMMER, K.; VINGRON, M. & von HAESELER, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 2002; 18 (3): 502-504.

SCHWEDE, T.; KOPP, J.; GUEX, N. & PEITSCH, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*. 2003; 31(13): 3381–3385.

SHIMIZU, T.; NAKATSU, T.; MIYAIRI, K.; OKUNO, T. & KATO, H. Active-Site Architecture of Endopolygalacturonase I from *Stereum purpureum* Revealed by Crystal Structures in Native and Ligand-Bound Forms at Atomic Resolution. *Biochemistry*. 2002, **41**: 6651-6659

SILVA, A. E. A. da. *Uma abordagem multi-objetivo e multimodal para reconstrução de árvores filogenéticas*. Campinas. Programa de Pós-graduação em Engenharia de Computação, 2007. Tese de doutorado, 158 p.

SILVA, A. *Participação do fungo Leucocoprinus gongylophorus na Produção de Enzimas Intestinais da Formiga Atta sexdens*. Rio Claro, Instituto de Biociências - UNESP, 1999. Dissertação de Mestrado. 104 p.

SILVA, A.; BACCI, M.JR; PAGNOCCA, F.C.; BUENO, O.C. & HEBLING, M.J.A. Starch metabolism in *Leucoagaricus gongylophorus*, the symbiotic fungus of leaf-cutting ants. *Microbiological Research*. 2006. 161(4): 299–303.

SINGER, R. *The Agaricales in Modern Taxonomy*, 4 ed. Koenigstein: Koeltz Scientific Books, 1986, p. 477.

SIQUEIRA, C. G. *Estudo do Potencial Metabólico do Fungo Leucocoprinus*

gongylophorus *Heim*, *Associado aos Ninhos de Atta sexdens rubropilosa Forel (Hymenoptera: Formicidae)*. Rio Claro, Instituto de Biociências - UNESP, 1997. Dissertação de Mestrado. 66 p.

SIQUEIRA, C. G.; BACCI JR., M.; PAGNOCCA, F. C.; BUENO O. C. & HEBLING M. J. A. Metabolism of plant polysaccharides by *Leucoagaricus gongylophorus*, the symbiotic fungus of the leaf-cutting ant *Atta sexdens* L. *Appl. Environ. Microbiol.* 1998; **64**: 4820.

STRIMMER, K. & HAESELER, A. von. Quartet puzzling: a quartet maximum likelihood method for reconstruction tree topologies. *Mol. Biol. Evol.* 1996; 13 (7): 964-969.

THARKUR, B.R.; SINGH, R.K. & HANDA, A.K. Chemistry and uses of pectin – A Review. *Critical in Food Science and Nutrition*. 1997; **37**: 47.

THOMPSON, J.D.; HIGGINS, D.G. & GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22 (22): 4673-4680.

THORNTON, J.W. Enzymatic degradation of polygalacturonic acids released from mechanical pulp during peroxide bleaching. *Tappi J.* 1994; 77(3): 161-167.

THORNTON, J.W.; ECKERMAN, C.S.; EKMAN, R.O. & HOLMBOM, B.R. Treatment of alkaline bleached mechanical pulp with pectinase. *U.S. Patent* 1996; 5, 487,812.

TURNER, P.; MAMO, G. & KARLSSON, E.N. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories*. 2007; 6(9): 1-23.

UENDO, M. & PASTORE, G.M. Pectinases: aplicações industriais e perspectivas. *Química Nova*. 2007; 30 (2): 388-394.

UNDERKOFER, L. A.; BARTON, R. R. & RENNERT, S. S. Production of microbial enzymes and their applications. *Appl Microbiol.* 1958 maio; 6(3): 212-221.

VAN SANTEN, Y. BENEN, J.A.E.; SCHRÖTER, K.H.; KALK, K.H.; ARMAND, S.; J.VISSER & DIJKSTRA, B.W. 1.68-Å crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis. *The Journal of Biological Chemistry*. 1999; 274(43): 30474–30480.

VIDAL, S.; DOCO, T.; WILLIAMS, P.; PELLERIN, P.; YORK, W.S.; O'NEIL, M.A.; GLUSHKA, J.; DARVILL, A.G. & ALBERSHEIM, P. Structural characterization of the pectic polysaccharide rhamnogalacturonan II: evidence for the backbone location of the aceric acid containing oligoglycosyl side chain. *Carbohydr. Res.* 2000; 326 (4): 277-294.

VRIEND, G. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.* 1990; **8**: 52--56.

WEBER, N.A. Pure cultures of fungi produced by ants. *Science*. 1955; **121**: 109.

WILLATS, W.G.T.; McCARTNEY, L.; MACKIE, W. & KNOX, P. K. Pectin: cell biology and prospects for functional analysis. *Plant Molecular Biology*.2001; **47**: 9–27.

WILSON, D.B. Cellulases and biofuels. *Current Opinion in Biotechnology*. 2009; **20**: 295–299.

YADAV, P. K.; SINGH, V. K.; YADAV, S.; YADAV, K. D. S. & YADAV, D. *In silico* Analysis of Pectin Lyase and Pectinase Sequences. *Biochemistry (Moscow)*, 2009; **74** (9): 1049-1055.

YANG, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS*. 1997; 13 (5): 555-556.

YODER, M. D.; KEEN, N. T.; & JURNAK, F. New domain motif: structure of pectate lyase C, a secreted plant virulence factor. *Science*. 1993; **260**: 1503-1507.

YOSHIDA, S.; SUZUKI, F.; TSUKIBOSHI, T. & SHINOHARA, H. *Mycol. Res*. 2004; **108** (12): 1407–1414.

## Apêndice A – Tabela das 63 seqüências de Endo-PG de fungos e 38 de Exo-PG com seus respectivos motivos conservados.

Seqüências	Nome	Motivos				
		1	2	3	4	5
ABF51154.1	EndoPG [Alternaria alternata]			x		x
EED53217.1	EndoPG [Aspergillus flavus NRRL3357]	x	x	x	x	x
CAK40265.1	EndoPG-D [Aspergillus niger]	x	x	x	x	x
CAK47667.1	EndoPG-B [Aspergillus niger]	x	x	x	x	x
BAH14957.1	EndoPG-1 [Athelia rolfsii] precursor	x	x	x	x	x
AAV84618.1	EndoPG-6 [Botryotinia fuckeliana]	x	x	x	x	x
AAV84616.1	EndoPG-4 [Botryotinia fuckeliana]	x	x	x	x	x
CAM91338.1	EndoPG-2 [Botryotinia fuckeliana]	x	x	x	x	x
CAM91053.1	EndoPG-2 [Botrytis fabae]	x	x	x	x	x
CAQ03431.1	EndoPG-4 [Botrytis fabae]	x	x	x	x	x
BAA96351.1	EndoPG [Chondrostereum purpureum]		x			x
AAM00249.1	EndoPG [Chrysosporthe cubensis]	x	x	x		x
CAA71247.2	EndoPG [Claviceps purpurea]	x	x	x	x	x
ABL01533.1	EndoPG [Colletotrichum lupini var. setosum]precursor	x	x	x	x	x
CAE46647.1	EndoPG [Diaporthe helianthi]	x	x	x	x	x
ABF50868.1	EndoPG [Emericella nidulans]	x	x	x	x	x
ABQ11246.1	EndoPG [Fusarium nygamai]	x	x	x	x	x
BAE97111.1	EndoPG [Fusarium oxysporum f. sp. lagenariae]	x	x	x	x	x
ABQ11234.1	EndoPG [Fusarium sacchari]	x	x	x	x	x
BAB97386.1	EndoPG-1 [Galactomyces geotrichum]	x	x	x	x	x
ABQ11247.1	EndoPG [Gibberella circinata]	x	x	x	x	x
ABQ11240.1	EndoPG [Gibberella intermedia]	x	x	x	x	x
AAA74586.1	EndoPG [Gibberella moniliformis]	x	x	x	x	x
CAA64727.1	EndoPG [Glomerella lindemuthiana]	x	x	x	x	x
CAM07141.1	EndoPG [Hypocrea lixii]	x	x	x	x	x
AAF70169.1	EndoPG-1 [Leptosphaeria maculans]	x	x	x	x	x
ADD73449.1	EndoPG-1 [Phomopsis mangiferae]		x			x
EDN63471.1	EndoPG [Saccharomyces cerevisiae YJM789]	x	x	x	x	x
AAC14466.1	EndoPG [Sclerotinia sclerotiorum]	x	x	x	x	x
ACL80126.1	EndoPG-1 [Thanatephorus cucumeris]		x			x
BAG72132.1	EndoPG [Venturia nashicola]	x	x	x	x	x
EY20327.1	EndoPG [Verticillium albo-atrum VaMs.102]	x	x	x	x	x
EED50415.1	ExoPG [Aspergillus flavus NRRL3357]					x
EED47623.1	ExoPG [Aspergillus flavus NRRL3357]					x
EED46018.1	ExoPG [Aspergillus flavus NRRL3357]					x
EDP55950.1	ExoPG [Aspergillus fumigatus A1163]					x
EAL86831.1	ExoPG [Aspergillus fumigatus Af293]					x
EAL84906.1	ExoPG [Aspergillus fumigatus Af293]					x
ABD61564.1	ExoPG-B [Aspergillus niger]					x
ABD61562.1	ExoPG-X [Aspergillus niger]					x
AAF05088.1	ExoPG [Botryotinia fuckeliana]					x
ABF50895.1	ExoPG [Emericella nidulans]					x
BAE97166.1	ExoPG [Fusarium oxysporum f. sp. lycopersici]					x
EFY92040.1	ExoPG [Metarhizium acridum CQMa 102]					x
EEA27533.1	ExoPG [Penicillium marneffei ATCC 18224]					x
EEA22673.1	ExoPG [Penicillium marneffei ATCC 18224]					x
AAP83024.1	ExoPG [Sclerotinia sclerotiorum]					x
EED23059.1	ExoPG [Talaromyces stipitatus ATCC 10500]					x
EED15423.1	ExoPG [Talaromyces stipitatus ATCC 10500]					x
EY20199.1	ExoPG [Verticillium albo-atrum VaMs.102]					x
EY16298.1	ExoPG [Verticillium albo-atrum VaMs.102]					x
BAB32924.1	EndoPG [Alternaria alternata]	x	x	x	x	x
EDP49013.1	EndoPG [Aspergillus fumigatus A1163]	x	x	x	x	x
CAK47020.1	EndoPG-A [Aspergillus niger]	x	x	x	x	x
BAH29710.1	EndoPG-2 [Athelia rolfsii] precursor	x	x	x	x	x
ADZ99366.1	EndoPG-1 A [Bispora sp. MEY-1]	x	x	x	x	x
AAV84617.1	EndoPG-5 [Botryotinia fuckeliana]	x	x	x	x	x
CAM91235.2	EndoPG-3 [Botryotinia fuckeliana]	x	x	x	x	x
CAQ03438.1	EndoPG-1 [Botrytis fabae]	x	x	x	x	x
CAQ03435.1	EndoPG-3 [Botrytis fabae]	x	x	x	x	x
CAM32882.1	EndoPG-5 [Botrytis fabae]	x	x			x
BAE80700.1	EndoPG [Chondrostereum purpureum]		x			x
AAA79885.1	EndoPG [Cochliobolus carbonum]	x	x	x	x	x
CAC14022.1	EndoPG [Colletotrichum gloeosporioides f. sp. malvae]	x	x	x	x	x
AAB36616.1	EndoPG [Cryphonectria parasitica]	x	x	x	x	x
ABF50893.1	EndoPG [Emericella nidulans]	x	x	x	x	x
ABQ11248.1	EndoPG [Fusarium anthophilum]	x	x	x	x	x
BAE97018.1	EndoPG [Fusarium oxysporum f. sp. fragariae]	x	x	x	x	x
BAE97133.1	EndoPG [Fusarium oxysporum f. sp. radidis-lycopersici]	x	x	x	x	x
ABQ11241.1	EndoPG [Fusarium subglutinans]	x	x	x	x	x
BAB61793.1	EndoPG-1 [Galactomyces geotrichum]	x	x	x	x	x
ABQ11236.1	EndoPG [Gibberella fujikuroi]	x	x	x	x	x
ABW96528.1	EndoPG [Gibberella moniliformis]	x	x	x	x	x
ABQ11245.1	EndoPG [Gibberella thapsina]	x	x	x	x	x
CAA61552.1	EndoPG [Glomerella lindemuthiana]	x	x	x	x	x
AAR84199.1	EndoPG [Kluyveromyces marxianus]	x	x	x	x	x
EAW19942.1	EndoPG [Neosartorya fischeri NRRL 181]	x	x	x	x	x
EDU44300.1	EndoPG [Pyrenophora tritici-repentis Pt-1C-8FP]precursor	x	x	x	x	x
AAM22187.1	EndoPG [Sclerotinia sclerotiorum]	x	x	x	x	x
CAD62160.1	EndoPG [Sclerotinia sclerotiorum]	x	x	x	x	x
ABH12113.1	EndoPG [Thanatephorus cucumeris]		x			x
BAG72133.1	EndoPG [Venturia pirina]	x	x	x	x	x
EED57938.1	ExoPG [Aspergillus flavus NRRL3357]					x
EED49587.1	ExoPG [Aspergillus flavus NRRL3357]					x
EED47501.1	ExoPG [Aspergillus flavus NRRL3357]					x
EDP47682.1	ExoPG [Aspergillus fumigatus A1163]					x
EDP52870.1	ExoPG [Aspergillus fumigatus A1163]					x
EAL85450.1	ExoPG [Aspergillus fumigatus Af293]					x
ABD61565.1	ExoPG-C [Aspergillus niger]					x
ABD61563.1	ExoPG-A [Aspergillus niger]					x
EAU32536.1	ExoPG [Aspergillus terreus NIH2624]precursor					x
ABF50897.1	ExoPG [Emericella nidulans]					x
AAO61898.1	ExoPG [Emericella nidulans]					x
BAE97090.1	ExoPG [Fusarium oxysporum f. sp. radidis-lycopersici]					x
EFY99815.1	ExoPG [Metarhizium anisopliae ARSEF 23]					x
EEA25131.1	ExoPG [Penicillium marneffei ATCC 18224]					x
ACA48699.1	ExoPG [Rhizopus oryzae]					x
AAP83023.1	ExoPG [Sclerotinia sclerotiorum]					x
EED19995.1	ExoPG [Talaromyces stipitatus ATCC 10500]					x
EY21174.1	ExoPG [Verticillium albo-atrum VaMs.102]					x
EY19557.1	ExoPG [Verticillium albo-atrum VaMs.102]					x





```

EED57938.1 -----MLITKTA 7
EDP52870.1 -----MLITNPA 7
ABD61565.1 -----MSV 3
BAH29710.1 -----MARVS 5
BAH14957.1 -----MPRVS 5
BAA96351.1 -----MTSFS 5
BAE80700.1 -----MPSFS 5
ACL80126.1 -----MHYL 4
ABH12113.1 -----MRFSL 5
AAV84618.1 -----
AAM22187.1 -----
ADD73449.1 -----
CAM91235.2 VATTLATVAKSTSAAAVSGVAASSAAAVPVASGSAASSYVRSYVKSVTASVSVQAHKA 94
CAQ03435.1 VATTLATVAKSTSAAAVSGVAASSAAAVPVASGSAASSYVRSYVKSVTASVSVQAHKA 94
CAD62160.1 -----MQLASATTSVTALT 14
ABQ11248.1 -----
ABQ11247.1 -----
ABQ11241.1 -----
ABW96528.1 -----
ABQ11245.1 -----
ABQ11246.1 -----
ABQ11236.1 | -----
ABQ11240.1 -----
ABQ11234.1 -----
BAE97018.1 -----
AAA74586.1 -----
CAK40265.1 NCDYKTKHLSSACSLGTDGSGAPASTPAAPAVSSSSKVGKAPAVAATRTSAIPTTFHTT 120
ABF51154.1 -----
BAB32924.1 -----MVAL 4
AAF70169.1 -----MVAL 4
EDU44300.1 -----
ABF50893.1 -----MF 2
CAM07141.1 -----MTKL 4
CAE46647.1 -----MSCSPSTHSLTHIHSLELYSRPNFLTPTSLEFFNNHFLPTT 41
AAV84616.1 -----MPSTKSMAMLLSANTLA 18
CAQ03431.1 -----MPSTKSMAMLLSANTLA 18
AAV84617.1 -----MVK 3
CAM32882.1 -----
EDP49013.1 -----MHF 3
EAW19942.1 -----MHF 3
CAK47667.1 -----MHF 3
ABF50868.1 -----MHF 3
EED53217.1 -----MQL 3
BAB97386.1 -----MLFS 4
BAB61793.1 -----MLFS 4
CAM91338.1 -----MVHIT 5
CAM91053.1 -----MVHIT 5
CAQ03438.1 -----
AAC14466.1 -----MVHILS 6
AAM00249.1 -----MLSLVLL 7
AAB36616.1 -----MFSTLLL 7
BAG72132.1 -----MMNTFS 6
BAG72133.1 -----MMNTFS 6
ABL01533.1 -----MVSSL 5
CAA61552.1 -----MVSYL 5
BAE97111.1 -----MFSST 5
BAE97133.1 -----MFSST 5
EY20327.1 -----MLTNLI 6
CAC14022.1 -----
CAA64727.1 -----
AAA79885.1 -----M 1
CAA71247.2 | -----MYSLL 5
CAK47020.1 -----MPSAKPL 7
ADZ99366.1 AAPHRFHDGQWKDIPFKFPSAEFFPGSGIAPPGRHGGFPGFPIGTGFPTGFSGSGMPI 80
AAR84199.1 -----MLFSN 5
EDN63471.1 -----MISA 4
ACA48699.1 -----MVQFL 5
EDP47682.1 -----MKFSYSFVQVVTLTLLSLSPSVEGFTRSRNDA-----CKPN 35
EAU32536.1 -----MKFLHTVAQTATLLSLGASVEGFNRARND-----CKPH 35
ABF50895.1 -----MKLSHLLTSAVSVLSLGLTVEGHFRRSRNDA-----VGPK 35
EED50415.1 -----MKLTQATTLTLLSLGLSLPVEGFSLRSTNA-----VGPK 33
ABD61562.1 -----MRLTHVLSHTLGLLALGATAEAFSRSREAA-----CSPK 34
EER22673.1 -----MRFSTLQLLLAATTVLAVPSIDERSPSISRKNQ-----VGPK 38
EED15423.1 -----MKFNFSLKLLVASTALAAPSEIERTTAVSRNDI-----IGPK 38
AAF05088.1 -----MHFQLSTILSIASA AVASPTPQDDSDWQALARRGHPKHKARSARNVQFHPK 52
BAE97166.1 -----MLFNVISAVAVSALVSSVGTSSRNSALPKRPHVEAAP-----YGTG 43
EY19557.1 -----MIFGQQLLVGLLNLSSALVSAHSSPQSIPIRFSIEAAP-----YNTG 43
AAP83023.1 -----
ABD61563.1 -----MKLPILVTLFITLPALCVSSKTPSAPTISAYP-----KSPG 36
ABF50897.1 -----MGFKRTIGLLGLLALDQVSVLAQGRPTFAKRPDIQP-----NPIQP 44
EED46018.1 -----MKFFLATLFAVAVSIAVDRLIPGAQIIPESD-----TRALL 37
EAL84906.1 -----MKFFALAAALFASTVNSIAVDGLIPGARVIPAND-----VVALK 38
AAO61898.1 -----MRLHFLPLVALCATTASSLAFDSSTSQPPGAQVSVND-----AAGLK 43
ABD61564.1 -----MYLLPLTLFLTAAGVSIIPRPLIPGAQIVPASS-----TADLR 39
BAE97090.1 -----MKVSSFFSTLSLFAVGLAEADLPAVPRPISIE-----FRAXH 38
EY21174.1 -----MAETSPGEIISTRGRR-----RNQTP 20
EY20199.1 -----MHLKSVAVTLLAFFSNASLAIELPAGVPRPISIE-----FREKH 38
EY16298.1 -----MRFQVLSFVAVATAAVEAVSLPPGVPRNVLE-----FRDKH 38
EED49587.1 -----
EED23059.1 -----MRLTLRSLLLAPVA 15
EED47623.1 -----MFLSRVFC 8
EAL86831.1 SRCIGRSATIKVTRTRDNHERYAVVIGGSHESYARQSCQPIVVDIMYLANILPWVLLPLA 95
AAP83024.1 -----NAYIRLFGKFKSIKMRSFHLAIALAGSLFFD 30
EY20240.1 -----MRLTLTLLAALQWLVAVAVDKWNVLDHSDVLAEQWLPDTEIS 43
EY99815.1 -----MRLTLTLLAALQWLVAVAVDKWNVLDHSDVLAEQWLPDTEIS 43
EER27533.1 -----MYWLRNLLPLTICSTAFAA 19
EED19995.1 -----MYWLTNLAPIAWICATALAA 20
EDP55950.1 -----MRPSIISLVALSFLG 16
EED47501.1 -----MISLNSIFLLSVGLSRAAPSR 22
EAL85450.1 -----MHKMLYPRNLALFSLLSAAPSQ 25
EER25131.1 -----MFLLDGCFIFHQTDGSGMSGLKCKHDLQCAHIVSS 36

```

EED57938.1	FLAFLSSVPLAHGAGGN-----S	26
EDP52870.1	LLGILASLVPLALGAP-----N	24
ABD61565.1	FKASFLFLLSSSLVHGVP-----H	22
BAH29710.1	TLLGASMLIAGAYAAP-----HANS	25
BAH14957.1	TVLGASFLASAVSAVPY-----ANSTE	28
BAA96351.1	SLFKLLILASATAAAP-----	21
BAE80700.1	SLIKHLALLSLVAAAP-----	21
ACL80126.1	SFAALAFAPILLATATP-----	20
ABH12113.1	GFLATLPTALVAASPA-----	21
AAV84618.1	-----MPKNSQISG-----LLALTL	15
AAM22187.1	-----MHRDFSIPG-----LLALTL	15
ADD73449.1	-----	
CAM91235.2	TTGVVASTSVPAVSSAAG-----ASTAVI	118
CAQ03435.1	TTGVVASTSVPAVSSAAG-----ASTAVI	118
CAD62160.1	HVAVTTSASSTAVASSAS-----AASAVI	38
ABQ11248.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11247.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11241.1	-----MVRNIAIAALLP-----AAWALP	18
ABW96528.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11245.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11246.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11236.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11240.1	-----MVRNIAIAALLP-----AAWALP	18
ABQ11234.1	-----MVRNIAIAALLP-----AAWALP	18
BAE97018.1	-----MVRNIAIAALLP-----AAFAS-	17
AAA74586.1	-----MVRNIVSR LCS-----QLFALP	17
CAK40265.1	AVRVLSTSAAAA VTSQA-----VLP AQ S	144
ABF51154.1	-----	
BAB32924.1	TLGFVATALVASSVAAPA-----PAIT-PAPK	30
AAF70169.1	TLGFVATALVASSVAAPA-----PEIDGAHNA	31
EDU44300.1	-----	
ABF50893.1	YALGPLALFAFATEVMAT-----PVAYPMTTA	29
CAM07141.1	SLLLGALAASVCVQSHAL-----PSPTVTQA	31
CAE46647.1	KMVKSGSLIAAVLAASVM-----ASPTITNG	68
AAV84616.1	PALANPIAAPMPTAAPSP-----VEVESAFVE	45
CAQ03431.1	PALANPIAAPMPTAAPSP-----VEVESAFVE	45
AAV84617.1	FSACLLGGLSALASALP-----AAPAFTAA	30
CAM32882.1	-----	
EDP49013.1	FQSSSLVAATMGAALVAAA-----P	22
EAW19942.1	FQSSSLVAATMGAALVAAA-----P	22
CAK47667.1	LQNAVVAATMGAALAAAA-----P	22
ABF50868.1	LQNSLIAAAMGAALVAAA-----P	22
EED53217.1	LQSSVIAATVGAALVAAV-----P	22
BAB97386.1	TTAILAMAAMVAAP-----A	20
BAB61793.1	KSILLSMAAALVAAPTE-----T	23
CAM91338.1	SLISFLASTALVSAAPGS-----A	24
CAM91053.1	SLISFLASTALVSAAPGS-----A	24
CAQ03438.1	-----VSAAPAP-----A	8
AAC14466.1	SALSLRLRGAASVSAAPAP-----A	25
AAM00249.1	AALLPLIQAVPAPAVTPA-----A	26
AAB36616.1	AALLPLIQAVPAPAVTPA-----A	26
BAG72132.1	ALAILAYVAAAAPALRAV-----D	25
BAG72133.1	AFAILAYVAAGAPAFRAV-----D	25
ABL01533.1	LALGALAATAIATAP-----	19
CAA61552.1	FVLGALASVAIASPV-----	20
BAE97111.1	VLLSGLVASVLAAP-----	19
BAE97133.1	VLLSGLVASVLAAP-----	19
EY20327.1	LALGALASTAANPLPAK-----F	25
CAC14022.1	--MRVLPMIAGFAALASA-----A	17
CAA64727.1	--MRVLPMIAGFAALASA-----A	17
AAA79885.1	VAYALTSMLLSAGALVAA-----A	20
CAA71247.2	MLLLPVWTGSFVSAAPAV-----S	24
CAK47020.1	FCLATLAGAALAAPAPS-----R	25
ADZ99366.1	IPGSLSGSSSVLPTTSSAALSTSVVKSSNITSSLSAAPPSTTASSSSSNSTGGASAGE	140
AAR84199.1	TLLIAAASALLAEASF-----	21
EDN63471.1	NSLLISTLCAFATATP-----	20
ACA48699.1	SLSSVAFFLLSLG-----	20
EDP47602.1	HPFRPLPPSQPRTKTCHV-----	53
EAU32536.1	HPFRPLPASTPRTKTCHV-----	53
ABF50895.1	RPFKPLPYSHPRKVKVCHV-----	53
EED50415.1	PPFRPLPASTPRNKTCQV-----	51
ABD61562.1	KPFKPLPTSSSRDKTCHV-----	52
EAA22673.1	CAFKPLPETAKRNKVKVCHV-----	56
EED15423.1	NSFKPLPETCKRNKVKVCHV-----	56
AAF05088.1	SLDTPAPSPAARTKTCTYV-----	70
BAE97166.1	KAPPASPARSRKDFCYVN-----	61
EY19557.1	RALPVSPTRREDRYCHVP-----	61
AAF83023.1	-----	
ABD61563.1	NFKPAS--GRQNSTSNVC-----	52
ABF50897.1	YKAIPLHSRRNSHKVCYV-----	62
EED46018.1	RVGGHHDKYHDRRTITIR-----	55
EAL84906.1	KAGAYHKHHHRRVTIIR-----	56
AAO61898.1	RIGAHHPKYHDRRTVTIIR-----	61
ABD61564.1	AIGAQHKYDPDRETVTIIR-----	57
BAE97090.1	PYELP-VKRSRHKVYTIIR-----	55
EY21174.1	KRGAR-RGLVPEIREIR-----	37
EY20199.1	PYRARSTKGDCKRVTRIR-----	56
EY16298.1	PYKNP--KPGCRKIVRIR-----	54
EED49587.1	TANVDNQGTCTLYPESL-----	31
EED23059.1	LGYYVNEGTTYQIYPEAL-----	33
EED47623.1	VLAALACVSAARETCVVP-----	26
EAL86831.1	VIADATTGQSHRRTCVVK-----	113
AAF83024.1	STNGDPSVHSPRATCTVQ-----	48
EFY92040.1	LLMLRIAEQDGRKLCILR-----	61
EFY99815.1	LLALRIAEQGRKLCILR-----	61
EAA27533.1	AASKGGVWQ--KGVKCNV-----	35
EED19995.1	AASKGGVWK--KGNVCTV-----	36
EDP55950.1	LAYAGGVEEDKRDNVCTV-----	34
EED47501.1	SETSPDRTIKFRACCTPT-----	40
EAL85450.1	VERSPDVLKPRAVCTPT-----	43
EAA25131.1	TALPPSKTIEKRAICTPG-----	54

EED57938.1 SSPDARGRRCCVRS----SNGTADDSPEVSRVFAQCATNSVIVFQEGVDYNIHQPIKATN 82  
EDF52870.1 QPQARSKRCVPISSYASSHGTTADDSVAVSAFAQCAENSIVIVFQEGVDYNIHFPIKATN 84  
ABD61565.1 SSRASRSQCCVPSKYQASNGMADDSVAVSQAFQCATDSDVIFEEGVNINIHFQPIATN 82  
BAH29710.1 TEVIGKRCTATIKS-----ASDITSDNLKCTTINIES---FTMTAGETMA 67  
BAH14957.1 FVRIGKRCTSTISS-----LSDVTDSELETTINIKS---FVSPAGKTFD 70  
BAA96351.1 ----SKRATCTVKS-----VDDAK-DIAGCSAVTLNG---FTVPAGNTLV 58  
BAE80700.1 ----SKRATCTVKS-----VADSN-DIAGCAAVTIDS---FTVPAGSTID 58  
ACL80126.1 ----VSRCTGTIAS-----LDDVA-AAQKCTVTITKG---FTVPAGKTFE 57  
ABH12113.1 ----GKRCTGTISS-----LNDVA-AAQKCTTININA---FTVPAGQTEA 58  
AAV84618.1 AGACTAQTACTASV-----YSQIAPCVASSTAIVLNN---VFAPSGSSID 57  
AAM22187.1 AGVCTAQTACTASV-----YSQIAPCVASSTAIVLNN---YAPSGSSID 57  
ADD73449.1 -----  
CAM91235.2 TSAPAAPAGCTATA-----YADIADI IASCTNIVLDN---ISAPASSTID 160  
CAQ03435.1 TSAPAAPAGCTATA-----YADIADI IASCTNIVLDN---ISAPASSTID 160  
CAD62160.1 TPAPSAPAGCTATA-----YADIAFVVASCTNIVLDN---ISAPASSTID 80  
ABQ11248.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLNG---FQVPTGKQLD 60  
ABQ11247.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLNG---FQVPTGKQLD 60  
ABQ11241.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLNG---FQVPTGKQLD 60  
ABW96528.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLAG---FQVPTGKQLD 60  
ABQ11245.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLAG---FQVPTGKQLD 60  
ABQ11246.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLTG---FQVPTGKQLD 60  
ABQ11236.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLAG---FQVPTGKQLD 60  
ABQ11240.1 SSSLQERDACTVTD-----YSGLATAVSSCTNIVLAG---FQVPTGKQLD 60  
ABQ11234.1 STQLKERDACTVTD-----YSGLATAVSSCTNIVLAG---FQVPTGKQLD 60  
BAE97018.1 --TLPKRPDPCSVTD-----YSGLATAVSSCTNIVLNG---FQVPTGKALD 57  
AAA74586.1 SSSLQERDPCSVTE-----YSGLATAVSSCTNIVLNG---FQVPTGKQLD 59  
CAK40265.1 SVAGNGGTTCTVTE-----YASISSAVASCSNILLSN---INAPASSTID 186  
ABF51154.1 -----VLSD---VAVPSGTTLD 14  
BAB32924.1 FEVVKRASSCTFSG-----SNGAAEASKSQSSCATMVLSD---VAVPSGTTLD 75  
AAF70169.1 FVTAKRADSCFTFG-----ANGAAEASKASKSCTMVLSD---LAVPAGTTLD 76  
EDU44300.1 -----MTLSS---LDVPAGTTLD 15  
ABF50893.1 SPTLAKRDSCTFSG-----SDGAASASRSQTDCAITLSD---ITVPSGTTLD 74  
CAM07141.1 PELEDRATCTFSG-----ANGASSASKSQKSCATIVLSN---VAVPSGTTLD 76  
CAE46647.1 DEVMKRASCTFSG-----ASGAAQAMKEQASCSITLKN---LQVPAGTTLD 113  
AAV84616.1 RNIKRAATCTFSG-----SLGYSSASKSKASCSTIILKE---LTVPGGVTTLD 90  
CAQ03431.1 RNIKRAATCTFSG-----SLGYSSASKSKASCSTIILKE---LTVPGGVTTLD 90  
AAV84617.1 PDLDRRATCTFSG-----SGGASSASKSKSCTIILSA---LAVPAGTTLD 75  
CAM32882.1 -----  
EDP49013.1 AADLETRGSCFTFS-----TSALKSGKASCSTITLQN---IAPVAGETLD 64  
EAW19942.1 AADLETRGSCFTFS-----TSALKSGKASCSTITLHN---IAPVAGETLD 64  
CAK47667.1 LE---KRSCFTFS-----ASAASKSGKSCSTITLND---IAPVAGETLD 60  
ABF50868.1 AADLDARSSCTFST-----ASAASKSGKSCSTITLKS---IQVPAGETLD 64  
EED53217.1 VE---LEARDSCFTFS-----AADAKSGKSCSTITLNS---IEVPAGETLD 63  
BAB97386.1 DGLHARASCTFKD-----AKSAIAGKKSCTNIVLQD---IAPVAGETLD 62  
BAB61793.1 EVNLRYARDSCVFST-----ADKAIADKKSCTITLKN---IAPVAGKTLD 65  
CAM91338.1 PADLDRRAGCTFST-----AATAIASKTCSTIILDS---VVVPAGTTLD 66  
CAM91053.1 PADLDRRAGCTFST-----AATAIASKTCSTITLNS---VVVPAGTTLD 66  
CAQ03438.1 PTAAPNPAEALAAIEQRGTACTFAGAGAAAASKSCTIVLSA---LSVPSGTTLD 64  
AAC14466.1 PTAAPNPAEALAAVEKRAGSCTFSDPAVPLPAIKSKASCATIVISA---VAVPSGTTLD 81  
AAM00249.1 QLVDRAKSGSCTFTD-----AAVSKSCTSCATITLNS---IAPVAGTTLD 68  
AAB36616.1 HLEDRASKSCTFTD-----AAVSKSCTSCATITLNN---IAPVAGTTLD 68  
BAG72132.1 SAIEERASSCTFTT-----AAAAISGKKSCTITLNS---IAPVAGTTLD 67  
BAG72133.1 SAIEERASSCTFTT-----AAAAISGKKSCTITLNS---IAPVAGTTLD 67  
ABL01533.1 ---LDARASCTFTD-----AAAAIKGKASCTIILNG---IVVPAGTTLD 58  
CAA61552.1 ---PELKRASCTFTD-----AASAIKKGKASCTIILNN---IAPVAGTTLD 61  
BAE97111.1 ---ALEPRAGSCTFTD-----AATAIKNKAGCSTITLNN---IAPVAKTTLD 60  
BAE97133.1 ---ALEPRAGSCTFTD-----AATAIKNKAGCSTITLNN---IAPVAKTTLD 60  
EY20327.1 EVLESRACTACTFTN-----TADLAKSKKTCDSITLNN---IAPVAGTTLD 67  
CAC14022.1 FAELDTRASCFTFD-----AKTAMSKKTSCTDIVLNG---IKVPAGQTTLD 59  
CAA64727.1 FAELDTRASCFTFD-----AKTAMSKKTSCTDIVLNG---IKVPAGQTTLD 59  
AAA79885.1 PSLGDRARGDCTFTD-----AATAIKNKASCSNIVISG---MTVPAGTTLD 62  
CAA71247.2 RAAALSGADCTFSD-----VAQAIQKKTSCSTITLNR---LKVPAKTTLD 66  
CAK47020.1 VSDFTKRSTCTFTD-----AATASEKTSCTSDIVLKD---ITVPAGETLN 67  
ADZ99366.1 VMSGASBSCTFST-----ASDAISGKKSCTNIVLSG---ISVPAGTTLD 182  
AAR84199.1 ---LEKRDSCTLSG-----KTAG--GGLSNCAVTVNN---VEVPAGTTLD 59  
EDN63471.1 ---LSKRDSCTLTG-----SSLSSLSTVKKCSIVIKD---LTVPAQTTLD 60  
ACA48699.1 IHDVAAAANCVVAP-----SSGDAANIIVSAFNSCENN---GGVTFTKGV 62  
EDP47682.1 ---VSNHGKDDSKN-----IMKALHKCNNGGKVVFDAN---KVYVTGTL 93  
EAU32536.1 ---RSHGDGTDAAAF-----VLSALRKNNGGKVVFDAD---KEYTIGTAL 93  
ABF50895.1 ---RSHGDGRDDSAF-----ILSALKSCNNGGKVVFAEE---KEYTIGTAL 93  
EED50415.1 ---QSGDGTDDAPY-----ILAALKQCNNGGKVVFAED---KEYTIGTAL 91  
ABD61562.1 ---RSHGDGSDSDY-----ILSALHQCNNGGKVVFEDE---KEYIIGTAL 92  
EED22673.1 ---KTHGKDDSDY-----ILAALKSCNNGGKVVFD---KDYTIGTAL 94  
EED15423.1 ---RSHGDGKDDSDN-----ILDALRTCNNNGKVVIFD---QYTIIGRAL 94  
AAF05088.1 ---KSSNGTDDSPA-----IMSALHECNNGGHVIFSKG---VTYIIGTAM 110  
BAE97166.1 ---PGKGNVDDAPS-----ILKAFKKNKGG-TIVLVD---QKYSIASPL 99  
EY19557.1 ---ASCNNGDDAPR-----LLAAIKCNNGG-TVVLVD---KVYRIASPL 99  
AAP83023.1 -----  
ABD61563.1 ---EVKFNQTDAAAPG-----ILAAAHTCNNGG-TVFLPP---GDFVVATAL 91  
ABF50897.1 ---KPSLNGDDAGR-----VEAALRRCNNGGTVLVDK---EYSICTPL 100  
EED46018.1 ---PSRNDTDDVSRD-----FLWGIKIHANHGGRLLKKG---EKYVIGRKL 95  
EAL84906.1 ---SSSDEDDVSAD-----FLWGIKIHANHGGRLLKQG---KKYVIGKKL 96  
AAO61898.1 ---SSHNDTDDVSAD-----FLWGIKIHANHGGRLLKKG---QKYVIGRKL 101  
ABD61564.1 ---ASRNALDDVSSD-----FLWGLKQANHGGRLLKQG---ETYVIGKKL 97  
BAE97090.1 ---HSKNNDNDVSSD-----FYKGLKKNKGGTLYLPKG---QTFVIGKPL 95  
EY21174.1 ---ASNNDTDDVADA-----FLKGLHDANHGGRLLKQG---ETFVIGKPL 77  
EY20199.1 ---PSKHDDDISDD-----FLEGIKQANNGGTLHLPKQ---ELFVIGKPL 96  
EY16298.1 ---ASEDDEDDVSAE-----FKKGVSEANNGGTLVLPKG---ETFVIGKVL 94  
EED49587.1 ---THQQVDDSDPS-----IQQAFDMCGTNGTIVFSE---NVFHVNTVL 70  
EED23059.1 ---THNGQAVDDAPS-----IHQAFDLOGINGTIVFTN---NTFHNSTL 72  
EED47623.1 ---AGGSNATDDAPA-----IVHAFKCKGRGGKVVFKP---TTHYIGSVM 65  
EAL86831.1 ---AGGSEVDDTPA-----ILNAPRRCKGKGRVFNQ---TTHVNSVM 152  
AAP83024.1 ---PSLN-GTDDAPA-----ILSAFKTCGKDGTVVFLN---HTYINSIM 86  
EFY92040.1 ---PKSR-GRDNAPM-----IMKALNDDCREHSLVYLPK---EVNINSFM 100  
EFY99815.1 ---PKSL-KRDNAPM-----IMKTLNEVCREKSLVYMPG---EVNINSFM 100  
EED27533.1 ---TALGKRRDDVFN-----ILEAFSRCKGKGTIVFPED---QNYWIATRL 75  
EED19995.1 ---TALGKRRDDVFN-----ILEAFSSCKGKGTIVFPED---ENYWIATRL 76  
EDF55950.1 ---KANGKQDDVFN-----LLRAFNECGNGGTVIFPED---QSYWIGARL 74  
EED47501.1 ---AGSSSTDDVPA-----IQEAITSCGDDGTVIIPAD---TYYLNSVL 80  
EAL85450.1 ---AGGSPSIDVPA-----IRKAIASCNGGTVIFPAG---STYYLNSVL 83  
EED25131.1 ---GYGNASKDDTPA-----IASAIATCGKGGIVIPSG---TNYNRSTL 94

EED57938.1	LSNVEIQMHGNLHLPQNISA-----VRDIVNA-----	109
EDP52870.1	LSNVEIRVLGNLHLPQDITA-----VQNIIVKS-----	111
ABD61565.1	LSNVEIRMHGNLHLPQNITA-----VQNIIVSD-----	109
BAH29710.1	LTNLAGSTVNNLLGEVTFG-----FREWAGP-----	93
BAH14957.1	ISNLADCTVNNLLGDVTFG-----YKEWDGP-----	96
BAA96351.1	LN-PDKGATVTMAGDITFA-----KTLLDGP-----	83
BAE80700.1	MK-PDSGATVTMAGDLTFA-----QTLLDGP-----	83
ACL80126.1	LS-LLDNTVVNMEGDVKFG-----VANWAGP-----	82
ABH12113.1	IA-ALDGTVVNMLGDVKFG-----VANWAGP-----	83
AAV84618.1	LTGVKAGTTITFAGKTTFGF-----TNDSSFE-----	84
AAM22187.1	LTGVKAGTKITFAGKTTFGF-----TNDSSFD-----	84
ADD73449.1	-----	-----
CAM91235.2	LQKLQDGSVTVFSGTTSFGT-----TADSDFD-----	187
CAQ03435.1	LQKLQDGSVTVFSGTTSFGT-----TADSDFN-----	187
CAD62160.1	LQKLKDGTTVTFSGKTTFGT-----TSDDSFN-----	107
ABQ11248.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11247.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11241.1	LSKLKAGTVTFKGKATFAT-----TADNDFD-----	87
ABW96528.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11245.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11246.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11236.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11240.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
ABQ11234.1	LSKLKAGTTVTFKGKATFAT-----TADNDFD-----	87
BAE97018.1	LSKLKAGTVTFKGKATFAT-----TADNDFD-----	84
AAA74586.1	LSSLQNDSTVTFKGKATFAT-----TADNDFN-----	86
CAK40265.1	LTGLQGAAVIFAGETTFGD-----YDSDSFD-----	213
ABF51154.1	LSSLADCTVIVFEGTTTWG-----YSEWKGFP-----	40
BAB32924.1	LSSLADGTTVIFEGTTTWG-----YSEWKGFP-----	101
AAF70169.1	LSSLADGTTVIFQGTTFWG-----YKEWAGP-----	102
EDU44300.1	LSKLADGTTVIFEGTTTFG-----FKEWEGP-----	41
ABF50893.1	LSDLDDTTVIFEGTTSWE-----YEWDGFP-----	100
CAM07141.1	LSDLNDGTTVIFQGTTFWG-----YKEWSGP-----	102
CAE46647.1	LTGVKANTRVIFEGETTFG-----YKEWEGP-----	139
AAV84616.1	MTDLPDDTTVIFQGETSFA-----FKEWEGP-----	116
CAQ03431.1	MTDLPDDTTVIFQGETSFA-----FKEWEGP-----	116
AAV84617.1	LTGLTKGTTVIFEGITTFG-----YEWSGFP-----	101
CAM32882.1	-----	-----
EDP49013.1	LTGLKAGTTVVFQGTTFFG-----YKEWEGP-----	90
EAW19942.1	LTGLKGTGTVVFQGTTFFG-----YKEWEGP-----	90
CAK47667.1	LTGLKGTGTVIFEGETTFG-----YKEWKGFP-----	86
ABF50868.1	LTGLKSGATVIFEGETTFG-----YKEWKGFP-----	90
EED53217.1	LTGLNDGTTVIFSGETTFG-----YKEWEGP-----	89
BAB97386.1	LTGLNKGTVVIFAGTTTFG-----YKEWVGP-----	88
BAB61793.1	LSNLNKGTTVIFQGTTFFG-----YKEWTFP-----	91
CAM91338.1	LTGLTKGTVVIFQGTATFG-----YSEWEGP-----	92
CAM91053.1	LTGLNTGTVVIFQGTTFFG-----YSEWEGP-----	92
CAQ03438.1	LTGLNSGTVVIFEGTTTFG-----YEWSGFP-----	90
AAQ14466.1	LTGLKSGTHVVFEGTTTFG-----YEWSGFP-----	107
AAM00249.1	LTKLNSGTVVIFEGTTSFG-----YKEWGLP-----	94
AAB36616.1	LTKLNSGTVVIFAGTTSFG-----YKEWEGP-----	94
BAG72132.1	LTGLTKGTVVIFSGTTSFG-----YKEWEGP-----	93
BAG72133.1	LTGLTKGTVVIFSGTTSFG-----YKEWEGP-----	93
ABL01533.1	MTGLKSGTTVTFQKTTTFG-----YKEWEGP-----	84
CAA61552.1	MTGLKSGTHVVSFGKTTFG-----YKEWEGP-----	87
BAE97111.1	LTKLNDGTHVIFQKTTTFG-----YAEWEGP-----	86
BAE97133.1	LTKLNDGTHVIFQKTTTFG-----YAEWEGP-----	86
EY20327.1	LSSLKSGAHVVFQKTTTFG-----YKEWEGP-----	93
CAC14022.1	LTGLKDGTRVTFKGTTFFG-----YKEWEGP-----	85
CAA64727.1	LTGLRDGTVVTFKGTTFFG-----YKEWAVL-----	85
AAA79885.1	LTGLKSGATVTFQGTTFFG-----YKEWEGP-----	88
CAA71247.2	LSDLNDGTRVIFTGRTTFE-----YQEWEGP-----	92
CAK47020.1	LKDLNDGTTVTFEGTTTWE-----YEWDGFP-----	93
ADZ99366.1	MTDLDGTTVTFEGTTTFG-----YEWDGFP-----	208
AAR84199.1	LTGLQDGTVNVFQVTFD-----YDEWVGP-----	85
EDN63471.1	LTGLSSGTTVTFEGTTTFQ-----YKEWSGP-----	86
ACA86699.1	TYNLNSVTVIQGLKHVNVNF-----AGTINLP-----	89
EDP47682.1	DMTFLKHIDLEVLGKIQFT-----NDTDYQAN-----SFKHG-----	126
EAU32536.1	DMTFLRHDLEILGRIQFT-----NDTDYQAH-----AFRHH-----	126
ABF50895.1	DLTFLKHVDLEILGRIQFT-----NDTDYQAN-----SFKHT-----	126
EED50415.1	DMTFLKHVDLEILGKIQFT-----PDTDYQAN-----SFKHT-----	124
ABD61562.1	NMTFLKHIDLEVLGKILFT-----NDTDYQAN-----SFKQG-----	125
EAA22673.1	DIQWLKHIDLDITGTIQFT-----NDTNYWINN-----AFQIV-----	127
EED15423.1	DLQFLKHIDLDITGTMKFT-----NDTNYWINN-----AFQIT-----	127
AAF05088.1	DWTFLKHIDLDIQGELLFS-----DDTAYWEAN-----SFKFV-----	143
BAE97166.1	DLTFLAHVDVITGEVNFK-----SDPYWADH-----SFKYD-----	132
EY19557.1	DLTFLRHDIDVITGEIHPAD-----EDVYYWAEN-----SFKYK-----	133
AAP83023.1	-----	-----
ABD61563.1	DLTFLNNDFAIWNITFK-----KDIDLWTTQ-----AFQYT-----	124
ABF50897.1	DLRFLKHVDVALTGKVEFC-----PELEFWQON-----VFQPH-----	133
EED46018.1	DLTFLDDIEVQLDGEL-----KFTDDVSYQEN-----NFFYD-----	128
EAL84906.1	DLTFLKIDIEVQLDGELKVLWPKILDVLAGVNTIEKFTNDVPYQAN-----NFFYD-----	148
AAO61898.1	DLTFLDNVEVQLDGELK-----FTDDVPYQEN-----NFFYD-----	134
ABD61564.1	DLTFLDNIEVQLEGEIQ-----FTNNITYQAN-----NFFYD-----	130
BAE97090.1	DLTFLNDIHVHLEGEIKF-----TNDTEYQKN-----AYKHP-----	128
EY21174.1	DLTFLNNDVHVRDLGEIKF-----TNDTPFQAN-----SFRHP-----	110
EY20199.1	DLTFLNDIHVRDLGKILF-----TDDVPQWAT-----AFKHP-----	129
EY16298.1	DLTGLSDIHVRDLGEIKVG-----FMGIDICIRLL-----MACSSRIM-----	133
EED49587.1	NTNLLNCDVHLRGLRFS-----ANVPYWRTHAISVV-----	103
EED23059.1	NTNLLNCDVHLRGLRFS-----TNTPYKSHVYSVI-----	105
EED47623.1	NI TWLEDVDIDLGGTLLWG-----TNIQYWLNHS LPVG-----	98
EAL86831.1	NTSGLNCDIDLGYTLLWG-----TNSYWLNSLSLDVG-----	185
AAP83024.1	NTTGLSNCDIDYGTLLWS-----TNITYWLNQSMAPG-----	119
EY202040.1	ITTNMRDKVLLVLAGRLLWS-----GDVDHVLVDSMPIG-----	133
EY99815.1	VTTNMVDVKIILTGRLOWS-----GNIDHVLVRSMPIG-----	133
EAA27533.1	N-PIVSNVVVIDRGVWTF-----DALDYWRNHSYPIA-----	107
EED19995.1	N-PVSNVVVIEWRGVWTF-----DDFDYWRNHSYPIA-----	108
EDP55950.1	N-PVLNDVSVQWRGKWTFS-----DNFDYWRNHSFPVA-----	106
EED47501.1	DFKGCSCNDFQVEGLLQFT-----SSTDYWNGK-----	108
EAL85450.1	DLAGCSNCDIQVEGLKFS-----GSTYWGK-----	111
EAA25131.1	DFAGCFGCELQLKGLTKVS-----NDYQCWSY-----	122

EED57938.1 -----GTSTWFTLEGPRVDWTGPE--DVNNGWI 135  
EDF52870.1 -----GQSTWFTFQGPVRVDWTGAD--DIKNGWI 137  
ABD61565.1 -----GTSTWFTLEGPKVDWIGPE--DVNNGWI 135  
BAH29710.1 -----LFEIGASSDSGFTTFNNGN-----GYSL 115  
BAH14957.1 -----LFGIGTKSKTATWHFNNGN-----GHTL 118  
BAA96351.1 -----LFTIDGTG----INFVGA-----DHIF 101  
BAE80700.1 -----LFTIEGTN----IQFVGA-----NHKF 101  
ACL80126.1 -----LFSVSGKG----ITFNNGN-----GHTF 100  
ABH12113.1 -----LFSITGNK----LVFNNGN-----GHTL 101  
AAV84618.1 -----PIKLGSGG---ITVTA-----EPDAII 103  
AAM22187.1 -----PIKLGSGG---IHVTG-----APGAVI 103  
ADD73449.1 -----  
CAM91235.2 -----PIVVKGTD---ITITG-----AAGHVI 206  
CAQ03435.1 -----PIVVKGTD---ITITG-----AAGHVI 206  
CAD62160.1 -----PIVVKGKN---ITLGT-----APGHVI 126  
ABQ11248.1 -----PIVISGSG---ITITG-----ASGHVI 106  
ABQ11247.1 -----PIVISGSG---ITITG-----ASGHVI 106  
ABQ11241.1 -----PIVISGSG---ITITG-----ASGHVI 106  
ABW96528.1 -----PIVISGSG---ITITG-----ASGHVI 106  
ABQ11245.1 -----PIVISGSG---ITITG-----ASGHVI 106  
ABQ11246.1 -----PIVISGSG---ITITG-----ASGHVI 106  
ABQ11236.1 | -----PIVISGSG---ITITG-----ASGHVI 106  
ABQ11240.1 -----PIVISGSG---VTITG-----ASGHVI 106  
ABQ11234.1 -----PIVISGSG---ITITG-----ASGHVI 106  
BAE97018.1 -----PIVISGNG---ITITG-----ASGHVI 103  
AAA74586.1 -----PIVISGSN---ITITG-----ASGHVI 105  
CAK40265.1 -----PIVISGTD---VTITG-----EEGHVI 232  
ABF51154.1 -----LLDIQGGK---ITVKG-----AEGSVL 59  
BAB32924.1 -----LLDIQGGK---ITVKG-----AEGSVL 120  
AAF70169.1 -----LLSIGGNK---ITVKG-----ASGSL 121  
EDU44300.1 -----LLSIGGNK---IVVKG-----AAGSSL 60  
ABF50893.1 -----LLQIKGNG---ITIKG-----ADGAKL 119  
CAM07141.1 -----LLQIEGND---ITIQG-----ASGSL 121  
CAE46647.1 -----LLSIAGTG---LTVEG-----ASGSL 158  
AAV84616.1 -----LFAVSGNN---IKVAGS---NADTAIL 137  
CAQ03431.1 -----LFAVSGKN---ITVAGS---NADTAIL 137  
AAV84617.1 -----LVSVSGTD---ITVTQ-----TTGAYL 120  
CAM32882.1 -----  
EDP49013.1 -----LISASGTS---ITIKQ-----NPGAKI 109  
EAW19942.1 -----LISASGTS---ITIKQ-----NPGGKI 109  
CAK47667.1 -----LISMSGTD---ITVKG-----ASGAKI 105  
ABF50868.1 -----LISMSGDK---ITVKG-----ASGAKI 109  
EED53217.1 -----LISVSGTN---IKVQQ-----ASGAKI 108  
BAB97386.1 -----LISVSGDS---ITVNQ-----ASGGKI 107  
BAB61793.1 -----LIKVSGDT---ITVKG-----TSGAKI 110  
CAM91338.1 -----LISISGQD---IVVTG-----ASGNKI 111  
CAM91053.1 -----LISISGQD---IVVTG-----ASGNKI 111  
CAQ03438.1 -----LFSVSGTD---ITVKG-----ASGSKL 109  
AQC14466.1 -----LVSVSGTD---ITVTG-----ASGSL 126  
AAM00249.1 -----LISVSGTD---IEVTG-----ASGHVI 113  
AAB36616.1 -----LISVSGTD---IEVTG-----ASGHVI 113  
BAG72132.1 -----MISIAGTG---IKVSG-----ASGHVI 112  
BAG72133.1 -----MISIAGTG---IKVSG-----ASGHVI 112  
ABL01533.1 -----LISFSGTN---INING-----ASGHSI 103  
CAA61552.1 -----LISFSGSN---VVIDG-----ASGHSI 106  
BAE97111.1 -----LISFTGNN---LLIEG-----AAGHSI 105  
BAE97133.1 -----LISFTGNN---LLIEG-----AAGHSI 105  
EY20327.1 -----LISFTGNN---LLIEG-----AAGHSI 105  
CAC14022.1 -----LIAVGGKK---VAVVG-----ASGSLI 104  
CAA64727.1 -----LIAVGGKK---VAVVG-----ASGSLI 104  
AAA79885.1 -----LISVSGTN---IKVVG-----ASGHTI 107  
CAA71247.2 | -----LIAVSGTN---IKVSG-----SPGNII 111  
CAK47020.1 -----LLRISGKD---ITVTQ-----SSDAVL 112  
ADZ99366.1 -----LISFSGTD---ITIQG-----ADGHVI 227  
AAR84199.1 -----LVSVSGNN---IKVVG-----KSGHLL 104  
EDN63471.1 -----LISISGSK---ISVVG-----ASGHTI 105  
ACA48699.1 -----ARSSSPKNDHYIELKGDNVQVFGGGTV 117  
EDP47682.1 -----FQNATTFPQLGGEDVNVYG-----GGTL 149  
EAU32536.1 -----FQNATTFPQLGGEDVNVYG-----SGTL 149  
ABF50895.1 -----FQNATTFPQLGGEDVNVYG-----GGTL 149  
EED50415.1 -----FQNATTFPQLGGEDVNVYG-----GGEL 147  
ABD61562.1 -----FQNATTFPQLGGEDVNMVG-----GGTI 148  
EEA22673.1 -----FQNSTTFPQLGGEDVNIYG-----GGTI 150  
EED15423.1 -----FQNSTTFPQLGGEDVNIYG-----GGTI 150  
AAF05088.1 -----FQNATSFPPKLGEDVFMVG-----GGTL 166  
BAE97166.1 -----FQNMSFWKIGGKDIHIYGDLD--TKGESLL 160  
EY19557.1 -----FQNQSVFWKLGEDVNIYGDLD--GNDQSVI 161  
AAP83023.1 -----  
ABD61563.1 -----FQTASLFWRFGGNNVNIYG----DGKGV 149  
ABF50897.1 -----FQNASSWVWGGEDIHLYG----AGTGV 158  
EED46018.1 -----FQKSIIFWRWGGDKIKIFG----TGLL 151  
EAL84906.1 -----FQKSIIFWRWGGEDIKIFG----SGVL 171  
AAO61898.1 -----FQKSIIFWRWGGHDVRIFG----RGTL 157  
ABD61564.1 -----FQKSIIFWRWGGQDKIKIFG----SGVL 153  
BAE97090.1 -----FQNSIMFWKGGKNIKLYG----KGV 151  
EY21174.1 -----FQNSIMFWKGGKDKIYG----EGVL 133  
EY20199.1 -----FQTNLMFWKGGKFKIYG----DGV 152  
EY16298.1 -----SSTGRTRGTPSSAPLCSVSIKLSRKAFLLSADNMKGKGGHDKIYG----EGVI 186  
EED49587.1 -----LQDQVTAWLFCGHNVSYFG----EGG-FY 127  
EED23059.1 -----YQNQSTAWLFGGINT----- 120  
EED47623.1 -----YQNQSTAFILGGNNVRIIDG----HGIGTL 123  
EAL86831.1 -----YQNQSTAWVIGGKNIRFNG----HGYGTL 210  
AAP83024.1 -----YQNQSTAWLWGGNNINVNG----YGYGTL 144  
EY20240.1 -----FQNQSTAWHFGGDNVIFDG----FGVGT 158  
EY99815.1 -----FQNQSTVWHFGGHNVLFDFG----FGIGTL 158  
EEA27533.1 -----FQNHAGFVLTGDHITIEG----HGTGQI 132  
EED19995.1 -----FQNHAGFILTGDHITIKG----HGTGQI 133  
EDF55950.1 -----FQNHAGFLLVDS--TSPL-----MDTVPE 129  
EED47501.1 -----TAMISVSDIDGLKLRSVTG----SGVI 131  
EAL85450.1 -----TAMLNIDMINGLRLRSLTG----SGVI 134  
EEA25131.1 -----KAILVSDITGATITSTG----AGVI 145

EED57938.1 KSYGQAWWDANPPNGT-----GISGRPHLMSYKTSQATIKYFRSGKPIAWNM 182  
EDP52870.1 NSYGAQAWWDANPANS-----SFPNRPHLMSYKTSQASIKNFRSRKPIAWNV 184  
ABD61565.1 DSYGQAWWDANPAGSS-----GIDNRPHLMSFKSSQATMKYFRSRKPIAWNV 182  
BAH29710.1 NCQGASVYWDGKGTNGG-----VTKPHPMVKIYVKGSGTFKD-VTVLNSPAQAVS 162  
BAH14957.1 NGQGQYWDGKGTNGG-----KTKPHPMVKIYHGTGSGFKD-VTVLNSPAQAVS 165  
BAA96351.1 DNGALYWDGKGTNGG-----THKPHFFLKI-KGSGTYKK-FEVLNSPAQAVS 147  
BAE80700.1 DNGALYWDGLGTNGG-----THKPHFFLKI-KGSGTYKK-FEVLNSPAQAVS 147  
ACL80126.1 DGQGPSYWDGQGGNGG-----VTKPHPMVKI-KGSGTYSN-VKVLNSPAHTYS 146  
ABH12113.1 DGQKAKYWDGLGSASG-----VTKPHPMVKI-KGSGTYSN-VKVLNSPAHTYS 147  
AAV84618.1 DNGQYVYWDGLGSNG-----GVKPNHFI AAKKLI GGS I ENLYIQNWPVHLEF 152  
AAM22187.1 DNGQYVYWDGLGSNG-----GVKPNHFI AAKKLI GGS I ENLYIQNWPVHLEF 152  
ADD73449.1 -----GLNIQNWPTHCFY 13  
CAM91235.2 DNGAAYWDGQGSNG-----GTDKPDHFFVVKDVVNG-VISNLNIQNWPTHCFD 254  
CAQ03435.1 DNGAAYWDGKGSNG-----GTDKPDHFFVVKDVVNG-VISNLNIQNWPTHCFD 254  
CAD62160.1 DNGNPAYWDGEGSNG-----GTDKPDHFFVVKDVI NG-VISNLNIQNWPTHCFD 174  
ABQ11248.1 DNGNQAYWDGQGSNNA-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11247.1 DNGNQAYWDGQGSNNA-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11241.1 DNGNQAYWDGQGSNNA-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABW96528.1 DNGNQSYWDGEGSNNK-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11245.1 DNGNQAYWDGEGSNNK-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11246.1 DNGNQAYWDSKGSNNK-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11236.1 DNGNQAYWDGQGSNNQ-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11240.1 DNGNQAYWDGEGSNNK-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
ABQ11234.1 DNGNQAYWDGQGSNNQ-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 156  
BAE97018.1 DNGNPAYWDGEGTNNK-----DNPKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 153  
AA74586.1 DNGNQAYWDGKGSNSN-----SNQKPDHFI VVKVTGTSKITNLNIQNWPHVCFD 155  
CAK40265.1 NNGEAYWDGEGSNG-----GQDKPDHFI VVKDMYNS-KIENLNIQNWPHVCFE 280  
ABF51154.1 NGDGARWWDGKGGNGG-----KTKPKFFSAHKLTDSTITG-ITIKNPPVQVVS 106  
BAB32924.1 NGDGARWWDGKGGNGG-----KTKPKFFSAHKLTDSTITG-ITIKNPPVQVVS 167  
AAF70169.1 NPDGARWWDGKGGNGG-----KDKPKFFAAHSLTDSITD-VNKNTPVQAVS 168  
EDU44300.1 DGQCALYWDGKGGDNG-----KVKPKFFAAHSLTGSISN-LLIKNPPVQVVS 106  
ABF50893.1 NPDGSRWWDGEGSNGG-----VTKPKFFYAHDLTDSITQ-LYIENTPVQAVS 166  
CAM07141.1 DPQGSRWWDGEGSNGG-----KTKPKFFAAHDLTPSSITN-LNIRNTPVQAVS 168  
CAE46647.1 NDLEKRYWDGKGSNGG-----KTKPKFFAAHMKMKSINN-ITIKNPPVQVVS 205  
AAV84616.1 NNGASVYWDGEGGSGG-----KTKPKFFQAHLTDSLIET-LTILNPPVQVVS 184  
CAQ03431.1 NNGASVYWDGKGGSGG-----KTKPKFFQAHLTDSLIET-LTILNPPVQVVS 184  
AAV84617.1 DGGGASVYWDGEGSNGG-----KTKPKFFYAHSLISSTIEN-LYFYNPPVQVMS 167  
CAM32882.1 -----  
EDP49013.1 DCDGARWWDGKGGNG-----GKKKPKFFSAHKLKNSNITG-LKVYNTPVHGFS 156  
EAW19942.1 DCDGARWWDGKGGNG-----GKKKPKFFSAHKLKNSNITG-LKVYNTPVHGFS 156  
CAK47667.1 NCDGARWWDGKGSNG-----GKTKPKFFQAHLKQSSITG-LKVYNTPVQVGS 152  
ABF50868.1 NCDGARWWDKGSNG-----GKTKPKFFSAHKLKNSKI QG-LKVYNTPVQVGS 156  
EED53217.1 DGDGSRWWDGKGGNG-----GKTKPKFFYAHKLDSSSITG-LQIYNTPVQVGS 155  
QCDGARWWDGKGSND-----GKIKPKFFYAHKMQNSNITG-LQVYNTPVQVGS 154  
BAB61793.1 DCAGSRWWDKGGNG-----GKKKPKFFAAHSLKNSKIEG-LKVYNTPVQVFS 157  
CAM91338.1 DGGGARWWDGLGSNVSA-----GKGVKPKFFSAHNLTGSSSITGLNPLNAPVQCS 163  
CAM91053.1 DGGGARWWDGLGSNVSS-----GKGVKPKFFSAHNLTGSSSITGLNPLNAPVQCS 163  
CAQ03438.1 DGQKAKYWDGKGTNG-----GKTKPKFFYAHSLKGSVSDINLVNTPVQVFS 157  
AAC14466.1 DNGAKAYWDGKGTNG-----GKTKPKFFYAHSLKGSVSDINLVNTPVQVFS 174  
AAM00249.1 DNGAARWWDGEGSNG-----GKTKPKMFFYAHSLKQSTISG-LNVKNTPVQVMS 160  
AAB36616.1 DNGAARWWDGEGSNG-----GKTKPKMFFYAHSLKQSTIHN-LNVKNTPVQVMS 160  
BAG72132.1 DNGAARWWDGKGSNG-----GKTKPKFFYAHKMDISTITG-LNKNHPVQVFS 159  
BAG72133.1 DNGAKAWWDGKGSNG-----GKTKPKFFYAHKMDISTITG-LNKNHPVQVFS 159  
ABL01533.1 DCQGSRWWDKGSNG-----GKTKPKFFYAHSLKSSNITG-LNVLNTPVQVFS 150  
CAA61552.1 DCQGSRWWDKGSNG-----GKTKPKFFYAHSLKSDITIRG-LHTLNTPVQVFS 153  
BAE97111.1 DCEGKRWWDGKGSNG-----GKKKPKFFSAHSLKNSNIKN-LNVLNTPVQVFS 152  
BAE97133.1 DCEGKRWWDGKGSNG-----GKKKPKFFSAHSLKNSNIKN-LNVLNTPVQVFS 152  
EEY20327.1 DADGARWWDGQGGNG-----GKTKPKLLAVKGLNNSIVKG-LNFKDQPVHGFS 159  
CAC14022.1 SCEGERWWDGKGGNG-----GKKKPKFFKVK-INDGSI TG-LKVNTPVHGFS 150  
CAA64727.1 SCEGERWWDGKGGNG-----GKKKPKFFKVK-INDGSI TG-LRVKNTPAHGFS 150  
AAA79885.1 DAAGQKAWWDGKGSNG-----GKTKPKFFYAHSLTSSISG-LNKNTPVQVFS 154  
CAA71247.2 DCEGQKAWWDGKGTNG-----GKTKPKFFYAHSLTSSISG-LNKNTPVQVFS 159  
CAK47020.1 DNGAKAWWDGEGTNG-----GKTKPKFFYAHDLTDSKISG-LYIKNTPVQVFS 159  
ADZ99366.1 DMEGQKAWWDGEGSNG-----GKTKPKGFFYAHDLTDSVIRG-LNKNTPVQVFS 274  
AAR84199.1 DGDGARWWDGKGDGSGK-----VKKPKFMSLKTGNSDVGG-LQIKNTPVQVFS 151  
EDN63471.1 DGQKAKWWDGLGDSGK-----VKKPKFVKLALTGTSKVTG-LNKNTPVQVFS 151  
ACA84699.1 NNGNQAYWDAQDHAP-----TVFRSLTNSKVSIG-INIINAPRAHLS 159  
EDP47682.1 DNGNQVWYDLYAE-----DALILRLPILFVIGLGGTIGPLKLYRSP-QWYQL 196  
EAU32536.1 DNGNQVWYDLYAE-----EPLTLRPLLGVIGLGGTIGPLKLYRSP-QWYQL 196  
ABF50895.1 DNGNQVWYDLYAE-----DPLILRLPILFVIGLGGTIGPLKLYRSP-QWYQL 196  
EED50415.1 NNGNQVWYDLYAE-----DALILRLPILVIGLGGTIGPLKLYRSP-QWYQL 194  
ABD61562.1 NNGNQVWYDLYAE-----DDLILRLPILMGIIGLGGTIGPLKLYRSP-QYHFE 195  
EEA22673.1 DNGNQVWYDMFAV-----NATLQRPILMGIIGLEGGSIGPLKLYRSP-SWYHI 197  
EED15423.1 DNGNQVWYDLFTQ-----NATLQRPILMGIIGLEGGSIGPLKLYRSP-SWYHI 197  
AAF05088.1 NNGNQVWYDYIAS-----NIYALRPVILIGI DGLKDSILSDLVLYRSP-QYHFE 213  
BAE97166.1 DGHGQAYWEEEMAV-----NKTLLRPLLTI EDAGHGLTMSNLRMRNPP-NWFFN 207  
EY19557.1 DGHGQAYWEEIRT-----NKSFLRPLMFAFGMHGAVMSNMRNPP-NWFFN 208  
AAP83023.1 -----SLGRPILFVVMGAVGASMSDLKMKSP-NWFFN 32  
ABD61563.1 DGAGQYVWWSAMAE-----DSSVMRPECLGT DGLHHAITISGLTMLNSP-NWFFN 196  
ABF50897.1 HNGNQVWYDAAAG-----NSSVRRPLLFTIDGWHGGSITGLKLRQSP-NWFFN 205  
EED46018.1 NNGNQRYWNEFAG-----QEILDPDNEYRPIFLTENATRISVEGITQLNSP-CWTNF 204  
EAL84906.1 NNGNQRYWNEFAG-----QEILDPNKNYRPIFLVTENATRISVEGITQLNSP-CWTNF 224  
AAO61898.1 NNGNQRYWNEFAG-----QEILDPDNTYRPIFLVAENATRISVEGITQLNSP-CWTNF 210  
ABD61564.1 NNGNQRYWNEFAG-----KQILVYVNFYRPIFLFDNATRISVEGITQLNSP-CWTNF 205  
BAE97090.1 NNGNQRYWNEFAG-----KEILDPTNAYLRPIFLFYAQNATNLDIQGVHFKDPS-CWTNF 204  
EY21174.1 NNGNQRYWNEFAG-----LEILDPTNAYLRPIFLFYAQNATNLDIQGVHFKDPS-CWTNF 186  
EY20199.1 DNGERWRWNEFAG-----LDLDPANNYRPIFLFYVGDANVQIEGIFKQDGP-IWHQL 206  
EY16298.1 EGQQRWVWNEFAGTGSILVSPANPNKYRPIFLFYAENTNLDVSGIHLKDS-CWNNF 245  
EED49587.1 NNGNQAWYNANRN-----ESNQAGRPMSTVYNSDTRLDVGLRVIQPP-FWATF 175  
EED23059.1 -----WYTENK-----QANQGRPISITFYNSNLLADGLTIFQPQ-FWATF 162  
EED47623.1 DNGNDTWWYQWIKQ-----PNTSNYPGRPHAITFNGLNTSIVRGLRFLRSQ-MCHHV 174  
EAL86831.1 NNGNQVWYDFVKS-----VSNYPRRPHALTVARASNSVFGQLRFLRSQ-MWTLFS 258  
AAP83024.1 DNGNQVWYDFVKG-----ISNYPRRPHALTIWNTSIVRGLRFLRSQ-MWYFF 192  
EY92040.1 DNGNQVWYDVAVS-----EGNLEHPRPILNFKRLSNSVIRGLRFLRSQ-MWMSA 206  
EY99815.1 DNGNQVWYDVAVS-----EGNLEHPRPILNFKRLSNSVIRGLRFLRSQ-MWMSA 206  
EEA27533.1 YGNGNAWYNVEQG-----VTQGRPMFFVFNWVSEVLVEDFTVKDPP-LWSLN 179  
EED19995.1 YGNGNAWYNVEQA-----VTQGRPMFFVFNWVSEVLVEDFTVKDPP-LWSLN 180  
EDP55950.1 ASMATAIRDIPQK-----RASHNLDRCSSFGTSMKLSMGTFFVKDPP-LWSLN 176  
EED47501.1 DNGNQVWYDFVKS-----DSSYKRPTLLYITGGSNIEVSGLRQKNPNVFNVS 179  
EAL85450.1 DNGNQVWYDFVKS-----DSSYKRPTLLYITGGSNIEVSGLRQKNPNVFNVS 182  
EEA25131.1 DNGNQDAYDAFAA-----DLSLRPILVYITGSSDITMSGFRVDPNPFVFSN 193

EED57938.1 KLHGEDIASV--HAIVDASSTGSGFP-----FNTDADFVQGTNRISDSIMYNGD 230  
EDP52870.1 KLGQDDIVTS--HAIVDATSTGGFP-----FNTDGFVDEGTNISITDSVMFNGD 232  
ABD61565.1 KLHGQDIVTS--HAIIDATSTGSGFP-----FNTDGFVDEGTNIQITDSIMYNGD 230  
BAH29710.1 VGN-TAAPTIVS--GITINNSAGSSKG-----HNTDCEFDVS-ASDVTITGNCTCENQ 209  
BAH14957.1 VGT-SAAFTIIS--GVTIDNSAGSSKG-----HNTDCEFDVS-ADDVITGNVCKNQ 212  
BAA96351.1 VGPTAHLITLD--GITVDDFAGDTKN-----LGHNTDGFVDS-ANNVITQNCIVKNO 197  
BAE80700.1 IGMTDGNIVFD--TVTVDSDAGDANS-----QGHNTDGFVDS-ADNVITKNCIVMNO 197  
ACL80126.1 ISN-PAKLVMS--KLTIKSDAGDAPNNSGKGAAGHNTDGFVDS-TTDLTIEDSTICNO 202  
ABH12113.1 VSN-PAALVMS--KLTIKSDAGDAPNNSGKGAAGHNTDGFVDS-TTDLTIEDSTIFNO 203  
AAV84618.1 ITGAVGLTIQ---NLVLNNTAGDAPNAASGTLAAAHNSDGFVSSSNTIKNQVFNQ 209  
AAM22187.1 ITGAVGLTVQ---NLVLNNTAGDAPNAASGTLAAAHNSDGFVSSSNTIIRNSSVYNO 209  
ADD73449.1 MTGNQHLTAQ---NILLNNTAGDAPNEASGTKAATHNSDGFVGSDDYVMSNIQVYNO 70  
CAM91235.2 ITGAQGLTVS---GLTLDNSAGDAPNSASGDDPAAHNSDGFDISGSDSVTLKDIVVKNO 311  
CAQ03435.1 ITGAQGLTVS---GLTLDNSAGDAPNSASGDKPAAHNSDGFDISGSDSVTLKDIVVKNO 311  
CAD62160.1 ITGAKGLTVS---GLTLDNSAGDAPNSASGSKAAAHNSDGFDISNSDSVTLKNIIVKNO 231  
ABQ11248.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11247.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11241.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABW96528.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11245.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11246.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11236.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11240.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
ABQ11234.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 213  
BAE97018.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 210  
AAA74586.1 ITGSSDLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 212  
CAK40265.1 IEMTEVLTIS---GLLIDNRAGDKPNKSKGGLPAAHNSDGFDISSDHVTLDNNHVYNO 337  
ABF51154.1 INCGDGLTIT---DMTIDASDGDKD-----EQGHNTDGFDISGSSNNVIDGAKVYNQ 154  
BAB32924.1 INCGDGLTIT---DMTIDASDGDKD-----EQGHNTDGFDISGSSNNVIDGAKVYNQ 217  
AAF70169.1 IASCDGLTIT---GMTIDDSGDDAN-----ELGHNTDGFIDGTSKNIIMIDGAKVYNQ 218  
EDU44300.1 INCGDGLTIT---DMTIDASAGDKG-----ALGHNTDGFIDGSSNNVIDGAKVYNQ 156  
ABF50893.1 INCGDGLTIT---DMTIDNSAGDDA-----G-HNTDGFIDGSSNNVIDGAKVYNQ 215  
CAM07141.1 VNGVGLTIT---GMTIDNSAGDSG-----G-HNTDGFIDGSSNNVIDGAKVYNQ 217  
CAE46647.1 INGCNDLTVN---RMHIDNKDGDGSK-----G-HNTDGFIDGSSNNVIDGAKVYNQ 254  
AAV84616.1 INGVSNLELA---YITVDASAGDSL-----G--KNTDAPDIGASDVTIIEHATVYNQ 232  
CAQ03431.1 INSVANLELA---YITVDASAGDSL-----G--KNTDAPDIGASDVTIIEHATVYNQ 232  
AAV84617.1 INSATDLTVS---GVTINAQDGDTD-----SLGANTDGFIDGSSNTVITGANVYNQ 217  
CAM32882.1 -----ANTDGFIDGSSNTVITGANVYNQ 25  
EDP49013.1 IQS-DHLTIK---DVLIDNSAGTKL-----GHNTDADFVGSSTYITIDGATVYNQ 203  
EAW19942.1 IQS-DHLTIK---DVLIDNSAGTKL-----GHNTDADFVGSSTYITIDGATVYNQ 203  
CAK47667.1 ILA-DHLTIT---DVTIDNSAGTSK-----GHNTDADFVGSSTYITIDGATVYNQ 199  
ABF50868.1 IQS-DHLTIS---DVTIDNSAGTSK-----GHNTDADFVGSSTYITIDGATVYNQ 203  
EED53217.1 IQS-DNLNIT---DVTIDNSAGTAE-----GHNTDADFVGSSTYITIDGATVYNQ 202  
BAB97386.1 ILS-DHLTIS---NILIDNRAGDKT-----NGGHNTDADFVGSSTYITIDHATVYNQ 203  
BAB61793.1 ISGSEQALAN---NILIDNKAGDSQ-----GGHNTDADFVGSSTYITIDGATVYNQ 206  
CAM91338.1 IGQSVGLS---LINITIDNSAGDAG-----SLGHNTDADFIDNLSQNIIFISGAIVKNO 213  
CAM91053.1 IGQSVGLS---LININIDNSAGDAG-----NLGHNTDADFIDNLSQNIIFISGAIVKNO 213  
CAQ03438.1 INGASGLN---LTNIHIDNSAGDAG-----ALGHNTDADFVGSSTYITISGANVYNQ 207  
AAQ14466.1 INSASGLT---LSGITIDNSAGNS-----LGHNTDADFVGSSTYITISGANVYNQ 222  
AAM00249.1 INSATDLN---IINVKMDNSAGAS-----KGHNTDADFVGSSENIYISGAVINNO 208  
AAB36616.1 INSATDLN---VIDVTMDNSAGAS-----KGHNTDADFVGSSENIYISGAVINNO 208  
BAG72132.1 VNGANNLV---IDSVTTDNSAGDTG-----KLGHNTDAYDVGSSNNVIVSNVAVKNO 209  
BAG72133.1 VNGADNLV---IDSVTTDNSAGDTG-----KLGHNTDAYDVGSSNNVIVSNVAVKNO 209  
ABL01533.1 INSATTLG---VYDVIIDNSAGDS-----AGGHNTDADFVGSSTGVYISGANVYNQ 199  
CAA61552.1 INGAANLG---VYDVSVDNSAGDS-----AGGHNTDADFVGSSTGVYISGADVKNQ 202  
BAE97111.1 INSVTNLG---VYGVHMDNSLGDG-----LGGHNTDADFVGSSTGVYISGAVVKNQ 201  
BAE97133.1 INSVTNLG---VYGVHMDNSLGDG-----LGGHNTDADFVGSSTGVYISGAVVKNQ 201  
EY20327.1 VNTVNNLQ---LIDVTLASAGDS-----KGGHNTDADFVGSSTGVYISGAVVKNQ 208  
CAC14022.1 INSVKGLK---VVNVFNNKDGDT-----KGGHNTDADFVGSSTGVYISGAVVKNQ 199  
CAA64727.1 INSVKGLKGRPSVNVLENNKDGDT-----KGGHNTDADFVGSSTGVYISGAVVKNQ 202  
AAA79885.1 INGVTLGT---LDRITIDNSAGDS-----AGHNTDADFVGSSTGVYISGANVYNQ 203  
CAA71247.2 ISHAENLG---IFNVNIDDSAGDT-----QGGHNTDGFIDGNSNGVITSGCSVRNO 208  
CAK47020.1 VE--SDNLVIE---DVTIDNSDGDSE-----GGHNTDGFIDSESTYITITGATVYNQ 207  
ADZ99366.1 VDMAQLQIT---DVTIDDSAGDVD-----NLGHNTDGFVQSGTGVYISGATVYNQ 324  
AAR84199.1 VNGCSDTVIH---DVTIDNSDGDKD-----SLGHNTDGFVGSNNVITENCHVYNQ 201  
EDN63471.1 INKCSDLTIS---DITIDIRDGD-S-----AGGHNTDGFVGSNNVLIQCGTVYNQ 200  
ACA84699.1 VTNSSQLVLD---NITLKTIVSTSKN-----LAKNTDALDVSSSQGITFQNSHLTVG 208  
EDP47682.1 VANSSDVIFD---GIDISGYSSS-----KNEAKNTDGDWYTRSDNIVIQNSVINNG 245  
EAU32536.1 VANSSTDVLFV---GIDISGYSSS-----KNTAKNTDGDWYTRSDNIVIQNSVINNG 245  
ABF50895.1 VANSSDVLFV---GIDISGYSSS-----ENEAKNTDGDWYTRSDNIVIQNSVINNG 245  
EED50415.1 VANSSDVLFV---GIDISGYSSS-----ENEAKNTDGDWYTRSDNIVIQNSVINNG 243  
ABD61562.1 VANSNVLFV---GIDISGYSSS-----DNEAKNTDGDWYTRSDNIVIQNSVINNG 244  
EED22673.1 VANSNVVFD---GLDISGYSTS-----VHVAKNTDGDWYTRSDNIVIQNSVINNG 246  
EED15423.1 VANSNVVFD---GIDISGYSSS-----GNVAKNTDGDWYTRSDNIVIQNSVINNG 246  
AAF05088.1 IANSSNVVYN---NINIAGGSSS-----KNPAKNTDGDWYTRSDNIVIQNSVINNG 262  
BAE97166.1 IINSTDVLIIS---DLDEAKSLN-----GVKIANSDGDWYTRSDRIVIQNSVIDNT 256  
EY19557.1 IANSSDILIS---NMHLNATSRD-----GVVIANSDGDWYTRSDRIVIQNSVIDNT 257  
AAP83023.1 ITGSQDVLIIS---GMDLKAVTSN-----SHPVKNSDGDWYTLSSHITQNSTILNT 81  
ABD61563.1 IANSTDILIS---NMTMLVESEIS-----DAPAKNTDGDWYTRSDNIVIQNSVIDNT 246  
ABF50897.1 IANSSDLIIS---DMDIFRSSSS-----EAWASNLGDWYTRSDNIVIQNSVINHD 254  
EED46018.1 FIQSKDVSFV---DVYIHAFSTNK-----SALPKNSDGFDSLNVGLTIVTNRVVDG 254  
EAL84906.1 FVRTKDISFD---NVFIHAYSTNA-----SALPKNTDGFDTLNVGLTIVTNRVVDG 274  
AAO61898.1 LVNSKDISFD---NVFINAYSTNA-----SAEPKNTDGFDSLNVGLTIVTNRVVDG 260  
ABD61564.1 FVRTNDVSFD---NVYIHAFSTNA-----SSDPANTDGMDSLNDVGVFTNMRVDG 255  
BAE97090.1 VVTSKDISFK---DVICTARSTNA-----TALPKNTDFFDSLNVENLVERAWVDG 254  
EY21174.1 VVTSKDISFK---DVACTAESNNA-----TALPKNTDFFDSLNVKVRVERAWVDG 236  
EY20199.1 IERTEGISYR---DVSCTARPRDQ-----TVRPANTDFFDSLNVKVRVERAWVDG 256  
EY16298.1 IVSSKNRYT---DVIATALSNNG-----SIIPKNTDFFMNTMNTSSVIERVWVVDG 295  
EED49587.1 VWASKNVFST---NLFVNATSDEWEG-----TMNTDGFDSWQSDNLLVENATIIMG 224  
EED23059.1 VWQSTVSLR---NLFVNATSDEWEG-----TVNTDGFDSWRSNLLVESATIING 211  
EED47623.1 EFDLILVNT---GNVVS-----FSDNTDGDATIRSSHITFNNWTVYNQ 215  
EAL86831.1 VIHSDHILFD---SIYVN-----STTNTDGNARSHTIFNNWTVYNQ 298  
AAP83024.1 QSCYLY----- 198  
EY92040.1 ISHSRNVLT---DIYVNSTNSRWS-----TLNTDGCDTISSDNITFRRWYVANG 255  
EY99815.1 ISHSRNVLELS---DIYVNSTNSRWS-----TLNTDGCDTISSDNITFRRWYVANG 255  
EED27533.1 IMNGTMMVFN---NITCNATAVNAPYG-----TNWVQNTDGFDTMDANNIELNGTYQGG 233  
EED19995.1 IMNGTMMVFN---NITCNATAVNAPYG-----TNWVQNTDGFDTMDANNIELNGTYQGG 232  
EDP55950.1 IINGTMMRFN---SIYCNATSVDPYQ-----HNWVRNTDGFDTMDANNIQLNGLTYQGG 229  
EED47501.1 KGDTSNAQFT---SLTMDATSNSD-----NLPKNTDADFIDGASTYVITSSVAITND 228  
EAL85450.1 KGDTHVTFK---NLRMDATSNSQ-----NPPKNTDGFIDGASTHVTISSVSVTND 231  
EED25131.1 TGSRRNIEYA---SLTMIASAKST-----NAPKNTDGFVGTSTFTIHNVYVSNQ 242



EED57938.1 A-----IAVGSDSHNVFERNTIGYQS-HGMSIGSLGKDA-SAFANITNLRFEVD 278  
 EDP52870.1 A-----IAVNTPSHNVFARNTIGYQS-HGMSIGSLGKDP-TDFANITNLRFEVD 280  
 ABD61565.1 A-----IAVGADSHDLFTRNTIGYQT-HGMSIGSLGKDP-TDFANISNIRFDDV 278  
 BAH29710.1 D-----CLAINSGSN-IVFSDNKCSCGG-HGISIGSLAAGK-----TVSNVVKVTGN 252  
 BAH14957.1 D-----CLAVNKGSN-IEFSDNKCSCGG-HGISIGSVKSDH-----NVESVTISGN 255  
 BAA96351.1 D-----CIAINDGNN-IRFENNQCSCGG-HGISIGSIATGK-----HVSNNVVIKGN 240  
 BAE80700.1 D-----CIAINDGKN-IQFLNNQCSCGG-HGISVSGSIATGK-----HVSSVTIKDN 240  
 ACL80126.1 D-----CIAINKGSN-IIFQRNSCTGG-HGISIGSISTGA-----TVQNVQILNN 245  
 ABH12113.1 D-----CIAINQGSN-IIFQRNCTCGG-HGISIGSVSTGA-----GCQQHPDLEQ 246  
 AAV84618.1 D-----CVAVTSGNN-ITIDGLYCSGG-HGLSIGSVGGKS---NNNVTNITFKNS 254  
 AAM22187.1 D-----CVAVTSGNN-ITIDGLYCSGG-HGLSIGSVGGKS---NNNVTNITFKNS 254  
 ADD73449.1 D-----CVAVTSGTQ-ITVSNLFCSCGG-HGLSIGSIGGKS---NNTVDCVLFEDS 115  
 CAM91235.2 D-----CVAVTSGSN-ILVTGLTCSGG-HGLSIGSVGGKS---NNTVSGVTFSDS 356  
 CAQ03435.1 D-----CVAVTSGSN-ILVTGLTCSGG-HGLSIGSVGGKS---NNTVSGVTFSDS 356  
 CAD62160.1 D-----CVAVTSGSN-ILVTGMACSOGG-HGLSIGSVGGKS---NNTVSGVTFSDS 276  
 ABQ11248.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABQ11247.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABQ11241.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABW96528.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---DNVVDGVQFLNS 258  
 ABQ11245.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABQ11246.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABQ11236.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABQ11240.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 ABQ11234.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 258  
 BAE97018.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---NNVVDGVQFLNS 255  
 AAA74586.1 D-----CVAVTSGTN-IVVSNMYCSGG-HGLSIGSVGGKS---DNVVDGVQFLNS 257  
 CAK40265.1 D-----CVAVTSGSS-IVVDNLYCYGG-HGLSIGSIGGKS---NNTVDCVTFSDS 382  
 ABF51154.1 D-----CVAVNSGTE-ITFKNGLCSGG-HGLSIGSVGGRR---DNTVDTVTFSDS 262  
 BAB32924.1 D-----CVAVNSGTE-ITFKNGLCSGG-HGLSIGSVGGRR---DNTVDTVTFSDS 262  
 AAF70169.1 D-----CVAVNSGTE-ITFKNGLCSGG-HGLSIGSVGGRR---DNTVDTVTFSDS 262  
 EDU44300.1 D-----CVAVNSGTE-ITFKNGLCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 201  
 ABF50893.1 D-----CVAVNSGTS-ITFSGGTCSCGG-HGLSIGSVGGRR---DNTVDTVTFSDS 260  
 CAM07141.1 D-----CVAVNSGTE-ITFSGGTCSCGG-HGLSIGSVGGRR---DNTVDTVTFSDS 262  
 CAE46647.1 D-----CVAVNSGTE-ITFSGGTCSCGG-HGLSIGSVGGRR---DNTVDTVTFSDS 299  
 AAV84616.1 D-----CVAVNSGSN-IIFKNKYCSGG-HGLSIGSVGGRR---NNTVDTVTFSDS 277  
 CAQ03431.1 D-----CVAVNSGSN-IIFKNKYCSGG-HGLSIGSVGGRR---NNTVDTVTFSDS 277  
 AAV84617.1 D-----CVAVNSGTE-ITFSGGTCSCGG-HGLSIGSVGGRR---DNVDTVTFSDS 262  
 CAM32882.1 D-----CVAVNSGTE-ITFSGGTCSCGG-HGLSIGSVGGRR---DNVDTVTFSDS 70  
 EDP49013.1 D-----CLAVNSGEH-ITFTNGYCNCGG-HGLSIGSVGGRR---NNVVDVTFSDS 248  
 EAW19942.1 D-----CLAVNSGEH-ITFTNGYCNCGG-HGLSIGSVGGRR---NNVVDVTFSDS 248  
 CAK47667.1 D-----CLAINSGEH-ITFTNGYCNCGG-HGLSIGSVGGRR---DNVDTVTFSDS 244  
 ABF50868.1 D-----CLAINSGEH-ITFTNGYCNCGG-HGLSIGSVGGRR---DNVDTVTFSDS 248  
 EED53217.1 D-----CLAINSGEH-ITFTNGYCNCGG-HGLSIGSVGGRR---DNVDTVTFSDS 247  
 BAB97386.1 D-----CLAINSGDH-ITFQNGYCSGG-HGLSIGSVGGRR---NNVVDVTFSDS 248  
 BAB61793.1 D-----CIAVNSGNN-IKFLNGYCSGG-HGLSIGSVGGRR---NNVVDVTFSDS 251  
 CAM91338.1 D-----CVAVNSGTE-ITFTGGNCSGG-HGLSIGSVGGRRSGTGANDVKDVRFLSS 261  
 CAM91053.1 D-----CVAVNSGTE-ITFTGGNCSGG-HGLSIGSVGGRRSGTGANDVKDVRFLSS 261  
 CAQ03438.1 D-----CLAINSGTG-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 252  
 AQC14466.1 D-----CLAINSGTG-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 267  
 AAM00249.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 252  
 AAB36616.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 253  
 BAG72132.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 254  
 BAG72133.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 254  
 ABL01533.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 244  
 CAA61552.1 D-----CLAVNSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 247  
 BAE97111.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 246  
 BAE97133.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 246  
 EY20327.1 D-----CLAINSGTN-ITFTGGNCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 253  
 CAC14022.1 D-----CLAINSGTR-ITFENGYCYGS-HGLSIGSVGGRR---DNVVDVTFSDS 244  
 CAA64727.1 D-----CLAINSGTR-ITFENGYCYGS-HGLSIGSVGGRR---DNVVDVTFSDS 247  
 AAA79885.1 D-----CVAVNSGSD-IHVTCNQCSCGG-HGLSIGSVGGRR---DNVVDVTFSDS 248  
 CAA71247.2 D-----CVAVNSGSD-IHVTCNQCSCGG-HGLSIGSVGGRR---DNVVDVTFSDS 253  
 CAK47020.1 D-----CVAVNSGEN-IYFSGGTCSCGG-HGLSIGSVGGRR---DNVVDVTFSDS 252  
 ADZ99366.1 D-----CLAINSGTD-ISFTGGYCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 369  
 AAR84199.1 D-----CIAVNSGTG-VYFKNNYCSGG-HGLSIGSVGGRR---DNVVDVTFSDS 246  
 EDN63471.1 D-----CIAVNSGSD-IKFMNNYCYNG-HGLSIGSVGGRR---DNVVDVTFSDS 245  
 ACA86699.1 D-----CLAVNAGVTNITLNNVNCVGG-HGLSIGSVGGRR---ATENVNLRFLNS 255  
 EDP47682.1 DSRNKLTLEYADCVSFKPNSTNIIVQNLHCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 303  
 EAU32536.1 D-----CVSFKPNSTNIIVQNLHCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 293  
 ABF50895.1 D-----CVSFKPNSTNIIVQNLHCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 293  
 EED50415.1 D-----CVSFKPNSTNIIVQNLHCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 291  
 ABD61562.1 D-----CVSFKPNSTNIIVQNLHCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 292  
 EEA22673.1 D-----CVSFKPNSTNIIVQNLHCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 294  
 EED15423.1 D-----CVAFKPNSTNIIVQNLHCNCS-HGISVSLGQYA-GETDIVQNVLVYNI 294  
 AAF05088.1 D-----CVSFKPNSTNIIVQNLFCNCS-HGISVSLGQYK-GEVDIVQNVLVYNI 310  
 BAE97166.1 D-----CVSFKPNSTNIIVQNLFCNCS-HGISVSLGQYK-GETDIVQNVLVYNI 304  
 EY19557.1 D-----CVSFKPNSTNIIVQNLVCHGS-HGISVSLGQYA-GEVDIVQNVLVYNI 305  
 AAP83023.1 D-----CVSFKPNSSIIIVQNLACTGS-HGISVSLGQYK-GVNDIVQNVLVYNI 129  
 ABD61563.1 D-----CVSFKPNSTQIIVQNLDCGTS-HGISVSLGQYK-GETDIVQNVLVYNI 294  
 ABF50897.1 D-----CVSFKPNSTNIIVQNLHCNCS-HGISVSLGNYP-YQYDIVQNVLVYNI 302  
 EED46018.1 D-----CFSPKPNTNIIIVQNLCCNNT-HGSMGSIQYK-GVMDIIEHAYIENV 302  
 EAL84906.1 D-----CLSPKPNTNIIIVQNLWCNCT-HGSMGSIQYK-GVLDIIEHAYIENV 322  
 AAO61898.1 D-----CFSPKPNTNIIIVQNLWCNCT-HGSMGSIQYK-GVLDIIEHAYIENV 308  
 ABD61564.1 D-----CFSPKPNTNIIIVQNMWCNNT-HGSMGSIQYK-GEMDIENVIENV 303  
 BAE97090.1 D-----CFSPKSNATNIIIVQNVCHGS-HGQSIQYK-GEMSVKDVVIENV 302  
 EY21174.1 D-----CFSPKSNATNIIIVQNVCHGS-HGQSIQYK-GEMSVKDVVIENV 284  
 EY20199.1 D-----CFSPKSNATNIIIVQNVCHGS-HGQSIQYK-GEMSVKDVVIENV 304  
 EY16298.1 D-----CFSPKSNSSDLYVNTMYCNGT-HGQYSARIKIWI-AGADIGTYGVNVNT 343  
 EED49587.1 D-----CIAAKGNTNIIIVQNVCHGS-AGMTIIGSIQYK-EMPDYENVVTFDGV 272  
 EED23059.1 D-----CVAAGKNTNIIIVQNVCHGS-AGMTIIGSIQYK-EMPDYENVVTFDGV 259  
 EED47623.1 D-----SISLKGNTDIIITNSNLYNG-LGTAIGSIQYK-DQFETIERLKVSNV 263  
 EAL86831.1 D-----SISLKGNTDIIITNSNLYNG-LGTAIGSIQYK-DQFETIERLKVSNV 346  
 AAP83024.1 D-----GIALKGNSSFIYIYDSVYFNG-OGTAIGSMGQFE-GKYEYINNLYARNL 303  
 EY92040.1 D-----GIALKGNSSFIYIYDSVYFNG-OGTAIGSMGQFE-GKYEYINNLYARNL 303  
 EY99815.1 D-----AIALKGNSSFIYIYDSVYFNG-OGTAIGSMGQFE-GKYEYINNLYARNL 303  
 EEA27533.1 D-----CVAIKPRSYNVRVSNVACHGG-NGIAIGSLGQYLEDASVKNVVDNVEI 281  
 EED19995.1 D-----CIAIKPRSYNVRVSNVACHGG-NGIAIGSLGQYLEDASVKNVVDNVEI 282  
 EDP55950.1 D-----CIAIKPRSYNVRVSNVACHGG-NGIAIGSLGQYLEDASVKNVVDNVEI 278  
 EED47501.1 D-----CVAFKPGANYTVENVCSTGS-HGISVSLGKSS---DDTVQNVYARNI 274  
 EAL85450.1 D-----CVAFKPGANYTVENVCSTGS-HGISVSLGKSS---DDTVQNVYARNI 277  
 EEA25131.1 D-----CITFKPGANYTVENVCSTGS-HGISVSLGKSS---DDTVQNVYARNI 289

EED57938.1 TVIDA-LYAAAFKSWTGGQ-----GLVKNVTKNIRVYVNFPIFVTQSY 322  
EDP52870.1 TVIDA-LYAAAFKSWSGGR-----GLVKNVWKNIRTFNVTFFPIFVTQSY 324  
ABD61565.1 TVVDG-LYAAAFKSWSGGT-----GLVKNVWNNIRVFNVTFFPIFVTQSY 322  
BAH29710.1 -TVSNSANGLRIKTIYVATD-----ASVSDITYESNTVSGISD---YGVVI 294  
BAH14957.1 -TVTDSANGLRIKTIYVATD-----AAVTNIVYKNDTVSGISK---YGVVI 297  
BAA96351.1 -TVTRSMYGVRIKRAQRTATS-----ASVSGVTYDANTVSGIAK---YGVLI 282  
BAE80700.1 -TVTNSMYGFVIVVQASATS-----ASVSGVTYDANTVSGIAK---YGVLI 282  
ACL80126.1 -QIINNDQALRIKTKADATS-----ASVPGITFSGNTATGTRK---YGVIV 287  
ABH12113.1 TRWSTTKALRIKTKSDATN-----ASVTNVVFNNGTATGTRK---YGVIV 289  
AAV84618.1 ELVNS-SNGARIKSNSETT-----GFISNITYSNIKLNTNIDT---YGIDV 295  
AAM22187.1 QIVNS-SNGARIKSNSETT-----GFISNITYSNIKLNTNIDT---YGIDV 295  
ADD73449.1 EQVNS-SNGVRIKSN-----GFISNITYSNIKLNTNIDT---YGIDV 129  
CAM91235.2 TITNS-QNGCRIKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 397  
CAQ03435.1 TITNS-QNGCRIKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 397  
CAD62160.1 TITNS-QNGCRIKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 317  
ABQ11248.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11247.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11241.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABW96528.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11245.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11246.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11236.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11240.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
ABQ11234.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 299  
BAE97018.1 QIVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 296  
AAA74586.1 QVNS-QNGCRIKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 298  
CAK40265.1 QVNS-ENGCRKSNSETT-----GTIENVVTSNIKMSNISD---YGIDV 423  
ABF51154.1 -----  
BAB32924.1 EVTKS-VNGVRIKAVVGT-----GKINKVYEDITLSEISK---YGVLI 303  
AAF70169.1 VVTKS-TNGIRVKAQVGT-----GKINKVYEDITLSEISK---YGVLI 304  
EDU44300.1 QVTKS-TNGIRVKAQVGT-----GKINKVYEDITLSEISK---YGVLI 242  
ABF50893.1 TVSNS-VNGIRIKASAGDT-----GKINKVYEDITLSEISK---YGVLI 301  
CAM07141.1 QVTKS-TNGIRVKAQVGT-----GKINKVYEDITLSEISK---YGVLI 303  
CAE46647.1 TVQNS-DNGIRIKAKKDT-----GNISNIVYEDITLSEISK---YGVLI 340  
AAV84616.1 TVTKS-VQIRIKTIAGDT-----GTVNNVYEDITLSEISK---YGVLI 318  
CAQ03431.1 TVTKS-VQIRIKTIAGDS-----GTVNNVYEDITLSEISK---YGVLI 318  
AAV84617.1 EIKAS-QNGIRIKTIAGDT-----GTVSGITYSGITLSEISK---YGVIV 303  
CAM32882.1 EIKAS-QNGIRIKTIAGDT-----GTVSGITYSGITLSEISK---YGVIV 111  
EDP49013.1 QVNS-QNGARIKTVYGAT-----GSVTGVKFDVSLKGIK---YGVIV 289  
EAW19942.1 QVNS-QNGARIKTVYGAT-----GSVTGVKFDVSLKGIK---YGVIV 289  
CAK47667.1 KIVNS-QNGVRIKTVYKAT-----GTVNVVYEDITLSEISK---YGVIV 285  
ABF50868.1 KVVDS-QNGVRIKTVYKAT-----GTVNVVYEDITLSEISK---YGVIV 289  
EED53217.1 KVVNS-QNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 288  
BAB97386.1 QVVNS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 289  
BAB61793.1 QVVNS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 292  
CAM91338.1 TVQKS-TNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 302  
CAM91053.1 TVQKS-TNGVRIKTVYDTR-----GTVNVVYEDITLSEISK---YGVIV 302  
CAQ03438.1 TVKNS-ANGVRIKTVYDTR-----GTVNVVYEDITLSEISK---YGVIV 293  
AAQ14466.1 TVKNS-ANGVRIKTVYDTR-----GTVNVVYEDITLSEISK---YGVIV 308  
AAM00249.1 KIITTSQNGVRIKTVYDAA-----GTVNVVYEDITLSEISK---YGVIV 294  
AAB36616.1 KIINS-QNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 294  
BAG72132.1 TIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 295  
BAG72133.1 TIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 295  
ABL01533.1 KIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 285  
CAA61552.1 KIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 288  
BAE97111.1 SIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 287  
BAE97133.1 SIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 287  
EY20327.1 KIINS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 294  
CAC14022.1 TIEKA-DNGIRIKTIAGDT-----GTVNVVYEDITLSEISK---YGVIV 285  
CAA64727.1 TIEKA-DNGIRIKTIAGDT-----GTVNVVYEDITLSEISK---YGVIV 288  
AAA79885.1 TIANS-DNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 289  
CAA71247.2 KVTKS-MYGVRIKTIAGAE-----GTVNVVYEDITLSEISK---YGVIV 294  
CAK47020.1 TVSDS-ENGVRKTIYVATD-----GTVNVVYEDITLSEISK---YGVIV 293  
ADZ99366.1 TVTAS-ENGVRKTIYVATD-----GTVNVVYEDITLSEISK---YGVIV 410  
AAR84199.1 QIVNS-DNGLRIKTIQKAT-----GTVNVVYEDITLSEISK---YGVIV 287  
EDN63471.1 HVINS-DNGLRIKTIQKAT-----GTVNVVYEDITLSEISK---YGVIV 286  
ACA48699.1 SCTNC-QNGVRIKTVYDAT-----GTVNVVYEDITLSEISK---YGVIV 299  
EDP47682.1 SMYNA-SDGARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 360  
EAU32536.1 SMYNA-SDMARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 350  
ABF50895.1 SMYNA-SDMARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 350  
EED50415.1 SMYNA-SDMARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 348  
ABD61562.1 SMFNA-SDMARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 349  
EAA22673.1 SMFNA-SDGARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 351  
EED15423.1 SMFNA-SDGARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 351  
AAF05088.1 SLHNA-TDGARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 367  
BAE97166.1 SMSNA-SDGARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 361  
EY19557.1 SMANA-SDGARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 362  
AAP83023.1 TLSNA-SDAARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 188  
ABD61563.1 SMTDA-SDARIKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 350  
ABF50897.1 TMANT-TAARLKVVWPGVSSAMSEDLQ--GGGLGSRVKNVYDQYIENVDWAEIVTQCY 361  
EED46018.1 TLLNG-QNGARLKAWAGQDVG-----YGRINNVYKNIHVENTDNPVLDQCY 349  
EAL84906.1 TLLNG-ENGARLKAWAGQDVG-----YGRINNVYKNIHVENTDNPVLDQCY 369  
AAO61898.1 TLLNG-ENGARLKAWAGQDVG-----YGRINNVYKNIHVENTDNPVLDQCY 355  
ABD61564.1 TLLNG-QNGARLKAWAGQDVG-----YGRINNVYKNIHVENTDNPVLDQCY 350  
BAE97090.1 WMLNG-QHARLKAWAGQDVG-----YGFINDVFRNFWGNNYTAIFDQCY 349  
EY21174.1 WMLNG-QHARLKAWAGQDVG-----YGFINDVFRNFWGNNYTAIFDQCY 331  
EY20199.1 WMLNG-QHARLKAWAGQDVG-----YGFINDVFRNFWGNNYTAIFDQCY 329  
EY16298.1 FKYD-----QALIPPTL-----DKSVSHKRRRNFVWARMYDGIILDSY 383  
EED49587.1 TIKDA-MGAYIKTWQCTRIPTPSNGDW--GGGLGLVKNVTFRDFVMDNVG-LPIQVAC 329  
EED23059.1 RCINS-STGAYIKTWQCTRIPTPSNGDW--GGGLGLVKNVTFRDFVMDNVG-LPIQVAC 316  
EED47623.1 NFHNT-LHAIYVKTWTADQNGYPPNG--GGGLGLYISDVTAKDLKATSLRGAFAISQ 319  
EAL86831.1 TYGT-LHSLYFKTWTGEQVNYPPNG--GGGLGLYISDVTAKDLKATSLRGAFAISQ 402  
AAP83024.1 -----  
EY92040.1 TFINN-AHVYLYKTVWAGVSRGFPNG--GGGLGLATNITMEDIKIRLRQPPFFSWQ 359  
EY99815.1 TFINN-AHVYLYKTVWAGVSRGFPNG--GGGLGLATNITMEDIKIRLRQPPFFSWQ 359  
EAA27533.1 IGVNK-DMHNAAYIKTYMGRVLPQSSYE-SAGLPNGGGWGTVENILFSNFVHQASAGPA 330  
EED19995.1 IRYNE-DMHNSAYIKTYMGRVLPQSSYE-SAGLPNGGGWGTVENILFSNFVHQASAGPA 349  
EDP55950.1 APIQT-DYCLHAQLRVYKN---MGRLS-GAPKLRKRWSPQWRGLQCYTSGGTSV 330  
EED47501.1 TMINS-SKAAGIKTYPSGGHGH-----LSTVKNATFEDFVGDGDFYAFQIQSCY 322  
EAL85450.1 TMIES-TKAAGIKTYPSGGHGH-----LSTVKNATFEDFVGDGDFYAFQIQSCY 325  
EAA25131.1 SASGC-SKAAGIKTYPSGGHGH-----LSTVKNATFEDFVGDGDFYAFQIQSCY 337

EED57938.1 YDQSVSRDG-VDESSVMMEDFTWEDFSGSINSYQPGDGSVCATDPCWYNAGLPLNKHTEA 381  
EDP52870.1 SDQASASRPGTIDPFSSVMMEDFTWSDFGSINTYHPGDGSCVTDPCWYNVGLPLNKHTEA 384  
ABD61565.1 SDQASRSRGTNASSAVMMEDFTWSDFAGSINTYQPGDGSVSDPCWYNVGLPLNKHTEA 382  
BAH29710.1 EQDYENGSPGTGTPSNGVTLGPVTFSG-----TNTVEVNSGAEQ 332  
BAH14957.1 EQDYENGSPGTGTPSNGVTLGPVTFEG-----TNKVSNSGAEK 335  
BAA96351.1 SQSYPP--DDVGNPGTGAPFSDVNFTGG-----ATTIKVNNAAAT 319  
BAE80700.1 SQSYPP--ADEGKAGTGGPISGINFTGG-----ATNIKVNVTGSKR 319  
ACL80126.1 DQGYPP--TTLGAPNGVKISGINFTGS-----TNNIAVTSSAQ 324  
ABH12113.1 DQGYPP--TTLGTPGNVAMVSGISFG-----TNNIAVTSDAQ 324  
AAV84618.1 QQDYLNNGGPTGPTNGVVIENILFE-----NVVGTAAASARN 332  
AAM22187.1 QQDYLNNGGPTGNPTNGVLIQNILFQ-----NVTGTAASARN 332  
ADD73449.1 -----  
CAM91235.2 QQDYLNNGGPTGPTNGVVKISGITFS-----SVTGTDDAYN 434  
CAQ03435.1 QQDYLNNGGPTGKPTNGVVKISGITFS-----NVTGTATDDAYN 434  
CAD62160.1 QQDYLNNGGPTGPTNGVTISNIAFS-----GVTGTTSSNAYN 354  
ABQ11248.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABQ11247.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABQ11241.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABW96528.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABQ11245.1 QQDYLNNGGPTGKPTNGVVKISNINFT-----KVTGTVASSAQ 336  
ABQ11246.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABQ11236.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABQ11240.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
ABQ11234.1 QQDYLNNGGPTGKPTNGVVKISNIKFT-----KVTGTVASSAQ 336  
BAE97018.1 QQDYLNNGGPTGKPTNGVVKISGKIFI-----KVTGTVASSAQ 333  
AAA74586.1 QQDYLNNGGPTGKPTNGVVKISGKIFI-----KVTGTVASSAQ 335  
CAK40265.1 QQDYENGGATGDPNTNGVKIENISFV-----NVKGTMSD-GKD 459  
ABF51154.1 -----  
BAB32924.1 EQNYDGGDLHGADTGVPI TALTDNV-----TGG---VSSSGY 339  
AAF70169.1 EQNYDGGDLHGPTDGTGPIITGLVLDV-----KGGGAVANGFN 343  
EDU44300.1 EQNYDGGDLHGHTAGTGPITGLTLQNI-----SCTGAVASSGFD 281  
ABF50893.1 EQNYDGGDLGDEVTSGIPITDLTIENI-----SGSGAVSDGYN 340  
CAM07141.1 EQNYDGGDLHGSPSTSGIPITNLVQNI-----SGSGVLSANN 342  
CAE46647.1 EQNYDGGDLHGEPNKLPI SHVTVNGL-----HGAGAIASGGHN 379  
AAV84616.1 EQNYDGGDLHGDPPTTGVPI TDLTIKNI-----SCTGAVSSSGY 357  
CAQ03431.1 EQNYDGGDLHGDPSTGVPI TGLTIKNI-----SGTGAVSSSGY 357  
AAV84617.1 NQAYDG--TSGEPTNGVAISKFILENI-----TGTVESTATNL 140  
CAM32882.1 NQAYDG--TSKPTNGVAISNFILENI-----TGTVEST---- 343  
EDP49013.1 QQDYENKPTGKPTNGVKVSDITFE-----KVTGVTSSATD 326  
EAW19942.1 QQDYENKPTGNPTNGVKVSDITFE-----KVTGVTSSATD 326  
CAK47667.1 EQDYENGSPGTGPTNGVVKVEDITFK-----KVTGSKSSGTD 322  
ABF50868.1 EQDYENGSPGTGPTNGVVEVEDITFK-----KITGSDSSATR 326  
EED53217.1 EQDYENGSPGTGPTNGIKVSDITFD-----KVTGTVESDAD 325  
BAB97386.1 QQDYKNGGPDGKPTNGVVKISGITFS-----NIHGSIKSSGK 326  
BAB61793.1 QQDYENGSPGTGNPTNGVSVTNI TFK-----NVKGNVLAAGIN 329  
CAM91338.1 QQDYQNGSPTGTPTAGVPI TGLTMN-----NVHGNVVG-GQN 338  
CAM91053.1 QQDYQNGGPTGPTTGVPI TGLTVN-----NVHGTVVG-GQN 338  
CAQ03438.1 QQDYKNGSPTGTPTGGVPI INDVTFSS-----NVKGTVASSGATN 330  
AAQ14466.1 EQDYENGSPGTGKPTSGVPI TGVTLSS-----NVHGTVSSSATN 345  
AAM00249.1 EQDYENGSPGTGPTTGVPI TGLTVS-----KVTGVSASSATD 331  
AAB36616.1 EQDYENGSPGTGPTTGVPI TGLTVS-----KVTGVSASSATD 331  
BAG72132.1 EQDYENGSPGTGPTTGVPI TDLTVS-----GVKGTVASSGNT 332  
BAG72133.1 EQDYENGSPGTGKPTTGVPI TGLTVS-----GVKGTVASSGNT 332  
ABL01533.1 EQDYENGSPGTGPTTGVPI TGLTVS-----KITGVSASSGNT 322  
CAA61552.1 EQDYENGSPGTGKPTSGVPI SGLTVS-----KISGVSASSAT 325  
BAE97111.1 EQDYENGSPGTGPTAGVPI TDVTIN-----KVTGTVKSSGTD 324  
BAE97133.1 EQDYENGSPGTGPTAGVPI TDVTIN-----KVTGTVKSSGTD 324  
EY20327.1 EQDYENGSPGTGPTAGVPI TDLTIN-----KVVGTVASKGNT 331  
CAC14022.1 QQDYENGSPGTGKPTAGVPI TGVTA-----NLKGTVASKGNT 322  
CAA64727.1 QQDYENGSPGTGKPTAGVPI TGVTIK-----NVKGSVASKGNT 325  
AAA79885.1 EQDYLNNGGPTGKPTTGVPI TGVTLK-----NVAGSVTGSGET 326  
CAA71247.2 QQDYKNGKPTGNPTTGVPI TNLTVN-----KVTGNVPGGAN 331  
CAK47020.1 EQDYENGSPGTGTPSNGVTSIDVTL-----DITGSDSDAVE 330  
ADZ99366.1 EQDYENGSPGTGTPSNGVTVELTLE-----DITGSDSSGANP 447  
AAR84199.1 ETDYNSNGTGTGTPGSKVPI TNFEVD-----GLTGSVDSAYR 324  
EDN63471.1 EGDYLNKGTGTTATGCVPI SMLVMK-----DITGSDNSTAKR 323  
ACA48699.1 CDQNESYCNNGNDASSLSITGV TISG-----ITGSDSSAGNPI 337  
EDP47682.1 GQKNLTLCNEYPSNLTISDIHFKNFRG-----TSSGKRDPNVGT 399  
EAU32536.1 GQKNQTLNQNYSNLTISDVHKNMVG-----TSSGKRDPNVGT 389  
ABF50895.1 GQKNMTLCNEYPSNLTISDVYISNMYG-----TSSARDPNIGT 389  
EED50415.1 GQKNQTLNENPNLSNLTISDVYFNDLTG-----VTSGKNDPNVGT 387  
ABD61562.1 GQKNMTLCNEYPSNLTISDVHKNFRG-----TSSGSDPYVGT 388  
EAA22673.1 GQSNQTLNDEYPSKLTITDIHFKNFSG-----VTSTKKNKIAT 390  
EED15423.1 GQSNQALNDEYPSNLTISDIHFKNFSG-----VTSTKKNKIAT 390  
AAF05088.1 GQKNLTLCLEFPSPLTITDI IFRNFTG-----KTSTKYQPQIGT 406  
BAE97166.1 GQKNQTLNCFEPANLTIEDVIMKNFWG-----TVSTKYDPRAGS 400  
EY19557.1 GQKNQTLNCFEPANLTIEDVIMKNMWG-----TSSSLDPQAGT 401  
AAP83023.1 MQTT--ANCNAYPTKMI IQDVVFKNFVG-----VASSKHDPKVG 226  
ABD61563.1 ESKNQTMCDYSYPSKLVIEDVLFKDFKG-----TSSKVDPEIGE 389  
ABF50897.1 GAINASECLDHPGSLVITNLVLFKNMNG-----TSNGANDPVAG 400  
EED46018.1 FDEAAECAQYPSQVNVNII LFENISG-----TSSGKNKGVVAD 388  
EAL84906.1 FNINATQCAAYPSRVNFTNIVFEDIY-----TSSGKRGKVAD 408  
AAO61898.1 FNIPADECASYPQVNVNII VFENVY-----TSSGAEGNVVAE 394  
ABD61564.1 FDIATTECAKYPSAVNII NII LFENIW-----SSSGKDKIVAD 389  
BAE97090.1 FNINATTCAYPSRMI NII NII LFENFTG-----YSSGKYGNAVAK 388  
EY21174.1 FQINSTCAQYPSKVDISNVLFEFTG-----TSSGKYGRAVAR 370  
EY20199.1 -----N 330  
EY16298.1 FNKSVEENKYP-----395  
EED49587.1 VYSAGSNK--SCNTSTLQIEDVKNWNI-----RGTSS--RPNIAS 364  
EED23059.1 MYSESSG--ACNTSKLQIEDVSWWNI-----TGTS--RYNVVA 350  
EED47623.1 TRFSGAPGEGCNTSEFQVRDI SVLNL-----SGTT--ESNRVA 356  
EAL86831.1 TTFPPAP--SNCTSSKFQIRDLRIEDV-----TGSS--TTDVVA 437  
AAP83024.1 -----  
EY92040.1 ENYSGWAG--RDCNSSKFKMSHVKNRVN-----TGTVKNKVDKVG 397  
EY99815.1 ENYSGWAG--RDCNSSKFKMSRVTKWNV-----TGTVKNKVDKVG 397  
EAA27533.1 ITQDTGNDGTFKGTSLIAISNVAFVNF-----TGYLNCVS--RA 376  
EED19995.1 ITQDTGNDGTFKGTSLMAISNVAFVNF-----TGYLNNVANNRI 379  
EDF55950.1 -----  
EED47501.1 GE--DDTYCEENPGDAVLEGI VVKGFTG-----TTSKEDPVPVAN 360  
EAL85450.1 GE--SESYCESNPGNAI LQGI VVKGFSG-----TSSGKYDPVAVAN 363  
EAA25131.1 NS--NSSYCAEYPTASLTDVYFKDFSG-----TSSGKHGVSVAN 375

EED57938.1 LVIECNTDKSKCNFVTKNIQLYQVLEPA-----SVICMKATAELNPNLGFN 428  
EDP52870.1 IVLECHTESSCKNFRTEGIRLHPQSKDSP-----SVICMKATAELNPKLGEF 431  
ABD61565.1 LIIECHTAQSCKNFVTDNIQLYQVLEPA-----SVICMNATAALNPLDGLFT 429  
BAH29710.1 VYVLCGSK-CTGTWNNWGLTTSGGKAGSS-----NYKISGYTV----- 370  
BAH14957.1 VYVLCGSK-CTGTWNNWGLTTSGGKAGSS-----NYDHISNFKL----- 373  
BAA96351.1 VTVECGN--CSGNWNNWQLTVTGGKAGTI-----KSKDAKITGGQYLADQPA 364  
BAE80700.1 VSVNCAA--CSGTWNNWQLTVTGGKAGTI-----NSDQAQISGGKYLDLDDSS 364  
ACL80126.1 VAVNCGTG-CTGTWNNWGLTVTGGKASDS-----KYRYSVGVKGSF----- 364  
ABH12113.1 VAVNCGSK-CTGTWNNWGLKVTGGNAG-K-----VYNYKNIKAGTY----- 363  
AAV84618.1 YVYVLCGEGS-CSNIKFSGVKITGGQKESS-----CNPATGCPA----- 370  
AAM22187.1 YVYVLCGSGS-CNNITFSGVKITGGAVAST-----CNPSTGCPK----- 370  
ADD73449.1 -----  
CAM91235.2 YVYVLCGSGS-CSDITFTDVSISGGGKTST-----CNPSTGCPG----- 472  
CAQ03435.1 YVYVLCGSGS-CSDITFTDVSISGGGKTST-----CNPSTGCPG----- 472  
CAD62160.1 YVYVLCGSGS-CSNFKFTDVSISGGGKTSS-----CNFPSSGCPA----- 392  
ABQ11248.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABQ11247.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABQ11241.1 WYILCGVGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABW96528.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPSTNTCPS----- 374  
ABQ11245.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABQ11246.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABQ11236.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABQ11240.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
ABQ11234.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 374  
BAE97018.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 371  
AAA74586.1 WYILCGDGS-CSGFTFSGNAITGGGKTSS-----CNPPTNTCPS----- 373  
CAK40265.1 YVYVLCGSGS-CSNFVFTDVIDTGGSDDS-----CNPSSGCP----- 495  
ABF51154.1 -----  
BAB32924.1 DVVVTGCGKSGCTGWTWTVGDVDTGG-KTYD-----KCSNVPSVTKCS----- 379  
AAF70169.1 VVVTGSGSCTGKWNKVVITGGSKVYG-----NCTNVPSVTKCS----- 384  
EDU44300.1 VVVTGSGSCTGWTWNSVSVVDTGG-KTYG-----SCTNVPSVTKCS----- 321  
ABF50893.1 IVIVCG-DDACSNWTSNVDVEVDTGG-EDYG-----SCENVPVAVASCST----- 380  
CAM07141.1 IAVVCG-SGACSSWTSNVDVDTGG-KKYG-----SCQNVPSVATC----- 380  
CAE46647.1 MAIVCG-SGCSANWTKVDKVVSGG-KKYD-----SCKNIPGVVSGSC----- 418  
AAV84616.1 VVIVCG-STSCSWSWTSNVAVDTGG-KTYS-----SCQNVPSVAVGSCS----- 397  
CAQ03431.1 VVIVCG-STSCSWSWTSNVAVDTGG-----SCLNVPTSGGVVSCSL----- 380  
AAV84617.1 IECGSG---SCTDWTWTVDVSVTGG-KTSS-----SCLNVPTSGGVVSCSL----- 380  
CAM32882.1 -----  
EDP49013.1 IYILCGSGS-CTNWTWNSGNSVTGG-KKSS-----SCKNVPAAGASC----- 364  
EAW19942.1 IYILCGSGS-CTNWTWNSGNSVTGG-KKSS-----KCKNVPAAGASC----- 364  
CAK47667.1 IYILCGSGS-CKDWTWNSGVDITGG-KKSS-----KCKNVPSGASCSD----- 362  
ABF50868.1 VNILCGSGS-CKDWTWNSGVDITGG-KKSS-----KCKNVPSGASC----- 364  
EED53217.1 IYILCGSGS-CTDWTWNSGNSVTGG-KTSS-----KCNVPTGASC----- 363  
BAB97386.1 AYLLCGSGS-CSNFTWKNINLTGG-KTSG-----ACKNVPSPATCNI----- 366  
BAB61793.1 YVYVLCGSGS-CSNWTWNSVNIIGG-KTSD-----GCKNYPNATCKN----- 369  
CAM91338.1 TYILCAK---CSGWTWNVAVDTGG-TVKK-----ACAGIPTGASC----- 374  
CAM91053.1 TYILCAN---CSGWTWNVAVDTGG-TAKK-----TCAGIPTGASC----- 374  
CAQ03438.1 VYVLCVK---CSRWSW-VSVSGG-K-----KAGLPSGVK----- 351  
AAQ14466.1 VYVLCVK---CSGWTW-VNVDTGG-KTST-----KAGLPSGVK----- 380  
AAM00249.1 VYILCGKGS-CSGWKWSGNSVTGG-KKSS-----SCKNIPSGASC----- 369  
AAB36616.1 VYILCGKGS-CSGWKWSGNSVTGG-KKSS-----SCKNIPSGASC----- 369  
BAG72132.1 IYILCGKGS-CSDWTWNSVAVDTGG-KKST-----ACLNIPEGSGGATC----- 372  
BAG72133.1 IYILCGKGS-CSGWKWSVAVDTGG-KKST-----ACLNIPEGSGGATC----- 372  
ABL01533.1 VYILCASGA-CSNWKWSGNSVTGG-KKST-----KCSNIPSGGAAC----- 362  
CAA61552.1 VYILCAS---CTNWKWSGNSVTGG-KKSS-----KCTGIPSGGAAC----- 363  
BAE97111.1 VYILCAN---CKNWTWNNKVTGG-KTST-----KCKGIPSGASC----- 360  
BAE97133.1 VYILCAN---CKNWTWNNKVTGG-KTST-----KCKGIPSGASC----- 360  
EY20327.1 VYILCAKGA-CSNWKWSGNSVTGG-GKPK-----SCQNVPSPAKCS----- 370  
CAC14022.1 VYVLCAKGA-CSNWKWSGNSVTGG-KSST-----ECSGIPSGGAK----- 362  
CAA64727.1 VYVLCAKGA-CSDWKWSGNSVTGG-KSST-----ECSGIPSSGAK----- 365  
AAA79885.1 IYVLCGKGS-CSGNWNSGNSVTGG-KKSS-----SCLNVPSGASC----- 364  
CAA71247.2 VYILCGSSSSCKGWTWTDNKTVDG-ATPK-----ANSVAVAGVVL----- 370  
CAK47020.1 IYILCGDGS-CSDWTMSGIDITGG-ETSS-----DCENVPAGASCQ----- 370  
ADZ99366.1 YVYVLCAS---CSDWTWSSVDVDTGG-SSSD-----KCEGEPGVCNFC----- 484  
AAR84199.1 VKILVAGAS---KWTWTKVDITGG-CSPG-----SCTGIPSGGAF----- 362  
EDN63471.1 VKILVKNAT---NWQNSGNSVTGG-SSYS-----GCSGIPSGGASC----- 361  
ACA48699.1 VDIDCSTNPTCTDFTISGVTVNKASNTKK-----NVCINLQGGSSGFPICSQ----- 383  
EDP47682.1 IVCSSP--NVCSDIYAENIDVKSPPKGT-----NFVCTNVKSLLDVNCA----- 442  
EAU32536.1 IVCSSP--DVCSDIYVTVNVTSPSGTD-----DYICTNVDESLLDVCNCSG----- 434  
ABF50895.1 IVCSSP--DVCSNIYVENIDVSPSGTN-----DFICTNVDESLLQVNTS----- 434  
EED50415.1 IICSSP--DVCSGIHATNIDVKSPPGDS-----GFVCTNVDEEFLDVCASS----- 433  
ABD61562.1 IVCSSP--DTCSDIYTSNINVTSPDGTN-----DFVCDNVDESLLSVNCTAS----- 434  
EAA22673.1 LVCSPP--NVCSDIYAENINVVSPSGYN-----QAVCTNVDDSLLLNCTA----- 434  
EED15423.1 LVCSPP--NVCSNIFAENINVVSPSGYN-----QAVCTNVDDSLLLNCTK----- 434  
AAF05088.1 FACSSP--SVCNNIVASGIDVLSPKGVD-----EAYCLNVNETGLDVTCTTGF----- 452  
BAE97166.1 LVCSAP--DRCSNIVAEDIKVPVSKKPP-----VYDQCNIDTSTLDTICRDP----- 447  
EY19557.1 LVCSAP--DRCSNIRAENITINVPKPP-----VYACVNVDEGLDLDTCVKPA----- 448  
AAP83023.1 LVCSA--SSCINISAQINITTPSKKKA-----TWTCTNVKSLLLGNCV----- 270  
ABD61563.1 LTCSSP--DVCHNITVQDINVTSPSGDS-----TFTCNMGNLSLEDTCA----- 434  
ABF50897.1 LICGSA--DSCDNIRAENVTLTSSGQPS-----EWQCRMYDEEELLDGGVGC----- 447  
EED46018.1 LVCSPN--AVCSDIQLKNINLTSPAGSPP-----EIVCEGVQDGVGVEQASAD----- 435  
EAL84906.1 LTCSPN--AVCSGIRLKNIHILTSFAGSPP-----VIVCDGIQDGVGVEQSSN----- 455  
AAO61898.1 LICSPN--AICEDIKLGKINLTTPGGERG-----VVVCDGISGGVVEQSSSE----- 440  
ABD61564.1 LVCSPD--AVCTNITLNSVNLTSFPG-TA-----EIVCDDIQGGIGVDCVSDS----- 435  
BAE97090.1 LTCSTSKAVCENIKFKDFNKTCPGDDP-----VFLCDGVKIDGVDCVSAT----- 437  
EY21174.1 LTCSSSPNAVCENIRFKDFNVQSPAAARP-----SSCATASRTVSAPRVSRRA----- 419  
EY20199.1 LTCSTSPDAVCENIRFRNFAVRSFPGGPA-----TAICDGVGTGDTLPCYAAG----- 379  
EY16298.1 -SGMDITNIRFEELHRLHVGRLRQCRGPP-----LLLDGRLGRLRQHRHQLQR----- 443  
EED49587.1 SIYCSDEVP--CPNMSFENVNITSLNATRGQAYDITDIQYELFQCTNVVQNSSGIPCNQA----- 423  
EED23059.1 SLYCSDPVP--CPGIIFDNVTFESVNHITLGLPMWNTMQDEVYQCANIDGQNSSGIPCNV----- 409  
EED47623.1 SLQCSAVAP--CTDIALSGINLELKNGT-----ASEYLCGNVENQK--GFECTGA----- 403  
EAL86831.1 SFQCSAVPK--CTEIQIQNVNQLANGTM-----ADEYLCGNVEGTR--GFNCTGK----- 484  
AAP83024.1 -----  
EY92040.1 SFQCSAAGGDDFEVADIDISVDGGGK-----LDKWHCENVHGK--GFTCNDP----- 445  
EY99815.1 SFQCSAAGGDDFEVSDIDVVDGGDK-----LDNWHCENVHGK--GFTCNDP----- 445  
EAA27533.1 ASVCSNVHPCYNIDFQNVLLPSQNATA-----YGNVTCYSIESGGVRLSGC----- 428  
EED19995.1 ASVCSNVHPCYNIDFQNVLLPSQNVSA-----YGNVTCYSIESGGVHLLGC----- 425  
EDP55950.1 SKVCSNVHPCYNIEFDNVVLYSGKNASS-----PRTGSCKYTANGVHGLAGCWL----- 381  
EED47501.1 LNCGSK---GTCDVITISGFVKAAPSADK-----ILCGNTPSGLVTCSSGAS----- 405  
EAL85450.1 LNCGAR---GTCDVMSAFSVKAPSADK-----VLCNTPSGLVTCSSGAS----- 408  
EAA25131.1 LDCSAD---MTCVYVENFVSKPPSGTAE-----VLCANFDGSPGVCTPGAS----- 420

EED57938.1	CSNGTFTSA-----	437
EDP52870.1	CKNGTFVPH-----	440
ABD61565.1	CKNGTYSPLSN-----	440
BAH29710.1	-----	
BAH14957.1	-----	
BAA96351.1	SNDIEEMPAQDPNDPEDPDTAMQEAEEAEAAAGNSTTSG-----	403
BAE80700.1	ETS-----	367
ACL80126.1	-----	
ABH12113.1	-----	
AAV84618.1	-----	
AAM22187.1	-----	
ADD73449.1	-----	
CAM91235.2	-----	
CAQ03435.1	-----	
CAD62160.1	-----	
ABQ11248.1	-----	
ABQ11247.1	-----	
ABQ11241.1	-----	
ABW96528.1	-----	
ABQ11245.1	-----	
ABQ11246.1	-----	
ABQ11236.1	-----	
ABQ11240.1	-----	
ABQ11234.1	-----	
BAE97018.1	-----	
AAA74586.1	-----	
CAK40265.1	-----	
ABF51154.1	-----	
BAB32924.1	-----	
AAF70169.1	-----	
EDU44300.1	-----	
ABF50893.1	-----	
CAM07141.1	-----	
CAE46647.1	-----	
AAV84616.1	-----	
CAQ03431.1	-----	
AAV84617.1	-----	
CAM32882.1	-----	
EDP49013.1	-----	
EAW19942.1	-----	
CAK47667.1	-----	
ABF50868.1	-----	
EED53217.1	-----	
BAB97386.1	-----	
BAB61793.1	-----	
CAM91338.1	-----	
CAM91053.1	-----	
CAQ03438.1	-----	
AAC14466.1	-----	
AAM00249.1	-----	
AAB36616.1	-----	
BAG72132.1	-----	
BAG72133.1	-----	
ABL01533.1	-----	
CAA61552.1	-----	
BAE97111.1	-----	
BAE97133.1	-----	
EY20327.1	-----	
CAC14022.1	-----	
CAA64727.1	-----	
AAA79885.1	-----	
CAA71247.2	-----	
CAK47020.1	-----	
ADZ99366.1	-----	
AAR84199.1	-----	
EDN63471.1	-----	
ACA48699.1	-----	
EDP47682.1	-----	
EAU32536.1	-----	
ABF50895.1	-----	
EED50415.1	-----	
ABD61562.1	D-----	435
EED22673.1	-----	
EED15423.1	-----	
AAF05088.1	LGFN-----	456
BAE97166.1	SARDTTNG-----	455
EY19557.1	GDRDTSQG-----	456
AA83023.1	-----	
ABD61563.1	-----	
ABF50897.1	PA-----	449
EED46018.1	-----	
EAL84906.1	STTRRSVDLARSCLKYRA-----	472
AAO61898.1	-----	
ABD61564.1	VTR-----	438
BAE97090.1	AGKAALANKCVAPQASASPFKVRKFNG-----	464
EY21174.1	TRASRAQGVHIASGGAWSAVVGVWFDKTRSWTAPASLRGQTGCAKKVKPAGGGYSPCI-----	479
EY20199.1	SABAKAALRDTCIGEKSGYVKGPDWADGGPGF-----	410
EY16298.1	QVAVRAASPSSSATGIGDDLGLPLRAL-----	470
EED49587.1	APSNFSQNIYGNVDSGLATALSDHLGN-----	451
EED23059.1	APDSFGGVSHNVQ-----	422
EED47623.1	VCEGGSATGEC-----	414
EAL86831.1	ACEKPSATGGC-----	495
AA83024.1	-----	
EFY92040.1	PPQKSG-----	451
EFY99815.1	PPQKSG-----	451
EED27533.1	-----	
EED19995.1	-----	
EDP55950.1	SWREIVEIVSSSLRRNTDNIPSLECYQRHREPCRV-----	416
EED47501.1	G-----	406
EAL85450.1	G-----	409
EED25131.1	G-----	421

EED57938.1	-
EDP52870.1	-
ABD61565.1	-
BAH29710.1	-
BAH14957.1	-
BAA96351.1	-
BAE80700.1	-
ACL80126.1	-
ABH12113.1	-
AAV84618.1	-
AAM22187.1	-
ADD73449.1	-
CAM91235.2	-
CAQ03435.1	-
CAD62160.1	-
ABQ11248.1	-
ABQ11247.1	-
ABQ11241.1	-
ABW96528.1	-
ABQ11245.1	-
ABQ11246.1	-
ABQ11236.1	-
ABQ11240.1	-
ABQ11234.1	-
BAE97018.1	-
AAA74586.1	-
CAK40265.1	-
ABF51154.1	-
BAB32924.1	-
AAF70169.1	-
EDU44300.1	-
ABF50893.1	-
CAM07141.1	-
CAE46647.1	-
AAV84616.1	-
CAQ03431.1	-
AAV84617.1	-
CAM32882.1	-
EDP49013.1	-
EAW19942.1	-
CAK47667.1	-
ABF50868.1	-
EED53217.1	-
BAB97386.1	-
BAB61793.1	-
CAM91338.1	-
CAM91053.1	-
CAQ03438.1	-
AAC14466.1	-
AAM00249.1	-
AAB36616.1	-
BAG72132.1	-
BAG72133.1	-
ABL01533.1	-
CAA61552.1	-
BAE97111.1	-
BAE97133.1	-
EY20327.1	-
CAC14022.1	-
CAA64727.1	-
AAA79885.1	-
CAA71247.2	-
CAK47020.1	-
ADZ99366.1	-
AAR84199.1	-
EDN63471.1	-
ACA48699.1	-
EDP47682.1	-
EAU32536.1	-
ABF50895.1	-
EED50415.1	-
ABD61562.1	-
EAA22673.1	-
EED15423.1	-
AAF05088.1	-
BAE97166.1	-
EY19557.1	-
AAF83023.1	-
ABD61563.1	-
ABF50897.1	-
EED46018.1	-
EAL84906.1	-
AAO61898.1	-
ABD61564.1	-
BAE97090.1	-
EY21174.1	Y 480
EY20199.1	-
EY16298.1	-
EED49587.1	-
EED23059.1	-
EED47623.1	-
EAL86831.1	-
AAF83024.1	-
EFY92040.1	-
EFY99815.1	-
EAA27533.1	-
EED19995.1	-
EDP55950.1	-
EED47501.1	-
EAL85450.1	-
EAA25131.1	-

Os resíduos grifados em amarelo são referentes aos 8 resíduos conservados.  
Em azul os resíduos de Gly e Ser envolvidos no reconhecimento do substrato.

**Apêndice C – Resultado da avaliação feito pelo programa WHAT IF para o modelo 3D do *L. gongylophorus*.**

Número	AA	Valor	Número	AA	Valor
1	GLY ( 23)	1,608	168	ILE ( 190)	3,844
2	THR ( 24)	0,108	169	GLN ( 191)	-0,740
3	ILE ( 25)	2,830	170	ASN ( 192)	-2,254
4	ASN ( 26)	-2,627	171	SER ( 193)	2,176
5	SER ( 27)	4,822	172	VAL ( 194)	1,540
6	MET ( 28)	-3,105	173	ILE ( 195)	4,385
7	ALA ( 29)	0,186	174	MET ( 196)	-1,453
8	ASP ( 30)	-1,143	175	ASN ( 197)	-0,690
9	VAL ( 31)	-1,557	176	GLN ( 198)	-3,987
10	ALA ( 32)	-2,567	177	ASP ( 199)	-2,272
11	ALA ( 33)	-2,487	178	ASP ( 200)	-0,430
12	ALA ( 34)	-0,712	179	CYS ( 201)	-1,126
13	VAL ( 35)	-3,937	180	LEU ( 202)	2,872
14	ALA ( 36)	-0,936	181	ALA ( 203)	4,497
15	CYS ( 37)	-0,479	182	ILE ( 204)	2,205
16	THR ( 38)	-1,857	183	ASN ( 205)	-1,101
17	THR ( 39)	-2,056	184	ARG ( 206)	-2,113
18	VAL ( 40)	3,231	185	GLY ( 207)	0,203
19	ASN ( 41)	2,596	186	ALA ( 208)	0,853
20	LEU ( 42)	3,004	187	ASN ( 209)	-0,707
21	ASN ( 43)	-0,432	188	ILE ( 210)	3,733
22	SER ( 44)	-2,985	189	VAL ( 211)	2,241
23	PHE ( 45)	2,556	190	PHE ( 212)	2,949
24	GLN ( 46)	-4,470	191	LYS ( 213)	1,439
25	VAL ( 47)	2,527	192	ASN ( 214)	-1,459
26	PRO ( 48)	-1,719	193	ASN ( 215)	2,366
27	ALA ( 49)	-1,790	194	HIS ( 216)	0,276
28	GLY ( 50)	-2,416	195	CYS ( 217)	2,467
29	GLN ( 51)	-4,930	196	SER ( 218)	2,569
30	THR ( 52)	-1,433	197	GLY ( 219)	-1,635
31	LEU ( 53)	4,050	198	GLY ( 220)	0,922
32	ALA ( 54)	0,041	199	HIS ( 221)	-2,782
33	LEU ( 55)	5,484	200	GLY ( 222)	2,038
34	LYS ( 56)	-4,035	201	ILE ( 223)	2,369
35	LEU ( 57)	-0,884	202	SER ( 224)	1,244
36	LEU ( 58)	-3,462	203	VAL ( 225)	1,997
37	THR ( 59)	-2,655	204	GLY ( 226)	-1,465
38	GLY ( 60)	0,103	205	SER ( 227)	-0,537
39	THR ( 61)	1,146	206	ILE ( 228)	2,242
40	THR ( 62)	2,555	207	ASN ( 229)	-2,348
41	VAL ( 63)	6,553	208	SER ( 230)	-1,477
42	ASN ( 64)	0,698	209	GLY ( 231)	-0,398
43	MET ( 65)	-1,172	210	VAL ( 232)	0,292

44	ASN ( 66)	-3,339	211	THR ( 233)	-1,069
45	GLY ( 67)	-0,616	212	VAL ( 234)	2,965
46	ASP ( 68)	0,058	213	ASN ( 235)	0,744
47	ILE ( 69)	7,639	214	GLY ( 236)	1,217
48	THR ( 70)	4,056	215	ILE ( 237)	3,682
49	PHE ( 71)	1,058	216	THR ( 238)	2,107
50	ALA ( 72)	-1,803	217	ILE ( 239)	3,662
51	HIS ( 73)	-4,117	218	THR ( 240)	1,009
52	GLU ( 74)	-3,368	219	GLY ( 241)	-0,604
53	ASN ( 75)	-3,885	220	ASN ( 242)	2,537
54	TRP ( 76)	-4,151	221	THR ( 243)	1,902
55	ALA ( 77)	-3,959	222	ILE ( 244)	1,118
56	GLY ( 78)	-3,824	223	THR ( 245)	0,084
57	PRO ( 79)	0,137	224	ASN ( 246)	-4,149
58	LEU ( 80)	-0,640	225	ASN ( 247)	2,489
59	PHE ( 81)	3,704	226	ASP ( 248)	-1,461
60	GLN ( 82)	1,271	227	GLN ( 249)	-2,239
61	ILE ( 83)	4,142	228	ALA ( 250)	-0,436
62	SER ( 84)	-0,309	229	LEU ( 251)	1,232
63	GLY ( 85)	1,348	230	ARG ( 252)	1,195
64	SER ( 86)	-0,681	231	ILE ( 253)	4,256
65	SER ( 87)	-2,189	232	LYS ( 254)	1,640
66	ILE ( 88)	2,924	233	THR ( 255)	2,716
67	THR ( 89)	1,735	234	ASP ( 256)	1,453
68	PHE ( 90)	0,754	235	ALA ( 257)	0,026
69	ASN ( 91)	2,394	236	ALA ( 258)	-2,401
70	GLY ( 92)	-1,958	237	ALA ( 259)	-1,081
71	ASN ( 93)	-3,198	238	THR ( 260)	-2,753
72	GLY ( 94)	-0,960	239	GLY ( 261)	-2,839
73	HIS ( 95)	-0,289	240	SER ( 262)	0,632
74	THR ( 96)	2,026	241	THR ( 263)	1,627
75	PHE ( 97)	4,027	242	VAL ( 264)	4,218
76	ASP ( 98)	1,490	243	THR ( 265)	-0,118
77	GLY ( 99)	-2,756	244	ASN ( 266)	-1,074
78	ASN ( 100)	-1,446	245	ILE ( 267)	3,647
79	GLY ( 101)	-1,389	246	THR ( 268)	2,755
80	PRO ( 102)	-1,405	247	TYR ( 269)	1,802
81	LEU ( 103)	-3,631	248	SER ( 270)	0,827
82	TYR ( 104)	-3,644	249	GLY ( 271)	-0,377
83	TRP ( 105)	-3,796	250	ASN ( 272)	1,283
84	ASP ( 106)	-1,908	251	THR ( 273)	0,825
85	GLY ( 107)	-2,706	252	ALA ( 274)	1,269
86	LEU ( 108)	-3,677	253	THR ( 275)	0,503
87	GLY ( 109)	0,406	254	GLY ( 276)	0,403
88	GLY ( 110)	-3,344	255	CYS ( 277)	4,681
89	ASN ( 111)	<b>-5,380</b>	256	LYS ( 278)	-2,020
90	GLY ( 112)	-2,676	257	ARG ( 279)	-2,456



91	GLY ( 113)	-0,708	258	PHE ( 280)	1,361
92	VAL ( 114)	-2,302	259	GLY ( 281)	1,038
93	THR ( 115)	-4,272	260	VAL ( 282)	4,891
94	LYS ( 116)	-3,488	261	ILE ( 283)	1,205
95	PRO ( 117)	-2,631	262	ILE ( 284)	1,224
96	ALA ( 118)	-2,767	263	ASP ( 285)	0,646
97	PRO ( 119)	-0,308	264	GLN ( 286)	0,109
98	MET ( 120)	0,315	265	SER ( 287)	-1,094
99	MET ( 121)	-0,005	266	TYR ( 288)	-2,428
100	LYS ( 122)	2,444	267	PRO ( 289)	-2,629
101	ILE ( 123)	-2,527	268	SER ( 290)	0,112
102	SER ( 124)	-1,410	269	THR ( 291)	-3,565
103	MET ( 125)	-5,696	270	ILE ( 292)	-5,487
104	SER ( 126)	0,217	271	GLY ( 293)	-3,792
105	GLY ( 127)	1,078	272	THR ( 294)	-4,420
106	THR ( 128)	1,294	273	PRO ( 295)	-1,652
107	PHE ( 129)	3,629	274	GLY ( 296)	-2,783
108	SER ( 130)	0,617	275	SER ( 297)	-1,136
109	ASN ( 131)	-2,167	276	GLY ( 298)	0,141
110	VAL ( 132)	1,404	277	VAL ( 299)	-0,910
111	LYS ( 133)	2,720	278	LEU ( 300)	0,220
112	VAL ( 134)	4,582	279	LEU ( 301)	3,519
113	LEU ( 135)	0,540	280	SER ( 302)	2,894
114	ASN ( 136)	-2,085	281	ALA ( 303)	0,962
115	SER ( 137)	-0,870	282	VAL ( 304)	0,183
116	PRO ( 138)	-2,343	283	ASN ( 305)	3,286
117	ALA ( 139)	0,241	284	PHE ( 306)	0,453
118	ARG ( 140)	1,134	285	VAL ( 307)	-2,566
119	THR ( 141)	-1,768	286	SER ( 308)	-3,716
120	TYR ( 142)	-5,644	287	PRO ( 309)	-1,868
121	SER ( 143)	0,399	288	GLN ( 310)	-3,172
122	VAL ( 144)	-1,336	289	THR ( 311)	2,330
123	SER ( 145)	0,199	290	SER ( 312)	1,825
124	ASN ( 146)	-2,295	291	ILE ( 313)	7,057
125	PRO ( 147)	-0,664	292	THR ( 314)	0,074
126	ALA ( 148)	-2,420	293	VAL ( 315)	-0,513
127	PRO ( 149)	-3,217	294	ASN ( 316)	-0,969
128	LEU ( 150)	-2,866	295	SER ( 317)	-3,556
129	VAL ( 151)	-2,729	296	GLY ( 318)	-3,985
130	MET ( 152)	3,249	297	ALA ( 319)	-2,905
131	ASP ( 153)	0,564	298	GLU ( 320)	-2,117
132	GLY ( 154)	0,825	299	ARG ( 321)	-1,203
133	LEU ( 155)	3,931	300	VAL ( 322)	3,435
134	THR ( 156)	2,228	301	ALA ( 323)	0,123
135	ILE ( 157)	2,705	302	VAL ( 324)	-0,042
136	ASP ( 158)	-0,471	303	ASN ( 325)	0,045
137	ASN ( 159)	2,063	304	CYS ( 326)	1,889

138	SER ( 160)	-3,172	305	GLY ( 327)	-1,271
139	LEU ( 161)	-1,544	306	VAL ( 328)	-2,210
140	GLY ( 162)	-2,710	307	GLY ( 329)	-1,600
141	ASP ( 163)	-2,529	308	SER ( 330)	0,713
142	GLN ( 164)	-2,754	309	CYS ( 331)	0,288
143	PRO ( 165)	-0,962	310	THR ( 332)	0,007
144	ASN ( 166)	-5,047	311	GLY ( 333)	-0,488
145	SER ( 167)	-3,553	312	THR ( 334)	-2,075
146	ASN ( 168)	-2,095	313	TRP ( 335)	-0,623
147	SER ( 169)	-3,907	314	ASN ( 336)	0,888
148	ASN ( 170)	-4,857	315	TRP ( 337)	-0,617
149	GLY ( 171)	-1,262	316	SER ( 338)	-0,256
150	LEU ( 172)	-4,204	317	ASN ( 339)	0,071
151	ALA ( 173)	-2,815	318	LEU ( 340)	3,381
152	ALA ( 174)	-1,198	319	LYS ( 341)	0,633
153	GLY ( 175)	-2,258	320	VAL ( 342)	-0,353
154	HIS ( 176)	-3,806	321	SER ( 343)	0,379
155	ASN ( 177)	-0,019	322	GLY ( 344)	-2,508
156	THR ( 178)	-0,992	323	GLY ( 345)	-1,943
157	ASP ( 179)	1,172	324	THR ( 346)	-1,401
158	GLY ( 180)	0,236	325	THR ( 347)	-3,880
159	PHE ( 181)	2,291	326	GLY ( 348)	-1,420
160	ASP ( 182)	-2,184	327	LYS ( 349)	-0,273
161	ALA ( 183)	1,167	328	ILE ( 350)	-2,559
162	SER ( 184)	-3,025	329	THR ( 351)	-0,542
163	THR ( 185)	-2,411	330	ASN ( 352)	-1,112
164	GLU ( 186)	-1,832	331	PHE ( 353)	-4,101
165	ASN ( 187)	-2,193	332	ASN ( 354)	-3,811
166	LEU ( 188)	1,192	333	GLY ( 355)	-2,962
167	THR ( 189)	1,252			

Número: refere-se ao número sequencial do resíduo na proteína; AA: resíduo. Os aminoácidos em destaque são aqueles que apresentaram piores valores.

## Apêndice D – Escore resultante do alinhamento entre Endo-PG de fungos de estrutura 3D conhecida e a sequência de PG do *L. gongylophorus*

Sequence 1: ADV30326.1	361 aa
Sequence 2: 1K5C A	335 aa
Sequence 3: 1HG8 A	349 aa
Sequence 4: 1IA5 A	339 aa
Sequence 5: 1CZF A	362 aa
Sequence 6: 1NHC A	336 aa
Sequence 7: 2IQ7 A	339 aa

Sequences (1:2) Aligned.	Score: 50
Sequences (1:3) Aligned.	Score: 32
Sequences (1:4) Aligned.	Score: 38
Sequences (1:5) Aligned.	Score: 32
Sequences (1:6) Aligned.	Score: 34
Sequences (1:7) Aligned.	Score: 37
Sequences (2:3) Aligned.	Score: 32
Sequences (2:4) Aligned.	Score: 36
Sequences (2:5) Aligned.	Score: 35
Sequences (2:6) Aligned.	Score: 38
Sequences (2:7) Aligned.	Score: 39
Sequences (3:4) Aligned.	Score: 44
Sequences (3:5) Aligned.	Score: 41
Sequences (3:6) Aligned.	Score: 40
Sequences (3:7) Aligned.	Score: 46
Sequences (4:5) Aligned.	Score: 58
Sequences (4:6) Aligned.	Score: 61
Sequences (4:7) Aligned.	Score: 61
Sequences (5:6) Aligned.	Score: 59
Sequences (5:7) Aligned.	Score: 63
Sequences (6:7) Aligned.	Score: 61