UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

ATÍLIO TOMAZINI JÚNIOR

CARACTERIZAÇÃO ESTRUTURAL DA GLICOSILHIDROLASE
DE *Xanthomonas campestris* CONTENDO OS DOMÍNIOS:
CATALÍTICO DA FAMÍLIA GH5 E SIMILAR À EXPANSINA.

São Carlos

2012

ATÍLIO TOMAZINI JÚNIOR

CARACTERIZAÇÃO ESTRUTURAL DA GLICOSILHIDROLASE
DE *Xanthomonas campestris* CONTENDO OS DOMÍNIOS:
CATALÍTICO DA FAMÍLIA GH5 E SIMILAR À EXPANSINA.

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular

Orientador: Prof. Dr. Igor Polikarpov

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2012

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Tomazini Júnior, Atílio

Caracterização estrutural da Glicosilhidrolase de
Xanthomonas campestris contendo os domínios:
catalítico da família GH5 e semelhante à Expansina. /
Atílio Tomazini Júnior; orientador Igor Polikarpov -
versão corrigida -- São Carlos, 2012.
143 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Física Aplicada Biomolecular) -- Instituto de Física
de São Carlos, Universidade de São Paulo, 2012.

1. Glicosilhidrolase. 2. Expansina. 3.
Xanthomonas campestris. 4. Biomassa. 5.
Biocombustíveis. I. Polikarpov, Igor, orient. II.
Título.

A meus pais Atílio e Fátima,
pelo apoio, confiança e amor.

Dedico.

AGRADECIMENTOS

A Deus, pela vida e pelas pessoas que Ele colocou em meu caminho;

Aos meus pais Atílio e Fátima, aos meus irmãos Daniela e Maico e minhas sobrinhas Yasmin e Helena, pelo apoio e carinho em todos os momentos;

Ao Prof. Dr. Igor Polikarpov, pela confiança, pelas discussões sempre construtivas e pelo apoio na realização deste trabalho;

A Helena Judith minha avó, Márcia e Fabiana em nome das quais agradeço a todos meus familiares;

Aos amigos Lucas Leme (pela amizade), Luisa (pela presença e companheirismo), Priscilla (Pucca), Ana Paula (Coro), Alessandro, Ártemis, Dimas, Ericka e Itamar;

Aos amigos de graduação: Caio (pela amizade), Ivan (por todo o apoio acadêmico e amizade), Geraldo (pela prestatividade constante), Rodrigo (pelas conversas sempre produtivas), Hilde (pelo contágio de felicidade), Wagner, B2, Luciano, Paulinha, Ricardo, Udi, Erica, Cauê, Ila, Alexandre e Wally;

Aos amigos Ivoney, Vera, Marina e Raquel;

Aos meus iniciadores na vida científica Joci Neuby Alves Macedo (pelo apoio constante), Andressa Patrícia A. Pinto (pelo apoio constante), Derminda Isabel de Moraes (Bel) (pelo apoio constante), Prof. Ana Paula Ullian de Araújo, Prof. Ricardo de Marco, Prof. Nelma Segnini Bosolam e Prof. Leila Beltramini ;

Ao pesquisador Sasha (Alexander) Popov e Sean McSweeney pela oportunidade do estágio no E.S.R.F em Grenoble;

A Silvio Athayde pela dedicação, atenção e apoio em nome do qual agradeço a todos do departamento acadêmico;

A Erika Pellegrini e Daniele de Sanctis em nome dos quais agradeço a todos do PSB em Grenoble;

Aos meus alunos de iniciação científica Luciano Dolce e Anna Naice que co-orientei e com os quais aprendi muito;

A Elizabeth Conti pelo apoio constante nos assuntos acadêmicos desde a graduação;

Aos amigos de sala Vanessa, Evandro (com o qual pude ensinar e aprender muita coisa), César, Mi, Eliane, Andre, Fernando Maluf, Fernando Bachega e Jademilson;

Aos colegas de laboratório Ana, Amanda, Fernanda, Marco, Antônio, Renata, Fran, Tavin, Vivi, Nicoli, Thais, Mariana, Paola, Heline, Marisa, Camila, Malu, Flávio, Livia, Bruno, Daiana, Carol, Kelven e Evandro;

Aos técnicos do Grupo de Biotecnologia Molecular Maria, Livia, Josimar e João e do grupo de cristalografia Suzana, Humberto e Simone;

As secretárias Mariana, Luciana e Éster;

A Daniele Santini, Cristina Silva em nome das quais agradeço a todos os funcionários do IFSC;

Aos amigos do Grupo de Biofísica Molecular Prof. Sérgio Mascarenhas;

Ao professor Richard Charles Garrett em nome do qual agradeço a todos os docentes do IFSC;

Ao Prof. Antonio José da Costa Filho (Jabah) pela prestatividade;

A Ana, Marco, Amanda e Larissa Textor pelo auxílio com os dados de cristalografia e ao Prof. Mario de Oliveira Neto pelo auxílio com os dados de SAXS e a Bruno Matsuyama pelo auxílio com os dados de DLS;

Às funcionárias do serviço de Biblioteca e Informação do IFSC-USP, pela prestatividade em todos os momentos em especial Neusa e Cris;

À Universidade de São Paulo e o Instituto de Física de São Carlos, por oferecer toda a estrutura necessária para realização deste trabalho;

Ao CNPQ, pela concessão da bolsa de mestrado;

Ao LNLS e E.S.R.F, pela infra-estrutura para realização dos experimentos;

Enfim, a todos que contribuíram de forma direta ou indireta para a realização deste trabalho.

"O sucesso dos que triunfam é saber começar tudo de novo"

Manuel Angel Pan Martinez

RESUMO

TOMAZINI JUNIOR, A. **Caracterização estrutural da Glicosilhidrolase de *Xanthomonas campestris* contendo os domínios: catalítico da família GH 5 e domínio similar a Expansina**. 2012. 143 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

A demanda global por energia tem acompanhado o desenvolvimento de novas indústrias, o que exige um constante aprimoramento e busca por novas fontes de energia. Com a conscientização da população a respeito da necessidade de fontes de energia renovável se abriu uma expectativa otimista para o uso da biomassa como fonte de energia.

Uma variedade de microrganismos, como bactérias e fungos produzem enzimas que podem ser usadas para a despolimerização de biomassa celulósica em monômeros de glicose. Dentro deste contexto, estudamos uma celulase (XCC3535) de *Xanthomonas campestris*. Esta proteína é um híbrido natural com dois domínios (um domínio glicosilhidrolase e um domínio semelhante à expansina) que podem possuir funções complementares. Acredita-se que expansinas e proteínas semelhantes a expansinas possam ajudar na degradação da celulose cristalina, participando na ligação ao substrato, seguida pela formação de complexos enzimáticos ou, possivelmente, associadas à superfície para degradação.

Este trabalho foi iniciado com análises de bioinformática da sequência para a proteína Glicosilhidrolase permitindo três construções para clonagem : uma completa contendo ambos os domínios (catalítico e semelhante a Expansina), outro com apenas o domínio catalítico GH 5, e um terceiro com apenas o domínio semelhante a Expansina. Obtemos por expressão recombinante em *Escherichia coli* Rosetta proteína das três construções.

A amostra protéica pura foi utilizada para estudos por espalhamento dinâmico de luz (DLS), espalhamento de raios-X a baixo ângulo (SAXS), dicróismo circular (CD) e cristalização para uso aplicando difração de raios-X. Assim, nosso principal objetivo neste estudo foi a descrição da Glicosilhidrolase XCC3535, permitindo assim a geração de conhecimento para o entendimento da função desta proteína e de seus domínios.

Palavras-chave: Glisossilhidrolase. Expansina. *Xanthomonas campestris*. Biomassa. Biocombustíveis.

ABSTRACT

TOMAZINI JUNIOR, A. **Structural characterization of Glycosyl hydrolase by *Xanthomonas campestris* containing: GH5 catalytic domain and Expansin-like domain.** 2012. 143 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

Global demand for energy has grown with the development of new industries that requires constant improvement and search for new sources of energy. Nowadays, the growing awareness of the population about the need for renewable energy have opened an optimistic expectation for the use of biomass as energy source.

A variety of microorganisms like bacteria and fungi produce enzymes that can be used for the depolymerization of cellulosic biomass into glucose monomers. Within this context, we study one cellulase (XCC3535) from *Xanthomonas campestris* organism. This protein is a natural hybrid with two domains (a glicosyl hidrolase domain and an expansin-like domain) which seem to have complementary functions. It is believed that expansins help in the degradation of crystalline cellulose by participating in substrate binding, followed by formation of enzyme complexes or possibly attaching to cell surface.

This work was initiated using bioinformatics analysis of the protein sequence and three constructs were made: a full length containing both domains (catalytic and expansin), another with only the catalytic domain, and a third one with only the expansin-like domain. We were able to overexpress these recombinant proteins in *Escherichia coli* Rosetta strain.

The pure protein sample was used for studied by Dynamic Light Scattering (DLS), Small angle X-ray scattering, Circular Dichroism (CD) and Crystallization for X-ray diffraction. Thus, our main objective in this study was structurally describe these proteins, thereby allowing the generation of knowledge for understanding function of these proteins.

Keywords: Glicosyl hydrolase. Expansin . *Xanthomonas campestris*. Biomass. Biofuels.

LISTA DE FIGURAS

FIGURA 1 – ESTRUTURA QUÍMICA DA CELOBIOSE ¹²	32
FIGURA 2 – ESTRUTURA QUÍMICA DOS PRECURSORES DA LIGNINA: (1) ÁLCOOL P-CUMARIL , (2) ÁLCOOL CONIFERIL , E (3)	33
ÁLCOOL SINAPIL.	33
FIGURA 3 –MATERIAL LIGNOCELULÓSICO, ADAPTADO ¹⁵	34
FIGURA 4 – MODO DE AÇÃO DAS ENZIMAS DO COMPLEXO CELULOLÍTICO. ADAPTADO 18	36
FIGURA 5 – A HEMICELULOSE MAIS PREVALENTE É O XILANO, AS ENZIMAS XILANOLÍTICAS ENVOLVIDAS EM SUA DEGRADAÇÃO ESTÃO INDICADAS NA FIGURA, ADPATADO ²²	37
FIGURA 6 – MODELO DE AÇÃO DAS EXPANSINAS NO AFROUXAMENTO DAS FIBRAS CELULÓSICAS . AS MICROFIBRILAS DE CELULOSE ESTÃO LIGADAS ENTRE SI POR GLICANAS (ESTRUTURAS EM AMARELO E VERMELHO) QUE PODEM PREGAR AS MICROFIBRILAS DA SUPERFÍCIE UMAS AS OUTROS. ACREDITASSE QUE A PROTEÍNA EXPANSINA (AZUL) (A) PERTURBA A INTERAÇÃO DAS GLICANAS ÀS MICROFIBRILAS SUPERFÍCIE (B) OU ENTRE SI (C) SOB O ESTRESSE MECÂNICO DECORRENTE DO TURGOR, A EXPANSINA PODE DESLOCAR-SE SOBRE A PAREDE DO POLÍMERO E INTERFERIR NOS PONTOS DE ADERÊNCIA ,ADAPTADO ²⁶	39
FIGURA 7 – REGIÕES CONSERVADAS NA SEQÜÊNCIA GÊNICA DA SUPERFAMÍLIA DE EXPANSINAS E A REGIÃO AMINO TERMINAL DE BAIXA CONSERVAÇÃO FORAM REMOVIDAS PARA O ALINHAMENTO, PORQUE ALGUMAS EXPANSINAS POSSUEM GRANDES SEQÜÊNCIA PEPTÍDEO SINAL E AMINO TERMINAL. RESÍDUOS CONSERVADOS ENTRE AS FAMÍLIAS ESTÃO SOMBREADAS, E A FRONTEIRA ENTRE OS DOIS DOMÍNIOS É INDICADO POR SETAS. RESÍDUOS ALTAMENTE CONSERVADOS QUE FAZEM PARTE DO SÍTIO CATALÍTICO DE PROTEÍNAS GH45 E QUE SÃO CONSERVADAS NO DOMÍNIO 1 DE ALGUNS FAMÍLIAS DE EXPANSINAS SÃO MOSTRADOS EM CÍRCULOS ACIMA DO ALINHAMENTO. AS FAMÍLIAS ALINHADAS FORAM: ALFA EXPANSINA, DESIGNADA EXPA , BETA EXPANSINA, DESIGNADA EXPB E A AS PROTEÍNAS SEMELHANTES A EXPANSINAS, PROTEÍNA SEMELHANTE A EXPANSINA DO TIPO ALFA, DESIGNADA EXLA (EXPANSIN-LIKE A) E PROTEÍNA SEMELHANTE A EXPANSINA DO TIPO BETA, DESIGNADA EXLB (EXPANSIN-LIKE B), ADPTADO ²⁷	41
FIGURA 8 – ESQUEMA DO PROCESSO PRODUTIVO DO ETANOL DE SEGUNDA GERAÇÃO.	42
FIGURA 9 – SINTOMAS DA PODRIDÃO NEGRA EM FOLHAS DE REPOLHO ³⁵	44
FIGURA 10 – MODELO ESTRUTURAL DE ALTA RESOLUÇÃO PARA A ENDOGLUCANASE DE <i>CLOSTRIDIUM THERMOCELLUM</i> (1CEC).....	46

FIGURA 11 – MODELO ESTRUTURAL DE ALTA RESOLUÇÃO PARA A EXPANSINA DE <i>ZEAMAYS</i> (2HCZ).	46
FIGURA 12 – AS ZONAS DE ALINHAMENTO DE SEQUÊNCIAS ⁴⁹ .	53
FIGURA 13 – REPRESENTAÇÃO ESQUEMÁTICA DO VETOR pSMT3 E QUADRO DE LEITURA DO VETOR pSMT3, POSSUI 5.6633PB, A REGIÃO DE MÚLTIPLA CLONAGEM ENCONTRA-SE APÓS O SUMO NA REGIÃO 3' (C-TERMINAL).	56
FIGURA 14 - REPRESENTAÇÃO ESQUEMÁTICA DAS CONDIÇÕES DA REAÇÃO DE AMPLIFICAÇÃO DOS GENES.	57
FIGURA 15 – ETAPAS DO PROCESSO DE CLONAGEM.	62
FIGURA 16 – ORIGEM DO EFEITO DO CD (A) A ESQUERDA (L) E DIREITA (R) COMPONENTES DE MESMO TAMANHO, LUZ CIRCULARMENTE POLARIZADA (II) COMPONENTES SÃO DE DIFERENTES MAGNITUDES E RESULTANDO EM LUZ ELÍPTICAMENTE POLARIZADA.	67
FIGURA 17 – ESPECTRO TÍPICO DAS PRINCIPAIS ESTRUTURAS SECUNDÁRIAS EM PROTEÍNAS, ILUSTRAÇÃO DE NORMA J. GREENFIELD	68
FIGURA 18 – AS PARTÍCULAS NUM LÍQUIDO SE MOVIMENTAM ALEATORIAMENTE E SUA VELOCIDADE DE MOVIMENTO É UTILIZADA PARA DETERMINAR O TAMANHO DAS PARTÍCULA, ADAPTADO ⁵⁷	70
FIGURA 19 - DISTRIBUIÇÃO DE TAMANHO MONODISPERSO E POLIDISPERSO, ADAPTADO ^{56; 57}	71
FIGURA 20 – ESQUEMA REPRESENTANDO O EXPERIMENT DE SAXS E A TRANSFORMADA DE FOURIER DO ESPAÇO REAL PARA O RECÍPROCO. ADAPTADO ⁵⁹	73
FIGURA 21 – PROCESSO APLICADO NO MÉTODO DE SUBSTITUIÇÃO MOLECULAR. ⁷⁰	79
FIGURA 22 – ARQUIVO DE SAÍDA DO PROGRAMA SIGNALP ⁴² , PARA AVALIAÇÃO DA SEQUÊNCIA DE XCC3535.	83
FIGURA 23 – ARQUIVO DE SAÍDA DO PROGRAMA DOMMAING FISH ⁴⁴ PARA AVALIAÇÃO DO DOMÍNIO DA SEQUÊNCIA XCC3535.	85
FIGURA 24 – DOMÍNIOS CONSERVADOS: OBTIDOS A PARTIR DO CDD, ANOTAÇÃO FUNCIONAL ESPECÍFICA, PARA AVALIAÇÃO DO DOMÍNIO DA SEQUÊNCIA XCC3535.	86
FIGURA 25 – ALINHAMENTO MÚLTIPLO UTILIZANDO O PROGRAMA CLUSTAL, SEQUÊNCIAS ALINHADAS 3AXX (ENDOGLUCANASE <i>PYROCOCCUS HORIKOSHU</i>), GI 188575103 (GLICOSILHIDROLASE <i>XANTHOMONAS ORYZAE</i>) E A CELLULASE DE <i>XANTHOMONAS CAMPESTRIS</i> (PROTEÍNA DE ESTUDO).	87
FIGURA 26 – ALINHAMENTO MÚLTIPLO UTILIZANDO O PROGRAMA CLUSTAL, SEQUÊNCIAS ALINHADAS CELLULASE DE <i>XANTHOMONAS CAMPESTRIS</i> (PROTEÍNA DE ESTUDO), 2BH0	

(PROTEÍNA SIMILAR À EXPANSINA DE <i>BACILLUS SUBTILIS</i>) E 2HCZ (EXPANSINA DE <i>ZEAMAYS</i>).....	88
FIGURA 27 - IDENTIFICAÇÃO DOS DOMÍNIOS PARA XCC3535. EM VERDE TEMOS A SEQUÊNCIA PEPTÍDEO SINAL QUE FOI REMOVIDA, EM VERMELHO A REGIÃO DELIMITADA COMO DOMÍNIO UM, EM AZUL O DOMÍNIO DOIS E EM ROXO O LINKER.....	89
FIGURA 28 – ALINHAMENTO UTILIZANDO O CLUSTAL DO DOMÍNIO CATALÍTICO(GLI) COM A SEQUENCIA PARA 2ZUN	92
FIGURA 29 – CORREÇÃO DO ALINHAMENTO, EM CÓDIGO PIR.....	92
FIGURA 30 – GRÁFICO ENERGIA EM FUNÇÃO DO MODELO PARA O DOMÍNIO CATALÍTICO.....	93
FIGURA 31 – DIAGRAMA DE RAMACHANDRAN PARA O MODELO DO DOMÍNIO CATALÍTICO.....	94
FIGURA 32 – SOBREPOSIÇÃO ENTRE A ESTRUTURA GERADA PARA O DOMÍNIO CATALÍTICO (VERDE) E O MODELO UTILIZADO 2ZUN (ROSA), EM DESTAQUE OS LOOPS QUE FORAM MODELADOS.	95
FIGURA 33 - AMPLIFICAÇÃO DAS ORF DOS GENES, GLICOSILHIDROLASE 1707 PB, DOMÍNIO CATALÍTICO 954 PB E DOMÍNIO EXPANSINA 618PB. RESULTADO ANALISADO EM GEL DE AGAROSE 1%, TAE 1X. M: MARCADOR DE MASSA MOLECULAR GENE RULER 1Kb LADDER PLUS (FERMENTAS LIFE SCIENCES).	96
FIGURA 34 - CLIVAGEM DO VETOR DE SUBCLONAGEM pGEM-T: pGEM-T LINEARIZADO 3000PB, GLICOSILHIDROLASE 1707 PB, DOMÍNIO CATALÍTICO 954 PB E DOMÍNIO SEMELHANTE À EXPANSINA 618PB. RESULTADO ANALISADO EM GEL DE AGAROSE 1%, TAE 1X. M: MARCADOR DE MASSA MOLECULAR GENE RULER 1Kb LADDER PLUS (FERMENTAS LIFE SCIENCES).....	97
FIGURA 35 - CLIVAGEM DO VETOR DE CLONAGEM OU EXPRESSÃO pSMT3 5.6633PB: pSMT3 LINEARIZADO, GLICOSILHIDROLASE 1707 PB, DOMÍNIO CATALÍTICO 954 PB E DOMÍNIO SEMELHANTE À EXPANSINA 618PB. RESULTADO ANALISADO EM GEL DE AGAROSE 1%, TAE 1X. M: MARCADOR DE MASSA MOLECULAR GENE RULER 1Kb LADDER PLUS (FERMENTAS LIFE SCIENCES).....	97
FIGURA 36 –RECORTES DO CROMATOGRAMA DE SEQUENCIAMENTO DA ORF PARA GLICOSILHIDROLASE, NA FIGURA S2 E S3 REPRESENTAM AS SEQUÊNCIAS NO SENTIDO DIRETO E INVERSO RESPECTIVAMENTE, GLICOSILHIDROLASE A SEQUÊNCIA DEPOSITADA NO NCBI E CONTIG A SEQUÊNCIA CONSENSO DO ALINHAMENTO DAS SEQUÊNCIAS AMPLIFICADAS.	98
FIGURA 37 - A. EXPRESSÃO DA GLICOSILHIDROLASE. NA 1ª LINHA ESTÃO OS MARCADORES DE PESO MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-	

3-FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE PÂNCREAS BOVINO (24kDa), INIBIDOR DE TRIPSINA DE SOJA(20,1kDa) E A-LACTALBUMINA DE LEITE BOVINO (14,2kDa). A PRÓXIMA COLUNA REPRESENTA O EXTRATO DE PROTEÍNAS ANTES DA INDUÇÃO, A SEGUNDA LINHA O EXTRATO DE PROTEÍNAS PÓS INDUÇÃO. AS 3ª E 4ª COLUNAS REPRESENTAM, RESPECTIVAMENTE, PÉLETE DA LISE CELULAR E SOBRENADANTE DA LISE CELULAR. B. EXPRESSÃO DO DOMÍNIO CATALÍTICO. NA 1ªCOLUNA ESTÃO OS MARCADORES DE PESO MOLECULAR O MESMO JÁ DESCRITO. A PRÓXIMA COLUNA REPRESENTA O EXTRATO DE PROTEÍNAS ANTES DA INDUÇÃO, A SEGUNDA COLUNA O EXTRATO DE PROTEÍNAS PÓS INDUÇÃO. AS 3ª E 4ª COLUNAS REPRESENTAM, RESPECTIVAMENTE, PÉLETE DA LISE CELULAR E SOBRENADANTE DA LISE CELULAR. C. EXPRESSÃO DO DOMÍNIO SEMELHANTE À EXPANSINA. NA 1ª COLUNA ESTÃO OS MARCADORES DE PESO MOLECULAR JÁ DESCRITOS. A PRÓXIMA COLUNA REPRESENTA O EXTRATO DE PROTEÍNAS ANTES DA INDUÇÃO, A SEGUNDA LINHA O EXTRATO DE PROTEÍNAS PÓS INDUÇÃO. AS 3ª E 4ª COLUNAS REPRESENTAM, RESPECTIVAMENTE, PÉLETE DA LISE CELULAR E SOBRENADANTE DA LISE CELULAR. 100

FIGURA 38 - A. PURIFICAÇÃO DA GLICOSILHIDROLASE. NA 1ª COLUNA ESTÃO OS MARCADORES DE PESO MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-3-FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE PÂNCREAS BOVINO (24kDa), INIBIDOR DE TRIPSINA DE SOJA(20,1kDa) E A-LACTALBUMINA DE LEITE BOVINO (14,2kDa). A PRÓXIMA COLUNA REPRESENTA O VOID CELULAR, A SEGUNDA A LAVAGEM DA COLUNA. A 3ª A ELUIÇÃO EM 300MM DE IMIDAZOL E A 4ª LINHA A ELUIÇÃO EM 500MM DE IMIDAZOL. 101

FIGURA 39 - A. PURIFICAÇÃO DO DOMÍNIO CATALÍTICO. NA 1ª COLUNA ESTÃO OS MARCADORES DE PESO MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-3-FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE PÂNCREAS BOVINO (24kDa), INIBIDO DE TRIPSINA DE SOJA(20,1kDa) E A-LACTALBUMINA DE LEITE BOVINO (14,2kDa). A PRÓXIMA COLUNA REPRESENTA O VOID CELULAR, A SEGUNDA A LAVAGEM DA COLUNA. AS 3ª E 4ª COLUNA REPRESENTAM, RESPECTIVAMENTE, ELUIÇÃO A 300MM E 500MM DE IMIDAZOL. 101

FIGURA 40 - A. PURIFICAÇÃO DA EXPANSINA. NA 1ª COLUNA ESTÃO OS MARCADORES DE PESO MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-3-FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE PÂNCREAS BOVINO (24kDa), INIBIDO DE TRIPSINA DE SOJA(20,1kDa) E A-LACTALBUMINA DE LEITE BOVINO (14,2kDa). A PRÓXIMA COLUNA REPRESENTA O VOID CELULAR, A

SEGUNDA A LAVAGEM DA COLUNA. AS 3ª E 4ª LINHAS REPRESENTAM, RESPECTIVAMENTE,
ELUIÇÃO A 100MM E 500MM DE IMIDAZOL..... 102

FIGURA 41 - *A*. CROMATOGRAMA DA EXCLUSÃO MOLECULAR DA GLICOSILHIDROLASE. *B*
FRAÇÕES DO PICO DO CROMATOGRAMA. NA 1ª COLUNA ESTÃO OS MARCADORES DE PESO
MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-3-
FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE
PÂNCREAS BOVINO (24kDa), INIBIDO DE TRIPSINA DE SOJA(20,1KDa) E A-LACTALBUMINA
DE LEITE BOVINO (14,2kDa). DA COLUNA A SEXTA LINHA REPRESENTAM
RESPECTIVAMENTE OS VOLUMES DE 8,5;9;9,5;10;10,5; 11 E 11,5 ML. 102

FIGURA 42 - *A*. CROMATOGRAMA DA EXCLUSÃO MOLECULAR DO DOMÍNIO CATALÍTICO. *B*
FRAÇÕES DO PICO DO CROMATOGRAMA. NA 1ª COLUNA ESTÃO OS MARCADORES DE PESO
MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-3-
FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE
PÂNCREAS BOVINO (24kDa), INIBIDO DE TRIPSINA DE SOJA(20,1KDa) E A-LACTALBUMINA
DE LEITE BOVINO (14,2kDa). DA SEGUNDA A SEXTA COLUNA REPRESENTAM
RESPECTIVAMENTE OS VOLUMES DE 7,7.5,8,8.5 E 9 ML. 103

FIGURA 43 - *A*. CROMATOGRAMA DA EXCLUSÃO MOLECULAR DO DOMÍNIO EXPANSINA. *B*
FRAÇÕES DO PICO DO CROMATOGRAMA. NA 1ªCOLUNA ESTÃO OS MARCADORES DE PESO
MOLECULAR ALBUMINA BOVINA (66kDa), OVOALBUMINA (45kDa), GLICERALDEÍDO-3-
FOSFATO DESIDROGENASE (36kDa), ANIDRASE CARBÔNICA (29kDa), TRIPSINOGENIO DE
PÂNCREAS BOVINO (24kDa), INIBIDO DE TRIPSINA DE SOJA(20,1KDa) E A-LACTALBUMINA
DE LEITE BOVINO (14,2kDa). DA SEGUNDA A SEXTA COLUNA REPRESENTAM
RESPECTIVAMENTE OS VOLUMES DE,15,15.5,16, 16.5 E 17 ML. 104

FIGURA 44 - CURVA DE CALIBRAÇÃO DA SUPERDEX 75TM 10/300 PARA O CÁLCULO MASSA
MOLECULAR..... 105

FIGURA 45 - ESPECTRO DE DICROÍSMO CIRCULAR DA GLICOSILHIDROLASE EM TAMPÃO
FOSFATO. 106

FIGURA 46 - ESPECTRO DE DICROÍSMO CIRCULAR DA PROTEÍNA SIMILAR À EXPANSINA EM
TAMPÃO FOSFATO. 107

FIGURA 47 - SOBREPOSIÇÃO DO ESPECTRO DE DICROÍSMO CIRCULAR TEÓRICO E EXPERIMENTAL
PARA A GLICOSILHIDROLASE. 109

FIGURA- 48 SOBREPOSIÇÃO DO ESPECTRO DE DICROÍSMO CIRCULAR TEÓRICO E EXPERIMENTAL
PARA A PROTEÍNA SIMILAR À EXPANSINA EM TAMPÃO FOSFATO..... 109

FIGURA- 49 SOBREPOSIÇÃO DO ESPECTRO DE DICROÍSMO CIRCULAR TEÓRICO E EXPERIMENTAL PARA A PROTEÍNA SIMILAR À EXPANSINA EM TAMPÃO FOSFATO UTILIZANDO A BASE DE DADOS SP 43 E DESCONVOLUÇÃO NO PROGRAMA SELCON3.....	110
FIGURA 50 - ESPECTRO DE DICROÍSMO CIRCULAR DA GLICOSILHIDROLASE EM 10, 40 E 45°C	111
FIGURA 51 - ESPECTRO DE DICROÍSMO CIRCULAR DO DOMÍNIO SEMELHANTE EXPANSINA, EM 10, 45 E 50°C.	111
FIGURA 52 – DISTRIBUIÇÃO DO TAMANHO PELA INTENSIDADE PARA A AMOSTRA DE GLICOSILHIDROLASE	112
FIGURA 53 – DISTRIBUIÇÃO DO TAMANHO PELA INTENSIDADE PARA A AMOSTRA DO DOMÍNIO SEMELHANTE A EXPANSINA	113
FIGURA 54- OBTENÇÃO DO RAIOS DE GIRO PELO MÉTODO DE GUINIER: (A) GLICOSILHIDROLASE (B) DOMÍNIO SEMELHANTE À EXPANSINA.....	114
FIGURA 55 - FUNÇÃO DE DISTRIBUIÇÃO DE DISTÂNCIAS PARA OS DADOS EXPERIMENTAIS DA GLICOSILHIDROLASE (A) E PARA O DOMÍNIO SEMELHANTE À EXPANSINA (B).	115
FIGURA 56 - CURVA DE ESPALHAMENTO DE RAIOS-X A BAIXO ÂNGULO.(A) PARA A GLICOSILHIDROLASE DAM (DAMMIN) MODELO SIMULADO E EXP É A CURVA EXPERIMENTAL (B) PARA DOMÍNIO SEMELHANTE A EXPANSINA DAM (DAMMIN) MODELO SIMULADO, PDB MODELO DE ALTA RESOLUÇÃO E EXP É A CURVA EXPERIMENTAL.	116
FIGURA 57 - SOBREPOSIÇÃO DO MODELO DE CORPO RÍGIDO OBTIDO COM O PROGRAMA MASHA COM O DAM PARA A GLICOSILHIDROLASE. A POSIÇÃO INICIAL, B ROTAÇÃO DE 90° EM RELAÇÃO AO EIXO X PARA A POSIÇÃO INICIAL E C VISÃO LATERAL EM RELAÇÃO AO PRIMEIRO.....	117
FIGURA 58 - SOBREPOSIÇÃO DO DAM DO DOMÍNIO EXPANSINA COM O PDB. A POSIÇÃO INICIAL, B ROTAÇÃO DE 45° EM RELAÇÃO AO EIXO Y COM RELAÇÃO A POSIÇÃO INICIAL E C ROTAÇÃO DE 45° EM RELAÇÃO AO EIXO X E ROTAÇÃO DE 90° EM RELAÇÃO A Z COM RELAÇÃO A POSIÇÃO INICIAL, TOMANDO COMO COORDENADAS O PLANO CARTESIANO XYZ.	118
FIGURA 59 – CRISTAIS DO DOMÍNIO SEMELHANTE À EXPANSINA.	120
FIGURA 60 – EVOLUÇÃO DAS GOTAS DE CRISTALIZAÇÃO(<i>DROP POSITION</i>) A- 19-DEC-2011 , B- 23- DEZ-2011, C- 03-JAN-2012, D-25 JANE-2012 E E-17 FEC-2012, PARA A CONDIÇÃO 0.1 M DE ADA PH 6.5 E 12 %v/v MPD ,COM A PROTEÍNA EM MG/ML. FZ026332.	121
FIGURA 61 - PADRÃO DE DIFRAÇÃO PARA OS CONJUNTOS DE DADOS COLETADOS: (A) IFSC RESOLUÇÃO MÁXIMA 3,05Å ,(B) E.S.R.F-1,RESOLUÇÃO MÁXIMA 2,75Å E (C) E.S.R.F-2, RESOLUÇÃO MÁXIMA 2,5 Å.	122

FIGURA 62 - (A) REPRESENTAÇÃO DA UNIDADE ASSIMÉTRICA COM AS QUATRO MOLÉCULAS. (B) REPRESENTAÇÃO DO EMPACOTAMENTO CRISTALOGRAFICO.	124
FIGURA 63 - DIAGRAMA DE RAMACHANDRAN PARA O DOMÍNIO SEMELHANTE À EXPANSINA PROCESSADO COM O CONJUNTO ESRF-1.	125
FIGURA 64 - DIAGRAMA DO MAPA TOPOLÓGICO DA PROTEÍNA SEMELHANTE À EXPANSINA, AS SETAS REPRESENTAM AS FITAS-B, E O CILINDRO EM VERMELHO AS A-HÉLICES.	126
FIGURA 65 – MODELO CRISTALOGRAFICO DA PROTEÍNA SEMELHANTE À EXPANSINA (CADEIA B) COM A IDENTIFICAÇÃO DAS PRINCIPAIS ESTRUTURAS SECUNDÁRIAS PRESENTES NA MESMA, EM VERMELHO AS A-HÉLICES, EM AMARELO A FOLHAS-B E EM VERDE AS DEMAIS ESTRUTURAS SECUNDÁRIAS.	126
FIGURA 66 -MODELO CRISTALOGRAFICO DA PROTEÍNA SEMELHANTE À EXPANSINA OBTIDO COM OS QUATRO MOTIVOS NA UNIDADE ASSIMÉTRICA.	127
FIGURA 67 – EM DESTAQUE OS PRINCIPAIS AMINOÁCIDOS ENVOLVIDOS NO PROCESSO DE RECONHECIMENTO A POLISSACARÍDEOS (TRP-124, TRP-125, TYR-156).	128
FIGURA 68 – SUEPRFICIE ELETROSTATICA POTENCIAL SUPERFICIAL DE CONTATO DA PROTEÍNA DE CONTATO EM DESTAQUE OS PRINCIPAIS AMINOÁCIDOS ENVOLVIDOS NO PROCESSO DE RECONHECIMENTO A POLISSACARÍDEOS (TRP-124, TRP-125, TYR-156).	128
FIGURA 69 – ALINHAMENTO DO MODELO ESTRUTURAL DA EXPANSINA DE <i>ZEAMAYS</i> (2HCZ) EM VERMELHO COM O MODELO ESTRUTURAL DA PROTEÍNA SEMELHANTE À EXPANSINA DE <i>XANTHOMONAS CAMPESTRIS</i> EM VERDE.	129
FIGURA 70 – ALINHAMENTO MÚLTIPLO UTILIZANDO O PROGRAMA CLUSTAL PARA AS SEQUÊNCIAS DA EXPANSINA DE <i>ZEAMAYS</i> (2HCZ) COM A SEQUÊNCIA DA PROTEÍNA SEMELHANTE À EXPANSINA DE <i>XANTHOMONAS CAMPESTRIS</i>	129
FIGURA 71 – ALINHAMENTO DO MODELO ESTRUTURAL DA EXPANSINA DE <i>ZEAMAYS</i> (2HCZ) EM VERMELHO COM O MODELO ESTRUTURAL DA PROTEÍNA SEMELHANTE À EXPANSINA DE <i>XANTHOMONAS CAMPESTRIS</i>	130
FIGURA 72 – ALINHAMENTO MÚLTIPLO UTILIZANDO O PROGRAMA CLUSTAL PARA AS SEQUÊNCIAS DA EXPANSINA DE <i>ZEAMAYS</i> (2HCZ) COM A SEQUÊNCIA DA PROTEÍNA SEMELHANTE À EXPANSINA DE <i>XANTHOMONAS CAMPESTRIS</i>	130

LISTA DE TABELAS

Tabela 1 -	Projeções teóricas para a produção de biocombustíveis de segunda geração no Brasil. Adaptado ³	30
Tabela 2 -	Custos por litro em USD em função dos componentes para produção do etanol de segunda geração. Adaptado ⁴	30
Tabela 3 -	Principais componentes de materiais lignocelulósicos. Adaptado ¹¹	32
Tabela 4 -	Oligonucleotídeos utilizados para amplificação das ORFs	57
Tabela 5 -	Reação de amplificação para cada um dos genes.	57
Tabela 6 -	Reagentes usados nas reações de digestão do vetor de clonagem e do gene de interesse.	59
Tabela 7 -	Reagentes usados na reações de digestão do sítio XhoI do vetor de clonagem.	59
Tabela 8 -	Reagentes usados na reações de digestão do sítio BamHI do vetor de clonagem.	60
Tabela 9 -	Padrões utilizados na calibração da Superdex 75 10/300 GL (GE Healthcare)..	66
Tabela 10 -	Relação tipo de radiação aplicada, comprimento de onda e grupo de absorção.....	68
Tabela 11 -	Porcentagem de aminoácidos na estrutura primária da Glicosilhidrolase de <i>Xanthomonas campestris</i>	90
Tabela 12 -	Porcentagem de aminoácidos na estrutura primária do sítio catalítico da Glicosilhidrolase de <i>Xanthomonas campestris</i>	90
Tabela 13 -	Porcentagem de aminoácidos na estrutura primária do domínio semelhante à Expansina de <i>Xanthomonas campestris</i>	91
Tabela 14 -	Estimativas de estrutura secundária para Glicosilhidrolase.	108
Tabela 15 -	Estimativas de estrutura secundária para o domínio semelhante à Expansina	108
Tabela 16 -	Estrutura secundária para o modelo cristalográfico (Modeller) para Glicosilhidrolase e para o modelo cristalográfico (difração de raios-X em monocristal) do domínio semelhante à Expansina	108

Tabela 17 -	Estrutura secundária para o modelo cristalográfico (Modeller) para Glicosilhidrolase e para o modelo cristalográfico (difração de raios-X em monocristal) do domínio semelhante à Expansina.	110
Tabela 18 -	Parâmetros estruturais da Expansina de <i>Xanthomonas campestris</i> obtidos por SAXS	116
Tabela 19 -	- Estatísticas de coleta dos dados de difração de raios-X, do domínio semelhante à Expansina..	123
Tabela 20 -	Análise do Coeficiente de Matthews	123
Tabela 21 -	Estatísticas finais para a estrutura domínio semelhante à Expansina.	123
Tabela 22 -	Estatísticas validação da estrutura-Programa Polygon.	125
Tabela 23 -	Estatísticas PROCHECK	125

LISTA DE ABREVIATURAS E SIGLAS

ATCC	<i>American Type Culture Collection</i>
ATP	Adenosina-5'-trifosfato
BLAST	<i>Basic Local Alignment Sequence Tool</i>
CAZy	<i>Carbohydrate-Active enZymes Database</i>
CD	<i>Circular Dichroism</i>
D	Coeficiente de difusão translacional
DLS	<i>Dynamic Light Scattering</i>
$D_{máx}$	Dimensão máxima de uma proteína
DDBJ	<i>DNA Data Bank of Japan</i>
EBI	<i>European Bioinformatics Institute</i>
E.S.R.F	<i>European Synchrotron Radiation Facility</i>
ExpASy	<i>Expert Protein Analysis System</i>
I	Intensidade de espalhamento de raios-X
IPTG	Isopropil-tio-B-D-galactopiranosídeo
LB	Luria-Bertani
LNLS	Laboratório Nacional de Luz Síncrotron
MM, M	Massa molecular
NCBI	<i>National Center for Biotechnology Information</i>
ORF	Fase aberta de leitura, do inglês Open Reading Frame
$p(r)$	Função de distribuição de distâncias de centros espalhadores de raios-X
PCR	Reação em cadeia da polimerase
PDB	<i>Protein Data Bank</i>
Pi	Fosfato inorgânico
pI	Ponto isoelétrico
q	Vetor de espalhamento de raios-X
R_g	Raio de giro
SAXS	Espalhamento de Raios-X a baixo ângulo, do inglês <i>Small angle X-ray scattering</i>
Tm	<i>Melting Temperature</i>
Xcc	<i>Xanthomonas campestris</i>

SUMÁRIO

1	Introdução	29
1.1	Perspectivas para a produção de Etanol de Segunda Geração	29
1.2	Etanol	31
1.3	A biomassa lignocelulósica.....	31
1.4	Celulases	35
1.5	Hemicelulases	37
1.6	Biodegradação da Lignina	37
1.7	CAZy.....	38
1.8	Expansinas	39
1.9	Swoleninas	42
1.10	Produção de Etanol de Segunda Geração	42
1.11	O organismo <i>Xanthomonas campestris</i>	44
1.12	Modelos estruturais de alta resolução em GH5 e Expansinas.	45
2	Objetivos específicos	47
3	Materiais e Métodos.....	49
3.1	Estratégias Computacionais.	49
3.1.1	Seleção de um gene ou proteína a partir de sequências preditas depositadas em banco de dados. 50	
3.1.2	Análise da Sequência de Aminoácidos	51
3.1.3	Modelagem por Homologia	52
3.2	Tecnologia do DNA recombinante	54
3.2.1	Extração de DNA genômico de <i>Xanthomonas campestris</i> ATCC33913	54
3.2.2	Clonagem da sequência codificadora de glicosilhidrolase Xcc e de seus domínios catalítico e Expansin-Like.....	55
3.2.3	Transformação do vetor de subclonagem em células <i>E. coli</i> por choque térmico e mudança dos genes clonados para vetor de expressão.....	58
3.2.4	Sequenciamento de DNA e análise de sequências.....	60
3.2.5	Linhagens de bactérias utilizadas para a expressão das proteínas.	61
3.2.6	Expressão e purificação das proteínas recombinantes	62
3.3	Caracterização biofísica	66
3.3.1	Estimativa da massa molecular.....	66
3.3.2	Espectroscopia de Dicroísmo Circular.....	67

3.3.3	Espalhamento Dinâmico de Luz (DLS)	69
3.3.4	Espalhamento de Raios-X a Baixo Ângulo (SAXS)	72
3.3.5	A estrutura cristalina de uma proteína	76
3.3.6	Difração de Raios-X aplicado a Cristalografia de Proteínas	78
4	Resultados e Discussão	82
4.1	Estratégias Computacionais	82
4.1.1	Predição de genes a partir de banco de dados.	82
4.1.2	Análise da Sequência de Aminoácidos	82
4.1.3	Modelagem por Homologia	91
4.2	Tecnologia do DNA recombinante	96
4.2.1	Extração de DNA genômico de <i>Xanthomonas campestris</i> PV <i>campestris</i> ATCC33913	96
4.2.2	Clonagem da sequência codificadora de Glicosilhidrolase XCC e de seus domínios catalítico e semelhante à Expasina	96
4.2.3	Transformação de células de <i>E. coli</i> por choque térmico	97
4.2.4	Sequenciamento de DNA e análise de seqüências	98
4.2.5	Linhagens de bactérias e plasmídeo utilizados para a expressão das proteínas	99
4.2.6	Expressão e purificação das proteínas recombinantes	99
4.3	Caracterização biofísica	105
4.3.1	Estimativa da massa molecular	105
4.3.2	Espectroscopia de Dicroísmo Circular	106
4.3.3	Espalhamento Dinâmico de Luz (DLS)	111
4.3.4	Espalhamento de Raios-X a Baixo Ângulo (SAXS)	113
4.3.5	Ensaio de Cristalização	119
4.3.6	Difração de Raios-X	122
5	Conclusões	131
6	Perspectivas	133
	REFERÊNCIAS	135

1 Introdução

1.1 Perspectivas para a produção de Etanol de Segunda Geração

O desenvolvimento de uma matriz energética de base sustentável possui uma importância estratégica para qualquer nação industrializada ou em busca disso¹. Dessa maneira, embora o etanol a partir da cana de açúcar (etanol de primeira geração) seja um setor muito bem desenvolvido no Brasil a busca de novas fontes de energia renovável como a utilização da biomassa lignocelulósica para produção de etanol (etanol de segunda geração) é de extrema importância.

Materiais lignocelulósicos provindos da agricultura, de resíduos industriais e florestais constituem as principais fontes deste tipo de material. Para iniciar a utilização destes como fontes para biocombustíveis é necessária a disponibilização e conversão dos componentes celulósicos em açúcares fermentáveis.

Uma variedade de microrganismos, como bactérias e fungos podem disponibilizar os agentes promotores da despolimerização da biomassa celulósica em monômeros de glicose, o que abre uma expectativa otimista para as pesquisas na área². Entre os potenciais recursos bioenergéticos alternativos, os materiais lignocelulósicos se destacam como principal fonte de biocombustíveis e outros produtos de valor agregado².

Considerando a matriz energética brasileira, temos que cerca de 45% da oferta total de energia primária são derivados de fontes renováveis: biomassa (31%) e centrais hidrelétricas (13,8 %). A maior parte, no entanto, está no petróleo bruto (40,3 %), gás natural (7,6 %) e carvão (5,8 %)³.

A produção do etanol de primeira geração permitiu ao Brasil até 2004 a posição de maior produtor de biocombustíveis do mundo, marca esta ultrapassada então pelos EUA. Todavia, o país é o maior exportador mundial de etanol combustível, totalizando em 2008 aproximadamente 5,1 bilhões de litros de uma produção mundial de cerca de 27 bilhões de litros, sendo que, os principais destinatários destas exportações foram os Estados Unidos e Europa, que juntos representaram um total de aproximadamente 4,1 bilhões de litros em 2008³.

Avaliando as perspectivas da produção do etanol de segunda geração pelo Brasil, um estudo feito pela Agência Internacional de Energia tomando os seguintes parâmetros: preço médio para cada tonelada de bagaço (50% de massa bruta) entre 3 e 8 USD, com preços máximos entre 13.7 e 31.12 USD para cada tonelada, distância média de transporte da colheita (bagaço) à transformação de 19 km e utilizando um modelo de cálculo desenvolvido por esta agência (IEA modelo de mobilidade), propôs os custos de produção de combustíveis de segunda geração apresentados na tabela 1 ³. Numa abordagem em nível mundial apresentamos a tabela 2 referente a custos por componente na produção do etanol de segunda geração⁴.

Tabela 1 – Projeções teóricas para a produção de biocombustíveis de segunda geração no Brasil, adaptado³.

Matéria prima	Preço da matéria prima USD/t _{FM}	Petróleo 60USD/bbl		Projeção a longo prazo (USD/lge)	
		Hoje (USD/lge)		BTL-diesel	LC-etanol
		BTL-diesel	LC-etanol	BTL-diesel	LC-etanol
Palha e folha de cana-de-açúcar	14-31	0,64-0,76	0,65-0,80	0,4-0,5	0,39-0,52
Bagaço	4-8	0,57-0,60	0,55-0,60	0,34-0,37	0,3-0,34
Resíduos de madeira	16-23	0,65-0,80	0,67-0,86	0,41-0,54	0,4-0,57
Petróleo 120USD/bbl					
Palha e folha de cana-de-açúcar	14-31	0,82-0,93	0,75-0,9	0,51-0,61	0,42-0,56
Bagaço	4-8	0,74-0,78	0,66-0,70	0,45-0,48	0,34-0,38
Resíduos de madeira	16-23	0,83-0,98	0,77-0,96	0,52-0,65	0,44-0,61

A tabela projeta duas situações uma em que o valor do barril de petróleo seja de 60 dólares e outra a 120 dólares e calcula a relação de custo. Dos termos utilizados na tabela temos: **t_{FM}** corresponde a matéria prima bruta total, **BTL** combustível sintético fabricado a base de biomassa **LC** lignocelulósico, **Lge** equivalente energético comparado a um litro de gasolina e **Bbl** (Oil Barril) barril de petróleo.

Tabela 2 – Custos por litro em USD em função dos componentes para produção do etanol de segunda geração, adaptado⁴.

Custo por componente (\$)	2001	2005	2012	2020
Material celulósico	0,21	0,21	0,13	0,09
Enzima	0,85	0,11	0,08	0,03
Conversão	0,44	0,31	0,21	0,15
Total de custos por litro	1,50	0,63	0,41	0,27
Total de custo por equivalente litro de gasolina *	2,46	1,03	0,68	0,50

*0,43L de etanol equivale 0,264L de gasolina, em termos de energia/calor.

A partir dos dados apresentados na tabela 1, observamos que os custos de BTL-Diesel e etanol lignocelulósico seriam atualmente entre 0,55 -0,86/lge USD³. Mesmo com as projeções de redução de custo em longo prazo teremos os combustíveis de primeira geração com custos inferiores, contudo os investimentos nesta fonte de energia são justificáveis devido à grande quantidade de biomassa lignocelulósica disponível para este processo, constituindo-se numa fonte alternativa de energia.

1.2 Etanol

A fermentação etanólica é o processo de maior utilização microbiana. A produção industrial de etanol utiliza o melaço de cana de açúcar, amido ou fécula que é hidrolisado enzimaticamente e fermentado em batelada pela levedura *Saccharomyces cerevisiae* para produzir etanol. São derivados deste processo gás carbônico e baixas concentrações de metanol, glicerol, etc. Etanol não necessita de alta purificação para ser utilizado como combustível⁵.

O Brasil possui tecnologia para o motor a combustão com etanol, e foi o pioneiro na sua utilização em larga escala como combustível automotivo. Na década de 1980, cerca de 95% dos veículos produzidos no Brasil eram equipados com motores a combustão a álcool. Atualmente, a maior parte do etanol produzido no Brasil é utilizado como aditivo à gasolina, com uma porcentagem de 24-26%. Entretanto, o uso do álcool hidratado como combustível automotivo está crescendo atualmente devido à crescente popularidade dos veículos que podem utilizar o álcool quanto a gasolina ou a mistura em diferentes proporções (tecnologia Flex de combustível)⁶.

1.3 A biomassa lignocelulósica

A principal fonte de matéria prima vegetal para combustíveis é a parede celular vegetal. Microrganismos anualmente hidrolisam 10^{11} toneladas de biomassa vegetal, principalmente a parede celular, que contém o equivalente energético de 640 bilhões barris de petróleo⁷.

A biomassa lignocelulósica é constituída por celulose, hemicelulose, lignina, pectinas e outros componentes (Tabela 3)⁸.

A combinação da celulose e da hemicelulose é chamada de holocelulose e sua porcentagem em massa no material lignocelulósico seco representa de 65 a 70 %^{9; 10}.

Tabela 3 – Principais componentes de materiais lignocelulósicos. Adaptado ¹¹.

Constituinte	Angiospermas (%)	Gimnospermas (%)
Celulose	40 a 50	40 a 50
Hemicelulose	25 a 35	25 a 30
Lignina	20 a 25	25 a 35
Pectina	1 a 2	1 a 2
Amido	Traços	Traços

A celulose é o principal componente da biomassa lignocelulósica, este homopolissacarídeo é constituído por unidades repetitivas de D-glicopirranose que estão unidas por ligações β -(1,4) glicosídicas (Figura 1), esta unidade repetitiva formada por dois açúcares é nomeada celobiose.^{9; 10}

FIGURA 1 – Estrutura química da celobiose ¹².

O grau de polimerização “degree of polymerization” (DP) consiste no número médio de unidades de açúcar presentes na celulose, experimentos de espalhamento de luz indicam que a celulose de madeira tem em média um DP de 9.000–10.000 com maior grau de polimerização de 15.000. Em média um DP de 10.000 pode corresponder a um comprimento linear de aproximadamente 5 μ m na madeira ¹⁰.

A organização das moléculas de celulose ocorre de maneira aleatória com a formação de ligações de hidrogênio de maneira intra- e intermolecular, tendo como efeito o aumento da densidade de empacotamento, levando à formação das regiões cristalinas. As regiões com baixo empacotamento são nomeadas de amorfa ou não cristalina¹⁰.

A polimerização e cristalinidade da celulose dependem da espécie vegetal originária da mesma, estes fatores irão interferir diretamente no processo hidrolítico (ácido ou enzimático)⁸.

A acessibilidade e não acessibilidade referem-se à disponibilidade da celulose para a água, microrganismos, etc. A superfície cristalina da celulose será acessível mas o restante da estrutura cristalina não será. A maior parte da celulose não cristalina está acessível, mas parte da mesma está coberta com hemicelulose e lignina que a torna inacessível. A acessibilidade ou não à celulose é importante uma vez que serão fatores determinantes no processo de desfibrilação, modificação química, extração e interação com microrganismos¹⁰.

A hemicelulose é o componente mais simples do material lignocelulósico, sua concentração está entre 25 e 35% sendo que a mesma pode ser facilmente hidrolisada para açúcares fermentáveis⁸. A hemicelulose é um sacarídeo do tipo heteropolímero, formado por pentoses (D-xilose e D-arabinose), hexose (D-manose, D-glucose e D-galactose) e açúcares ácidos.¹³ Hemicelulose provinda de gimnospermas possui como principal componente a manose enquanto que a provinda de angiospermas possui majoritariamente xilanos⁸.

A lignina permite às plantas um suporte estrutural e impermeabilizante que proporciona resistência ao ataque de microrganismos e ao estresse oxidativo. Estas propriedades da lignina podem ser atribuídas a sua natureza amorfa, insolubilidade em água e inatividade óptica⁸. A lignina é um complexo polimérico de fenil-propanóides (p-cumaril, coniferil e álcool sinapil)⁸ figura 2, que estão associados à hemicelulose formando um complexo carboidrato-lignina que é resistente à hidrólise mesmo sob condições de desfibrilação. Não há evidências de que a lignina apresente associação à celulose¹⁰.

FIGURA 2 – Estrutura química dos precursores da lignina: (1) álcool p-cumaril , (2) álcool coniferil , e (3) álcool sinapil.

O precursor menos presente na lignina de gimnospermas ou angiospermas é o álcool p-cumaril, o álcool coniferil é o precursor predominante para a lignina de gimnospermas e para a lignina de angiospermas tanto o álcool coniferil quanto o álcool sinapil são importantes em igual quantidade¹⁴.

Quanto aos componentes de menor quantidade presentes na biomassa lignocelulósica de angiospermas e gimnospermas podemos citar: pectinas, amidos e proteínas. A pectina é um polímero apolissacarídeo formado pelas unidades de repetição de D-galacturônico com ligações glicosídicas α -(1 $\rightarrow$ 4). As pectinas são encontradas em maior concentração nas células da parede do parênquima onde podem ser encontradas como aglutinante. São encontradas pequenas frações de L-Arabinofuranose and D-galactopiranosose em pectinas¹⁰. O principal polissacarídeo de reserva em vegetais é o amido, pequenas quantidades dele podem ser encontradas nos vasos da parede celular. Ocorrem como grânulos e são compostos por D-glucopiranosose unidas por ligações α -(1 $\rightarrow$ 4) (amilose) ou α -(1 $\rightarrow$ 4) com ramificações sobre cada unidade glucopiranosil por ligação α -(1 $\rightarrow$ 6) (amilopectina). Amilose ocorre como uma hélice no estado sólido devido à configuração alfa em polímeros e a amilopectina é altamente ramificada¹⁰.

O pré-tratamento de qualquer biomassa lignocelulósica (figura 3) antes da hidrólise enzimática é fundamental. Sendo que, o principal objetivo deste processo é diminuir o grau de cristalinidade da celulose o que aumenta a hidrólise da celulose por celulasas. Alguns métodos de pré-tratamento do material lignocelulósico são apresentados no quadro 1.¹³

FIGURA 3 – Material lignocelulósico, adaptado¹⁵.

QUADRO 1 – Métodos de pré-tratamento de biomassa lignocelulósica. Adaptado¹³.

Método	Exemplo
Auto-hidrólise	Água quente, pressão de vapor, explosões a vapor, explosão super-crítica na presença de CO ₂
Tratamento ácido	Ácido (H ₂ SO ₄) diluído ou concentrado.
Tratamento alcalino	Hidróxido de sódio, cal, amônia e peróxido de hidrogênio alcalino.
Solvente orgânico com água	Metanol, etanol, butanol e fenol

1.4 Celulases

Para que a celulose seja hidrolisada e metabolizada, celulases extracelulares devem ser produzidas e estas podem estar livres ou associadas à celulose¹⁶. Estas enzimas do complexo celulolítico pertencem às O-glicosídeo hidrolases (EC 3.2.1.-) cuja ação é a hidrólise das ligações glicosídicas entre dois ou mais carboidratos e entre carboidratos e não carboidratos. A nomenclatura das enzimas adotada pela IUBMB é baseada no tipo de reação e na especificidade ao substrato¹⁶. Os componentes do sistema de celulases foram classificados primeiramente de acordo com o modo de ação e recentemente baseado nas propriedades estruturais. Estas enzimas estão separadas em três grandes grupos: endoglucanases; exoglucanases e β -glicosidases¹⁶.

Endoglucanases (EnG) ou 1,4- β -D-glucano-4-glucanohidrolase (EC 3.2.1.4) atuam internamente e de maneira aleatória em regiões amorfas da celulose, gerando oligossacarídeos de vários graus de polimerização (DP) e consequentemente novos sítios na extremidade das cadeias¹⁶, sendo um redutor (quando a glicose possui uma hidroxila heterosídica livre) e um não redutor (quando a hidroxila heterosídica da molécula da extremidade participa de ligação com a glicose adjacente)¹⁷.

Exoglucanases (ExG) incluem 1,4- β -D-glucano glucanohidrolase (celodextrinas) (GH) (EC 3.2.1.74) e 1,4- β -D-glucano celobiohidrolases (celobiohidrolases) (CBH) (EC 3.2.1.91) e atuam de forma processiva em extremidades reduzidas e não reduzidas, liberando glicose (GH) ou celobiose (CBH) como principais produtos. As exoglucanases podem também atuar sobre a celulose microcristalina causando desestratificação de sua estrutura¹⁶, esta mudança promove aumentos na taxa de hidrólise da celulose, por tornar amorfas as regiões cristalinas do polímero, deixando-o mais exposto às celulases¹⁷. Quanto ao mecanismo de ação as CBH podem ser divididas em dois grupos: CBHI enzimas que hidrolisam terminais redutores (R) e CBHII enzimas que não hidrolisam terminais redutores (NR), um terminal redutor contém

uma hidroxila livre ligada ao carbono um, estas enzimas sofrem inibição pelo próprio produto de hidrólise (celobiose)¹⁷.

A β -glicosidase (BG), ou β -glicosídeo gluco-hidrolase (EC 3.2.1.21) possui a capacidade de hidrolisar celodextrinas e celobioses em glicose¹⁶.

Uma característica comum da maioria das celulases é a presença de um módulo estrutural que inclui a região catalítica e um outro módulo responsável pela ligação a carboidratos (CBMs). O CBM liga-se à superfície da celulose, facilitando a hidrólise e trazendo o sítio catalítico para próximo ao substrato. A presença dos CBMs é particularmente importante para iniciar a processividade em exoglucanases¹⁶.

As celulases possuem acentuada atividade coletiva, tanto que a soma das atividades individuais das enzimas é inferior que a ação conjunta das mesmas, este fenômeno é conhecido por sinergismo¹⁶. A Figura 4 ilustra a ação sinérgica entre exoglucanases, endoglucanase e β -glicosidase na hidrólise da fibra celulósica e o quadro 2 indica a ação das diferentes celulases em diferentes substratos.

FIGURA 4 – Modo de ação das enzimas do complexo celulolítico. Adaptado¹⁸.

QUADRO 2 – Ação de celulase sobre diferentes substratos. Adpatado ¹⁹.

Enzima	Celulose cristalina	Celulose amorfa	CM-celulose	Celoligo-sacarídeos	Celobiose
Exoglucanase	Baixa	Muito ativa	Nula	Ativa	Nula
Endoglucanase	Nula	Muito ativa	Muito ativa	Ativa	Nula
β -glicosidase	Nula	Nula	Nula	Ativa	Ativa

1.5 Hemicelulases

O mecanismo de ataque enzimático em hemiceluloses torna-se complexo devido à sua característica heteropolissacarídica²⁰. Uma das principais hemiceluloses encontradas são as xilanas¹⁰. Para a completa hidrólise desta macromolécula várias enzimas estão envolvidas sendo a principal a endo-1,4- β -D-Xilanase ou endoxilanase (E.C 3.2.1.8), cuja ação está na clivagem aleatória do esqueleto de arabinoxilana produzindo oligossacarídeos de xilose²¹. Para a completa hidrólise são necessárias enzimas desramificadoras dos grupos laterais ligados à cadeia principal de xilana: β -1,4-Xilosidase (E.C 3.2.1.37), α -D-Glicuronidase (E.C 3.2.1.139), β -L-Arabinofuranosidase (E.C 3.2.1.55) e Acetilxilana-esterase (E.C 3.1.1.72). A figura 5 ilustra a ação sinérgica de enzimas sobre a hemicelulose ²⁰.

FIGURA 5 – A hemicelulose mais prevalente é o xilano, as enzimas xilanolíticas envolvidas em sua degradação estão indicadas na figura, adaptado ²².

1.6 Biodegradação da Lignina

Um complexo sistema enzimático extracelular de baixa especificidade está envolvido na biodegradação da lignina, produzidos por fungos que vivem na madeira, entre os quais *Phanaerochatae chrysosporium* e entre as bactérias que são capazes de degradar a lignina temos os actinomicetos. Entre as enzimas envolvidas nesta degradação temos lignina peroxidase, manganês peroxidase e lactase²⁰.

1.7 CAZy

O banco de dados *Carbohydrate-Active Enzyme* (Cazy) descreve as famílias de proteínas estruturalmente relacionadas à atividade catalítica ou de ligação a carboidratos que sejam capazes de degradar, modificar ou criar ligações glicosídicas²³. Este banco de dados utiliza a sequência de aminoácidos para correlacionar os mecanismos de ação e enovelamentos protéicos em famílias distintas.

As famílias protéicas do CAZy são normalmente designadas através de um formulário simples, incluindo a classe ou categoria do módulo e um número que reflete a ordem da criação familiar dentro do mesmo grupo. O CAZy abrange os seguintes grupos de proteínas: Glicosilhidrolase (GH), Glicosiltransferase (GT), Liasas polissacarídeas (PL), Esterase carboidrato (CE) e Módulo de ligação a carboidrato (CBM)²³. Destes grupos destacaremos dois o das GH e o CBM.

O grupo das GH apresenta mais de 120 famílias, nomeadas respeitando a ordem cronológica de identificação das mesmas, dentre as quais temos a família 5 (GH5). A GH5 é uma das maiores famílias de glicosilhidrolases e encontra-se amplamente distribuída em bactérias e eucariotas (especialmente em fungos e plantas), apresentando como especificidade já descrita de: endoglucanase, exoglucanases, β -glicosidase, 1,6-galactanase, 1,3-manase, 1,4-xilanase e endoglicoceramidase. O mecanismo de ação destas enzimas usa o clássico modelo de Koshland de duplo deslocamento sendo que os resíduos catalíticos do processo são dois glutamatos, estes dois aminoácidos são a marca desta família de glicosilhidrolases.

Os CBM são sequências de aminoácidos que possuem a capacidade de ligar-se a carboidratos, não apresentam atividade enzimática, por si só, mas são potencializadores da mesma prolongando a interação com substratos²³. Os módulos de ligação a carboidrato na maioria das vezes constituem-se num domínio estrutural e funcional de proteína com atividade catalítica. No entanto, ocasionalmente, podem estar presentes em formas isoladas ou tandem não acoplados com uma enzima. Cerca de 7% das entradas Cazy contém pelo menos um módulo CBM na cadeia peptídica²³. Os CBMs estão classificados em 64 famílias, sendo que a família 63 descreve a proteína semelhante à expansina de *Bacillus subtilis* EXLX1²⁴.

1.8 Expansinas

Expansinas são proteínas que foram descritas primeiramente em vegetais como responsáveis pelo afrouxamento da rede celulósica da parede celular durante o crescimento²⁵. Estudos subsequentes indicaram duas atividades físicas evidentes decorrentes da ação desta proteína: Extensão irreversível da parede celular e estimulação do seu relaxamento das paredes (“stress relaxation”) durante um largo intervalo de tempo ($< 0,1$ a > 100 s). Estas atividades são dependentes do pH e nenhuma atividade enzimática foi detectada para as Expansinas sendo seu mecanismo de ação não elucidado até o momento²⁵.

No entanto, a maioria das evidências sugere uma ação não-hidrolítica das Expansinas: ampliando as cavidades da parede celular a partir das ligações de polissacarídeos e perturbando ligações não-covalentes de celulose na parede celular. Apesar da falta de atividade hidrolítica de Expansinas, alguns estudos têm mostrado que elas melhoraram a hidrólise enzimática da celulose cristalina por celulasas. Neste modelo, acredita-se que elas agem como um zíper na abertura da reticulação celulósica levando a um descolamento das microfibrilas, que por sua vez, melhoram a acessibilidade da celulose, acelerando a ação de celulasas. Na figura 6 temos um modelo de como seria esse processo²⁶.

FIGURA 6 – Modelo de ação das expansinas no afrouxamento das fibras celulósicas. As microfibrilas de celulose estão ligadas entre si por glicanas (estruturas em amarelo e vermelho) que podem pregar as microfibrilas da superfície umas as outras. Acredita-se que a proteína expansina (azul) (a) perturba a interação das glicanas às microfibrilas da superfície (b) ou entre si (c) sob o estresse mecânico decorrente do turgor, a expansina pode deslocar-se sobre a parede do polímero e interferir nos pontos de aderência. Adaptado²⁶.

Tratando-se da nomenclatura destacaremos alguns pontos, o termo Expansinas discutido até o momento refere-se à superfamília Expansina, que por sua vez é dividida em famílias. Entre essas famílias temos a das alfa Expansinas, designadas EXPA, e as beta

expansinas, designadas EXPB, e a das proteínas semelhantes à expansinas, que por sua vez podem ser divididas nas famílias de proteínas semelhante à expansina do tipo alfa, designada EXLA (EXPANSIN-LIKE A) e proteína semelhante à expansina do tipo beta, designada EXLB (EXPANSIN-LIKE B), ou ainda proteína semelhante à expansina do tipo X, designada EXLX (EXPANSIN-LIKE X) quando a proteína ocorre em organismos não vegetais, o produto protéico tem homologia com os domínios de expansinas.

Desta maneira, as expansinas ou proteínas semelhantes à Expansinas são divididas principalmente em duas famílias: alfa e beta. As alfa-expansinas podem ser diferenciadas das demais pela presença de uma grande inserção e uma deleção no domínio 1, as proteínas semelhantes a alfa-proteínas por sua vez apresentam um exclusivo motivo de conservação (CDRC) na extremidade amino-terminal do domínio 1 e seu domínio 2 apresenta uma extensão de aproximadamente 17 aminoácidos não encontrados em outras famílias de Expansinas²⁷. Já as beta-Expansina ou proteínas semelhantes à beta-Expansinas são aquelas relacionadas filogeneticamente com o grupo pólen alergênico de gramíneas. Todos os membros da superfamília Expansina possuem dois domínios: um N-terminal que é relacionado distantemente ao domínio catalítico de glicosilhidrolases da família 45 (GH45) e um C-terminal que está relacionado distantemente ao domínio pólen alergênico em gramíneas^{27; 28}, estas análises podem ser observadas na figura 7.

FIGURA 7 – Regiões conservadas na sequência gênica da superfamília de expansinas e a região amino terminal de baixa conservação foram removidas para o alinhamento, porque algumas expansinas possuem grandes sequência peptídeo sinal e amino terminal. Resíduos conservados entre as famílias estão sombreadas, e a fronteira entre os dois domínios é indicado por setas. Resíduos altamente conservados que fazem parte do sítio catalítico de proteínas GH45 e que são conservadas no domínio 1 de alguns famílias de expansinas são mostrados em círculos acima do alinhamento. As famílias alinhadas foram: alfa expansina, designada EXPA, beta expansina, designada EXPB e as proteínas semelhantes a expansinas, proteína semelhante a expansina do tipo alfa, designada EXLA (EXPANSIN-LIKE A) e proteína semelhante a expansina do tipo beta, designada EXLB (EXPANSIN-LIKE B), adaptado²⁷.

1.9 Swolleninas

A proteína swollenina foi isolada pela primeira vez e caracterizada em fungos celulolíticos saprofíticos *Trichoderma reesei*. Ela recebeu esta nomeação Swollenina devido à sua capacidade para inchar (“swell” em inglês) as fibras do algodoeiro (*Gossypium hirsutum*) sem produzir quantidades detectáveis de açúcares redutores²⁹. A proteína tem na parte N-terminal uma região de ligação a carboidratos, módulo 1 domínio (CBD) com função de ligação a celulose, e ligado a este domínio covalentemente na região C-terminal o domínio 2 uma região com homologia a proteínas semelhantes a Expansinas uma vez que a mesma possui identidade com o domínio pólen alergênico de Expansinas²⁹. Isto sugere que as Swolleninas provavelmente desempenham um papel semelhante ao das Expansinas na desruptura da rede celulósica.

1.10 Produção de Etanol de Segunda Geração

A produção de etanol a partir de material lignocelulósico compreende os seguintes passos: hidrólise da hemicelulose, hidrólise da celulose, fermentação, separação do resíduo lignina, recuperação e concentração do etanol e manuseio de águas residuais³⁰. A figura 8 mostra de maneira simplificada este processo de produção de etanol a partir de material lignocelulósico baseado na hidrólise enzimática.

FIGURA 8 – Esquema do processo produtivo do etanol de segunda geração.

Alguns dos fatores mais importantes para a redução de custos na produção do etanol estão na obtenção da maior quantidade possível de etanol a partir de um material celulósico, ou seja, alta produtividade, integração do processo para redução de custos e demanda energética. Parte da lignina pode ser utilizada na geração de calor e eletricidade para o processo, são necessários investimentos na redução energética do processo de produção de etanol³⁰.

Os principais passos deste processo são: hidrólise da hemicelulose e celulose em açúcares fermentáveis e a fermentação de todos os açúcares³⁰. Para que tenhamos um rendimento enzimático maior alguns pontos devem ser atingidos: Maior estabilidade térmica: aumentando-se a temperatura em que estas enzimas podem funcionar, melhoramos drasticamente a taxa de hidrólise celulósica; Melhorar a interação domínio de ligação a celulose: Uma melhor interação do domínio de ligação à celulose conduzirá a uma interação mais eficaz entre as enzimas celulasas e a superfície insolúvel da biomassa; Melhor interação sítio ativo ao substrato: modificações nas sequências de aminoácidos do sítio ativo que tragam ganhos podem levar a ganhos radicais na atividade catalítica da enzima³¹; Redução de agentes não vinculados ao processo. Enzimas que são adsorvidas pela lignina tornam-se disponíveis para hidrólise, desta maneira modificações genéticas da enzima para ajustar sua carga de superfície podem minimizar essa interação indesejada³¹.

O processo enzimático mostra-se como o mais atrativo, contudo a conversão da celulose em açúcares é muito lenta a menos que a biomassa tenha passado por algum processo de pré-tratamento, caso este tenha sido feito de maneira eficiente traz um grande número de vantagens: alta recuperação de todos os carboidratos, maior degradação da celulose e, por conseguinte, maior hidrólise enzimática, leva uma baixa demanda energética e possui um baixo capital de investimento operacional³⁰.

O desenvolvimento de organismos fermentadores capazes de atuar no processo fermentativo em temperaturas superiores a 50 °C pode reduzir o custo das celulasas a metade. Isto porque os organismos fermentadores possuem desempenho desejável em temperaturas entre 30-33 °C. Utilizando celulasas de *T. reesei*, conseguiu-se uma duplicação na taxa de hidrólise da biomassa para cada aumento de 20 °C na temperatura de sacarificação, logo a associação de celulasas em temperaturas ótimas de atividade e organismos conversores de açúcares fermentais em etanol em temperaturas de atividade ideais trará redução nos custos deste processo³¹.

Dessa maneira, os fatores apresentados somam-se na redução de custos e aplicabilidade da geração do etanol provindo da fermentação de açúcares oriundos de materiais celulósicos.

1.11 O organismo *Xanthomonas campestris*

Xanthomonas campestris pv. *Campestris* pertence à família *Xanthomonadaceae* que é uma família de bactérias *gram* negativas pertencentes à ordem *Xanthomonadales*. Eles são tipicamente caracterizados como organismos ambientais e são encontrados no solo e água, bem como em tecidos vegetais. Muitas *Xanthomonadaceae*, principalmente as espécies dos gêneros *Xanthomonas* e *Xylella*, causam doenças em plantas e apenas uma, a *Stenotrophomonas maltophilia*, é conhecida por ser um patógeno humano³².

Este organismo é responsável pela doença podridão negra, que afeta crucíferas (figura 9), como *Brassica* e *Arabidopsis*. A bactéria coloniza os vasos do xilema da planta, escurecendo-os. Com o desenvolvimento da doença, as folhas tornam-se amarelas e podem apresentar necrose. Em alguns casos, observa-se subdesenvolvimento, murcha, queda prematura de folhas e apodrecimento das plantas afetadas. Temperaturas entre 28 e 30°C e presença de água de irrigação, de chuva ou de condensação são favoráveis para a penetração da bactéria³³. A penetração ocorre através de aberturas naturais (estômatos e hidatódios) ou por ferimentos provocados na superfície da parte aérea³³. Esta bactéria é cultivada comercialmente para produzir a goma xantana exopolissacarídeos, que é usada para controlar a viscosidade e como um agente estabilizador em muitos setores³⁴.

FIGURA 9 – Sintomas da podridão negra em folhas de repolho ³⁵.

Xanthomonas campestris pv. *campestris* (cepa ATCC33913) tem um único cromossomo circular compreendendo: 5.076.187 pb com 4181 códigos gênicos para proteína³². A proteína de estudo desta dissertação foi identificada a partir de uma busca no NCBI tendo como banco de dados o genoma da cepa ATCC33913 e como modelo para esta busca a sequência BSEXLX1 (PDBid2bh0)²⁸. Como resultado para esta busca obtivemos várias sequências de celulosas entre as quais a glicosilhidrolase de *Xanthomonas campestris* (XCC 3535), selecionada devido: a disponibilidade do organismo em nosso laboratório de pesquisa, importância econômica deste organismo e presença de dois domínios identificados para esta proteína: domínio catalítico de endoglucanase e domínio semelhante ao de expansinas. Dessa maneira, o entendimento do possível sinergismo conferido por estes domínios para esta proteína motivou este trabalho.

1.12 Modelos estruturais de alta resolução em GH5 e Expansinas.

Para as enzimas da GH5 temos um grande número de modelos estruturais de alta resolução disponíveis, sendo que, o primeiro modelo (figura 10) proposto foi para a endoglucanase CelC de *Clostridium thermocellum* em 1995. Os membros da GH5 formam um Clan (grupo de família de proteínas que apresentam significativa similaridade na sua estrutura terciária) em que possuem o enovelamento (α/β)₈ TIM barril com dois resíduos de ácido glutâmico essenciais no processo de atividade, estando separados um do outro em aproximadamente 200 resíduos.

As expansinas por sua vez, possuem um número reduzido de estruturas disponíveis sendo: para a EXLX1 de *Bacillus subtilis* e para a expansina de *Zea mays* (figura 11). Estruturalmente estas proteínas possuem dois domínios um com identidade a GH45 e o outro com identidade a pólen alergênico, sendo o motivo estrutural reconhecido como de ligação a carboidratos.

FIGURA 10 – Modelo estrutural de alta resolução para a endoglucanase de *Clostridium thermocellum* (1CEC).

FIGURA 11 – Modelo estrutural de alta resolução para a expansina de *Zea mays* (2HCZ).

2 Objetivos específicos

- Análises de bioinformática: seleção do gene de interesse dentro do contexto de hidrólise e disponibilização de polissacarídeos para fermentação bem como análise de domínios conservados.
- Clonagem do gene: extração do DNA genômico, amplificação por PCR, clonagem em vetor de propagação, clonagem em vetor de expressão e sequenciamento.
- Purificação das proteínas: superexpressão em *Escherichia coli*, purificação por métodos cromatográficos.
- Caracterização estrutural: espalhamento dinâmico de luz, CD, espalhamento de raios-X a baixo ângulo (SAXS), cristalização e estudos de difração de raios-X.

3 Materiais e Métodos

3.1 Estratégias Computacionais.

Nas últimas duas décadas observamos um crescimento acelerado da genômica e proteômica, isto se deve aos recentes avanços no processo de sequenciamento gênico em larga escala disponibilizando uma grande quantidade de informação para ser estudada, ferramentas computacionais e capacidade computacional de processamento de dados e também como consequência da difusão de informação permitida com World Wide Web ("web") que permitiu um acesso rápido e “universal” ao conhecimento gerado³⁶.

Como consequência deste processo temos o crescimento das comunidades de investigação virtuais organizadas em bases especializadas e grupos de pesquisa que disponibilizam ferramentas, bases de dados, descritores científicos, atualizações em áreas específicas do conhecimento com a finalidade de geração de conhecimento.

Sendo exemplo de organizações que prestam tais serviços: NCBI com sede nos Estados Unidos (<http://www.ncbi.nlm.nih.gov/>), EBI organização européia (<http://www.ebi.ac.uk/>) e o DDBJ do Japão (<http://www.ddbj.nig.ac.jp/>), estes três grandes sites apesar de possuírem ferramentas para proteômica, filogenia e outras áreas de estudos bioquímicos possuem como foco principal a genômica³⁶. A disponibilização de sequências protéicas e ferramentas para o seu estudo têm como principais bases de dados ExPASy Molecular Biology Server (SWISS-PROT) (<http://expasy.ch/>), Protein Information Resource (PIR) (<http://pir.georgetown.edu/>), Protein Research Foundation (PRF) (<http://www.prf.or.jp/index-e.html>) e o UniProt (<http://www.uniprot.org/>).

3.1.1 Seleção de um gene ou proteína a partir de sequências preditas depositadas em banco de dados.

A seleção de um gene ou proteína a partir de banco de dados tem como princípio o alinhamento de sequências, com isso podemos selecionar um alvo para estudo. Entre os pacotes de programas de domínio público que permitem o estudo de sequências de polipeptídios ou polinucleotídeos temos o BLAST que realiza alinhamento de sequências aos pares e o CLUSTAL que é capaz de realizar o alinhamento de múltiplas sequências.

O pacote BLAST utiliza um algoritmo desenvolvido para alinhamento de sequências de nucleotídeos ou aminoácidos de acordo com a similaridade existente entre a sequência alvo (query) e as sequências depositadas no banco de dados (subject)³⁷, cujo número máximo pode ser definido pelo usuário. Ao final de cada alinhamento, o BLAST gera uma lista de sequências similares à sequência alvo e classifica cada alinhamento individual em termos estatísticos utilizando dois valores: o “escore” e o “E-value”³⁷.

O escore é um valor que indica o número de semelhanças entre duas sequências. O “E-value” é um valor estatístico que indica a expectativa de que outra sequência com escore idêntico ou maior ao valor obtido em um determinado alinhamento seja encontrada dentro do banco de dados analisado³⁷, isto significa que se em um alinhamento com um valor E igual a 5 é insignificante em termos estatísticos, enquanto que um alinhamento com E igual a 0,01 é significativo e para um E igual a 1×10^{-20} determina uma confiança tão elevada que as sequências em comparação são homólogas³⁸.

O programa CLUSTAL³⁹ consegue realizar alinhamento de múltiplas sequências. A entrada para este programa (input) é um arquivo contendo todas as sequências de peptídeos ou de DNA a serem alinhadas^{38; 39}. O CLUSTAL num primeiro momento encontra alinhamentos aos pares de todas as sequências introduzidas, permitindo com isso que o programa determine as relações das sequências introduzidas com base em seus escores de similaridade, de modo a produzir uma árvore filogenética grosseira, chamada dendograma³⁸. Depois, iniciando com o alinhamento de pares mais elevados, em ordem crescente de parentesco com as sequências adicionadas, neste processo o programa introduz lacunas nas sequências, conforme a necessidade para o alinhamento^{38; 40}.

Utilizando dos programas descritos identificamos o gene para a Glicosilhidrolase de *Xanthomonas campestris* XCC3535 para estudo desta dissertação.

3.1.2 Análise da Sequência de Aminoácidos

A partir da identificação da sequência gênica de trabalho, de código XCC3535 iniciamos o seu estudo. O primeiro passo foi conhecer sua classificação predita no banco de dados de celulasas utilizando o website CAZy (<http://www.cazy.org/>)²³.

O banco de dados do CAZy descreve as famílias de proteínas relacionadas com catálise (enzimas) e ligação (ou domínios funcionais) a carboidratos que sejam capazes de degradar, modificar ou criar ligações glicosídicas.

Estas proteínas foram divididas nas seguintes famílias: Glicosidehidrolases (GHs) (ou Glicosidases e Transglicosidases) responsáveis pela hidrólise e rearranjo das ligações glicosídicas, Glicosiltransferases (GTs) responsáveis pela formação de ligações glicosídicas, Polissacarídeo Liases (PLs) responsáveis pela clivagem de ligações glicosídicas de maneira não hidrolíticas, Carboidrato Esterase (CEs) responsáveis pela hidrólise das ligações ésteres em carboidratos e Módulos de Ligação a Carboidrato (CBMs) que aderem-se a carboidratos²³.

Um aspecto original do CAZy é sua tentativa de abranger todas as proteínas que atuam sobre carboidratos de todos os organismos e divididas de acordo com sua função. De acordo com o CAZy, a XCC3535 trata-se de uma proteína da família GH5 e família 53 de módulos de ligação a carboidratos⁴¹.

A representação de uma proteína como uma sequência de letras representativa de seus aminoácidos (estrutura primária) possibilita o estudo de propriedades pós-traducionais utilizando de ferramentas específicas, um exemplo disso é o estudo de regiões peptídeo sinal.

Dessa maneira, utilizamos a ferramenta SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) para avaliar a predição de sequência peptídeo sinal, que discute a probabilidade de sequências peptídeo sinal para procariontes *gram* positivos e *gram* negativos e para eucariontes. Assim, ao utilizar a ferramenta de predição temos que lembrar que, *Xanthomonas campestris* é uma bactéria *gram* negativa. O arquivo de saída desta ferramenta lança os seguintes *Scores* avaliando a sequência de resíduos: *Score S* AA pertencer a sequência sinal, *Score C* probabilidade de ser um sítio de clivagem, *Score Y* deriva da combinação de S e C⁴².

O uso de ferramentas de predição de domínios (parte da cadeia polipeptídica que pode se enovelar independentemente para formar uma estrutura compacta e estável⁴³) permite uma pré-caracterização das proteínas. Entre as ferramentas disponíveis para tal procedimento usou-

se o *Dommaing Fish*⁴⁴ (http://bmm.cancerresearchuk.org/~3djigsaw/dom_fish) e o CDD^{45; 46} (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>). Estes geram uma lista de domínios conservados reconhecidos e atribuem a cada um deles um escore e um valor de E^{44} , o que permitiu a identificação de dois domínios um envolvido na atividade catalítica com identidade a Glicosilhidrolases e outro de proteína semelhante a Expansina.

O próximo passo foi o cálculo teórico dos parâmetros físico-químicos da proteína e de seus domínios identificados, utilizando a ferramenta *ProtParam*⁴⁷ do servidor *ExPASy* (<http://www.expasy.org/proteomics>), a partir de sua sequência de aminoácidos.

Entre os parâmetros computados pelo *ProtParam* estão a massa molecular, o pI teórico, composição por aminoácido e o coeficiente de extinção molar. O pI de uma proteína corresponde ao valor de pH em que a mesma se apresente eletricamente neutra, isto é, quando o número de cargas positivas for igual ao número de cargas negativas, neste pH a molécula não migra quando submetida a eletroforese (diferença de campo elétrico). O coeficiente de extinção indica quanto de luz em um determinado comprimento de onda (280 nm para o estudo de proteínas) uma determinada proteína é capaz de absorver.

3.1.3 Modelagem por Homologia

Como ferramentas teóricas para determinação de estruturas tridimensionais de proteínas temos os seguintes métodos de modelagem: Homologia, *Ab Initio*, *Threading* e Aprendizado de Máquina. Destas, a melhor sucedida é a modelagem por homologia. Este método baseia-se em padrões gerais observados em nível molecular e no processo de evolução⁴⁸.

A nível molecular a estrutura de uma proteína é determinada unicamente por sua sequência de aminoácidos, conhecendo a sequência poderíamos, pelo menos em teoria, determinar a estrutura. Quanto ao processo de evolução a estrutura é mais estável a mudanças do que aquelas associadas à sequência de aminoácidos e muda muito mais lentamente do que mudanças associadas a sequência, de modo que sequências semelhantes adotam enovelamentos praticamente idênticos e sequências ainda que distantemente relacionadas podem apresentar estruturas semelhantes, esta relação está ilustrada na figura 12⁴⁹.

FIGURA 12 – As zonas de alinhamento de sequências⁴⁹.

Para a aplicação do método de modelagem por homologia utilizamos a ferramenta de modelagem Modeller v9.8. Como proteína para modelagem utilizamos o domínio catalítico da proteína glicosilhidrolase seguindo os seguintes passos:

Passo 1: Identificação do modelo e alinhamentos iniciais: Este é o passo mais importante no processo, uma vez que, o processo de modelagem por homologia não pode recuperar de um mal alinhamento;

Passo 2: Correção do alinhamento;

Passo 3: Geração do backbone (cadeia principal): Foram gerados 100 modelos;

Passo 4: Modelagem do *Loop*;

Passo 5: Modelagem da cadeia lateral;

Passo 6: Otimização do modelo: minimização de energia;

Passo 7: Validação do modelo, utilizando os seguintes programas: Procheck, Errat, Verify_3D, Prove e Dope. O Procheck verifica a qualidade estereoquímica de uma estrutura de proteína e o programa Errat consiste num algoritmo de verificação da estrutura protéica que é especialmente adequado para avaliar o progresso da construção de modelos cristalográficos e refinamento. O Verify 3D por sua vez analisa a compatibilidade de um modelo atômico (3D) com a sua própria sequência de aminoácidos (1D), Prove avalia o volume do modelo e o Dope é utilizado para avaliar a energia do modelo da proteína gerada através de várias interações pelo Modeller.

3.2 Tecnologia do DNA recombinante

Os ácidos nucleicos (DNA e RNA) constituem o ponto central para o armazenamento, transmissão e expressão da informação gênica. O crescimento da genética molecular desde 1970 tem sido baseado principalmente no desenvolvimento e aprimoramento de métodos para análise e manipulação do material genético. A engenharia genética tem aplicações práticas em muitos campos, como por exemplo, no desenvolvimento de coquetéis enzimáticos para hidrólise, objeto de estudo de nosso grupo de pesquisa, possibilitando com o uso desta técnica alterações nas sequências gênicas a partir de mutações sítio dirigidas, criação de híbridos, estudo de proteínas específicas. Neste contexto, faremos a expressão heteróloga da proteína XCC3535, bem como de seus domínios (catalítico e semelhante a expansina) em *Escherichia coli*.

3.2.1 Extração de DNA genômico de *Xanthomonas campestris* PV *campestris* ATCC33913

O cultivo e extração do DNA genômico do organismo ATCC33913 foi realizado pelo então mestrando Flávio Rodolfo Rosseto que gentilmente cedeu uma alíquota para que pudéssemos realizar os procedimentos de amplificação do gene de interesse. A cepa ATCC33913 foi disponibilizada pelo Prof. Dr. Shaker Chuck Farah (IQ/USP) para estudos das celulasas deste organismo.

Os procedimentos para crescimento do Xcc foram: a partir de um estoque glicerinado a 20% de células do organismo, mantido a -80°C, uma alíquota foi transferida para uma placa de petri LB (Tryptona 10g/L, Extrato de levedura 5 g/L e NaCl 10 g/L e Ágar 15 g/L) contendo 100 µg/mL de ampicilina (antibiótico de resistência do Xcc), que foi mantida a 30°C por três dias. A partir desta placa preparou-se um pré-inóculo de 10 mL de meio 2XTY (Tryptona 16g/L, Extrato de levedura 10 g/L e NaCl 10 g/L) contendo 100 µg/mL de ampicilina que foi mantido a 30°C, 250 rpm por 16 horas⁵⁰. De posse deste pré-inóculo foi utilizado o Wizard® Genomic DNA Purification Kit, do fabricante Promega para extração do

DNA genômico do organismo e o equipamento espectrofotômetro NanoDrop 1000 (Thermo Scientific), para quantificação da amostra de DNA extraída.

3.2.2 Clonagem da sequência codificadora de glicosilhidrolase Xcc e de seus domínios catalítico e semelhante à Expansina

De posse das sequências gênicas para clonagem: proteína Glicosilhidrolase (Xcc3535) e de seus domínios catalítico e semelhante a Expansina, fizemos o desenho dos oligonucleotídeos específicos (*primers*) para uso da técnica de PCR. No processo de desenho dos oligonucleotídeos tomamos como parâmetros os plasmídeos ou vetores que gostaríamos de trabalhar tanto para as etapas de subclonagem quanto de clonagem e a própria sequência gênica para clonagem. Utilizando o software de análises de sequências *Gene Runner* identificamos os sítios de clivagem natural para enzimas conhecidas presentes na sequência gênica da proteína Xcc3535 e de posse desta informação partimos para a seleção dos sítios de clivagem nos vetores de clonagem e desenhos dos oligonucleotídeos.

Foram utilizados os seguintes vetores, para subclonagem o pGEM®-T do fabricante Promega e para clonagem o vetor pSMT3 (figura 13).

O vetor pSMT3 foi gentilmente disponibilizado pelo pesquisador Lawrence Shapiro da Universidade de Columbia NY – USA. O pSMT3 é baseado no vetor pET: usa a T7 polimerase, possui 5.6633pb codifica um cauda de 6-histidinas e uma pequena proteína chamada SUMO. A SUMO é alvo de uma protease chamada ULP1 com especificidade pela extremidade C-terminal de SUMO, deixando apenas um serina extra no N-terminal da proteína de interesse e o pSMT3 apresenta resistência ao antibiótico canamicina. A proteína SUMO (*Small Ubiquitin-like Modifier*) é originária de *Saccharomyces cerevisiae* e é altamente solúvel. Logo o uso desta proteína expressa juntamente com a proteína de interesse aumenta a solubilidade da mesma.

FIGURA 13 – Representação esquemática do vetor pSMT3 e quadro de leitura do vetor pSMT3, possui 5.6633pb, A região de múltipla clonagem encontra-se após o SUMO na região 3' (C-terminal).

No processo de desenho dos oligonucleotídeos foram avaliadas as seguintes propriedades: a região de anelamento deveria conter no mínimo 28 nucleotídeos da sequência gênica para PCR de “pesca” genômica, a porcentagem de GC em sua sequência deveria ser de pelo menos 50%, a extremidade 3' de um primer deve terminar com G ou C de preferência com os dois ou com um deles repetido, evitando a presença de A ou T isto se deve ao número de ligação de hidrogênio efetuadas que são fundamentais para o anelamento do primer ao DNA amplificado, o Tm da fita direta e inversa devem ser próximos variando no máximo de 2°C.

Os oligonucleotídeos utilizados para amplificação da ORF dos genes de interesse por meio de PCR foram sintetizados pela empresa Sigma-Aldrich Brasil Ltda e são apresentados na tabela 4.

TABELA 4 – Oligonucleotídeos utilizados para amplificação das ORFs.

SÍTIO RESTRIÇÃO	OLIGONUCLEOTÍDEOS
Glicosilhidrolase F BamH1	AGGATCCATGCAGGTCAGTACGCAAGCTCCCC
Glicosilhidrolase R Xho1	TCTCGAGTAACTACGGAACTGCACATGCCCCGG
Domínio catalítico Glicosil F BamH1	AGGATCCATGCAGGTCAGTACGCAAGCTCCCC
Domínio catalítico Glicosil R Xho1	ACTCGAGTAATAAGCCGTTGTTGGCATTGCCCCGTTTC
Domínio Expansina F BamH1	AGGATCCCCGTGGAATAGTACTTTCACCGGCAC
Domínio Expansina R Xho1	ACTCGAGTAACTAGCCGTCAGAACGGTCTTTCAGC

F refere se a fita direta (do inglês forward) e R refere se a fita reversa.

A amplificação dos fragmentos gênicos de interesse foi realizada utilizando o método de reação em cadeia da polimerase (PCR), a partir dos oligonucleotídeos descritos na tabela 4 e do DNA genômico extraído de *Xanthomonas campestris* e dos oligonucleotídeos, nas condições da tabela 5, empregando-se as condições descritas no ciclo esquematizado na figura 14.

TABELA 5 - Reação de amplificação para cada um dos genes.

Reação de amplificação	Volume (μL)
Tampão (20mM Tris-HCl pH 8,4, 50mM KCl)	5
dNTP [10mM] (N= A, G, C, T)	1
MgCl ₂ [25mM]	5,28
Oligonucleotídeo [100 pmol/μL] (5' do gene)	1
Oligonucleotídeo [100 pmol/μL] (3' do gene)	1
DNA genômico	1,4
Taq DNA Polimerase (5 U/μL)	0,36
H ₂ O milli-Q	37,96

FIGURA 14 - Representação esquemática das condições da reação de amplificação dos genes.

O produto da amplificação foi analisado em gel de agarose 0.8% com 1 % com brometo de etídio (0,5 μg/mL), em tampão TAE 1X (8 mM Tris-acetato pH 8,0, 0,2 mM

EDTA). Os fragmentos de gene amplificados da Glisossilhidrolase e de seus domínios foram purificados a partir do gel de agarose utilizando-se o kit *Wizard® SV Gel Clean-Up System* (PROMEGA), de acordo com as condições do fabricante com exceção da eluição final que foi feita com 50 µL de água deionizada purificada em sistema Milli-Q e a quantificação da amostra foi realizada no equipamento espectrofotômetro NanoDrop 1000 (Thermo Scientific).

Em seguida os amplificados passaram pela etapa de clonagem em vetor pGEM-T (Promega), esta etapa tem como função o aumento do número de cópia do amplificado para então subclonagem no vetor de expressão, este processo aumenta as chance de ligação neste último. Dessa maneira, o purificado do gel de agarose foi submetido à reação de ligação ao vetor pGEM-T, seguindo protocolo do fabricante *Cloning PCR Products with pGEM®-T and pGEM®-T Easy Vectors*.

3.2.3 Transformação do vetor de subclonagem em células *E. coli* por choque térmico e mudança dos genes clonados para vetor de expressão.

Os produtos de ligação resultantes com o vetor pGEM-T descrito na seção 3.2.2, foram introduzidos em células de propagação *Escherichia coli* DH5α (Novagen) químio-competentes por transformação do tipo choque térmico. O volume de ligação foi adicionado ao das células de propagação, mantidas em repouso no gelo por 20 minutos. Em seguida foi feito um choque térmico a 42°C por 1 minuto, seguido de incubação no gelo por 5 minutos. Após esta etapa, foram adicionados 250 µL de meio LB, mantendo-se a amostra sob agitação por 1 hora, 200 rpm a 37°C. Em seguida, 100 µL do produto de transformação foram adicionados à placa de Petri contendo 25 mL de meio LB sólido e antibiótico ampicilina em concentração de 50 µg/mL, 0.1mM IPTG e 40 µg/ml X-gal. A placa foi incubada durante 16 horas a 37°C.

Colônias selecionadas: brancas e isoladas foram inoculadas em meio LB com antibiótico ampicilina em concentração de 50 µg/mL e incubadas durante 16 horas a 37 °C, em seguida submetidas a uma reação de extração de plasmídeo utilizando Wizard®Plus Minipreps DNA Purification System (Promega), seguindo o protocolo descrito pelo mesmo com exceção da etapa de eluição final que foi realizada com 50 µL de água deionizada pelo

sistema Milli-Q, as amostras foram quantificadas no equipamento espectrofotômetro NanoDrop 1000 (Thermo Scientific).

O próximo passo consistiu na análise de restrição dos sítios de clonagem para confirmar a ligação do gene de estudo (procedimento realizado para as três construções) ao vetor seguindo a reação descrita na tabela 6, mantida por 3 horas a 37°C. Os resultados foram verificados a partir de gel de agarose 0.8 %.

TABELA 6 - Reagentes usados nas reações de digestão do vetor de clonagem e do gene de interesse.

Reação de digestão	Volume (µL)
Plasmídeo p-GEM-T com o gene	2,0
Enzima <i>Bam</i> HI (1 U)	1,0
Enzima <i>Xho</i> I (1 U)	1,0
Tampão Tango(10X) <i>Fermentas</i>	2,0
H ₂ O MilliQ	2,0

Feita a confirmação que os genes amplificados foram ligados procedemos com a clivagem em maior volume do vetor de propagação para então inserirmos o gene amplificado no vetor de expressão. Para esta etapa optamos pela clivagem separada de cada sítio para evitar problemas decorrentes da ação conjunta de duas enzimas que apesar de possuírem atividade comum em tampão Tango (Fermentas) possuem uma atividade ótima em tampões diferentes.

Para tanto, procedemos com a clivagem do sítio para *Xho*I descrito na tabela 7, reação feita para cada uma das três construções separadamente. A reação de clivagem foi mantida a 37°C por 5 horas.

TABELA 7 - Reagentes usados na reação de digestão do sítio *Xho*I do vetor de clonagem.

Reação de digestão sítio para <i>Xho</i> I	Volume (µL)
Plasmídeo p-GEM-T com o gene	30,0
Enzima <i>Xho</i> I (1 U)	6,0
Tampão Tango(10X) <i>Fermentas</i>	20,0
H ₂ O MilliQ	44,0

Ao final do tempo de clivagem adicionamos 50 µL de água Milli-Q em cada reação levando a um volume final de 150 µL, então utilizamos o Wizard® SV Gel and PCR Clean-Up System (Promega) seguindo o protocolo do fabricante com as seguintes alterações adicionamos 150 µL de solução de ligação de membrana aos 150 µL de reação e seguimos o

protocolo até etapa de eluição realizada com 40 μ L de água Milli-Q e procedemos com a clivagem do segundo sítio seguindo a reação descrita na tabela 8, reação mantida a 37 °C por 4 horas. O resultado das restrições foram verificados e confirmados a partir de gel de agarose 0.8 % no qual purificamos os fragmentos de DNA clivados seguindo o protocolo Wizard® SV Gel and PCR Clean-Up System (Promega) com eluição final em 50 μ L de água Milli-Q e a quantificação da amostra foi realizada no equipamento espectrofotômetro NanoDrop 1000 (Thermo Scientific).

TABELA 8 - Reagentes usados nas reações de digestão do sítio BamHI do vetor de clonagem.

Reação de digestão sítio para XhoI	Volume (μ L)
Plasmídeo p-GEM-T com o gene	40,0
Enzima <i>Bam</i> HI (1 U)	7,0
Tampão Green(10X) <i>Fermentas</i>	10,0
H ₂ O MilliQ	43,0

Procedemos então com a ligação dos genes para Glicosilhidrolase e seus dois domínios isolados em vetor pSMT3, de acordo com o protocolo fornecido pelo fabricante da enzima de ligação T4 DNA ligase (Fermentas), com a reação mantida a 10°C por 16 horas.

O produto de ligação resultante do vetor pSMT3 com os genes de interesse, foram introduzidos em células de propagação *Escherichia coli* DH5 α químico-competentes seguindo o mesmo procedimento para quimio transformação descrito no início desta seção, com diferença apenas no antibiótico utilizado nas placas de Petri que foi a canamicina em concentração de 50 μ g/mL. A placa foi incubada durante 16 horas a 37°C. Colônias providas deste processo foram selecionadas, crescidas em meio LB e seu plasmídeo analisado seguindo os métodos descritos nesta seção para clivagem dos sítios enzimáticos, confirmando que os genes amplificados eram do tamanho esperado.

3.2.4 Sequenciamento de DNA e análise de sequências

Os plasmídeos recombinantes foram submetidos ao sequenciamento dos nucleotídeos pelo método de terminação de cadeia por di-deoxinucleotídeos, utilizando o equipamento *ETDye Terminator* (*Applied Biosystems*), procedimento este gentilmente realizado pela técnica Andressa Patrícia Alves Pinto no laboratório de Biofísica Sergio Mascarenhas do Instituto de Física de São Carlos da USP.

O sequenciamento foi realizado a partir dos DNAs plasmidiais pSMT3 com o gene ou seja, para as três construções. Utilizamos os seguintes nucleotídeos para o sequenciamento: pSMT3-Glicosilhidrolase, oligonucleotídeos glicosilhidrolase F (descrito na tabela 4) para a fase direta e “terminator” para a fase reversa, pSMT3- domínio Glicosilhidrolase, oligonucleotídeos glicosilhidrolase F (descrito na tabela 4) para a fase direta e Domínio catalítico Glicosil R XhoI (descrito na tabela 4) para a fase reversa e para o pSMT3- domínio semelhante à Expansina , oligonucleotídeos Domínio Expansina F BamH1 (descrito na tabela 4) para a fase direta e “terminator” para a fase reversa. As amostras de pSMT3 com os insertos para as três construções estavam a 400ng e os oligonucleotídeos utilizados estavam a 3.3mMol. Este procedimento também foi realizado com os insertos em pGEM-T, utilizando os oligonucleotídeos específicos para este vetor.

Os resultados do sequenciamento foram analisados pelo programa ContigExpress (*Invitrogen*).

3.2.5 Linhagens de bactérias utilizadas para a expressão das proteínas.

A seleção das linhagens de bactéria de expressão têm como ponto de partida a associação proteína solúvel e quantidade de proteína expressa. Para estes testes, escolhemos duas linhagens bacterianas para trabalho a *E. coli* Rosetta(DE3)pLysS e a *E. coli* BL 21, estas linhagens de bactérias são deficientes em proteases que possam degradar as proteínas durante o processo de purificação, logo proteínas expressas nestas linhagens tornam-se mais estáveis. A Rosetta™ é uma derivada da BL21 e designada por aprimorar a expressão de proteínas eucarióticas que contenham códons raramente vistos em *E. coli*⁵¹.

A partir de células químio competentes e das linhagens de *E. coli* Rosetta(DE3)pLysS e a *E. coli* BL 21 procedemos com as transformações por choque térmico do plasmídeo pSMT3 com os insertos gênicos de interesse: pSMT3-glicosilhidrolase, pSMT3-domínio-catalítico-glicosilhidrolase e pSMT3-domínio-semelhante à Expansina. Seguimos o protocolo de transformação por choque térmico descrito na seção 3.2.3, a diferença está no fato que os 100 µL da reação de transformação que cresceu durante 1 hora a 37 °C e 250 rpm foram adicionados a uma placa de Petri contendo 25 mL de meio LB sólido e antibiótico canamicina em concentração de 50 µg/mL para a *E. coli* BL 21 e para a *E. coli* Rosetta(DE3)pLysS adicionamos ainda 34 µg/mL de cloranfenicol . A placa foi incubada durante 16 horas a 37°C.

Colônias providas deste processo foram selecionadas e iniciamos os testes de expressão. O quadro da figura 15 ilustra os passos seguidos no processo de clonagem gênica.

FIGURA 15 – Etapas do processo de clonagem. Adaptado⁵².

3.2.6 Expressão e purificação das proteínas recombinantes

As proteínas de trabalho foram purificadas pelo método de cromatografia por afinidade e gel filtração ou exclusão molecular. Na cromatografia por afinidade utilizamos a coluna Ni-NTA (Qiagen)⁵³ e para a cromatografia de exclusão molecular utilizamos o sistema

Äkta Purifier 10 (Amershan-Pharmacia) utilizando coluna Superdex 75 10/300 GL (GE Healthcare).

As colunas e os tampões utilizados foram selecionados levando em consideração às características particulares de cada uma das construções protéicas. A pureza da preparação protéica foi analisada por SDS-PAGE (*Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis*) em gel de poliacrilamida 15%, sob condições desnaturantes⁵⁴. Os géis foram corados com *Coomassie Brilliant Blue R-250* e o conteúdo protéico determinado pelo método de *Bradford*⁵⁵.

Dessa maneira podemos dividir o processo de expressão e purificação das proteínas em seis etapas.

Primeira etapa: Superexpressão dos genes:

A partir da seleção de clones provindos da reação de transformação descrita na seção anterior iniciamos a superexpressão dos genes para Glicosilhidrolase, domínio Catalítico e domínio semelhante a Expansina. Fizemos um inóculo de 10mL de LB com 50 $\mu\text{g.mL}^{-1}$ de canamicina e 34 $\mu\text{g.mL}^{-1}$ de cloranfenicol para cada uma das construções e cultivamos a 37°C, 100RPM por 16 horas.

Após este período de crescimento, as colônias foram transferidas para 1 litro de meio LB, contendo os mesmos antibióticos a 37°C e 160 rpm. Nesta etapa acompanhamos a $\text{DO}_{600\text{nm}}$ até 0,6 quando reduzimos a temperatura a 18°C, aguardamos 30 minutos, e aplicamos o agente indutor IPTG (foi adicionado para concentração final de 1mmol.L^{-1}).

Segunda etapa: Lise celular:

As células foram incubadas por mais 16 h, após este período foram centrifugadas a $4700 \times g$ por 10 min, então descartamos o sobrenadante e ressuspendemos o precipitado com 25mL do tampão A (100mM de NaCl, 50 mM de Tris(Base), 2mM de B- mercaptoetanol, 5% de glicerol a pH7.5) para a Glicosilhidrolase, tampão B (300mM NaCl, 50mM Tris, 5% glicerol, pH 7.5) para o domínio Catalítico e tampão C (20mM NaCl, 10mM Tris, 5% glicerol, pH 7.5) para o domínio Expansina.

O ressuspensão foi lisado utilizando o sonicador por 6 minutos (30''on e 45''off), amplitude 15%. Centrifugamos o lisado ($6500 \times g$, 30min), e recuperamos o sobrenadante, sendo a expressão verificada por meio de SDS-PAGE 15%.

Terceira etapa: Purificação por Afinidade Química e Diálise:

Para isto, aplicamos em uma coluna cromatográfica contendo 5mL de resina Ni-NTA *Superflow* (quiagen), previamente equilibrada com 5 volumes de tampão, o sobrenadante do processo de lise (25mL) e deixou-se incubando sob leve agitação por 60 min a 10°C.

Após este período procedemos com o protocolo de purificação por afinidade química: deixamos passar pela coluna a primeira fração nomeada não ligado (*void*) que foi descartada, em seguida, a coluna foi lavada com 20 vezes o volume de resina com o de tampão da proteína com 10mM de imidazol e em seguida a proteína recombinante foi eluída com um gradiente de imidazol em tampão, típico para cada uma das construções.

A diálise do imidazol presente na amostra foi realizada por troca de tampão utilizando o concentrador *Vivaspin 500* (GE Healthcare), num processo de diluição do imidazol da amostra.

Quarta etapa: Clivagem da His-Tag (Sumo) e Segunda purificação por afinidade:

Feita a troca do tampão, colocamos a amostra para clivagem da His-tag de purificação (cauda de histidina e SUMO), para cada 100mg de proteína utilizamos 0.1mg de enzima ULP1, deixando sob agitação a 8°C por 12 horas. Após este período, injetamos a amostra na coluna Ni-NTA *super flow* (Quiagen), previamente equilibrada com 5 volumes de tampão da proteína, deixamos incubando sob leve agitação por 60 min a 10°C. Após este período, deixamos passar pela coluna o void e o coletamos. Com o uso do concentrador *Vivaspin 500* (GE Healthcare) concentramos a amostra e então aplicamos numa coluna de gel filtração.

Quinta etapa: Purificação por Cromatografia de Exclusão molecular:

A cromatografia de exclusão molecular (gel filtração) foi utilizada para separar frações do Tag (His-Sumo) que não tenham sido retidas na etapa anterior de purificação e também para conter eventuais contaminantes. Um volume de 500µL foi aplicado na coluna Superdex 75 (HR 10/30, GE Healthcare), pré-equilibrada com o tampão da proteína e acoplada a um sistema *Äkta purifier* (GE Healthcare).

A cromatografia foi realizada com eluição isocrática de 1,3 volume da coluna, sob uma taxa de 0.5mL/min e monitorada pela absorbância de 280 nm. Frações de 500µL dos picos de interesse foram coletadas e analisadas por SDS-PAGE 15%. De posse das frações

correspondentes aos picos do cromatograma da proteína utilizamos o concentrador *Vivaspin 500* (GE Healthcare) para concentrar a amostra e realizar estudos biofísicos da mesma.

Sexta etapa: Quantificação das soluções protéicas pelo método de Bradford⁵⁵:

O conteúdo das soluções protéicas foi determinado pelo método de *Bradford*⁵⁵, utilizando o reagente de Bradford (Bio-Rad), com albumina soro bovina como padrão, o protocolo de quantificação utilizado foi o seguinte: Branco (800 μ L de água Milli-Q e 200 μ L do reagente de Bradford), amostra (795 μ L água Milli-Q, 5 μ L solução protéica e 200 μ L reagente de Bradford), a quantidade de solução protéica utilizada variou em função da concentração da mesma a fim de não ultrapassar valores de absorbâncias maiores que 1 aferidos a 595 nm no equipamento Multiskan® Spectrum (Thermo Scientific). A quantidade protéica foi determinada em função da curva de absorbância padrão obtida para a proteína albumina soro bovina.

3.3 Caracterização biofísica

A caracterização biofísica das proteínas recombinantes de *Xanthomonas campestris* foi feita empregando-se espectroscopia de dicroísmo circular para investigação de sua estrutura secundária, espalhamento dinâmico de luz, espalhamento de Raios-X a baixo ângulo e cristalografia.

3.3.1 Estimativa da massa molecular.

Proteínas-padrão com diferentes massas moleculares (tabela 9) foram analisadas nas mesmas condições da cromatografia por exclusão molecular das proteínas de trabalho. Determinando-se o volume de eluição de cada padrão e conhecendo-se sua massa molecular, determinamos uma curva padrão de calibração da coluna cromatográfica por exclusão molecular Superdex 75 10/300 GL (GE Healthcare).

TABELA 9 – Padrões utilizados na calibração da Superdex 75 10/300 GL(GE Healthcare).

Proteína	K _{av}	MM (kDa)
Conalbumina	0.029570	75,0
Anidrase Carbônica	0.165323	29,0
Ovoalbumina	0.074597	44,0
Ribonuclease A	0.282930	13,7

O volume morto da coluna (V_o) foi determinado aplicando o corante Blue Dextran pela coluna, as indicações de fluxo, pressão, concentração da amostra foram seguidas de acordo com especificações do fabricante para a coluna. Obtido o volume de eluição de cada proteína (V_e), foi possível calcular os valores de K_{av}, estes valores estão relacionados aos perfis de migração de cada molécula, segundo a equação 1.

$$K_{av} = \frac{V_e - V_o}{V_t - V_o} \quad (1)$$

sendo que V_t corresponde ao volume total da coluna. Tomando como referência a massa molar e seus respectivos K_{av}s, obteve-se o gráfico log(MM) versus K_{av}, com estes dados foi possível calcular as massas moleculares, a partir da equação da reta, obtida por regressão linear dos dados.

3.3.2 Espectroscopia de Dicroísmo Circular

O Dicroísmo circular (CD) refere-se a uma técnica em que a absorção diferencial das componentes da luz circularmente polarizada é o objeto de estudo. Esta absorção diferenciada advém das características naturais de moléculas assimétricas (um cromóforo) que tornam a luz plano polarizada em elípticamente polarizada ao passar através delas.

Na prática, a radiação é dividida em duas componentes polarizadas através da passagem por um modulador (normalmente um cristal piezelétrico, quartzo) na presença de um campo elétrico alternado (50 kHz). O modulador irá transmitir cada uma das componentes. Se ao passar pela amostra, as componentes não forem absorvidas, a combinação das componentes poderá gerar uma onda plana novamente. Caso contrário, se uma das componentes for absorvida, o resultado será uma onda elípticamente polarizada, como ilustrado na figura 16⁵⁶.

FIGURA 16 – Origem do efeito do CD (A) a esquerda (L) e direita (R) componentes de mesmo tamanho, luz circularmente polarizada (II) componentes são de diferentes magnitudes e resultando em luz elípticamente polarizada.

O instrumento de CD não recombina as componentes mas detecta as mesmas separadamente, sendo que o equipamento irá apresentar a diferença de absorbância das componentes dado pela equação 2, ou através da elipticidade em graus (θ), dado pela equação 3, onde b e a são as componentes dos eixos da elipse⁵⁶.

$$\Delta A = A_L - A_R \quad (2)$$

$$\theta = \tan^{-1} \left(\frac{b}{a} \right) \quad (3)$$

Existe uma relação numérica entre ΔA e θ (θ em graus), apresentada na equação 4.

$$\theta = 32.98\Delta A \quad (4)$$

Na prática a maior parte dos trabalhos envolvendo amostras biológicas aplicando CD são da ordem de 10 miligraus (isto quer dizer que a diferença na absorbância da luz circularmente polarizada é da ordem de 3×10^{-4} unidades de absorbância), na tabela 10 temos a relação entre tipo de radiação aplicada, comprimento de onda e grupo de absorção comuns para amostras biológicas⁵⁶.

TABELA 10 – Relação tipo de radiação aplicada, comprimento de onda e grupo de absorção.

Radiação	Comprimento de onda (nm)	Grupo de Absorção
UV distante	180 – 240	Ligações peptídicas
UV próximo	250 – 290	Cadeias laterais aromáticas
Visível, Ir	300 – 1400	Ligantes

As ligações peptídicas são os principais cromóforos responsáveis pela absorção na região do ultravioleta (UV) distante (195-250 nm), fato decorrente das transições eletrônicas fracas do tipo $n-\pi^*$, centradas em 220 nm e transições fortes do tipo $\pi-\pi^*$ próximas à 190 nm. Na figura 17, temos a ilustração dos espectros esperados para cada tipo de estrutura secundária aplicando o CD⁵⁶.

FIGURA 17 – Espectro típico das principais estruturas secundárias em proteínas, ilustração de Norma J. Greenfield⁵⁷.

O espectro de Dicroísmo Circular da Glicosilhidrolase e do domínio similar à Expansina foram obtidos utilizando um espectropolarímetro *JASCO J-815* (Toquio, Japão), equipado com um sistema de controle de temperatura (*PELTIER*) do Laboratório de Biofísica Sérgio Mascarenhas, sendo as medidas gentilmente auxiliadas pela técnica Andressa Patrícia Alves Pinto.

As medidas foram realizadas com uma média de 8 acumulações, sensibilidade de 100 miligraus, velocidade de varredura de $100\text{nm}\cdot\text{min}^{-1}$, em uma cubeta de quartzo retangular de caminho óptico de 1 mm para medidas de CD no UV distante (190-250nm), largura de banda de 1nm e tempo de resposta de 0,5s. Os efeitos de contribuição do solvente foram subtraídos das curvas originais e os dados foram tratados aplicando a Transformada Rápida de Fourier (FFT- Fast Fourier Transform).

Os espectros de CD foram convertidos em elipticidade molar. A deconvolução do espectro de Dicroísmo Circular e determinação das frações de estrutura secundária foram feitas com o software CDPro.

3.3.3 Espalhamento Dinâmico de Luz (DLS)

Espalhamento dinâmico de Luz (DLS) é uma importante técnica experimental aplicada na ciência e na indústria, sendo também conhecida como espectroscopia de correlação de fótons (PCS). Esta técnica mede as flutuações de luz espalhada por uma amostra, os objetos espalhadores (proteína) estão em movimento aleatório através da amostra, logo o nome espalhamento dinâmico de Luz. A velocidade de movimento das partículas é relacionada ao tamanho das mesmas, como ilustrado na figura 18^{58; 59}.

FIGURA 18 – As partículas num líquido se movimentam aleatoriamente e sua velocidade de movimento é utilizada para determinar o tamanho das partículas, adaptado⁵⁹

Quando a Luz incide sobre uma amostra material, ela pode ser absorvida ou espalhada. A dispersão ocorre pela interação dos elétrons com a radiação incidente, a qual induz esses elétrons a serem acelerados de maneira oscilante que por sua vez reemitem a radiação. Desta maneira, os elétrons tornam-se fontes de radiação em movimento Browniano. e consequentemente emitem luz⁶⁰.

O movimento Browniano das partículas em solução está relacionado tanto ao coeficiente de difusão das moléculas, quanto a seu tamanho e forma, sendo que partículas menores por movem-se mais rapidamente que partículas grandes possuindo coeficiente de difusão (D) maior^{58; 59}. Considerando a dispersão de partículas esféricas, com viscosidade η , sob temperatura constante T , temos que o coeficiente de difusão D é inversamente proporcional ao raio hidrodinâmico r_h das partículas (o raio hidrodinâmico recebe este nome pelo fato de que a hidrodinâmica da solução é levada em conta), essa relação é conhecida como Stokes- Einstein e é representada pela equação 5, onde k é a constante de Boltzmann^{58; 59}.

$$D = \frac{k.T}{6\pi\eta r_h} \quad (5)$$

Existe uma relação entre a velocidade de flutuação da luz espalhada (intensidade da luz que atinge o detector irá flutuar devido ao movimento das partículas) e o coeficiente de difusão das partículas determinada pela função de auto-correlação. Para partículas pequenas, essa função de auto-correlação entre as intensidades diminui mais rapidamente com o tempo,

do que para partículas grandes^{58; 59}. A função de auto-correlação, $G(t)$, é apresentada pela equação 6.

$$G(t) = \langle I(t_0) \times I(t_0 + t) \rangle \quad (6)$$

Sendo $I(t_0)$ e $I(t_0 + t)$ as intensidades de luz espalhada nos instantes t_0 e $(t_0 + t)$, respectivamente. Para partículas esféricas e monodispersas, a função $G(t)$ é dada pela equação 7.

$$G(t) = Ae^{-2\Gamma t} + B \quad (7)$$

sendo A e B constantes e Γ a constante de decaimento da curva exponencial, descrita pela equação 8.

$$\Gamma = Dq^2 \quad (8)$$

sendo q o vetor de onda de luz espalhada dado pela equação 9,

$$q = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad (9)$$

onde λ é o comprimento de onda da luz espalhada e θ é o ângulo de espalhamento de luz.

A polidispersividade descreve a largura da distribuição do tamanho das partículas presentes numa solução, sendo o principal parâmetro observado na análise de proteínas, e é obtido do índice de polidispersividade (PdI), calculado a partir da análise acumulativa da intensidade medida no DLS, da sua função de autocorrelação^{58; 59}. Na análise acumulativa, um único tamanho de partícula é considerado e uma única exponencial é ajustada para a função de autocorrelação, na figura 19 temos um exemplo de sistemas monodisperso e polidisperso.

FIGURA 19 - Distribuição de tamanho monodisperso e polidisperso, adaptado ^{58; 59}.

Para o estudo do estado de oligomerização das proteínas: Glicosilhidrolase e de seu domínio semelhante a Expansina e também para avaliação das amostras utilizadas nos experimentos de espalhamento de raios-X a baixos ângulos, utilizamos a técnica de espalhamento dinâmico de Luz.

A técnica foi aplicada utilizando o equipamento Zetasizer μ V(Malvern) e cubeta de quartzo de 2 μ L. As medidas foram feitas a 20°C com a aquisição de 13 medidas por amostra. As características do tampão foram registradas e descontadas das medidas, sendo para amostra de Glicosilhidrolase 100mM de NaCl, 50 mM de Tris, 5 % de glicerol a pH 8 e para a amostra de proteína semelhante a Expansina 20mM NaCl, 10mM de Tris, 5% de glicerol a pH 8 . As amostras Glicosilhidrolase e domínio semelhante a Expansina nas concentrações de 3mg/mL, aplicadas foram previamente centrifugadas a 10.000 rpm, 10min. a 4 °C.

3.3.4 Espalhamento de Raios-X a Baixo Ângulo (SAXS)

O espalhamento de raios-X a baixo ângulo é uma ferramenta fundamental no estudo de macromoléculas biológicas. A principal vantagem deste método reside na sua capacidade de fornecer informação estrutural sobre sistemas parcialmente ou completamente desordenados.

SAXS permite estudar a estrutura de partículas em condições bem próximas as fisiológicas e analisar mudanças estruturais em resposta as variações de condições externas como, por exemplo, a mudança de temperatura.

Devido à exigência de bons cristais para cristalografia e baixa massa molecular para RMN, uma fração significativa de proteínas não podem utilizar estes dois métodos de alta resolução, o que torna o SAXS uma ferramenta extremamente importante para construção de um modelo estrutural⁶¹.

Para a obtenção de modelos tridimensionais, os experimentos de SAXS exigem uma solução homogênea, diluída em uma solução tampão sem aditivos especiais, monodisperso (sistemas de partículas idênticas) e com orientação aleatória de partículas em solução, levando a média ao espalhamento de uma única partícula.⁶¹

Raios-X são fótons com energia E com comprimento de onda dado por $\lambda = 1.256/E$, onde λ é expresso em nm e E em keV. Para estudos estruturais, os raios-X possuem energia

em torno de 10 keV onde temos (λ entre 0.10–0.15 nm). Quando um objeto é iluminado por uma onda plana monocromática $k_0 = |\mathbf{k}_0| = 2\pi/\lambda$, os átomos do objeto interagem com a onda incidente e tornam-se fontes de ondas esféricas⁶¹.

O processo de espalhamento tomando a primeira aproximação de Born pode ser descrito como a transformação do espaço “real” laboratório (objeto) em coordenadas $\mathbf{r}$ para o espaço recíproco pelo vetor $\mathbf{s} = \mathbf{k}_1 - \mathbf{k}_0$ (figura 20).

Seguindo as propriedades da transformada de Fourier (por exemplo, a reciprocidade entre dimensões no espaço real e recíproco, quanto menor for o tamanho 'real' maior será seu correspondente no espaço 'recíproco').

FIGURA 20 – Esquema representando o experimento de SAXS e a transformada de Fourier do espaço real para o recíproco. Adaptado⁶¹

A amplitude do espalhamento dos raios-X representa a transformada de Fourier da densidade de elétrons de um átomo (esférico) em função do momento transferido, $q = s = 4\pi\lambda^{-1}\sin(\theta)$ onde 2θ é o ângulo de espalhamento, e $f(0) = f_x$.

Assumindo que o solvente possui uma densidade de espalhamento constante ρ_q , a diferença de espalhamento para uma partícula singular em relação ao solvente será definido como $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_q$.

$$A(\mathbf{q}) = F[\Delta\rho(\mathbf{r})] = \int_V \Delta\rho(\mathbf{r}) \exp(i\mathbf{q}\mathbf{r}) d\mathbf{r} \quad (10)$$

A partir da integração em todo o volume. O experimento de espalhamento não pode medir diretamente a amplitude, mas sim a intensidade do espalhamento por uma única partícula $I_0(\mathbf{q})$ que é dado pelo produto da amplitude espalhada pelo seu complexo conjugado, numa dada direção ($\mathbf{q}$).

$$I_0(\mathbf{q}) = A(\mathbf{q})A(\mathbf{q})^* \quad (11)$$

A média sob todas as direções leva à fórmula de Debye

$$\langle e^{-i\mathbf{q}\cdot\mathbf{r}} \rangle_\Omega = \frac{\text{sen}(qr)}{qr} \quad (12)$$

Logo a equação (11), pode ser reescrita como:

$$I_0(q) = 4\pi \int_0^\infty \gamma(r) \frac{\text{sen}(qr)}{qr} r^2 dr \quad (13)$$

Desta maneira, obtemos a forma mais geral para o padrão de espalhamento de partículas, onde $\gamma(r)$ é conhecida como função de correlação ou função característica e pode ser obtida pela transformada inversa de Fourier, o fator de correlação também é escrito como $\gamma(r) = P(r)$.

$$P(r) = \gamma(r) = \frac{1}{2\pi^2} \int_0^\infty I_0(q) q^2 \frac{\text{sen}(qr)}{qr} dq \quad (14)$$

Avaliando o comportamento da intensidade do espalhamento para valores pequenas ($q \rightarrow 0$), fazendo a aproximação $q = r = 0$, o fator de Debye assume valor 1, dessa maneira as equações 13 e 14 passam a valer.

$$I_0(0) = \int_0^\infty 4\pi^2 \gamma(r) dr \quad (15)$$

$$P(0) = \gamma(0) = \frac{1}{2\pi^2} \int_0^\infty I(q) q^2 dq = V_p (\overline{\Delta\rho})^2 \quad (16)$$

Este resultado não é acessível experimentalmente, para tanto uma extrapolação até zero, chamada aproximação de Guinier permite a determinação de tais valores.

Admitindo regiões de baixo q , ou seja, $qr \ll 1$, o fator de Debye pode ser aproximado de acordo com a expansão de McLaurin

$$\frac{\text{sen}(qr)}{qr} \cong 1 - \frac{(qr)^2}{3!} + \dots \quad (17)$$

Substituindo (17) em (13) obtemos:

$$I_0(q) = 4\pi \int \gamma(r) \left(1 - \frac{(qr)^2}{6} + \dots\right) r^2 dr \cong I_0(0) \left(1 - \frac{q^2 r_g^2}{3}\right) \quad (18)$$

Sendo r_g , o raio de giro, obtido a partir da segunda normalização da distribuição de partícula.

Lembrando que a aproximação de Guinier é válida para baixos ângulos: $qr \ll 1$ e $e^{-x} \cong 1 - x$ e a equação (18) pode ser reescrita:

$$I_0(q) \cong I_0(0) \exp \left\{ -\frac{R_g^2 q^2}{3} \right\} \quad (19)$$

A equação (19) é conhecida como Lei de Guinier. Para um sistema monodisperso ideal, o gráfico de Guinier ($\ln(I(q))$ versus q^2) pode ser uma função linear, em que o intercepto na origem dá $I(0)$ e o coeficiente angular o raio de giro R_g . A linearidade do gráfico de Guinier é um teste para a avaliação da homogeneidade da amostra e desvios desta indica interferências advinda da interação interpartículas da amostra⁶¹.

A interpretação de dados de SAXS permite uma análise detalhada da estrutura global de proteínas, as quais podem ter seu tamanho variável entre poucos kDa e muitos MDa. A determinação direta do envelope molecular, a predição de estruturas quaternárias em oligômeros, a verificação de transições conformacionais que ocorrem por interação com ligantes, assim como, a visualização de processos de enovelamento e desenovelamento protéico podem ser realizadas⁶².

Neste trabalho, os resultados experimentais de espalhamento de raios-X a baixo ângulo foram obtidos para a Glicosilhidrolase em concentrações de 1; 3 e 6 mg/mL em tampão: 100 mM de NaCl, 50 mM de TRIS-HCl, 5% de glicerol e pH 8,0 e para o domínio semelhante a Expansina nas concentrações de 1; 2,8 e 6 mg/mL em tampão: 20mM de NaCl, 10 mM de TRIS-HCl, 5% de glicerol e pH 8,0. Foram coletados em colaboração com o Prof. Dr. Mario de Oliveira Neto. Os dados de SAXS foram coletados na linha SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP).

Dentre os parâmetros de coleta serão apresentados os referentes aos dados coletados e processados. Foi utilizado comprimento de onda $\lambda=0.148$ nm, detector bidimensional (2D) e a distância amostra-detector foi de 1495.45 mm, cobrindo um intervalo de transferência de momento de $0.01 \text{ \AA}^{-1} < q < 0.23 \text{ \AA}^{-1}$ para a proteína Glicosildrolase (concentração de 2,8 mg/mL) e 986.46 mm cobrindo um intervalo de transferência de momento de $0.015 \text{ \AA}^{-1} < q < 0.38 \text{ \AA}^{-1}$ para a proteína domínio semelhante a Expansina (concentração de 3 mg/mL).

Os resultados experimentais foram analisados utilizando o programa FIT2D⁶³, normalizando-se os dados pela intensidade do feixe e multiplicando-se pela absorção da amostra⁶⁴. Subtraiu-se o espalhamento da solução tampão foi subtraído da curva de espalhamento da solução de proteína. O raio de giro da proteína foi obtido utilizando-se a aproximação de Guinier⁶⁵:

A massa molecular da proteína foi estimada usando o programa *SAXS MoW*⁶⁶. O programa *GNOM*⁶⁷, foi utilizado para obter a curva de distribuição de distâncias $p(r)$ e a $D_{\text{máx}}$.

O programa *CRYSOI*⁶⁸ foi utilizado para a comparação deste resultado com estruturas de homólogas já conhecidas com a geração de uma curva experimental de SAXS para um modelo cristalográfico.

Utilizamos os programas *DAMMIN*⁶⁹ e *GASBOR* para a geração de modelos *ab Initio* de baixa resolução. A sobreposição do modelo de átomos dummy provenientes dos programas utilizados foram sobrepostas a modelos por homologia e cristalográfico, utilizando o SUPCOMB⁷⁰.

3.3.5 A estrutura cristalina de uma proteína

Processo de formação de cristais protéicos

O crescimento de cristais de proteínas é obtido por tentativa e erro, porém uma das principais regras para esse processo é a utilização de amostras protéicas puras. Quanto maior o grau de pureza de uma amostra, melhores são as chances de se obter cristais e cristais de qualidade, para os experimentos de difração com Raios-X⁷¹.

A obtenção dos cristais de proteína constitui num processo que passa por três etapas, nucleação, crescimento e cessação de crescimento. Utilizamos do método de cristalização por difusão de vapor, em particular aplicando a técnica da gota pendurada.

Uma pequena gota de proteína foi misturada com um volume igual da solução de cristalização (formada por tampão, sal e precipitante) e colocada sobre uma lamínula siliconizada. A lamínula foi invertida e selada sobre um reservatório contendo solução de cristalização. A diferença de concentração entre a gota e o reservatório determinou uma diferença de potencial químico, no sentido de um equilíbrio de concentração entre a gota e as soluções do reservatório e com a supersaturação da solução protéica têm a formação do cristal.

Varreduras sobre as condições de cristalização mais comuns em proteínas

Utilizando de sistemas automatizados de *Screening* de condições de cristalização é possível com uma pequena quantidade de proteína varrer várias condições que não seriam tão

facilmente obtidas, tanto pelo dispêndio de amostra protéica como preparo de condições de cristalização para teste.

IFSC

Utilizando o robô de cristalização Honeybee fizemos um *screening* de condições de cristalização. Para cada varredura de um kit de cristalização foram necessários 100 µL da proteína na concentração desejada.

Amostras da Glicosilhidrolase (puras sem HIS-SUMO) preparadas na concentração de 7,7mg/mL e 15mg/mL foram submetidas aos seguintes conjuntos de cristalização: Crystal screen HT- HR2-130 e o Crystal Screen Lite & PEG/Ion.

Amostras do domínio semelhante a Expansina (puras sem HIS-SUMO) na concentração de 20mg/mL e 40mg/mL foram submetidas ao seguinte conjunto de cristalização: CRYSTAL SCREEN HT- HR2-130.

Amostras do domínio Catalítico não foram objetos deste estudo pela dificuldade de se obter quantidade suficiente de proteína para realizar este procedimento.

EMBL HTLXLab (High Throughput Crystallisation Laboratory)

Em colaboração com o EMBL foi possível o uso do equipamento robô de cristalização Cartesian PixSys 4200 (Genomic Solutions, U.K.), para novos testes de cristalização.

Amostras de Glicosilhidrolase não foram submetidas a novos testes de varredura de condições de cristalização.

Amostras do domínio semelhante a Expansina (puras sem HIS-SUMO) na concentração de 20mg/mL e 40mg/mL foram submetida aos seguintes conjuntos de cristalização Crystal Screen Lite & PEG/Ion e Index Screen.

Amostras do domínio Catalítico (com HIS-SUMO) na concentração de 5mg/mL e 7 mg/mL foram submetidas aos seguintes conjuntos de cristalização: Crystal Screen I & II, Crystal Screen Lite & PEG/Ion, membFac & Natrix, QuickScreens Ammonium Sulfate, Sodium Malonate- Sodium formate, Grid screens PEG 6K, PEG/LiCl, MPD- Screen Mme e Index Screen.

Preparação do cristal para coleta.

O primeiro teste num procedimento de difração de raios-X é o teste da condição *crio*, para isso, aplicamos sobre o loop de fixação dos cristais a solução de cristalização, se na presença do fluxo de nitrogênio houver formação de gelo (observado a partir da difração da solução crio ou no próprio loop) será necessária a adição de aditivos (glicerol, etilenoglicol, etc.) para conter este processo. Outra característica destes aditivos está no possível fortalecimento da própria rede cristalina permitindo uma melhor difração do feixe de raios-X, sobre o Cristal.

3.3.6 Difração de Raios-X aplicado a Cristalografia de Proteínas

As fontes de raios-X comuns em laboratórios são do tipo tubo selado e anodo rotatório, sendo o último o preferencial para o estudo de proteínas por raios-x devido à alta intensidade do feixe. A radiação mais comum provinda de tubos possui comprimento de onda de 1.5410 Å. Este é o comprimento de onda característico K_{α} emitido pelo anodo de cobre sendo selecionado da distribuição espectral por um filtro ou, preferencialmente, com um monocromador⁷¹.

A alta intensidade da radiação por Raios-X de um Síncrotron permite coletas de dados extremamente rápidas. O feixe neste caso não é apenas intenso como também, altamente colimados, levando a “*spots*” menores e de maior intensidade ao detector quando comparando as fontes descritas anteriormente. Além disso, com a radiação Síncrotron, a resolução atingida é melhor do que com as anteriormente citadas, resultando em mais fótons para difração a altos ângulos permitindo mais detalhes a respeito da estrutura protéica⁷¹.

Para a resolução de uma estrutura protéica, o número de feixes difratados é extremamente alto, da ordem de 10^4 a 10^5 , logo ótimos detectores devem estar disponíveis como, por exemplo, o CCD.

Neste ano comemoramos o centenário da introdução do fenômeno de difração por Raios-X, antes disso apenas as propriedades externas dos cristais eram conhecidas. Os cristais constituem a principal ferramenta para a cristalografia, com seus planos de estrutura, simetria, grupos de espaço e unidades assimétricas.

A periodicidade cristalina nos permite extrair que apenas em determinadas orientações, um feixe de Raios-X pode difratar. O resultado de uma determinação de estrutura por Raios-X é uma densidade eletrônica e a equação fundamental para o cálculo desta grandeza é:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i] (hx + ky + lz) + i \alpha(hkl) \quad (20)$$

Nesta equação $\rho(x, y, z)$ é expresso como uma transformada de Fourier do fator de estrutura $F(hkl)$. O fator de estrutura é uma função dos fatores de espalhamento atômico f_j (denota a razão entre as amplitudes de espalhamento por um átomo e por um elétron isolado) e das posições xyz do átomos:

$$F(hkl) = \sum_i f_j \exp[2\pi i(hx + ky + lz)] \quad (21)$$

A amplitude do fator de estrutura é obtida a partir da intensidade de difração, sendo $I(hkl) = |F(hkl)|$. Os ângulos de fase $\alpha(hkl)$ são perdidos no processo de registro da informação (devido à falta de coerência dos Raios-X), não sendo obtido de forma simples, mas pode ser determinado por métodos como: substituição molecular, substituição isomorfa múltipla (MIR), dispersão anômala múltipla e métodos diretos.

Neste trabalho utilizamos a técnica de substituição molecular que determina as informações de fase de uma proteína não resolvida cristalograficamente utilizando as informações de fase de uma estrutura resolvida que apresente estrutura primária similar. O fluxograma da figura 21 ilustra os principais passos na resolução de uma estrutura de proteína a partir do método de substituição molecular:

FIGURA 21 – Processo aplicado no método de substituição molecular ⁷².

O coeficiente de correlação (CC) é um indicador estatístico:

$$CC = \frac{\sum |F_{obs}| |F_{calc}| - [\sum |F_{obs}|] [\sum |F_{calc}|] / N}{\left\{ \left[\sum |F_{obs}|^2 - \sum |F_{obs}|^2 / N \right]^2 \right\} \left\{ \left[\sum |F_{calc}|^2 - \sum |F_{calc}|^2 / N \right]^2 \right\}} \quad (22)$$

Este indicador assume valores próximos a 1 quando há correlação, e valores próximos a 0 quando não há correlação. Este é calculado usando-se os fatores de estrutura calculados (F_{calc}) e os fatores de estrutura observados (F_{obs}). Os F_{calc} são obtidos diretamente das coordenadas atômicas do modelo de busca, e os F_{obs} do experimento de difração de Raios X^{72;}
73.

As coletas de dados de difração de raios-X foram realizadas no Instituto de Física da USP em São Carlos IFSC, que disponibiliza um gerador um anodo rotatório modelo UltraX 18 (RIGAKU/MS) com detector tipo placa de imagem modelo Mar345dtb (MAR Research) e no E.S.R.F (European Synchrotron Radiation Facility) em Grenoble - França na linha ID23-1 sob a orientação do pesquisador Sasha Popov. A ID23-1 oferece a oportunidade de fazer medições altamente eficientes com um feixe monocromático com energias de 5 Kev a 20 Kev com um fluxo de $\sim 1-4 \text{ E12}$ fótons/segundo e com um feixe variável de tamanho de $10 \times 10 \mu\text{m}^2$ a $45 \times 30 \mu\text{m}^2$, equipada com um detector ADSC Q315R que possui uma área ativa de $315 \times 315 \text{ mm}^2$ (3072×3072 pixels). O tempo médio por imagem é de 1s a 0,3 s.

Dentre as diversas coletas de dados realizadas para o domínio semelhante a Expansina iremos discutir três em especial uma realizada no IFSC e duas coletas realizadas no E.S.R.F: intituladas ESRF-1 e ESRF2.

Na coleta realizada no IFSC, tempo de exposição de 420 s e $\lambda = 1.541870 \text{ nm}$, conjunto com 50 imagens (resolução máxima atingida de $3.03 \text{ \AA}$).

Na coleta ESRF-1, a distância amostra detector foi de 408.97mm do detector, ângulo de oscilação de 0.65° , tempo de exposição de 0,178s e $\lambda = 0,972 \text{ nm}$ para um conjunto com 83 imagens (resolução máxima atingida de 2.75), na coleta ESRF-2, coleta a distância amostra detector foi de 358.316 mm, com ângulo de oscilação de 0.1° , tempo de exposição de 0,178s $\lambda = 0,972 \text{ nm}$, na qual três conjuntos de 200 imagens foram unidas para processamento único (resolução máxima de $2.4 \text{ \AA}$).

Os dados coletados foram reduzidos e processados utilizando o pacote de programas XDS. As coordenadas atômicas de uma proteína homóloga com 68% de identidade foram utilizadas para aplicação do método da Substituição Molecular para a resolução da estrutura cristalográfica, utilizando para isto o programa Phaser_MR. Obtidas as fases iniciais, procedemos com o refinamento, buscando com isso a concordância entre os fatores de

estrutura calculado (F_c) e observado (F_o), que são correlacionados pelo R_{fator} de acordo com a equação:

$$R_{\text{fator}} = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|} \times 100\% \quad (23)$$

Onde k é um termo de escala entre as amplitudes dos fatores de estrutura medidos no experimento de difração F_{obs} e os mesmos fatores de estrutura calculados a partir do modelo F_{calc} .

Após a geração do modelo, para realizar o refinamento da estrutura, foi empregado o programa PHENIX⁷⁴, para os ajustes manuais do modelo à densidade eletrônica utilizou-se programa Coot.

As análises estereoquímicas foram feitas com o programa PROCHECK⁷⁵, cujos diagramas construídos indicam a qualidade do modelo. As imagens foram produzidas utilizando o programa Pymol (Sistema gráfico molecular, Delano).

4 Resultados e Discussão

4.1 Estratégias Computacionais.

Utilizando a base de dados do NCBI fizemos um levantamento de alvos para estudo tendo como objetivo, aumentar a eficiência da ação de Celulases. Para isto, optamos pelo estudo de proteínas auxiliares no processo de hidrólise de materiais lignocelulósicos como Swoleninas ou Expansina.

Selecionamos a sequência gênica descrita por Kim *et al*²⁸ para modelo nas buscas e fizemos um levantamento a respeito de quais organismos possuíam sequências homólogas preditas para a proteína Expansina. Como primeiro filtro nas buscas, colocamos organismos bacterianos envolvidos no processo de degradação a celulose ou a alguma doença diretamente relacionada a vasos lenhosos, retornando diversas sequências estatisticamente favoráveis utilizando a ferramenta BLAST.

4.1.1 Predição de genes a partir de banco de dados.

A proteína de estudo foi identificada a partir do banco de dados do NCBI como já descrito, tendo como modelo (*template*) BSEXLX1 (PDB id2bh0) retornando em relação a este a sequência da Celulase de *Xanthomonas campestris* (Glicosilhidrolase Xcc3535) com os seguintes valores: escore de 288 e E-value de 0, que são ótimos, pois, numa busca com esta ferramenta procuramos altos valores para escore e baixos para o E-value.

4.1.2 Análise da Sequência de Aminoácidos.

O primeiro passo na análise da sequência de aminoácidos do gene de trabalho XCC3535 foi conhecer a sua classificação predita no banco de dados do website CAZy (<http://www.cazy.org/>)²³. De acordo com o mesmo, a proteína de trabalho é classificada em duas famílias: Família 5 das Glicosilhidrolases (GH) e Família 53 dos módulos de ligação a carboidrato.

O segundo passo foi o estudo de regiões de peptídeo sinal. Uma sequência peptídeo sinal traduzida atua na orientação de uma proteína que será exportada para determinados locais, como por exemplo, o ambiente extracelular. Após a exportação, a proteína sinalizadora

é removida por ação de proteases, entretanto como se trata de uma expressão heteróloga (a técnica que será usada para expressar a proteína de trabalho) a secreção nem sempre acontece, dependendo da estrutura da proteína e da própria maquinaria do organismo, ou se acontece pode resultar em problemas na remoção da mesma caso seja necessária⁷⁶, dessa maneira evitar regiões que conferem tal atividade consiste numa estratégia no estudo de proteínas utilizando a técnica do DNA recombinante. Na figura 22 observamos o arquivo de saída do programa *SignalP*⁴², que identificou uma região até o resíduo 22 como de peptídeo sinal, pois apesar do Score S e C indicarem que a sequência apresentada não pertença a uma região deste tipo a combinação destes resultados possui valor característico de uma sequência peptídeo sinal, sendo que o resultado relevante nesta análise é a estatística S, selecionamos a região até o aminoácido 22 da sequência para ser excluído do gene em estudo para evitar problemas pós-traducionais relacionados com a mesma.

FIGURA 22 – Arquivo de saída do programa *SignalP*⁴², para avaliação da sequência de XCC3535.

O programa *Dommaing Fish* atribuiu dois domínios (figura 23) para a sequência, um ilustrado em vermelho (do resíduo 10 ao 370) com identidade de 95% com a proteína de

Xanthomonas oryza e também identidade sequencial de 38% com uma proteína de *Pyrococcus horikoshii* referente ao PDB 3axx, sendo que, este domínio foi classificado pertencente a Glicosilhidrolase da família 5. O outro domínio, ilustrado em cinza e azul (do resíduo 370 ao 550) composto por dois subdomínios característico de proteínas nomeadas como Expansinas, o domínio *Rare lipoprotein* (em cinza) com identidade de 38% a referida proteína (*Rare lipoprotein*) do organismo *Roseiflexus sp*, e o outro sub-domínio *Pollen allergen* (em azul) com identidade sequencial de 41% com a referida proteína (*Pollen allergen*) do organismo *Frankia sp*. O programa *Dommaing Fish* não apresentou PDB com identidade para esse dois subdomínios mas a partir da utilização do programa *Blastp* aplicando a sequência destes dois subdomínios finais foi possível identificar uma proteína com modelo cristalográfico PDB 3D3O com identidade sequencial de 68%.

Já o programa *CDD* atribui dois domínios para a sequência (figura 24). Um identificado como BgIC (endoglucanase) com alinhamento do resíduo 10 ao 290 e o outro domínio classificado como COG4305 (endoglucanase C-terminal) com alinhamento do aminoácido 356 a 569.

FIGURA 24 – Domínios conservados: obtidos a partir do CDD, anotação funcional específica, para avaliação do domínio da sequência XCC3535.

De posse das regiões conflitantes nomeadas como domínio, fizemos um levantamento de Glicosilhidrolases e Expansinas descritas na literatura. Utilizando o programa Clustal⁴⁰ realizamos múltiplos alinhamentos de maneira a confrontar quais seriam os limites de cada domínio, dentro da sequência XCC3535. Tomando como base as Glicosilhidrolases alinhadas 3AXX (Endoglucanase *Pyrococcus horikoshu*) e gi 188575103 (Glicosilhidrolase *Xanthomonas oryzae*) representadas na figura 25 foi possível a delimitação de uma região consenso para o domínio 1, nomeado de domínio catalítico. O mesmo procedimento foi realizado para a região conflitante predita como similar a expansina, para tal utilizamos 2BH0 (Proteína similar à Expansina de *Bacillus subtilis*) e 2HCZ (Expansina de *Zea mays*) cujo alinhamento esta representado na figura 26.

FIGURA 25 – Alinhamento múltiplo utilizando o programa Clustal, sequências alinhadas 3AXX (Endoglucanase *Pyrococcus horikoshu*), gi 188575103 (Glicosilhidrolase *Xanthomonas oryzae*) e a Cellulase de *Xanthomonas campestris* (proteína de estudo).

CLUSTAL 2.0.9 MULTIPLE SEQUENCE ALIGNMENT

FIGURA 26 – Alinhamento múltiplo utilizando o programa Clustal, sequências alinhadas Celulase de *Xanthomonas campestris* (proteína de estudo), 2BH0 (Proteína similar à Expansina de *Bacillus subtilis*) e 2HCZ (Expansina de *Zea mays*).

O resultado final confirmou o descrito pelo CAZy, com respeito a atribuição de duas famílias, que serão tratadas como domínios separados. Uma região predita como catalítica e uma de ação predita como semelhante à Expansina, que possui identidade de 68% com a

EXLX de *Bacillus subtilis*. Seguindo a nomenclatura para classificação de expansinas, a expansina identificada em *Xanthomonas campestris* seria uma *Expansin-likeX*, ou seja, proteínas que atuam da mesma maneira que expansinas, mas que não são do reino vegetal, sendo portando a nomenclatura desta proteína *Xanthomonas campestris* EXLX1 ou XcEXLX1. No quadro da figura 27, temos a identificação de cada região da proteína XCC3535.

FIGURA 27 - Identificação dos domínios para XCC3535. Em verde temos a sequência peptídeo sinal que foi removida, em vermelho a região delimitada como domínio um, em azul o domínio dois e em roxo o linker.

O último passo na análise das sequências de aminoácidos foi o cálculo teórico dos parâmetros físico-químicos da proteína inteira e de seus domínios identificados separadamente, utilizando a ferramenta *ProtParam*⁴⁷.

A Glicosilhidrolase de *Xanthomonas campestris* possui 569 aminoácidos e massa molecular de 60,17 kDa, com ponto isoelétrico teórico igual a 9,07. A tabela 11 mostra a porcentagem de aminoácidos na estrutura primária da proteína.

TABELA 11 - Porcentagem de aminoácidos na estrutura primária da Glicosilhidrolase de *Xanthomonas campestris*.

Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)
Ala (A)	13,5	Gly (G)	11,1	Pro (P)	6,3
Arg (R)	6,2	His (H)	1,6	Ser (S)	5,3
Asn (N)	4,4	Ile (I)	3,0	Thr (T)	5,4
Asp (D)	6,2	Leu (L)	9,8	Trp (W)	2,5
Cys (C)	0,7	Lys (K)	2,8	Tyr (Y)	1,9
Gln (Q)	5,1	Met (M)	1,8	Val (V)	7,9
Glu (E)	1,8	Phe (F)	2,8	Total	100

O coeficiente de extinção molar calculado é de $93390 \text{ M}^{-1}\text{cm}^{-1}$, considerando-se todos os resíduos de Cys reduzidos, valor que foi utilizado para medida de concentração das amostras de proteína, descrita na seção 3.2.6.

O domínio catalítico da Glicosilhidrolase de *Xanthomonas campestris* possui 318 aminoácidos e massa molecular de 33,27 kDa, com ponto isoelétrico teórico igual a 6,41. A tabela 12 mostra a porcentagem de aminoácidos na estrutura primária da proteína.

TABELA 12 - Porcentagem de aminoácidos na estrutura primária do domínio catalítico da Glicosilhidrolase de *Xanthomonas campestris*.

Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)
Ala (A)	16,7	Gly (G)	11,0	Pro (P)	5,3
Arg (R)	7,5	His (H)	1,3	Ser (S)	5,0
Asn (N)	4,4	Ile (I)	1,9	Thr (T)	3,8
Asp (D)	7,2	Leu (L)	9,4	Trp (W)	2,5
Cys (C)	1,3	Lys (K)	0,6	Tyr (Y)	0,9
Gln (Q)	6,0	Met (M)	2,2	Val (V)	9,8
Glu (E)	1,3	Phe (F)	2,8	Total	100

O coeficiente de extinção molar calculado é de $48470 \text{ M}^{-1}\text{cm}^{-1}$, considerando-se todos os resíduos de Cys reduzidos, valor que foi utilizado para medida de concentração das amostras de proteína, descrita na seção 3.2.6.

O domínio Expansina de *Xanthomonas campestris* possui 206 aminoácidos e massa molecular de 22,28 kDa, com ponto isoelétrico teórico igual a 9,13. A tabela 13 mostra a porcentagem de aminoácidos na estrutura primária da proteína.

TABELA 13 - Porcentagem de aminoácidos na estrutura primária do domínio semelhante à Expansina de *Xanthomonas campestris*.

Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)	Resíduo	Porcentagem (%)
Ala (A)	8,3	Gly (G)	11,2	Pro (P)	7,3
Arg (R)	3,9	His (H)	1,9	Ser (S)	5,3
Asn (N)	4,9	Ile (I)	4,4	Thr (T)	7,8
Asp (D)	4,4	Leu (L)	9,7	Trp (W)	2,4
Cys (C)	0,0	Lys (K)	4,9	Tyr (Y)	3,9
Gln (Q)	4,9	Met (M)	1,0	Val (V)	7,8
Glu (E)	2,9	Phe (F)	3,4	Total	100

O coeficiente de extinção molar calculado é de $39420 \text{ M}^{-1}\text{cm}^{-1}$, em 280 nm com a medida feita em água, valor que foi utilizado para medida de concentração das amostras de proteína, descrita na seção 3.2.6.

4.1.3 Modelagem por Homologia

Optamos pela modelagem do domínio catalítico isolado porque não encontramos uma estrutura modelo com identidade superior a 20 % para a proteína inteira. Então, optamos por obter um modelo para o domínio catalítico e de posse do modelo cristalográfico do domínio similar à Expansina (obtido por difração de monocristal), propor um modelo estrutural para a proteína inteira, que será útil na aplicação das técnicas de Dicroísmo Circular e SAXS e para o próprio conhecimento do possível formato da proteína de estudo.

A partir de alinhamentos iniciais utilizando o banco de dados do PDB, identificamos a estrutura 2ZUN (correspondente a endocelulase de *Pyrococcus horikoshii*) de maior identidade com a sequência de aminoácidos do domínio catalítico sendo esta de 28% ilustrado na figura 28.

```

CLUSTAL 2.0.12 multiple sequence alignment

22UN      MEGNTILKIVLICTILAGLFQGVVPVYAENTTYQTPTGIYYEVRGDTIYMINVTSGEETP 60
Gli      -----MQVSTQAP-----LVDATGQT 16
              :. : * :
              :. : .

22UN      IHLFGVNWFGFETPNHVVHGLWKRNWEDMLLQIKSLGFNAIRLPFCTESVKPGTQPIGID 120
Gli      LHIRGVTWPGFDRAGLAAGVMRNNTLAQLLDRMQASDINAVRVPVCAAVLQR-APVAAAE 75
              :* : * : * : * : * : * : * : * : * : * : * : * : * : * :
              :. : .

22UN      YSKNPDLRGLDSLQIMEKIIKKAGDLGIFVLLDYHRIGCTHIEPLWYTEDFSEEDFINTW 180
Gli      VAGDSTLRGLDSLQLLDVVHAASQRGMQVMFAFADGGCDDRAPLLG---AQQQAWTRGL 132
              :. : * : * : * : * : * : * : * : * : * : * : * : * : * :
              :. : .

22UN      IEVAKRFGKYWNVIGADLKNEPHSVTSPPAAYTDGTGATWGMGNPATDWNLAERIGKAI 240
Gli      VTLARRYGGNANVLGIDLGS-----SGYRN---ASWAGNAADQDWNRVASRAVARV 180
              :. : * : * : * : * : * : * : * : * : * : * : * : * : * :
              :. : .

22UN      LKVAPHWLIFVEGTQFTNPKTDSSYKWGYNAWWGGNLMVAKDYPVNLKPNKLVYSPHYVG 300
Gli      LAQAPRWVVGVEGVGSNAVCSDPERKAP-----GSNLQPFACVPLDIARRHLVLMPLKLAG 235
              * : * : * : * : * : * : * : * : * : * : * : * : * : * :
              :. : .

22UN      PDVYNQPYFGPAKGFPDNLPIWYHHFGYVKLELGYSVVIGEFGGKYGHGGDPRDVIWQN 360
Gli      PDRDTTDAF-AAPGFAQALPAMWQRDFG--QFAIDHAVVPVSLGGGLGDG-DPRDPAWQT 291
              ** : * : * : * : * : * : * : * : * : * : * : * : * : * :
              :. : .

22UN      KLVDMMIENKFCDFYWSWNPDSGDTGGILQDDWTTIWEDKYNNLKRLMDSCSKSSSTQ 420
Gli      ALSGYLANAGIRSAFLGSWETGNANNG----- 318
              * :. : : : * : * : * : * :
              :. : .

22UN      SVIRSTTPKTSNTSKKICGPAILIILAVFSLLLRRAPR 458
Gli      -----

```

FIGURA 28 – Alinhamento utilizando o Clustal do domínio catalítico(Gli) com a sequência para 22UN .

A correção do alinhamento está apresentada na figura 29.

```

>P1;22UN
structurex:22UN:34:A:458:A:::
-----QTPTGIYYEVRGDTIYMINVTSGEETP
IHLFGVNWFGFETPNHVVHGLWKRNWEDMLLQIKSLGFNAIRLPFCTESVKPGTQPIGID
YSKNPDLRGLDSLQIMEKIIKKAGDLGIFVLLDYHRIGCTHIEPLWYTEDFSEEDFINTW
IEVAKRFGKYWNVIGADLKNEPHSVTSPPAAYTDGTGATWGMGNPATDWNLAERIGKAI
LKVAPHWLIFVEGTQFTNPKTDSSYKWGYNAWWGGNLMVAKDYPVNLKPNKLVYSPHYVG
PDVYNQPYFGPAKGFPDNLPIWYHHFGYVKLELGYSVVIGEFGGKYGHGGDPRDVIWQN
KLVDMMIENKFCDFYWSWNPDSGDTGGILQDDWTTIWEDKYNNLKRLMD-----
-----*

>P1;gli
sequence:gli:1:A:318:A:::
-----MQVSTQAPLVDATGQT
LHIRGVTWPGFDRAGLAAGVMRNNTLAQLLDRMQASDINAVRVPVCAAVLQR-APVAAAE
VAGDSTLRGLDSLQLLDVVHAASQRGMQVMFAFADGGCDDRAPLLG---AQQQAWTRGL
VTLARRYGGNANVLGIDLGS-----SGYRNASWAGNAADQDWNRVASRAVARV
LAQAPRWVVGVEGVGSNAVCSDPERKAP-----GSNLQPFACVPLDIARRHLVLMPLKLAG
PDRDTTDAF-AAPGFAQALPAMWQRDFG--QFAIDHAVVPVSLGGGLGDG-DPRDPAWQT
ALSGYLANAGIRSAFLGSWETGNANNG-----
-----*

```

FIGURA 29 – Correção do alinhamento, em código Pir.

Preparado os arquivos, iniciamos a modelagem da cadeia principal utilizando o programa Modeller 9v8 e scripts escritos na linguagem Python. Foram gerados 100 modelos e na figura 30 abaixo apresentamos um gráfico dos modelos em função da energia, sendo que o modelo de menor energia foi o sétimo com energia de 2556.

FIGURA 30 – Gráfico de energia em função do modelo para o domínio catalítico.

A etapa seguinte foi a modelagem de loops, pois a superfície de um loop tende a participar nos contatos de um cristal, levando a uma mudança conformacional entre modelo e alvo de modelagem.

Visando melhorar a estrutura predita para o domínio catalítico, fizemos a modelagem de loop da região cujos loops encontravam-se em uma região de restrição no diagrama de Ramachandran. Para isso, selecionamos a região compreendida entre os loops de posição 110 a 123 e 151 a 167 e geramos 25 modelos com novos loops para esta região e escolhemos o modelo de menor energia. Em seguida avaliamos a cadeia lateral e otimizamos o modelo buscando a minimização de energia e a conformação correta da estrutura. Para avaliar a qualidade dos modelos aplicamos os programas de validação (Procheck, Errat, Verify_3D, Prove, Dope), durante o processo de obtenção do modelo. Apresentamos na figura 31 o diagrama de Ramachandran para o melhor modelo obtido, observamos que estereoquimicamente a estrutura é favorável com 90,9% dos resíduos em posição permitida, 8,7% em posição permitida, 0,4% em região generosamente permitida e nenhum aminoácido em região proibida.

FIGURA 31 – Diagrama de Ramachandran para o modelo do domínio catalítico.

Na figura 32 apresentamos a sobreposição entre a estrutura gerada para o domínio catalítico e o modelo utilizado 2ZUN. Observamos que, o modelo gerado pelo Modeller manteve a estrutura secundária e terciária em sua maior parte.

FIGURA 32 – Sobreposição entre a estrutura gerada para o domínio catalítico (verde) e o modelo utilizado 2ZUN (rosa), em destaque os loops que foram modelados.

4.2 Tecnologia do DNA recombinante

4.2.1 Extração de DNA genômico de *Xanthomonas campestris* PV *campestris* ATCC33913

O material genético do organismo *Xanthomonas campestris* pv *campestris* foi extraído utilizando o kit Wizard® Genomic DNA Purification (Promega) que permitiu uma amostra pura e de boa qualidade para os procedimentos de amplificação numa concentração de 351ng/uL.

4.2.2 Clonagem da sequência codificadora de Glicosilhidrolase XCC e de seus domínios catalítico e semelhante à Expasina

Utilizando oligonucleotídeos específicos, foi possível amplificar as ORFs de cada construção gênica de trabalho a partir do DNA genômico de *Xanthomonas campestris* pela técnica de PCR, e o resultado da amplificação foi analisado em gel de agarose 1% e apresentado na figura 33.

FIGURA 33 - Amplificação das ORF dos genes, Glicosilhidrolase 1707 pb, domínio Catalítico 954 pb e domínio Expansina 618pb. Resultado analisado em gel de agarose 1%, TAE 1X. M: marcador de massa molecular Gene Ruler 1Kb Ladder Plus (Fermentas Life Sciences).

4.2.3 Transformação de células de *E. coli* por choque térmico

A ligação do vetor de clonagem pGEM-T aos genes de estudo foi verificada por meio da reação de digestão com as enzimas de restrição *Bam*HI e *Xho*I e o resultado é mostrado na figura 34.

FIGURA 34 - Clivagem do vetor de subclonagem pGEM-T: pGEM-T linearizado 3000pb, Glicosilhidrolase 1707 pb, domínio Catalítico 954 pb e domínio semelhante à Expansina 618pb. Resultado analisado em gel de agarose 1%, TAE 1X. M: marcador de massa molecular Gene Ruler 1Kb Ladder Plus (Fermentas Life Sciences).

A ligação em vetor de expressão pSMT3 foi verificada por meio da reação de digestão com as enzimas de restrição *Bam*HI e *Xho*I e o resultado é mostrado na figura 35.

FIGURA 35 - Clivagem do vetor de clonagem ou expressão pSMT3 5.6633pb: pSMT3 linearizado, Glicosilhidrolase 1707 pb, domínio Catalítico 954 pb e domínio semelhante à Expansina 618pb.

Resultado analisado em gel de agarose 1%, TAE 1X. M: marcador de massa molecular Gene Ruler 1Kb Ladder Plus (Fermentas Life Sciences).

A análise de restrição do vetor de expressão pSMT3 para a Glicosilhidrolase, domínio catalítico e domínio semelhante à Expansina confirmaram a inserção do inserto no vetor.

4.2.4 Sequenciamento de DNA e análise de seqüências

O sequenciamento das ORF_s para Glicosilhidrolase, domínio catalítico e domínio semelhante à Expansina em vetor pSMT3 foi realizado pelo método de Sanger e as seqüências de nucleotídeos obtidas foram analisadas e alinhadas com a seqüência do banco de dados através do programa *ContigExpress* (a component of vector NTI Advance 11.0, Invitrogen Corporation) e o resultado do alinhamento das extremidades 5' e 3' para as três construções mostram-se condizentes com a seqüência depositada não apresentando nenhuma mutação, confirmando que a seqüência de trabalho é a mesma do organismo *Xanthomonas campestris*. Com intuito ilustrativo a figura 36, apresenta o início, meio e fim do alinhamento, da seqüência para Glicosilhidrolase de trabalho comparada com a depositada no NCBI.

FIGURA 36 –Recortes do cromatograma de sequenciamento da ORF para Glicosilhidrolase, na figura S2 e S3 representam as seqüências no sentido direto e inverso respectivamente, Glicosilhidrolase a seqüência depositada no NCBI e Contig a seqüência consenso do alinhamento das seqüências amplificadas.

4.2.5 Linhagens de bactérias e plasmídeo utilizados para a expressão das proteínas

As linhagens selecionadas para expressão das proteínas Glicosilhidrolase e seus domínios Catalítico e Expansina todos em vetor pSMT3, foram a *E. coli* Rosetta(DE3)pLysS (Novagen) e a *E. coli* BL 21 (Novagen). A transformação do vetor pSMT3+inserto nestas linhagens foi aferida por controle negativo da reação de transformação. Para garantir eficiência no processo, preparamos as células quimiocompetentes a partir das cepas originais do fabricante.

A partir da seleção de colônias pSMT3+inserto, em *E. coli* Rosetta(DE3)pLysS (Novagen) e em *E. coli* BL 21 (Novagen) fizemos testes de expressão a 37°C (4 horas de expressão) e 18°C (16 horas de expressão) a 160 rpm, avaliando a expressão das proteínas de trabalho tendo como foco a fração solúvel. Para a construção pSMT3+Glicosilhidrolase, a fração solúvel foi maior para a cepa *E. coli* Rosetta(DE3)pLysS (Novagen) e a melhor temperatura de expressão foi a 18°C por 16 horas, enquanto que para a construção pSMT3 domínio catalítico, a linhagem de melhor expressão foi *E. coli* Rosetta(DE3)pLysS (Novagen) na temperatura de 18°C e para pSMT3+Expansina, obtivemos uma boa fração solúvel para todas as duas linhagens bacterianas e para as duas condições de expressão. Também variamos a quantidade de agente indutor IPTG, de 0.1mM a 1mM e obtivemos melhores resultados a 1mM.

Desta maneira, tendo em vista as melhores condições de expressão padronizamos o seguinte: linhagem de trabalho *E. coli* Rosetta(DE3)pLysS (Novagen) a 18°C, 160 rpm por 16 horas e 1mM de IPTG $D.O_{595nm}=0.5$.

4.2.6 Expressão e purificação das proteínas recombinantes

A expressão da Glicosilhidrolase XCC3535 e de seus domínios isolados foi verificada utilizando a Eletroforese em gel de poliacrilamida (SDS-PAGE 15%), apresentado na figura 37. Notemos que a massa molar das proteínas estão acrescidas do Tag (Sumo), logo a massa molecular esperada das proteínas de trabalho serão: Glicosilhidrolase 60.17 kDa com o Sumo 13.8 kDa, massa molar resultante de 73.97 kDa, Domínio Catalítico 33.27 kDa com o Sumo 13.8 kDa, massa molar resultante de 47.07 kDa, Domínio semelhante à Expansina 22.28 kDa com o Sumo 13.8 kDa, massa molar resultante de 36.08 kDa.

FIGURA 37 - A. Expressão da Glicosilhidrolase. Na 1ª linha estão os marcadores de massa molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibidor de tripsina de soja (20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). A próxima coluna representa o extrato de proteínas antes da indução, a segunda linha o extrato de proteínas pós indução. As 3ª e 4ª colunas representam, respectivamente, pélete da lise celular e sobrenadante da lise celular. **B.** Expressão do domínio Catalítico. Na 1ª coluna estão os marcadores de massa molecular o mesmo já descrito. A próxima coluna representa o extrato de proteínas antes da indução, a segunda coluna o extrato de proteínas pós indução. As 3ª e 4ª colunas representam, respectivamente, pélete da lise celular e sobrenadante da lise celular. **C.** Expressão do domínio semelhante à Expansina. Na 1ª coluna estão os marcadores de massa molecular já descritos. A próxima coluna representa o extrato de proteínas antes da indução, a segunda linha o extrato de proteínas pós indução. As 3ª e 4ª colunas representam, respectivamente, pélete da lise celular e sobrenadante da lise celular.

Notamos que a Glicosilhidrolase possui uma boa fração solúvel, assim como a proteína semelhante à Expansina. Contudo, o domínio catalítico possui uma pequena fração solúvel, decorrência esta de problemas no enovelamento, possivelmente advindos da formação de ligações intermoleculares dissulfeto.

As frações do sobrenadante das proteínas foram purificadas por cromatografia de afinidade, realizada em resina Ni-NTA *super flow* (Quiagen), como descrito na seção 3.2.6. A proteína foi eluída da resina pela aplicação do tampão com diferentes concentrações de Imidazol, o que pode ser visualizado em gel SDS-PAGE 15%, como mostram as figuras 38, 39 e 40.

FIGURA 38 - Purificação da Glicosilhidrolase. Na 1ª coluna estão os marcadores de massa molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibidor de tripsina de soja(20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). A próxima coluna representa o *void* celular, a segunda a lavagem da coluna. A 3ª a eluição em 300mM de imidazol e a 4ª linha a eluição em 500mM de imidazol.

Notamos que a Glicosilhidrolase (60.17 kDa+13.8 kDa-Sumo) interage fortemente com a resina Ni-NTA *super flow* (Quiagen), exigindo altas concentrações de imidazol para ser eluída.

FIGURA 39 - Purificação do domínio Catalítico. Na 1ª coluna estão os marcadores de massa molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibidor de tripsina de soja(20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). A próxima coluna representa o *void* celular, a segunda a lavagem da coluna. As 3ª e 4ª coluna representam, respectivamente, eluição a 300mM e 500mM de imidazol.

Notamos que o domínio Catalítico (33.27 kDa+13.8 kDa-Sumo) também interage fortemente com a resina Ni-NTA *super flow* (Quiagen), exigindo altas concentrações de imidazol para ser eluída.

FIGURA 40 - Purificação da Expansina. Na 1ª coluna estão os marcadores de massa molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibidor de tripsina de soja (20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). A próxima coluna representa o *void* celular, a segunda a lavagem da coluna. As 3ª e 4ª linhas representam, respectivamente, eluição a 100mM e 500mM de imidazol.

Contrapondo as anteriores a Expansina (22kDa+14kDa-Sumo) possui fraca interação com a resina Ni-NTA *super flow* (Quiagen), exigindo baixas concentrações de imidazol para ser eluída. Após a purificação por coluna de afinidade, as amostras foram clivadas para eliminação do Tag (His-Sumo) e purificadas uma segunda vez por cromatografia de afinidade, contudo para eliminação de contaminantes do processo de clivagem, utilizamos da cromatografia de exclusão molecular, os resultados da aplicação desta técnica podem ser vistos nas figuras 41, 42 e 43.

A

B

FIGURA 41 - Cromatograma da exclusão molecular da Glicosilhidrolase. B Frações do pico do cromatograma. Na 1ª coluna estão os marcadores de massa molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibidor de tripsina de soja (20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). Da coluna a sexta linha representam respectivamente os volumes de 8,5; 9; 9,5; 10; 10,5; 11 e 11,5 mL.

As frações dos picos do cromatograma de exclusão molecular para a Glicosilhidrolase analisados por SDS-PAGE 15%, figura 41-B, mostram que a massa aparente da proteína está correta 60.17kDa, e que o grau de pureza atingido com a purificação não foi de 100%, contudo a amostra encontra-se em condições para as caracterizações biofísicas necessárias.

FIGURA 42 - Cromatograma da exclusão molecular do domínio catalítico. *B* Frações do pico do cromatograma. Na 1ª coluna estão os marcadores de massa molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibidor de tripsina de soja(20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). Da segunda a sexta coluna representam respectivamente os volumes de 7,7.5,8,8.5 e 9 mL.

As frações dos picos do cromatograma de exclusão molecular para o domínio catalítico analisados por SDS-PAGE 15%, figura 42-B, mostram que a massa aparente da proteína está correta 33.27 kDa, porém o rendimento do processo foi muito baixo.

FIGURA 43 - Cromatograma da exclusão molecular do domínio Expansina. B Frações do pico do cromatograma. Na 1ª coluna estão os marcadores de peso molecular albumina bovina (66kDa), ovoalbumina (45kDa), gliceraldeído-3-fosfato desidrogenase (36kDa), anidrase carbônica (29kDa), tripsinogênio de pâncreas bovino (24kDa), inibido de tripsina de soja(20,1kDa) e α -lactalbumina de leite bovino (14,2kDa). Da segunda a sexta coluna representam respectivamente os volumes de, 15, 15.5, 16, 16.5 e 17 mL.

As frações dos picos do cromatograma de exclusão molecular para a Expansina analisados por SDS-PAGE 15%, figura 43-B, mostram que a massa aparente da proteína está correta 22.28 kDa, e que o grau de pureza atingido com a purificação foi próximo a 100%, sendo que a amostra encontra-se em condições para as caracterizações biofísicas necessárias.

A quantidade protéica total de cada solução foi determinada em função da curva de absorbância padrão obtida para a proteína albumina soro bovina seguindo o método de *Bradford*⁵⁵. Após a primeira purificação na coluna de níquel, possuíamos 30mg, 6mg e 50mg de Glicosilhidrolase, domínio Catalítico e domínio semelhante à Expansina, respectivamente, quantidade esta provinda da expressão de um litro de cultura de *E. coli* Rosetta(DE3)pLysS (Novagen) a 18°C. Na etapa final pós clivagem e segunda purificação na coluna de afinidade e purificação por exclusão molecular chegamos a um rendimento de 24mg, 4mg e 30 mg de Glicosilhidrolase, domínio Catalítico e Expansina respectivamente para cada litro de cultura.

4.3 Caracterização biofísica

4.3.1 Estimativa da massa molecular

A curva de calibração para Massa Molecular (MM) está representada na figura 44.

FIGURA 44 - Curva de calibração da SUPERDEX 75TM 10/300 para o cálculo da massa molecular.

A partir da regressão linear da curva apresentada utilizando o programa Origin, obtemos a equações 24 para determinação da massa molecular.

$$K_{av} = (-0,356) \log(MM) + (0,68) \quad (24)$$

Os volumes de eluição para as proteínas de trabalho, apresentados na seção 4.2.6, foram: Glicosilhidrolase 16mL, Domínio Catalítico 9 mL e Domínio semelhante à Expansina 16mL. Aplicando estes valores na equação 24, obtemos que a massa molecular para cada proteína: Glicosilhidrolase 46 kDa, Domínio Catalítico 109 kDa e Domínio semelhante a Expansina 4 kDa.

Os resultados obtidos a primeiro momento poderiam ser tomados como inconsistentes, entretanto a Coluna Superdex 75 10/300 é composta por uma malha de agarose e dextrana que são açúcares e a proteína de trabalho Glicosilhidrolase possui um domínio específico de ligação a carboidratos, o domínio semelhante à Expansinas. Desta maneira, o retardo na eluição da proteína Glicosilhidrolase e do domínio semelhante à Expansina estão relacionados

não a problemas com a resolução da coluna empregada, mas sim devido a interações entre a matriz da coluna e a proteína, este atraso para a Glicosilhidrolase é de aproximadamente 0,5 mL na eluição e para o domínio semelhante a Expansina é de 4 mL, isto deve-se ao fato de que na Glicosilhidrolase o domínio Expansina corresponde a apenas um terço da massa total da estrutura logo seu efeito é minimizado pelo restante da proteína o que não acontece com o domínio semelhante à Expansina que interage fortemente. Quanto ao domínio catalítico, observamos que o mesmo é eluído antes do esperado o que pode evidenciar a formação de agregados ou um trímero devido à massa molecular aparente obtida com esta técnica.

4.3.2 Espectroscopia de Dicroísmo Circular

Foram realizados experimentos de espectroscopia de Dicroísmo Circular para a Glicosilhidrolase e domínio semelhante à Expansina; para o domínio Catalítico não obtivemos amostra suficiente para realizar o experimento. As medidas foram avaliadas em tampão Tris e tampão fosfato, sendo que, o melhor espectro obtido foi o coletado com as amostras protéicas em tampão fosfato. A amostra de Glicosilhidrolase estava a 0,3 mg/mL e 0,35 mg/mL em tampão tris e fosfato, respectivamente, da mesma forma estava o domínio semelhante à Expansina em concentração 0,28 mg/mL e 0,30 mg/mL.

No espectro obtido para a Glicosilhidrolase (figura 45), temos um pico negativo em 225 nm, devido a transições do tipo $n-\pi^*$ na ligação peptídica, além de outro pico negativo em 208 nm e uma banda positiva intensa em 190 nm, devido a transições $\pi-\pi^*$. Comparando com o espectro típico de estruturas secundárias em proteínas (figura 17), temos um espectro de Dicroísmo Circular provindo da mistura de hélices- α e folhas β .

FIGURA 45 - Espectro de Dicroísmo Circular da Glicosilhidrolase em tampão fosfato.

No espectro do domínio semelhante à Expansina (figura 46), temos um pico negativo em 208 nm devido a transições $n-\pi^*$ e um pico positivo em 226 nm cuja causa ainda não é muito entendida, entretanto discutimos duas possíveis razões para tal espectro: A primeira seria devido a presença de resíduos aromáticos contribuírem para uma região positiva do espectro entre 215 a 230 nm, e a estrutura da proteína semelhante à Expansina possui três aminoácidos característicos para sua atividade funcional formando um plano na superfície da estrutura como será discutido na seção de Difração de Raios-X. Segundo o que temos descrito na literatura em algumas proteínas, a interação específica entre certos aminoácidos aromáticos dão ganhos no sinal do UV distante⁷³, o que pode afetar a análise de estruturas secundárias por CD. Uma segunda causa advém do espectro de dicroísmo circular de regiões desordenadas que contribuem para um pico positivo em 220 nm^{74; 75}, a proteína em estudo possui uma alta porcentagem de estruturas secundárias deste tipo, logo a associação do espectro de CD de estruturas desordenadas associado ao de folhas β pode ter originado um espectro característico como o obtido.

FIGURA 46 - Espectro de Dicroísmo Circular da proteína similar à Expansina em tampão fosfato.

As tabelas 14 e 15 apresentam o teor de estrutura secundária calculada para a Glicosilhidrolase e para o domínio semelhante à Expansina, respectivamente. Foram utilizados os algoritmos ContinII, CDSSTR e SELCON3 com o banco de referência espectral *29-Soluble*. Nota-se que os valores de porcentagem para α -hélice e folhas β calculados pelos programas de desconvolução são muito próximos aos valores apresentados pelo modelo cristalográfico para Glicosilhidrolase (modelo este obtido pela união do modelo estrutural por homologia para o domínio catalítico obtido com o Modeller e domínio semelhante à Expansina obtido por difração de Raios-X monocristal) e para o modelo cristalográfico

(difração de Raios-X em monocristal, este modelo está descrito na última seção da dissertação) do domínio semelhante à Expansina, dados apresentados na tabela 16. Entretanto, uma análise da sobreposição dos espectros de dicroísmo teórico e experimental é necessária para validar este resultado na figura 47, temos a sobreposição para a Glicosilhidrolase. Na figura 48 para o domínio similar à Expansina, determinados tomando como base teórica os dados 29-*Soluble* e utilizando o programa de desconvolução SELCON3.

TABELA 14 - Estimativas de estrutura secundária para Glicosilhidrolase.

	Hélices- α (%)	Folhas- β (%)	Voltas (%)	Região Desordenada(%)
ContinII	17,2	33,5	21,3	28,1
CDSSTR	11,0	28,8	23	37,2
Selcon 3	14,6	29	20	36,4
Média	14,2	30,4	21,4	33,9

TABELA 15 - Estimativas de estrutura secundária para o domínio semelhante à Expansina.

	Hélices- α (%)	Folhas- β (%)	Voltas (%)	Região Desordenada(%)
ContinIII	3,9	48	20,5	27,4
CDSSTR	4,3	47,9	19,1	27,3
Selcon 3	4,8	45,1	21,4	27,4
Média	4,33	47	20,33	27,36

TABELA 16 - Estrutura secundária para o modelo cristalográfico (Modeller) para Glicosilhidrolase e para o modelo cristalográfico (difração de raios-X em monocristal) do domínio semelhante à Expansina.

	Hélices- α (%)	Folhas- β (%)	Voltas (%)	Região Desordenada(%)
Glicosilhidrolase	17,16	23,9		58,94
Expansina	3,9	38,8		57,8

Observamos que para a Glicosilhidrolase os valores teóricos e experimentais para a elipticidade molar em função do comprimento de onda apresentam valores próximos diferindo apenas na intensidade do sinal.

FIGURA 47 - Sobreposição do Espectro de Dicroísmo Circular teórico e experimental para a Glicosilhidrolase.

Para o domínio similar à Expansina os valores calculados e teóricos para a elipticidade molar em função do comprimento de onda não foram condizentes, uma das possíveis advém da carência de um maior número de espectro nos bancos de dados para uma melhor comparação, para tanto testamos outras duas bases de dados a SP 37 Soluble e a SP 43 Soluble, utilizando os programas de desconvolução já citados, cujas porcentagens de estrutura secundária predita estão na tabela 17, apresentamos na figura 49, o gráfico teórico e experimentais VS comprimento de onda advindos do banco SP 43 Soluble utilizando o programa SELCON3. Observamos que o gráfico da figura 49 é mais condizente com os dados do que o gráfico 48, contudo não descreve o perfil do espectro para a região problemática já descrita.

FIGURA- 48 Sobreposição do Espectro de Dicroísmo Circular teórico e experimental para a proteína similar à Expansina em tampão fosfato.

TABELA 17 - Estrutura secundária para o modelo cristalográfico (Modeller) para Glicosilhidrolase e para o modelo cristalográfico (difração de raios-X em monocristal) do domínio semelhante à Expansina.

Expansina	Hélices- α (%)	Folhas- β (%)	Voltas (%)	PP2	Região Desordenada(%)
SP 37 Soluble	2,0	33,9	8,7	11,9	57,8
SP 43 Soluble	3,3	38,2	21,5		37

FIGURA- 49 Sobreposição do Espectro de Dicroísmo Circular teórico e experimental para a proteína similar à Expansina em tampão fosfato utilizando a base de dados SP 43 e desconvoluído no programa SELCON3.

A estabilidade térmica das proteínas foi avaliada por CD, aplicando uma variação de temperatura de 10°C a 100°C apresentamos a avaliação da estabilidade da Glicosilhidrolase para as temperaturas de 10°C, 40°C e 50°C e a estabilidade para o domínio semelhante à Expansina nas temperaturas de 10°C, 45°C e 50°C. Na figura 50, temos o espectro da Glicosilhidrolase, observando a mudança no espectro entre 40°C e 45°C, correspondente a faixa da temperatura de desnaturação. Já na figura 51, temos o espectro da proteína semelhante à Expansina, observando a mudança no espectro entre 45°C e 50°C. As proteínas de estudo não apresentaram reversibilidade após a desnaturação. Os dados obtidos referentes às temperaturas de desnaturação são importantes para estudos cinéticos em que esta proteína possa vir a ser aplicada. O espectro de dicroísmo circular para a proteína similar à Expansina devido seu perfil diferenciado merece novos estudos e uma melhor caracterização para uma melhor compreensão do pico em 226 nm.

FIGURA 50 - Espectro de Dicroísmo Circular da Glicosilhidrolase em 10, 40 e 45°C

FIGURA 51 - Espectro de Dicroísmo Circular do domínio semelhante Expansina, em 10, 45 e 50°C.

4.3.3 Espalhamento Dinâmico de Luz (DLS)

A técnica de DLS foi utilizada para verificação da homogeneidade das soluções protéicas, avaliando-se com isto o estado oligomérico para aplicação da técnica de SAXS.

Para a amostra da Glicosilhidrolase, obtemos um raio de Stokes médio (raio hidrodinâmico) de (35 ± 6) Å e uma massa molecular média de (65 ± 12) kDa calculada pelo programa do equipamento Zetasizer μ V(Malvern), indicando a predominância do estado monomérico em solução, apresenta-se monodispersa, com coeficiente de polidispersão de

16,4% com a relação entre intensidade e distribuição do tamanho das partículas em solução apresentada na figura 52.

FIGURA 52 – Distribuição do tamanho pela intensidade para a amostra de Glicosilhidrolase.

Para a amostra da proteína semelhante à Expansina, obtemos um raio de Stokes médio (raio hidrodinâmico) de (22 ± 2) Å e uma massa molecular média de (22 ± 2) kDa calculada pelo programa do equipamento Zetasizer μ V(Malvern) como se a proteína fosse uma esfera com o raio de Stokes, indicando a predominância do estado monomérico em solução, apresenta-se monodispersa, com coeficiente de polidispersão de 9,9% com a relação entre intensidade e distribuição do tamanho das partículas em solução apresentada na figura 53.

FIGURA 53 – Distribuição do tamanho pela intensidade para a amostra do domínio semelhante à Expansina.

4.3.4 Espalhamento de Raios-X a Baixo Ângulo (SAXS)

As curvas de espalhamento da solução de Glicosilhidrolase e do domínio semelhante à Expansina bem como de seus tampões foram coletadas em 2 etapas de 300 segundos para monitoramento dos danos causados pela radiação à amostra, além da estabilidade do feixe. Análise sistemática foi realizada subtraindo os dados pelo ruído gerado pelo CCD no mesmo tempo de exposição, em seguida, normalizando-os pela intensidade do feixe incidente e multiplicando-os pela absorção da amostra. O espalhamento da solução tampão foi subtraído da curva de espalhamento da solução de proteína.

Finalmente os padrões bidimensionais da Glicosilhidrolase e do domínio Expansina foram integrados em 2θ . O R_g foi obtido por dois métodos distintos, utilizando a equação de Guinier⁷⁷ (figura 54) e pelo método de Transformada Inversa de Fourier implementado no programa Gnom⁶⁷, onde também obtivemos a função de distribuição de distâncias $p(r)$ e o $D_{m\acute{a}x}$ das proteínas.

FIGURA 54- Obtenção do raio de giro pelo método de Guinier: (A) Glicosilhidrolase (B) Domínio semelhante à Expansina.

Aplicando a aproximação de Guinier na curva de espalhamento ($qR_g < 1,3$) obtemos um raio de giro da Glicosilhidrolase e do domínio semelhante à Expansina iguais a $(36,7 \pm 0,7)$ Å e de $(20,2 \pm 0,6)$ Å, respectivamente, que estão de acordo com aqueles obtido pelo Gnom de $(36,8 \pm 0,05)$ Å e $(20,16 \pm 0,03)$ Å para Glicosilhidrolase e domínio semelhante à Expansina, respectivamente. Observamos que, no Gnom os pontos iniciais da curva de espalhamento foram eliminados devido à evidência de pequenos agregados na solução protéica medida.

Estes agregados não estavam presentes nas medidas de DLS feitas para confirmar o estado monomérico de ambas as proteínas, uma das causas da formação dos mesmos advém no intervalo de tempo entre o preparo da amostra até a coleta que foi de um dia. Os raios de giro gerados pelos dois métodos para ambas as proteínas confirmam os obtidos aplicando a técnica de DLS, descritos na seção anterior em que o raio de giro da Glicosilhidrolase determinado foi de (35 ± 6) Å e para o domínio semelhante à Expansina foi de (22 ± 2) Å.

Utilizando o servidor *SAXS MoW*⁷⁸ foi possível estimar a massa molecular da partícula em solução de Glicosilhidrolase e domínio Expansina como sendo de 83,5 kDa e 22,6 kDa, respectivamente. Estes valores indicam a predominância de monômeros em solução sendo o erro relativo para a massa molecular teórica 38,8 % e 1,2% para Glicosilhidrolase e domínio semelhante à Expansina, respectivamente.

A discrepância com o valor teórico obtido para a Glicosilhidrolase utilizando o servidor *SAXS MoW*⁷⁸ pode ser atribuído à flexibilidade estrutural da proteína⁷⁹ em estudo que possui um *linker* de 43 aminoácidos. O estado monomérico das proteínas foi confirmado aplicando a técnica de DLS, descrita na seção 3.1.3, na qual a massa molar das proteínas em

solução foi determinada em (65 ± 12) e (22 ± 2) kDa para Glicosilhidrolase e domínio semelhante a Expansina, que confirmam os valores teóricos para as mesma, estando a diferença no valor teórico e o experimental embutidos no próprio erro da técnica.

As distribuições de frequências de distâncias ($p(r)_s$) da Glicosilhidrolase e do domínio semelhante à Expansina são gaussianas deslocadas para esquerda. Sendo $D_{máx} = (125 \pm 5)$ Å para a Glicosilhidrolase e $D_{máx} = (65 \pm 5)$ Å para o domínio similar à Expansina. As $p(r)$ não contém parte negativa no eixo das abscissas, o qual indica ausência de efeitos de interferência produzidas por correlações espaciais inter-partículas (figura 55).

Os valores de R_g , D_{max} foram calculados utilizando o programa GNOM⁶⁷. Após a eliminação dos pontos de ajuste da curva de espalhamento obteve-se que os novos valores de q , para a Glicosilhidrolase são q_{Min} de $0,01932 \text{ Å}^{-1}$ e q_{Max} de $0,2291 \text{ Å}^{-1}$, logo uma resolução de $27,41 \text{ Å}$ e para o domínio semelhante à Expansina, q_{Min} de $0,04098 \text{ Å}^{-1}$ e q_{Max} de $0,2414 \text{ Å}^{-1}$, logo uma resolução de $26,03 \text{ Å}$.

A

B

FIGURA 55 - Função de distribuição de distâncias para os dados experimentais da Glicosilhidrolase (A) e para o domínio semelhante à Expansina (B).

A partir dos dados de SAXS foram obtidos modelos de átomos de *dummy* de maneira *ab initio* utilizando o programa Dammin⁸⁰. Para verificar a unicidade do modelo, simulações independentes partindo de diferentes parâmetros iniciais e sem imposição de simetria foram realizadas, chegando a resultados consistentes, para a Glicosilhidrolase figura 56A. A curva de espalhamento experimental sobre a curva simulada do modelo de átomos *dummy* se sobrepõe de maneira bem condizente, a mesma observação é válida para o domínio semelhante à Expansina figura 56B. A curva de espalhamento simulado das coordenadas dos modelos de átomo *dummy* e de alta resolução foi calculada utilizando o programa Crysol⁶⁸.

A

3

FIGURA 56 - Curva de espalhamento de raios-X a baixo ângulo.(A) Para a Glicosilhidrolase DAM (Dammin) modelo simulado e Exp é a curva experimental (B) Para domínio semelhante a Expansina DAM (Dammin) modelo simulado, PDB modelo de alta resolução e Exp é a curva experimental.

O modelo *ab initio* da Glicosilhidrolase obtido pelo modelo de átomos de *dummy* foi posicionado sobre o modelo cristalográfico por homologia para o domínio catalítico, obtido com o uso do programa Modeller⁸¹ e sobre o domínio semelhante à Expansina obtido por difração de monocristal, com o uso do programa de modelagem de corpo rígido Massha. O volume do envelope dos modelos de átomos de *dummy* explica bem os volumes dos domínios (figura 57). Para o domínio semelhante à Expansina utilizando o programa *supcomb20* o modelo *ab initio* foi posicionado sobre o cristalográfico da Expansina (figura 58). Os parâmetros estruturais obtidos a partir dos dados de SAXS, do DAM e do PDB podem ser visualizados na tabela 18.

TABELA 18 - Parâmetros estruturais da Expansina de *Xanthomonas campestris* obtidos por SAXS.

Parâmetros/Am	Glicosilhidrolase			Domínio Expansina		
	Exp. *	PDB.†	DAM.‡	Exp. *	PDB.†	DAM.‡
D_{máx} (Å)	125 ± 5	-	123 ± 5	52 ± 5	64,11	65
R_g (Å)	36,80 ± 0,045	-	36,6 ± 0,2	17,81 ± 0,02	17,65	20,2
Resolução (Å) €	27,03	-	27,41	16	2,75	26

* Exp., calculado a partir dos dados experimentais a 3 mg/ml para a Glicosilhidrolase e a 2,8mg/mL para o domínio semelhante à Expansina.

† PDB id: gerado por modelagem para a Glicosilhidrolase e para o domínio Expansina temos o próprio PDB,

‡ DAM, parâmetros do modelo de átomos dummy (DAM) gerado sem imposição de simetria no programa Dammin.

€ Resolução $2\pi/q_{máx}$.

FIGURA 57 - Sobreposição do modelo de corpo rígido obtido com o programa Masha com o DAM para a Glicosilhidrolase. **A** posição inicial, **B** rotação de 90° em relação ao eixo x para a posição inicial e **C** visão lateral em relação ao primeiro.

FIGURA 58 - Sobreposição do DAM do domínio Expansina com o PDB. **A** posição inicial, **B** rotação de 45° em relação ao eixo y com relação a posição inicial e **C** rotação de 45° em relação ao eixo x e rotação de 90° em relação a z com relação a posição inicial, tomando como coordenadas o plano cartesiano xyz.

Observando que os modelos de DAM apresentados nas figuras 57 e 58 explicam bem as estruturas sobrepostas englobando-as adequadamente.

O valor de χ , em média, foi cerca de 2,01 para Glicosilhidrolase e 2,30 para domínio semelhante a Expansina, para os 10 modelos, o que se mostra um bom ajuste dos dados no programa *DAMMIN*.

Por fim, como contribuição da aplicação desta técnica pode-se citar a obtenção de um envelope para a Glicosilhidrolase, o qual modula bem o modelo cristalográfico obtido para esta proteína, o que pode ser um indicativo que o modelo provindo do Modeller descreve adequadamente a proteína. A partir deste envelope fica evidente a presença de um linker conectando os domínios da Glicosilhidrolase o que evidencia que a técnica computacional aplicada para a definição de domínios foi válida. Para a proteína similar à expansina podemos concluir que a mesma apresenta-se estável em solução e possui a forma monomérica.

4.3.5 Ensaios de Cristalização

A partir da varredura inicial de condições de cristalização realizada no IFSC foi possível determinar uma condição de cristalização para a proteína semelhante à Expansina na condição: 0.1M de fosfato de sódio monobásico, 0.1M de fosfato de potássio monobásico, 0.1M de MES pH 6.5 e 2M de cloreto de sódio, com a proteína na concentração de 20mg/mL e 40 mg/mL, com formação de cristais em 24h para a concentração elevada de proteína e de três a quatro dias a de menor concentração. Para a Glicosilhidrolase não foi possível identificar nenhuma condição de cristalização. Uma das causas para isto pode estar relacionada ao *linker* de 43 aminoácidos que separa os domínios da mesma, o qual permite uma grande flexibilidade da estrutura interferindo diretamente no estado de equilíbrio ou arranjo protéico que é fundamental para o processo de cristalização.

A condição determinada para o domínio semelhante à Expansina foi reproduzida com soluções preparadas reproduzindo as especificações do fabricante (Hampton) para a condição de cristalização, aplicando-se a técnica da gota suspensa, descrita na seção 3.3.5. Obtivemos cristais bem formados e homogêneos como ilustrado na figura 59.

FIGURA 59 – Cristais do domínio semelhante à Expansina.

Com a varredura de condições de cristalização realizada no HTLXLab não obtivemos novas condições de cristalização para o domínio semelhante à Expansina. Já para o domínio Catalítico (com HIS-SUMO), foi possível determinar uma condição de cristalização (0,1 M de Hepes, 0,8 M de Sulfato de Amônio e pH 7,0) que está ilustrado na figura 60. Entretanto, a mesma ainda não foi reproduzida por dificuldades na obtenção da proteína domínio catalítico em quantidade suficiente para realizar os testes. A amostra foi purificada no Brasil e na França seguindo o mesmo protocolo e cepas de bactéria para expressão do mesmo tipo, com o mesmo vetor utilizado para a expressão nos dois países. Contudo, as condições de reprodutibilidade do experimento não foram alcançadas, e novos testes serão necessários para obtenção da proteína em quantidade suficiente para reprodução da condição de cristalização.

FIGURA 60 – Evolução das gotas de cristalização(*Drop position*) A- 19-Dec-2011, B-23- Dez-2011, C- 03-Jan-2012, D-25 Jane-2012 e E-17 Fec-2012, para a condição 0.1 M de ADA pH 6.5 e 12 %v/v MPD, com a proteína em mg/mL. FZ026332.

Na preparação do cristal da proteína semelhante à Expansina, para coleta, obteve-se como melhor condição de difração a aplicação de um gradiente de glicerol de 5 a 20% no cristal, o que permitiu um melhor padrão de difração, mesmo com a própria condição de cristalização sendo crio, o ganho na aplicação desta técnica foi sensível, nas coletas realizadas no IFSC tínhamos um primeiro momento uma resolução máxima de 3,5Å, com a aplicação da condição crio conseguimos aumentar essa resolução para quase 3 Å. Para os testes de difração no E.S.R.F, foi aplicado o gradiente de glicerol antes da exposição do cristal nas condição de difração.

4.3.6 Difração de Raios-X

As estatística de coleta de dados de difração de raios-X coletados no Instituto de Física da USP em São Carlos e no E.S.R.F na linha ID23-1, estão sintetizadas na tabela 19 e os padrões de difração na figura 61. Os dados anteriores são referentes ao domínio semelhante à Expansina.

TABELA 19 - Estatísticas de coleta dos dados de difração de raios-X do domínio semelhante à Expansina.

Coleta de Dados	IFSC	ESRF-1	ESRF-2
Grupo Espacial	P321	P321	P321
Lattice (Rede)	hP (hexagonal)	hP (hexagonal)	hP (hexagonal)
Parâmetros de Cella (a,b,c) (Å)	156,82 156,82 83,85	156,53 156,53 84,47	157,36 157,36 340,61
Comprimento de onda (Å)	1, 541870	0, 972420	0, 972420
Limite de resolução (Å)	35,58 (3,05)	52,85 (2,75)	52,85(2,5)
Completeza (%)	82,3(82,4)	99,4 (99,5)	99,2(99,2)
Multiplicidade	2,7	3,3	3,72
$\langle I/\sigma(I) \rangle$	13 (2,2)	15,94 (2,77)	9,99(2,49)
Mosaicidade (°)	0,43	0,2	0,2
Reflexões únicas	33463	30762	189408
$B_{\text{fator médio Wilson plot}}$ (Å ²)	75,3	56,3	45,3
R_{MERGE}	9,8 (47,3)	5,7 (49,3)	9,8 (32,1)

Os conjuntos de dados especificados acima foram obtidos a partir de cristais da proteína domínio Expansina. A amostra protéica para cristalização para IFSC e ESRF-1 foi produzida no IFSC e a de ESRF-2 no PSB em Grenoble.

FIGURA 61 - Padrão de difração para os conjuntos de dados coletados: (A) IFSC resolução máxima 3,05Å,(B) E.S.R.F-1,resolução máxima 2,75Å e (C) E.S.R.F-2, resolução máxima 2,5Å.

Apresentaremos as estatísticas de processamento e refinamento obtidos para o conjunto de dados IFSC e E.S.R.F-1. O conjunto ESRF-2 devido à mudança nas dimensões da cela unitária, ainda está em processo de busca por uma solução molecular. Os resultados e estatísticas da indexação, integração e escalonamento dos conjuntos, são apresentados na tabela 19.

O escalonamento foi determinado utilizando dez camadas de resolução. O cálculo do volume de Matthews para estimar a quantidade de cadeias na unidade assimétrica foi determinado após o escalonamento, sendo o resultado favorável para seis cadeias na unidade assimétrica, como apresentado na tabela 20. Entretanto ao utilizar o programa *Phaser MR* no pacote CCP4⁸², para se determinar as fases iniciais, tendo como modelo a estrutura 3D3O (66% de identidade sequencial com o domínio semelhante à Expansina) o mesmo não encontrou uma solução com seis unidades na unidade assimétrica e sim uma solução de quatro cadeias para a unidade assimétrica.

TABELA 20 - Análise do Coeficiente de Matthews

Nº moléc./unidade assimétrica	Coeficiente de Matthews	% Solvente	P(2.90)	P(tot)
4	3,38	63,66	0,10	0,08
5	2,71	54,57	0,35	0,31
6	2,25	45,48	0,44	0,47
7	1,93	36,40	0,09	0,13

O modelo inicial foi refinado através de modificações manuais utilizando o programa Coot⁸³, associado a ciclos de refinamento como o programa Phenix⁷⁴. As estatísticas finais estão apresentadas na tabela 21.

TABELA 21 - Estatísticas finais para a estrutura domínio semelhante a Expansina

Parâmetros	IFSC	ESRF-1
Rfactor	0,2469	0,2311
Rfree	0,3202	0,2687
R.M.S.D para o comprimento de ligação (Å)	0,0138	0,006
R.M.S.D para ângulos de ligação (°)	1,6441	1,027

O modelo cristalográfico do domínio semelhante à Expansina apresenta quatro cadeias na unidade assimétrica (figura 62 A), o empacotamento está apresentado a seguir (figura 62 B), a estrutura possui 823 aminoácidos e 12 moléculas de água.

Os parâmetros de refinamento do modelo cristalográfico obtido foram comparados com os parâmetros de 868 estruturas do PDB, resolvidas na mesma faixa de resolução, utilizando o programa *Polygon*⁸⁴ (tabela 22), os resultados obtidos foram favoráveis à estrutura em estudo.

O gráfico de Ramachandran (figura 63) é particularmente útil porque ele define os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis e orienta a avaliação e qualidade da estrutura⁴⁸. Analisando o mesmo percebe-se que não há aminoácidos em regiões não permitidas e aqueles que estão em regiões generosamente permitidas estão relacionados a regiões altamente flexíveis da estrutura: as estatísticas do diagrama de Ramachandran geradas pelo PROCHECK estão na tabela 23.

FIGURA 62 - (A) Representação da unidade assimétrica com as quatro moléculas. (B) Representação do empacotamento cristalino.

TABELA 22 – Estatísticas de validação do modelo cristalográfico - Programa Polygon.

Parâmetro\PDB	Polygon-média de 824 PDB,	Expansina
R _{work}	0,221	0,233
R _{free}	0,262	0,271
Média de B	50	57
RMSD ângulo	1,48	1,39
RMSD ligação	0,01	0,01
Clashscore	26,561	18,728

FIGURA 63 - Diagrama de Ramachandran para o domínio semelhante à Expansina processado com o conjunto ESRF-1.

TABELA 23 – Estatísticas PROCHECK

Conjunto de dados	% Resíduos em região favorável	% Resíduos em região permitida	% Resíduos em região generosamente permitida	% Resíduos em região proibida
IFSC	73.5%	21,5	3,3	1,8
ESRF-1	86,3	13,1	0,6	0

Com o uso das ferramentas disponíveis no PDBSum no site do EMBL-EBI, gerou-se um mapa topológico do domínio semelhante à expansina (figura 64) e também as análises estruturais no ProMotif⁸⁵.

FIGURA 64 - Diagrama do mapa topológico da proteína semelhante à Expansina, as setas representam as fitas- β , e o cilindro em vermelho as α -hélices.

Utilizando o programa ProMotif, foi listado os elementos de estrutura secundária: 3 folhas- β (β -sheets), 5 β -grampos (β -hairpins), 4 Ψ -laço (Ψ -loop), 15 fita (*strands*), 5 β -saliência (β -bulges), 2 hélices, 27 β -voltas (β -turn) e 2 γ -volta (γ -turn). Analisando a estrutura do domínio Expansina temos que a mesma é classificada como Duplo- Ψ β -barril⁸⁶, como observamos na figura 65. A estrutura cristalográfica obtida possui motivos unidades na unidade assimétrica e esta representada na figura 66.

FIGURA 65 – Modelo cristalográfico da proteína semelhante à Expansina (cadeia B) com a identificação das principais estruturas secundárias presentes na mesma, em vermelho as α -hélices, em amarelo as folhas- β e em verde as demais estruturas secundárias.

FIGURA 66 - Modelo cristalográfico da proteína semelhante à Expansina obtido com os quatro motivos na unidade assimétrica.

Estruturalmente a proteína semelhante à Expansina de *Xanthomonas campestris* é composta por dois domínios, o primeiro (D1, resíduos de 1 a 108) contém duas α -hélices e seis fitas- β formando um motivo conhecido por Duplo- Ψ β -barril de 6-fitas (*6-stranded Double- Ψ β -barrel*), comum em várias famílias de proteínas.

O segundo domínio (D2, resíduos 113 a 206) é formado por duas folhas- β de fitas anti-paralelas, formando um *Ig-like β -sandwich*. Domínios com esse tipo de enovelamento são característicos no processo de reconhecimento a polissacarídeos, sendo característico três aminoácidos aromáticos nestas estruturas (Trp-124, Trp-125, Tyr-156) que estão alinhados no plano da superfície do domínio D2 (figura 67)⁸⁷, observamos que estes aminoácidos estão localizados numa região flável (loop A aa 119 - 125 e loop B 155 - 157) o que pode ser um indicativo da relação estrutura função. Além disso, analisando-se o potencial superficial eletrostático gerado em condições de vácuo com o programa Pymol (figura 67) observamos

que estes aminoácidos localizam-se numa superfície hidrofóbica a qual favorece a interação da proteína com a cadeia polissacarídica.

FIGURA 67 – Em destaque os principais aminoácidos envolvidos no processo de reconhecimento a polissacarídeos (Trp-124, Trp-125, Tyr-156).

FIGURA 68 – Superfície eletrostática Potencial superficial de contato da proteína de contato. Em destaque os principais aminoácidos envolvidos no processo de reconhecimento a polissacarídeos (Trp-124, Trp-125, Tyr-156).

Apresentamos na figura 69 a sobreposição do modelo estrutural da proteína semelhante à Expansina de *Xanthomonas campestris* com a Expansina de *Zea mays* cuja identidade entre os resíduos é de 36%. Já na figura 69 apresentamos a sobreposição do modelo estrutural da proteína semelhante à expansina de *Xanthomonas campestris* com a proteína semelhante à expansina de *Bacillus subtilis* cuja identidade entre os resíduos é de 66%. A partir dos alinhamentos observamos que apesar da diferença entre a identidade da

sequência primária de 36% e 66% a estrutura secundária e terciária entre elas é bem conservada, comparando as duas sobreposições temos que o motivo estrutural tanto do domínio D1 quanto D2 é conservado.

FIGURA 69 – Alinhamento do modelo estrutural da Expansina de *Zea mays* (2HCZ) em vermelho com o modelo estrutural da proteína semelhante à Expansina de *Xanthomonas campestris* em verde.

FIGURA 70 – Alinhamento múltiplo utilizando o programa Clustal para as sequências da Expansina de *Zea mays* (2HCZ) com a sequência da proteína semelhante à Expansina de *Xanthomonas campestris*.

FIGURA 71 – Alinhamento do modelo estrutural da Expansina de *Zea mays* (2HCZ) em vermelho com o modelo estrutural da proteína semelhante à Expansina de *Xanthomonas campestris*.

CLUSTAL 2.0.9 MULTIPLE SEQUENCE ALIGNMENT

FIGURA 72 – Alinhamento múltiplo utilizando o programa Clustal para as sequências da Expansina de *Zea mays* (2HCZ) com a sequência da proteína semelhante à Expansina de *Xanthomonas campestris*.

Na figura 70, temos o destaque em azul dos aminoácidos fundamentais na estrutura de Expansinas como descrito na literatura⁸⁷ e na figura 72 o destaque em vermelho para os aminoácidos parentados na figura 67. Portanto, a partir do modelo estrutural da proteína semelhante à Expansina de *Xanthomonas campestris* foi possível confirmar a relação aminoácidos fundamentais e estrutura secundária e terciária.

5 Conclusões

Neste trabalho foram apresentados os resultados obtidos referentes às análises de bioinformática, clonagem, expressão, purificação e caracterização estrutural da proteína Glicosilhidrolase e de seus domínios catalítico da família GH 5 e domínio semelhante à Expansina de *Xanthomonas campestris*. Estas atividades foram desenvolvidas nos laboratórios de pesquisa do Instituto de Física de São Carlos, Laboratório Nacional de Luz Síncrotron em Campinas e no E.S.R.F em Grenoble na França.

A partir do emprego de ferramentas de Bioinformática definiu-se o objeto de trabalho, uma proteína híbrida natural (domínios GH5 e semelhante à expansinas) e fizemos predições de parâmetros estruturais que foram confirmados com a aplicação de técnicas biofísicas como cromatografia e DLS.

Quanto à caracterização estrutural, o emprego do espalhamento dinâmico de luz demonstrou que em solução as proteínas Glicosilhidrolase e o domínio semelhante à Expansina apresentam em estado monomérico, além de obtermos valores para a massa molecular e raio hidrodinâmico das mesmas, bem próximos aos teóricos, confirmando a aplicabilidade da técnica ao estudo destas proteínas.

Com o Dicroísmo Circular obtivemos as porcentagens de estrutura secundária para a Glicosilhidrolase condizentes com o previsto teoricamente e o mesmo foi observado para o domínio Expansina. Além disso, para o domínio semelhante à Expansina observamos o fenômeno da influência de aminoácidos aromáticos no desvio do plano da luz polarizada.

O espalhamento de raios-X a baixo ângulo (SAXS) permitiu propor um modelo condizente para o envelope da estrutura da Glicosilhidrolase e do domínio semelhante à Expansina.

A partir das técnicas de cristalização foi determinada a condição de cristalização do domínio catalítico e do domínio semelhante à Expansina. Sendo a condição referente ao domínio Expansina reproduzida e os cristais expostos a difração de raios-X, com a estrutura cristalográfica resolvida a 2,75 Å.

Desta maneira, os objetivos iniciais propostos para esta dissertação foram concluídos e os resultados obtidos confirmaram o proposto na literatura referente.

6 Perspectivas

O entendimento funcional da Glicosilhidrolase sem dúvida permitirá uma compreensão do porque da presença de um domínio como Expansina juntamente com o catalítico. Desta maneira testes cinéticos se fazem necessários, para avaliar a atividade desta proteína híbrida natural, bem como de seus domínios isolados, principalmente o domínio semelhante à Expansina. Portanto, este alvo de trabalho mostra-se promissor do ponto de vista de geração de informações para o entendimento do processo de hidrólise de celulose em monossacarídeos.

REFERÊNCIAS

- 1 HARRIS, P. V. et al. Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. **Biochemistry**, v. 49, n. 15, p. 3305-3316, 2010/04/20 2010. ISSN 0006-2960. Disponível em: < <http://dx.doi.org/10.1021/bi100009p> >. Acesso em: 3 Jun 2012.

- 2 KUMAR, R.; SINGH, S.; SINGH, O. V. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. **Journal of Industrial Microbiology & Biotechnology**, v. 35, n. 5, p. 377-391, May 2008. ISSN 1367-5435.

- 3 EISENTRAUT, A. **Sustainable production of second-generation biofuels**: potential and perspectives in major economies and developing countries. 2010. Disponível em: < http://www.iea.org/papers/2010/second_generation_biofuels.pdf >. Acesso em: 19 maio 2012.

- 4 CLIXOO. Future predictions of cellulosic ethanol production costs. **Clixoo**,Tamilnadu, India. Disponível em: < http://www.clixoo.com/includes/pdf/Prediction_of_Cellulosic_Ethanol_Production_Cost_in_Future.pdf > Acesso em: 19 maio 2012.

- 5 ANTONI, D.; ZVERLOV, V. V.; SCHWARZ, W. H. Biofuels from microbes. **Applied Microbiology and Biotechnology**, v. 77, n. 1, p. 23-35, Nov 2007. ISSN 0175-7598.

- 6 DA SILVA, G. P. et al. Ethanol fermentation of sucrose, sugarcane juice and molasses by *Escherichia coli* strain KO11 and *Klebsiella oxytoca* strain P2. **Brazilian Journal of Microbiology**, v. 36, n. 4, p. 395-404, Oct-Dec. 2005. ISSN 1517-8382.

- 7 RAGAUSKAS, A. J. et al. The path forward for biofuels and biomaterials. **Science**, v. 311, n. 5760, p. 484-489, Jan. 27 2006. ISSN 0036-8075.

- 8 JOSHI, B. et al. Lignocellulosic ethanol production: Current practices and recent developments. **Biotechnology and Molecular Biology Reviews**, v.6, n. 8, p. 172-182, 2011.

- 9 ROWELL, R. M. (Ed.). **Handbook of wood chemistry and wood composites**. Boca Raton: CRC Press, 2005.

10 ROGER, P. et al. Cell Wall Chemistry. In: Rowell, R. M. (Ed.). **Handbook of wood chemistry and wood composites**. Boca Raton: CRC Press, 2005.

11 MILLER, R. B. **Wood handbook**. Madison, WI: Forest Products Laboratory, 1999. (General Technical Report FPL ; GTR-113)

12 VAN RANTWIJK, F. Monosaccharides. Their Chemistry and Their Roles in Natural Products. P. Collins and R. Ferrier. **Recueil des Travaux Chimiques des Pays-Bas**, v. 115, n. 9, p. 420-420. Chichester, John Wiley, 1995. ISBN 0-471-95343-1 .

13 SAHA BADAL, C. Lignocellulose biodegradation and applications in biotechnology. In: SAHA, B.C.; HAYASHI, K. (Ed.). **Lignocellulose biodegradation**. Washington: American Chemical Society, 2004. p. 2-34. (ACS Symposium Series, v. 889).

14 ADLER, E. Lignin chemistry—past, present and future. **Wood Science and Technology**, v. 11, n. 3, p. 169-218, 1977. Disponível em: < <http://dx.doi.org/10.1007/BF00365615> >. Acesso em: 10 maio 2012.

15 DOE, U. **Genomic Science Program**: GenomicsGTL_Roadmap_lowres.pdf (objeto application/pdf).2005. Disponível em: < http://genomicscience.energy.gov/roadmap/pdf/GenomicsGTL_Roadmap_lowres.pdf >. Acesso em: 19 maio 2012.

16 LYND, L. R. et al. Microbial cellulose utilization: fundamentals and biotechnology. **Microbiology and Molecular Biology Reviews**, v. 66, n. 3, p. 506, Sep 2002.

17 CASTRO, A. M. D.; PEREIRA JUNIOR, N. Produção, propriedades e aplicação de celulasas na hidrólise de resíduos agroindustriais. **Química Nova**, v. 33, n. 1, p. 181-188, 2010.

18 SADDLER, J. N. Factors limiting the efficiency of cellulase enzymes. **Microbiological Sciences**, v. 3, n. 3, p. 84-87, Mar. 1986.

19 COUGHLAN, M. P. Mechanisms of cellulose degradation by fungi and bacteria. **Animal Feed Science and Technology**, v. 32, n. 1-3, p. 77-100, 1991. ISSN 0377-8401. Disponível em: < <http://www.sciencedirect.com/science/article/pii/037784019190012H> >. Acesso em: 20 maio 2012.

- 20 BON, E. P. S.; GÍRIO, F.; PEREIRA JUNIOR, N. Enzimas na produção de etanol. In: BOM, E. P. S.; FERRARA, M. A.; CORVO, M. L. (Ed). **Enzimas em biotecnologia – produção, aplicações e mercado**. Rio de Janeiro: Editora Interciência, 2008. p. 241-272. ISBN 9788571931893.

- 21 COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanase families and extremophilic xylanases. **Fems Microbiology Reviews**, v. 29, n. 1, p. 3-23, Jan 2005. ISSN 0168-6445.

- 22 SHALLOM, D.; SHOHAM, Y. Microbial hemicellulases. **Current Opinion of Microbiology**, v. 6, n. 3, p. 219-28, June 2003.

- 23 CANTAREL, B. L. et al. The carbohydrate-active enzymes database (CAZy): an expert resource for Glycogenomics. **Nucleic Acids Research**, v. 37, p. 233-238, Jan 2009. ISSN 1362-4962 .DOI: 10.1093/nar/gkn663

- 24 CAZy - carbohydrate active enzymes database: CBM63. Disponível em: < <http://www.cazy.org/CBM63.html> >. Acesso em: 10 maio 2012.

- 25 MCQUEEN-MASON, S.; DURACHKO, D. M.; COSGROVE, D. J. Two endogenous proteins that induce cell wall extension in plants. **Plant Cell**, v. 4, n. 11, p. 1425-33, Nov. 1992.

- 26 COSGROVE, D. J. Loosening of plant cell walls by expansins. **Nature**, v. 407, n. 6802, p. 321-326, 2000.

- 27 SAMPEDRO, J.; COSGROVE, D. J. The expansin superfamily. **Genome Biology**, v. 6, n. 12, p.242.1-242.12 , 2005, doi:10.1186/gb-2005-6-12-242.

- 28 KIM, E. S. et al. Functional characterization of a bacterial expansin from bacillus subtilis for enhanced enzymatic hydrolysis of cellulose. **Biotechnology and Bioengineering**, v. 102, n. 5, p. 1342-1353, Apr. 1 2009.

- 29 BROTMAN, Y. et al. Role of swollenin, an expansin-like protein from Trichoderma, in plant root colonization. **Plant Physiology**, v. 147, n. 2, p. 779-89, June 2008.

30 GALBE, M.; ZACCHI, G. Pretreatment of lignocellulosic materials for efficient bioethanol production. **Advances in Biochemical Engineering/ Biotechnology**, v. 108, p. 41-65, 2007.

31 SHEEHAN, J. The Road to bioethanol: a strategic perspective of the U.S. Department of Energy's National Ethanol Program. In: Himmel,M.E.; Baker,J.O; Saddler,J.N. **Glycosyl hydrolases for biomass conversion**. Washington. American Chemical Society, 2000. p.2-25. (ACS Symposium Series, v. 769).

32 PIERETTI, I. et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. **BMC Genomics**, v. 10, n. 1, p. 616, 2009-12-17 2009. ISSN 1471-2164.

33 MARINGONI, A.C .Doença das crucíferas. In: KIMATI, H. et al. **Manual de fitopatologia**. São Paulo - SP: Editora Agronômica Ceres Ltda., 1995- 1997. v.2. cap.28. ISBN ISBN 85-318-0008-0.

34 DA SILVA, A. C. et al. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. **Nature**, v. 417, n. 6887, p. 459-63, May 23 2002. ISSN 0028-0836.

35 MICHEREFF, S.J. Podridão negra. In: Del Ponte, E. M. . **Fitopatologia.net** - herbário virtual. 2007. Disponível em: <<http://www6.ufrgs.br/agronomia/fitossan/fitopatologia/ficha.php?id=177>>. Acesso em: 20 maio 2012.

36 BARNES, M. R.; GRAY, I. C. **Bioinformatics for geneticists**. England, John Wiley & Sons Ltd, 2003.

37 ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, Oct. 5 1990.

38 VOET, D.; VOET, J. G. **Bioquímica**. São Paulo: Artmed, 2006.

39 HIGGINS, D. G.; SHARP, P. M. Fast and sensitive multiple sequence alignments on a microcomputer. **Computer Applicatons in the Biosciences**, v. 5, n. 2, p. 151-3, Apr. 1989.

40 _____. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. **Gene**, v. 73, n. 1, p. 237-44, Dec. 15 1988.

41 CAZy - **Carbohydrate active enzymes database**. Disponível em: < <http://www.cazy.org/> >. Acesso em: 10/06/12.

42 PETERSEN, T. N. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, v. 8, n. 10, p. 785-786, 2011.

43 DEMARCO, R., **Biologia molecular computacional**. Disponível em: < <http://www.ifsc.usp.br/~rdemarco/FFI0760/dominios.pdf> > . Acesso em: 10 junho 2012.

44 CONTRERAS-MOREIRA, B.; BATES, P. A. Domain Fishing: a first step in protein comparative modelling. **Bioinformatics**, v. 18, p. 1141-1142, 2002.

45 SAYERS, E.; BRYANT S. Macromolecular Structure Databases. In: McEntyre J, Ostell J, editors. **The NCBI Handbook** [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002, c. 3. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21095/>. Acesso em: 10 junho 2012.

46 MARCHLER-BAUER, A. et al. CDD: a conserved domain database for the functional annotation of proteins. **Nucleic Acids Research**, v. 39, p. D225-229, Jan. 2011. DOI: 10.1093/nar/gkn845

47 Gasteiger, E., et al. . Protein Identification and Analysis Tools on the ExPASy Server, In: WALKER, J.M.: **The proteomics protocols handbook**, Humana Press, 2005. , p. 571-607, ISBN 978-1-58829-343-5.

48 SANTOS FILHO, O. A.; ALENCASTRO, R. B. D. Protein homology modeling. **Química Nova**, v. 26, n. 2, p. 253-259, 2003. ISSN 0100-4042. Disponível em: < http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0100-40422003000200019&lng=en&nrm=iso&tlng=pt > Acesso em: 10 maio 2012.

49 BOURNE, P. E.; WEISSING, H. **Structural bioinformatics**. New Jersey: Wiley-Liss, 2003. cap. 25, p.509-546. ISBN 0471201995.

50 ROSSETO, F. R. **Caracterização bioquímica, biofísica e estrutural da principal endoglucanase secretada por *Xanthomonas campestris* pv. *Campestris* ATCC33913**.p..94. Dissertação(Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

51 MERCK. **Como melhorar sua Expressão de Proteínas Recombinantes em *E.coli***. Disponível em: <http://www.ioc.fiocruz.br/peptideos2008/pdf/proteinas_recombinantes_1.pdf> Acesso em: 10 maio 2012.

52 KOOLMAN, J. R., K.H. **Color Atlas of Biochemistry**. Rüdigerstrasse 14, 70469 Stuttgart, Germany: Georg Thieme Verlag, 2005. ISBN 3-13-100372-3.

53 QIAGEN. The QIAexpressionist. **Qiagen**, 2003. Disponível em: <<http://www.qiagen.com/literature/handbooks/literature.aspx?id=1000146> > Acesso em: 10 maio 2012.

54 SCHAGGER, H.; VON JAGOW, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. **Analytical Biochemistry**, v. 166, n. 2, p. 368-79, Nov 1 1987.

55 BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, p. 248-54, May 7 1976.

56 KELLY, S. M.; PRICE, N. C. The application of circular dichroism to studies of protein folding and unfolding. **Biochimica et Biophysica Acta**, v. 1338, n. 2, p. 161-85, Apr. 4 1997. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9128135?dopt=Citation> >. Acesso em: maio 2012.

57 GREENFIELD, N. J. Analysis of circular dichroism data. In: **Methods in Enzymology**, v. 383, 2004. Disponível em: < [http://dx.doi.org/10.1016/S0076-6879\(04\)83012-X](http://dx.doi.org/10.1016/S0076-6879(04)83012-X) >. Acesso em: maio 2012.

58 ARZENŠEK, D. **Dynamic light scattering and application to proteins in solutions**. Seminar. University of Ljubljana. Ljubljana. 2010. Disponível em: <http://mafija.fmf.unilj.si/seminar/files/2009_2010/Dynamic_light_scattering_and_application_to_proteins_in_solutions.pdf> Acesso em: 10 maio 2012.

59 _____. Dynamic light scattering and application to proteins in solutions. University of Ljubljana. Ljubljana. 2010.

60 BERNE, B. J.; PECORA, R. **Dynamic light scattering: with application to chemistry, biology and physics**. 3rd. ed. Mineola: Dover Publications, 2000.

61 SVERGUN, D. I.; KOCH, M. H. J. Small-angle scattering studies of biological macromolecules in solution. **Reports on Progress in Physics**, v. 66, n. 10, p. 1735-1782, Oct 2003.

62 ASSIS, S. M. P. et al. Antagonism of yeasts to *Xanthomonas campestris* pv. *campestris* on cabbage phylloplane in field. **Review of Microbiology**, v. 30, n. 3, p. 191-195, 07/1999 2012.

63 HAMMERSLEY, A. P. **FIT2d: an introduction and overview**. Grenoble: E.S.R.F 1997.

64 ROSA E SILVA, I. **Estudos biofísicos de Selenofosfato Sintetase de *Escherichia coli* e investigação de seu papel na via de biossíntese de Selenocisteínas**. p.140 Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

65 SVERGUN, D. I. Small-angle X-ray and neutron scattering as a tool for structural systems biology. **Biological Chemistry**, v. 391, n. 7, p. 737-743, July 2010.

66 FISCHER, H.; DE OLIVEIRA NETO, M.; NAPOLITANO, H. B.; POLIKARPOV, I. AND A. F. CRAIEVICH, A. F., Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale: **Journal of Applied Crystallography**. v.43, pt 1, p.101-109, 2010. doi:10.1107/S0021889809043076.

67 SVERGUN, D. I. Determination of the regularization parameter in indirect transform methods using perceptual criteria. **Journal of Applied Crystallography**, v. 25, p. 495-503, Aug 1 1992.

68 SVERGUN, D.; BARBERATO, C.; KOCH, M. H. J. CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. **Journal of Applied Crystallography**, v. 28, p. 768-773, Dec 1 1995.

69 SVERGUN, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. **Biophysical Journal**, v. 76, n. 6, p. 2879-2886, June 01 1999.

70 KOZIN, M. B.; SVERGUM, D. I. Automated matching of high- and low-resolution structural models: **Journal of Applied Crystallography**, v.34, p. 33-41, 2001. DOI:10.1107/S0021889800014126.

71 DRENTH, J. **Principles of protein x-ray crystallography**. New York, Springer, 2007. ISBN 978-0-387-33334-2.

72 AZEVEDO JUNIOR, W. F. **Biologia estrutural**: solução do problema da fase. 2006. Disponível em: < http://azevedolab.dominiotemporario.com/doc/bioestrutural_9a.pdf >. Acesso em: 14 jun 2012.

73 ROSSMANN, M. G. The molecular replacement method. **Acta Crystallogr A**, v. 46 pt 2, p. 73-82, Feb 1 1990.

74 ADAMS, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. **Acta Crystallographica Section D: Biological Crystallography**, v. 66, n. 2, p. 213-221, 2010-02-01 2010.

75 LASKOWSKI, R. A. M., M.W MOSS, D.S THORNTON, J.M. PROCHECK: A program to check the stereochemical quality of protein structures: **Journal of Applied Crystallography**. 26: 283-291 p. 1993.

76 ALBERTS, B. *et al.* **Molecular Biology of the Cell**. New York: Garland science, Taylor & Francis Group 2008.

77 GLATTER, O.; KRATKY, O. **Small Angle X-ray Scattering**. New York Academic Press, 1982. ISBN 0-12-286280-5.

78 FISCHER, H. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. **Journal of Applied Crystallography**, v. 43, n. 1, p. 101-109, 2010-02-01 2010.

79 RAMBO, R. P.; TAINER, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. **Biopolymers**, v. 95, n. 8, p. 559-71, Aug 2011.

80 SVERGUN, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. **Biophysical Journal**, vol. 76, i.6 p. 2879–2886, 1999.

81 ESWAR, M. A. et al. Comparative Protein Structure Modeling With MODELLER. **Current Protocols in Bioinformatics**: John Wiley& Sons, Inc. Supplement 15, 5.6.1-5.6.30, 2006.

82 WINN, M. D. B. *et al.* Overview of the CCP4 suite and current developments. **Acta Crystallographica Section D: Biological Crystallography**, v. 67, n. 4, p. 235-242, 2011.

83 DEBRECZENI, J. E.; EMSLEY, P. Handling ligands with Coot. **Acta Crystallographica Section D :Biological Crystallography**, v. 68, n. Pt 4, p. 425-30, Apr. 2012.

84 URZHUMTSEVA L, A. P., ADAMS PD, URZHUMTSEV A. Crystallographic model quality at a glance: **Acta Crystallographica Section D: Biological Crystallography**, v.65, Pt 3, p. 297-300, 2009.

85 HUTCHINSON, E. G.; THORNTON, J. M. Promotif: a program to identify and analyze structural motifs in proteins. **Protein Science**, v. 5, n. 2, p. 212-20, Feb. 1996.

86 GARRATT, R. C.; ORENGO , C. A. **The protein chart**: Wienheim:WILEY-VCH Verlag Gmbh & Co. , 2008.

87 KERFF, F. et al. Crystal structure and activity of Bacillus subtilis YoaJ (EXLX1), a bacterial expansin that promotes root colonization. **Proceedings of the National Academy of Sciences of the United States of America (PNAS)**, v. 105, n. 44, p. 16876-81, Nov. 4 2008.