UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

GUSTAVO MACHADO ÁLVARES DE LIMA

Estudos Estruturais e Moleculares da Enzima Fosfopanteteinil Transferase de *Xanthomonas albilineans*: Alvo Molecular para o Desenvolvimento de Novos Agroquímicos para Cultura de Cana-de-Açúcar

São Carlos

2013

GUSTAVO MACHADO ÁLVARES DE LIMA

Estudos Estruturais e Moleculares da Enzima Fosfopanteteinil Transferase de *Xanthomonas albilineans*: Alvo Molecular para o Desenvolvimento de Novos Agroquímicos para Cultura de Cana-de-Açúcar

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. Rafael V. C. Guido

Versão Corrigida

A versão original está disponível na unidade que aloja o programa

São Carlos

2013

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)

Machado Alvares de Lima, Gustavo
Descoberta e Planejamento de Novos Agroquímicos para a Cultura da Cana-de-Açúcar / Gustavo Machado Alvares de Lima; orientador Rafael Victório Carvalho Guido - versão corrigida -- São Carlos, 2013.
121 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2013.

1. Agroquímica. 2. Modelagem molecular. 3. Triagem virtual. 4. Cana-de-açúcar. 5. Xanthomonas albilineans. I. Victório Carvalho Guido, Rafael, orient. II. Título.

Aos meus pais, João Carlos e Eliana.

Que minhas conquistas sejam as deles.

AGRADECIMENTOS

Aos meus pais João Carlos e Eliana, por sempre acreditarem em mim, pelo exemplo e pelo amor. Pelas noites mal dormidas, pelos fins de tarde que eu prometi e esqueci de telefonar. Ao meu irmão Guilherme, pelo exemplo de dedicação e garra e à minha irmã(zinha) Maria Clara, pelo amor e carinho sem iguais.

Ao Prof. Dr. Rafael V. C. Guido, pela confiança, amizade, orientação e pelas longas conversas, científicas ou não, que tornaram este trabalho possível.

À Lívia Manzine, Mamé e ao Prof Dr. Otávio H. Thiemann, por toda ajuda e suporte no início da minha vida científica, que levarei sempre comigo como uma lembrança boa.

Aos meus amigos da BIOMOL, Paula, Hilde, Ivan, Geraldo (kx), Wagner, Wally (Vict), Érica, Caio, Atílio, Ricardo, Rodrigo, Luciano e Cauê, pelos 7 anos de São Carlos, pela força e companheirismo indispensáveis nas horas difíceis e principalmente nas horas fáceis. Pelas biofeiras, que são uma saudade que eu gosto de ter.

Aos meus amigos de laboratório César Moises, Kadowaki, Jadinho, Postigo, Evandrinho, Tavin, Marina, Ana Cristina, Livia Maria, Daiana Evelin, Van Arnoldi, Fran e Maluzete, por todos os momentos inesquecíveis dentro e fora do lab.

À Hildinha, pelos cafés da manhã sagrados por tantos anos consecutivos, pelas horas que me ouvi reclamar e pelas risadas de doer a barriga. Ao Assirati, pelas tantas quartas feira no Trembão, por toda a amizade e esforço, fundamentais no final da graduação.

À Fernanda, pelo amizade e cumplicidade, pelas tantas orquestras regidas, por tantas longas explicações e discussões.

Ao Maluf, pela dedicação e ensinamentos indispensáveis, pelas tantas discussões, pela amizade. Por ter sido presente sempre.

Aos gordões Bacheguinha, Tonho, André doidão e Maluf, que são amigos incríveis, pesquisadores incríveis, pessoas incríveis. Pelas tantas hora em que a amizade falou mais alto.

Aos meus amigos do LQMC Wan Altei, Ivani, Mariana, Karina, Luma, Vanessa Kiraly, Gustavo Valente, Renatinha, Leonardo e César pelo apoio e amizade na sala 8.

À Dayane e Marina, que compartilharam alegrias e tristezas no laboratório, e me ensinaram mais do que eu à elas.

Aos meus professores do IFSC-USP, pela dedicação inquestionável e pela formação.

Aos funcionários da biblioteca, do serviço de graduação e pós-graduação e tantos outros do grupo de Cristalografia, pela prestatividade em diversos momentos.

À FAPESP e ao CNPq pelo apoio financeiro.

RESUMO

LIMA, G. M. A. **Estudos estruturais e moleculares da enzima fosfopanteteinil transferase de *Xanthomonas albilineans*: alvo molecular para o desenvolvimento de novos agroquímicos para cultura de cana-de-açúcar.** 2013. 121 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

A cana-de-açúcar é uma das principais fontes de energia renovável, constituindo a matéria-prima mais importante na busca por energia limpa e sustentável. Os benefícios ambientais provenientes da produção e do uso dos derivados de cana-de-açúcar fomentam o desenvolvimento de métodos e produtos que aumentem, de modo sustentável, a geração de bioenergia. Dentre os diversos fatores limitantes para o aumento da produção de cana-de-açúcar, destaca-se a ocorrência e a severidade de fitopatologias como a escaldadura das folhas. Essa doença, causada pela bactéria *Xanthomonas albilineans*, causa diminuição da produtividade, necessidade de reforma precoce dos canaviais e queda de qualidade do caldo extraído que determinam prejuízos econômicos significativos para os agricultores. Atualmente, não há alternativas disponíveis para o controle químico ou biológico dessa fitopatologia. Portanto, existe uma necessidade urgente de desenvolvimento de novas moléculas como defensivos agrícolas que sejam eficazes, seletivas, de baixo custo e impacto ao meio ambiente. A *X. albilineans* produz uma família de antibióticos e fitotoxinas conhecida como albicidinas. As enzimas envolvidas na biossíntese de albicidinas são alvos moleculares extremamente atrativos para o planejamento de novos agroquímicos. Entre as enzimas dessa via, destaca-se a fosfopanteteinil transferase (XaPPT, E.C. 2.7.8.7), uma enzima essencial para o desenvolvimento da *X. albilineans*. Essa dissertação está dividida em duas partes: i. estudos computacionais e ii. estudos experimentais. Na parte computacional foi desenvolvida uma nova ferramenta, denominada ViTaMIn para auxiliar o desenvolvimento de modelos tridimensionais (3D) de proteínas. A enzima *Xa*PPT foi utilizada como estudo de caso para validar o programa. Os resultados obtidos indicaram que ViTaMIn foi capaz de auxiliar a construção de um modelo 3D robusto da *Xa*PPT. Além disso, foi verificado que ViTaMIn é uma alternativa útil para usuários iniciante e experientes na modelagem molecular. O modelo de *Xa*PPT construído foi utilizado na triagem virtual para a

identificação de novos candidatos a inibidores. A estratégia incluiu a aplicação de filtros moleculares e estudos de docagem molecular que resultaram na seleção de 10 candidatos a inibidores da enzima-alvo. Na parte experimental, o cultivo de culturas de *X. albilineans* foi padronizado e estudos de biologia molecular com a *Xa*PPT conduzidos a partir da extração de DNA genômico da bactéria. Estratégias clássicas e modernas foram empregadas para a clonagem da *Xa*PPT. Os resultados obtidos indicaram a obtenção de proteína expressa na forma solúvel. Os trabalhos integrando estudos computacionais e experimentais apresentados nessa dissertação de mestrado significam importantes contribuições no desenvolvimento de bases científicas sólidas para o desenvolvimento de novos candidatos a agroquímicos para a cultura de cana de açúcar.

ABSTRACT

LIMA, G. M. A. **Structural and molecular studies of *xanthomonas albilineans* phosphopantetheinyl transferase enzyme: molecular target for new agrochemicals development for sugarcane cultures.** 2013. 121 P. Dissertação (Mestrado Em Ciências) – Instituto De Física De São Carlos, Universidade De São Paulo, São Carlos, 2013.

Sugarcane is a major source of renewable energy, representing the most important source for clean and sustainable energy. The environmental benefits collected from the production and use of sugarcane derivatives boost the development of new methods and products to improve, in a sustainable way, bioenergy generation. However, the occurrence and severity of plant diseases, such as leaf scald, hinder the productivity of sugarcane crops. Sugarcane leaf scald is a widespread and devastating disease caused by the bacteria *Xanthomonas albilineans*. The disease has a dramatic impact on crop productivity, including reduced yields and drop in quality of the juice. Currently, there is no chemical or biological treatment for disease control. Therefore, there is an urgent need for new effective and selective pesticides with low cost and environmental impact. The *X. albilineans* produces a family of antibiotics and phytotoxins known as albicidin. The enzymes involved in the biosynthesis of albicidin are attractive targets for the design of new agrochemicals. Among them, the phosphopantetheinyl transferase (*Xa*PPT, EC 2.7.8.7) enzyme plays an essential role to the development of *X. albilineans*. This dissertation is divided into two parts: i. computational studies and ii. experimental studies. In the computational part, we developed a new tool, named ViTaMIn to assist the development of three-dimensional models (3D) of proteins. The *Xa*PPT enzyme was employed as a case study to validate the program. The results indicated that ViTaMIn was capable of assisting the construction of a robust 3D model of *Xa*PPT. Moreover, Vitamin is a useful alternative for beginners and advanced modelers. The XaPPT model was used in virtual screening approach to identify new candidate inhibitors. The strategy included the application of molecular filters and molecular docking studies which afforded 10 inhibitor candidates for the target enzyme. In the experimental part, *X. albilineans* was cultured and standardized. On the basis of that, molecular biology studies were conducted on *Xa*PPT. Classical and modern strategies were employed to cloning *Xa*PPT. The results

indicated that protein is expressed soluble. The integration of computational and experimental studies presented in this dissertation are important contributions in the development of strong scientific basis for the design of new agrochemicals for sugarcane cultures.

LISTA DE FIGURAS

Figura 1 – Estrutura química 2D (A)e 3D (B) do dicloro-difenil-tricloroetano (DDT)..... 31

Figura 2 – Principais fatos da história da cana-de-açúcar no Brasil..... 35

Figura 3 – Variação da oferta interna das principais fontes energéticas no Brasil no período entre 2002 e 2011..... 36

Figura 4 – (A) Imagem representativa de cultura de *X. albilineans* cultivada em placa. (B) Destaque para uma colônia isolada de *X. albilineans*. (C) Imagem de microscopia eletrônica de transmissão de uma única célula de *X. albilineans* 26..... 38

Figura 5 – Sintomas crônicos (A–C) e agudos (D) da escaldadura das folhas em cana-de-açúcar..... 39

Figura 6 – (A) Estrutura 3D representativa do complexo PPT-coenzima A. (B) Estrutura química da Coenzima A (CoA). 1 = adenosina fosforilada; 2 = difosfato; 3 = ácido pantóico; 4 = β-Alanina; 5 = cisteamina. 47 Os grupos 3+4+5 formam a panteteína. (C) Reação bioquímica catalisada pela enzima fosfopanteteína transferase..... 42

Figura 7 – (A) Crescimento no período 1985-2013 do número de estruturas depositas no banco de dados de proteínas (PDB) 49. (B) Crescimento no período 1985-2013 do número de sequências depositas no banco de dados UniProtKB/Swiss-Prot 50..... 43

Figura 8 – Esquema das etapas envolvidas na triagem virtual de inibidores da XaPPT..... 55

Figura 9 – Sítio de ligação utilizado na triagem virtual com a enzima XaPPT. (A) Sítio de ligação (malha azul) explorado pelo programa Surflex. A molécula do ligante coenzima A está indicada como bastão laranja e o átomo de Mg2+ está indicado como esfera verde. (B) Mapeamento do sítio para identificação de interações favoráveis para grupos doadores (NH, bastão azul) e aceptores de ligação (C=O, bastão vermelho) de hidrogênio realizado pelo Surflex. (C) Mapeamento do sítio para identificação de

interações favoráveis para grupos hidrofóbicos (CH4, bastão laranja) realizado pelo Surflex. (D) Sítio de ligação explorado pelo programa Glide..... 56
Figura 11 – Integração do ViTaMIn com o PyMOL..... 60
Figura 12 – Etapas para modelagem comparativa no modo Basic Modeling. Os números em vermelho indicam as etapas do processo de construção dos modelos. 1. seleção da sequência da proteína alvo (formato FASTA); 2. identificação e selecição da estrutura homóloga (código PDB); 3. alinhamento das sequências alvo e molde; 4. otimização do alinhamento; 5. Seleção do número de modelos que serão construídos..... 62
Figura 13 – (A) Módulo integrado ViTaMIn-PyMOL para o refinamento das alças. Região em vermelho na interface do Pymol indica a região alvo do refinamento (resíduos 81-106). Os números em vermelho indicam as etapas do processo de refinamento de alças. 1. seleção da região para refinamento; 2. identificação do resíduo inicial e final da alça; 3. visualização com o módulo integrado no PyMOL; 4. seleção do número de modelos construídos e início do processo. (B) Exemplo representativo de gráfico de pontuação DOPE dos resíduos do modelo construído..... 64
Figura 14 – (A) Módulo integrado ViTaMIn-Pymol para a seleção e refinamento de mutações pontuais. Os números em vermelho indicam as etapas do processo de seleção e refinamento de mutações. 1. seleção dos resíduos que serão mutados; 2. seleção do número de modelos para construção e início do processo. (B) Detalhe da janela de seleção de resíduos para mutação..... 65
Figura 15 – Visão geral da janela de análise dos modelos moleculares..... 66
Figura 16 – Topologia representativa do domínio ACPS presente na XaPPT. 68
Figura 17 – Predição de estruturas secundárias da XaPPT..... 68
Figura 18 – Identificação de estruturas homólogas da XaPPT no servidor HHPred..... 69

Figura 19 – (A) Complexo ternário da BsPPT-CoA-Mg2+. A BsPPT está indicada como modelo de fitas, o ligante CoA está indicado como modelo de bastões, o íon Mg2+ está indicado como modelo de esfera e as interações polares estão indicadas como linhas tracejadas. (B) Sítio de ligação do Mg2+ hexacoordenado pelos resíduos Asp107, Glu109, Glu151, duas moléculas de água (esferas vermelhas) pelo substituinte difosfato da CoA. (C) Sítio de ligação do CoA. O sítio de ligação está indicado pela superfície de Connolly. (D) Resíduos do sítio de ligação da BsPPT e interações polares que estabilizam o modo de ligação do CoA..... 70

Figura 20 – Alinhamento estrutural do modelo construídos para a XaPPT (verde) pelo servidor I-Tasser com a proteína molde (ciano) BsPPT (PDB ID, 1QR0)..... 72

Figura 21 – Diagrama de Ramachandran para o melhor modelo construído pelo servidor I -Tasser empregando a estratégia de alinhavamento. Os pontos vermelhos correspondem aos resíduos localizados em regiões proibidas..... 73

Figura 22 – Pontuação DOPE dos modelos construídos pelo ViTaMIn. A seta vermelha indica o modelo com menor pontuação DOPE..... 74

Figura 23 – Alinhamento estrutural do modelo construídos para a XaPPT (vermelho) pelo ViTaMIn com a proteína molde (ciano) BsPPT (PDB ID, 1QR0)..... 74

Figura 24 – Diagrama de Ramachandran para o melhor modelo construído pelo método de restrições espaciais do Modeller com a interface ViTaMIn..... 75

Figura 25 – Comparação entre estrutura secundária predita da XaPPT pelo método PSIPRED e o verificado no modelo construído pelo método Stride..... 76

Figura 26 – (A) Sobreposição dos modelos construídos pelo I-Tasser (verde) e ViTaMIn (vermelho). Em destaque a região flexível dos resíduos de aminoácidos 200-220. (B) Representação esquemática da variação do fator B da BsPPT (PDB ID, 1QR0).

O calibre da representação e a temperatura da cor indicam a variação da deslocamento atômico. O espectro azul-vermelho representa o intervalo de 10-75 Å2, respectivamente. Regiões em azul e vermelho apresentam menor e maior deslocamento atômico, respectivamente..... 77

Figura 27 – Interação da Arg92 com os resíduos Leu189 e Ala190 da α-hélice..... 78

Figura 28 – Modo de ligação da CoA no sítio de ativo da XaPPT e BsPPT. (A) e (B) Destaque para as interações com o grupo 3-fosfato do anel ribose da adenosina. (C) e (D) Destaque para as interações entre o grupo difosfato e o íon Mg2+. (E) e (F) Destaque para as interações entre o grupo adenina e os resíduos do sítio de ligação..... 79

Figura 29 – (A) Estrutura química do candidato a inibidor ZINC08442194. (B) Sobreposição estrutural do modo de ligação predito para o composto ZINC08442194 pelo Glide (ciano) e Surflex-Dock (magenta). O sítio de ligação está representado pela superfície de Connolly e o átomo de Mg2+ está representado no modelo de esfera. (C) Modo de ligação do composto ZINC08442194 predito pelo Glide. (D) Modo de ligação do composto ZINC08442194 predito pelo Surflex-Dock. O sítio de ligação está representado em modelos de fitas e as ligações de hidrogênio estão ilustras em amarelo..... 88

Figura 30 – Mapa geral dos vetores LIC. As proteínas de fusão utilizadas estão descritas na Tabela 7..... 96

Figura 31 – Gel de eletroforese em agarose 1%. Reação de identificação molecular de *X. albilineans* através da amplificação de região gênica específica (288 bp). MK = Marcador de massa molecular..... 104

Figura 32 – Gel de eletroforese em agarose 1%. Extração de DNA genômico de *X. albilineans*..... 105

Figura 33 – Gel de eletroforese em agarose 1%. Reação de PCR para amplificação do gene *AlbXXI*..... 106

Figura 34 – Gel de eletroforese em agarose 1%. Reação de PCR dos clones recombinantes obtidos para as construções propostas..... 106
Figura 35 – Gel de eletroforese SDS-PAGE 15%. Teste de expressão para a construção XaPPT-pET28a. MK = Marcador de massa molecular 108

LISTA DE TABELAS

Tabela 1 – Principais características e funções das interfaces ViTaMIn, EasyModeller e PyMOD disponíveis para o Modeller..... 57
Tabela 2 – Comparação dos limites entre as propriedades físico-químicas de fármacos (Lipinski, fármaco-similar) e agroquímicos (Tice, agroquímico-similar) 80
Tabela 3 – Conjunto de 33 moléculas entre as 100 melhores pontuadas pelo Surflex-Dock e Glide..... 84
Tabela 4 – Moléculas identificadas na triagem virtual pelo consenso entre os resultados do Surflex-Dock e do Glide..... 87
Tabela 5 – Culturas de *X. albilineans* adquiridas da Coleção de Culturas Fitobactérias IBSBF do Instituto Biológico de Campinas..... 93
Tabela 6 – Reação de PCR utilizada nos estudos moleculares da *X. albilineans*..... 95
Tabela 7 – Características das construções gênicas da enzima XaPPT nos vetores de expressão empregados..... 96

LISTA DE ABREVIATURAS E SIGLAS

ACPS	Proteína carreadora de acila Sintase, do inglês, **A**cyl **C**arrier **P**rotein **S**ynthase
CTAB	Cetil trimetil brometo de amônio, do inglês, **C**etyl **T**rimethyl **A**mmonium **B**romide
DDE	**D**icloro-**D**ifenil-**E**tileno
DDD	**D**icloro-**D**ifenil-**D**icloroetano
DDT	**D**icloro-**D**ifenil-**T**ricloroetano
DOPE	Energia discreta e otimizada da proteína, do inglês, **D**iscrete **O**ptimized **P**rotein **E**nergy
EMBL	Laboratório europeu de biologia molecular, do inglês, **E**uropean **M**olecular **B**iology **L**aboratory
GPL	Licença pública geral GNU, do inglês, GNU **G**eneral **P**ublic **L**icense
GTk/GTk+	Ferramentas do programa de manipulação de imagens do GNU, do inglês, **G**NU Image Manipulation Program **T**ool**k**it
HTS	Ensaio em larga escala, do inglês, **H**igh-**T**hroughput **S**creening
IBSBF	Coleção de Culturas do Laboratório de Bacteriologia Vegetal do Instituto Biológico
IDE	Ambiente de desenvolvimento integrado, do inglês, **I**ntegrated **D**evelopment **E**nvironment
IPTG	**I**so**p**ropil β-D-1-**t**io**g**alactopiranosideo
LB	**L**uria-**B**ertani
LGPL	Licença pública geral menor GNU, do inglês, GNU **L**esser **G**eneral **P**ublic **L**icense
LIC	Clonagem independente de ligase, do inglês, **L**igase **I**ndependent **C**lonning
NUSA	Substância NusA
NRPS	Peptídeo não ribossomal sintase, do inglês, **N**on**r**ibosomal **P**eptide **S**ynthase
OPLS	Potenciais otimizados para simulações de líquido, do inglês, **O**ptimized **P**otentials for **L**iquid **S**imulations
PCR	Reação em cadeia da polimerase, do inglês, **P**olimerase **C**hain **R**eaction

PDB	Banco de dados de proteína, do inglês, **P**rotein **D**ata **B**ank
PEG	**P**oli-**e**tileno **G**licol
pH	**P**otencial **H**idrogeniônico
PIP	**P**rotetores **I**ncorporados em **P**lantas
PKS	Policetídeo sintase, do inglês **P**olyketide **S**ynthase
PPT	Fosfopanteteinil transferase, do inglês, **P**hospho**p**antetheinyl **T**ransferase
q.s.p.	**Q**uanto **s**eja **p**reciso para
RMN	**R**essonância **M**agnética **N**uclear
RMSD	Desvio quadrático médio padrão, do inglês, **R**oot **M**ean **S**quared **D**eviation
SUMO	do inglês, **S**mall **U**biquitin-related **Mo**difier
TRX	**T**iro**r**edo**x**ina
ULP-1	**U**biquitin-**l**ike-specific **p**rotease **1**
VS	Triagem virtual, do inglês, **V**irtual **S**creening
XaPPT	***Xanthomonas albilineans*** PPT
XAS	***Xanthomonas albilineans*** **S**elective
ZINC	**Zi**nc is **N**ot **C**ommercial

‘Pro Brasilia fiant eximia’

Lema do Estado de São Paulo

SUMÁRIO

LISTA DE FIGURAS **14**

LISTA DE TABELAS **20**

LISTA DE ABREVIATURAS E SIGLAS **22**

1 INTRODUÇÃO **29**

1.1 Cana-de-açúcar *30*

1.1.1 Agricultura e agroquímicos 30

1.1.2 Indústria agroquímica e o desenvolvimento de novos agroquímicos 32

1.1.3 Cana-de-açúcar e agronegócio 33

1.1.4 Fitopatologias da cana-de-açúcar 37

1.1.5 *Xanthomonas albilineans* e a escaldadura das folhas 37

1.1.6 Seleção de alvos moleculares para a escaldadura das folhas 40

1.1.7 Via de biossíntese de albicidinas e a enzima fosfopanteteinil transferase 41

1.2 Métodos computacionais *43*

1.2.1 Modelagem molecular de proteínas 43

1.2.1.1 Modelagem molecular comparativa 44

1.2.2 Desenvolvimento de programas computacionais, linguagem de programação 45

1.2.3 Ambiente gráfico e bibliotecas matemáticas 46

1.2.4 Planejamento Baseado em Estrutura e Triagem Virtual 47

2 OBJETIVOS **49**

3 MODELAGEM MOLECULAR **51**

3.1 Material e métodos *52*

3.1.1 Desenvolvimento do progama ViTaMIn (*Visual Table Modeller Interface*) – uma interface gráfica para o programa Modeller 52

3.1.2 Busca por estruturas homólogas e alinhamento sequencial 52

3.1.3 Construção, seleção e validação dos modelos moleculares da *Xa*PPT 53

3.1.4 Triagem virtual 54

3.1.5 Sítio de ligação 55
3.1.5.1 Base de dados de compostos.....57
3.2 Resultados e discussão.....58
3.2.1 ViTaMIn – Visual Table Modeller Interface 58
3.2.1.1 Modelagem comparativa.....61
3.2.1.2 Refinamento dos modelos 63
3.2.1.3 Análise dos modelos.....66
3.2.2 Estudo de caso: Enzima fosfopanteteinil transferase de *X. albilineans* 67
3.2.2.1 Estudo da sequência de aminoácidos da *Xa*PPT por bioinformática.....67
3.2.2.2 Identificação e análise estrutural de proteínas homologas à *Xa*PPT.....69
3.2.2.3 Construção, seleção e validação dos modelos de *Xa*PPT.....71
3.2.3 Triagem virtual de inibidores da *Xa*PPT.....79
3.2.3.1 Alvo molecular, sítio de ligação e base de dados de compostos 81
3.2.3.2 Docagem molecular e seleção de candidatos a inibidores da *Xa*PPT.....82
4 BIOLOGIA MOLECULAR.....91
4.1 Material e métodos.....92
4.1.1 Cultura de *Xanthomonas albilineans* 92
4.1.2 Extração do DNA genômico de *X. albilineans*.....93
4.1.3 Caracterização molecular de *X. albilineans*.....94
4.1.4 Clonagem da enzima *Xa*PPT.....95
4.1.4.1 Clonagem clássica nos vetores de expressão pET28a e pETSUMO.....96
4.1.4.2 Clonagem independente de ligação (LIC) nos vetores de expressão pETM11, pETTrx-1a, pETNus-1a 99
4.1.4.3 Testes de expressão da enzima XaPPT.....*100*
4.2 Resultados.....101
4.2.1 Identificação molecular de *X. albilineans*.....101
4.2.2 Clonagem da enzima *Xa*PPT.....102

4.2.3 Testes preliminares de expressão de *Xa*PPT 105

5 CONCLUSÕES E PERSPECTIVAS 107

REFERÊNCIAS.....111

CAPÍTULO 1

INTRODUÇÃO

1.1 Cana-de-açúcar

1.1.1 Agricultura e agroquímicos

Agricultura é o conjunto de técnicas utilizadas para cultivar plantas com o objetivo de obter alimentos, fibras, energia, matéria-prima. Os primeiros relatos de agricultura datam de 9.000 anos A.C. na região entre os rios Tigre e Eufrates onde atualmente localiza-se a Jordânia, Síria, Irã e Turquia. O período neolítico (10.200-2.000 anos A.C.) caracterizado pelo início do sedentarismo e surgimento da agricultura, foram fundamental na história do desenvolvimento humano com impacto marcante na transição cultural da colonização nômade para a permanente, criação e especialização de profissões e desenvolvimento de tecnologias [1].

A crescente demanda por alimentos e energia tem exigido índices cada vez maiores de produtividade agrícola. Nesse sentido, apesar do aumento significativo da produção nas últimas décadas, a agricultura mundial ainda enfrenta grandes desafios para suprir a necessidade de consumo das diversas populações. Entre os desafios para o aumento da produtividade, destaca-se a severidade das pragas e fitopatologias que, ainda hoje, são responsáveis por perdas substanciais na produção (e.g., média mundial de 36% e em alguns casos específicos podendo ultrapassar 80%) [2]. Diante disso, métodos alternativos, como a utilização de agroquímicos, também conhecidos como pesticidas ou defensivos agrícolas (e.g., pesticidas, herbicidas, antifúngicos), passaram a desempenhar função central no controle seletivo de pragas e fitopatologias.

A Organização das Nações Unidas para Alimentação e Agricultura (FAO, do inglês, *Food and Agriculture Organization*) define pesticida como “qualquer substância ou mistura de substâncias destinadas a prevenir, destruir ou controlar qualquer praga” [3]. Os pesticidas são divididos em três classes: i. pesticidas inorgânicos, ii. pesticidas orgânicos sintéticos e iii. biopesticidas. O uso dos primeiros pesticidas que se tem relato foi há 4.500 anos, pelos Sumérios, que empregavam compostos de enxofre para combater insetos [4]. No século XV, pesticidas inorgânicos, com destaque para o

arsênio e o mercúrio, começaram a ser utilizados no combate a pragas em lavouras. O primeiro pesticida orgânico sintético, dicloro-difenil-tricloroetano (**Figura 1**), popularmente conhecido pelo acrônimo DDT, foi sintetizado em 1939 pelo químico suíço Paul Hermann Müller. O trabalho sobre a descoberta do DDT e eficácia de suas propriedades inseticida nas lavouras rendeu ao Dr. Paul H. Müller o prêmio Nobel de Fisiologia ou Medicina em 1948 [5].

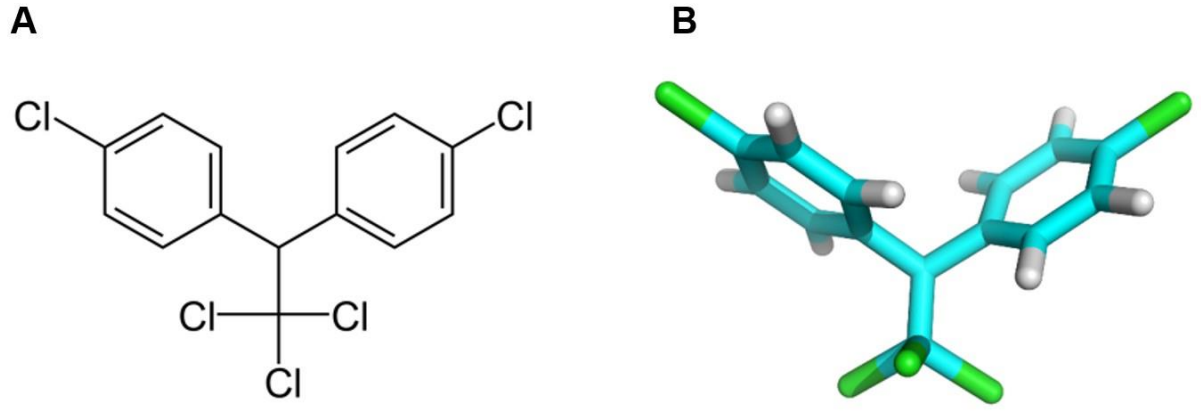


Figura 1 – Estrutura química 2D (**A**)e 3D (**B**) do dicloro-difenil-tricloroetano (DDT).

A acumulação no solo do DDT e de seus derivados DDE e DDD levou o pesticida a ser banido da utilização extensiva em 1992 [6]. Casos extremos mostram a permanência destas substancias em grande quantidade no solo por mais de 20 anos [7,8]. A utilização do DDT em outras aplicações, como no combate ao inseto vetor de doenças infecciosas, como por exemplo a malária, vem sendo discutida veementemente pelas Nações Unidas e, apesar de banida, o Tratado de Estocolmo assegura a utilização do DDT para o controle dos vetores [9].

O uso de biopesticidas em lavouras é mais recente e inclui: i. pesticidas microbiais – micro-organismos (e.g., bactérias, fungos ou vírus) que infectam e matam as pragas das lavouras (e.g., *Bacillus thuringiensis*, *Bacillus subtilis*); ii. bioprotetores (PIP, do inglês, *Plant-Incorporated-Protectors*) – produtos biológicos com propriedades pesticidas que são inseridos em plantas através de métodos biotecnológicos de engenharia genética (*e.g.*, toxina *Bt*); e iii. biopesticidas – moléculas isoladas de micro-organismos (e.g., feromônios) ou produtos naturais (e.g., espinosinas, quitina, quitosana, alcaloides, terpenoides e compostos fenólicos). Entre as vantagens dos biopesticidas destaca-se a alta eficiência e baixa toxicidade, contudo, apresentam

desvantagens como espectro de ação limitado e longo tempo de latência desde a aplicação até a obtenção de resultados satisfatórios [10].

1.1.2 Indústria agroquímica e o desenvolvimento de novos agroquímicos

As primeiras indústrias agroquímicas apareceram na segunda metade do século XIX, principalmente na Alemanha e Inglaterra. Atualmente, as indústrias agroquímicas constituem um segmento industrial de extremo valor, com vendas anuais da ordem de dezenas de bilhões de dólares [11]. Contudo, possuem uma imagem pública desfavorecida, devido às propriedades tóxicas de seus produtos. Diante disso, o paradigma moderno de desenvolvimento de novos agroquímicos inclui, como um dos principais aspectos, a preocupação sobre o impacto ambiental da molécula, bem como o aumento da eficácia. Nesse contexto, é importante mencionar que os agroquímicos desenvolvidos e lançados mais recentemente no mercado apresentam afinidade extremamente elevada por seus alvos biomoleculares. Portanto, concentrações significativamente menores de composto ativo são necessárias para obtenção do efeito desejado, consequentemente, reduzindo substancialmente os custos de produção bem como os danos ao meio ambiente [6]. Entre os avanços obtidos nas últimas décadas com as pesquisas em defensivos agrícolas destaca-se: i. redução de doses de aplicação (e.g., 87% para os fungicidas; 88% para os herbicidas e 94% para os inseticidas) e ii. redução da toxicidade (e.g., redução de 160 vezes para os inseticidas) [11,12].

A capacidade da indústria em destinar recursos para pesquisa e desenvolvimento (P&D) está diretamente relacionada à sua competência em gerar lucros através da venda de um conjunto atrativo de inovações tecnológicas. Nesse contexto, a indústria de defensivos agrícolas engloba atividades que se destacam entre as de mais elevada intensidade em pesquisa. De fato, a indústria de defensivos é a segunda que mais investe em P&D de novos produtos (~12% do faturamento), ficando atrás da indústria farmacêutica (13% do faturamento) [13].

Desde a concepção do projeto até a introdução de um único defensivo agrícola no mercado são investidos de 10–12 anos em P&D, com custos totais estimados da ordem de US$ 250–300 milhões [14,15]. Entretanto, defensivos agrícolas de grande sucesso são capazes de gerar rendimentos significativos da ordem de US$ milhões/ano. Por exemplo, entre os defensivos com maior sucesso de venda, cita-se o herbicida mesotriona (Callisto®, Syngenta), inibidor potente e seletivo da enzima 4-hidroxifenilpiruvato dioxigenase (HPPD, do inglês, *4-hydroxyphenylpyruvate dioxygenase* – EC 1.13.11.27). Os herbicidas seletivos da Syngenta geraram extraordinários US$ 2,6 bilhões com vendas no ano de 2011, sendo o Callisto®, disponível no mercado há mais de 10 anos, um dos contribuintes majoritários por esse faturamento, o qual apresentou aumento de 13% em relação a 2010 [16,17]. Os relatórios anuais de 2011 das duas maiores indústrias de defensivos agrícolas do mundo, Syngenta e Bayer, reportam faturamentos extraordinários de US$ 13,26 e 10,15 bilhões, respectivamente [16,18].

O Brasil consiste em um mercado-alvo extremamente atrativo para a indústria de agroquímicos, uma vez que é um dos maiores produtores agrícolas do mundo. O mercado nacional de agroquímicos, que correspondia a US$ 7,2 bilhões em 2009, expandiu consideravelmente e atualmente corresponde a US$ 8,5 bilhões [19]. O aumento na produção de grãos, algodão e cana-de-açúcar tem contribuído significativamente para essa expansão [19]. Dados do Centro de Estudos Avançados em Economia Aplicada (CEPEA), relativos ao peso de cada subsetor e insumos no agronegócio, estimam que o impacto financeiro dos agroquímicos nos custos da produção podem chegar a 18% do custo total de insumos e a 8% dos custos totais da agricultura [20]. Portanto, os defensivos agrícolas representam um mercado extremamente promissor que oferece diversas oportunidades de inovação tecnológica em uma área estratégica para o nosso País.

1.1.3 Cana-de-açúcar e agronegócio

A cana-de-açúcar (*Saccharum officinarum*), amplamente cultivada no Brasil e no mundo, é original do sudoeste asiático [21]. Posteriormente, foi levada à Índia, onde as

técnicas de extração e processamento do caldo de cana foram desenvolvidas e fizeram da planta um produto mais interessante economicamente. As tecnologias de produção de açúcar foram desenvolvidas pelos hindus e egípcios que por muito tempo mantiveram em segredo os detalhes de produção. Contudo, durante as guerras no século VII, os povos árabes descobriram, aperfeiçoaram e difundiram essas técnicas, fazendo da cana-de-açúcar uma das culturas mais cultivadas no mundo [21].

As plantas de cana-de-açúcar vieram para o continente americano no início do século XVI (**Figura 2**). No Brasil, as primeiras mudas foram trazidas da Ilha da Madeira, em 1532, por Martim Afonso de Souza, o qual foi também responsável pela instalação do primeiro engenho em São Vicente, no ano de 1533. A cultura da cana-de-açúcar prosperou substancialmente na região Nordeste, sendo o açúcar o principal produto do agronegócio brasileiro por mais de dois séculos. Na década 1970, teve início a transformação do setor com o Programa Nacional do Álcool (Proálcool). O Proálcool foi instituído pelo Decreto n° 76.593 em 1975, com a finalidade de expandir a produção do álcool etílico anidro e viabilizar seu uso como matéria-prima para a indústria química e como combustível adicionado à gasolina. Por meio desse programa, o Brasil poderia substituir parte do combustível derivado do petróleo e auxiliar o setor sucroalcooleiro através do deslocamento de parte da produção de açúcar para a produção de bioetanol [22]. Foi nessa época que as usinas passaram a ter maior incentivo para produção do etanol e, mais recentemente, à bioeletricidade, aos alcoolquímicos e à comercialização de créditos de carbono. Atualmente, a existência no exterior de políticas de fomento ao uso, produção e inserção de biocombustíveis, com destaque para o etanol, nas matrizes energéticas, oferece uma oportunidade ímpar para o Brasil, que dispõe de experiência e de tecnologias de produção adquiridas em quase um século de uso contínuo do bioetanol como combustível automotivo, e mais de 30 anos de existência do Proálcool [22]. Essa experiência, associada à grande disponibilidade de terras agrícolas com aptidão para o cultivo da cana-de-açúcar, impulsionam o desenvolvimento de novas tecnologias para o aumento de produção do setor.

A cana-de-açúcar é uma das principais culturas para agronegócio no contexto mundial. A planta é cultivada em mais de 100 países, onde representa uma importante fonte de geração de emprego. Apesar da difusão mundial, cerca de 80% da produção do planeta estão concentradas em dez países (i.e., Brasil, Índia, China, México,

Tailândia, Paquistão, Colômbia, Austrália, Indonésia e Estados Unidos) [23]. No Brasil, o agronegócio é um setor de destaque na economia. As exportações brasileiras do agronegócio atingiram em 2012 o maior valor da história (US$ 95,81 bilhões), apresentando uma expansão de, aproximadamente, US$ 1 bilhão em relação ao ano anterior (US$ 94,97 bilhões). O setor sucroalcooleiro, um dos cinco principais contribuintes para essa expansão, foi classificado em 2012 como o terceiro maior responsável pelas exportações do agronegócio brasileiro, no qual se destacou o etanol, cuja venda externa aumentou 46,6% (US$ 694 milhões) [24].

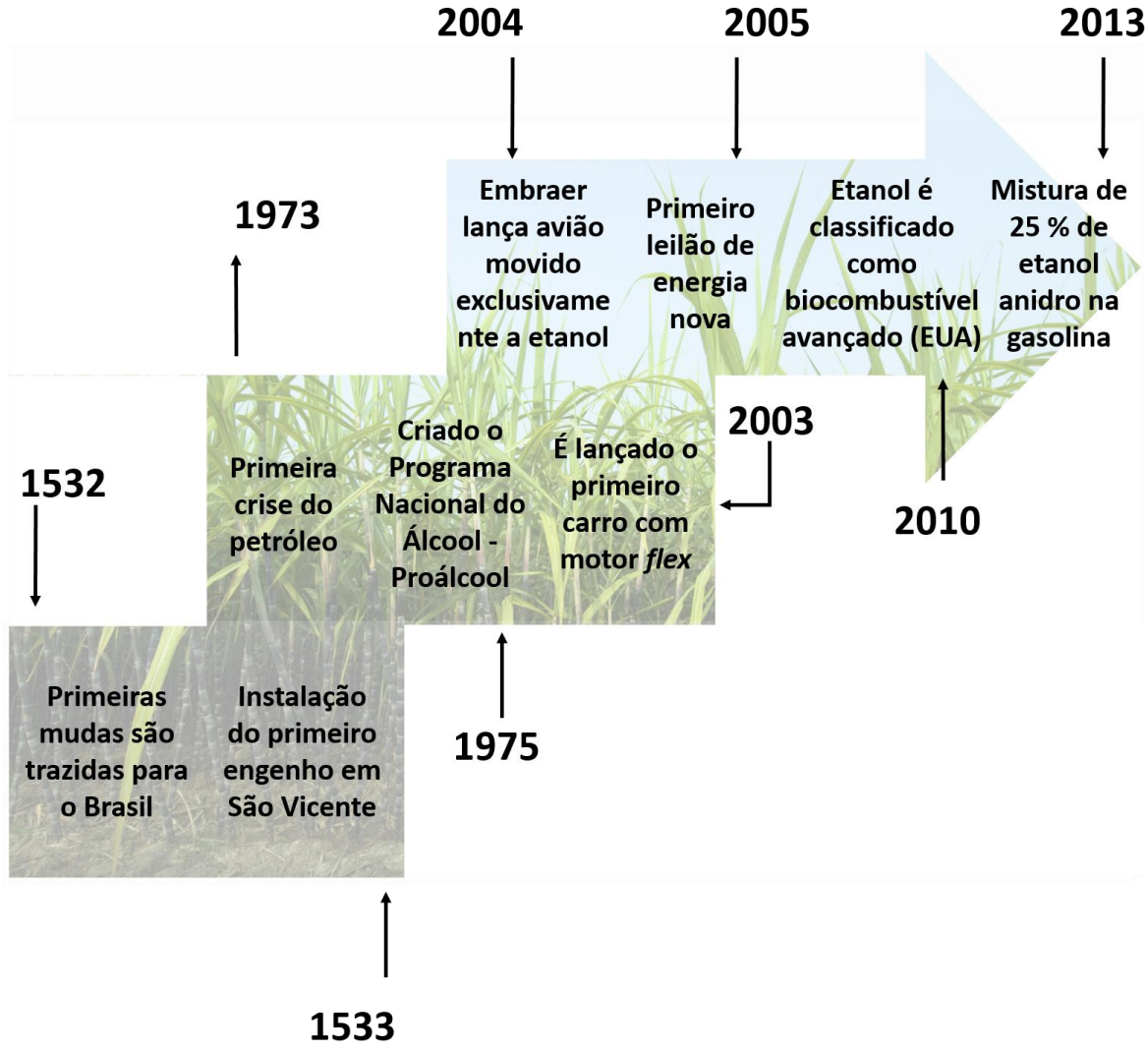


Figura 2 – Principais fatos da história da cana-de-açúcar no Brasil.

A cana-de-açúcar é uma das principais fontes de energia renovável, constituindo a matéria-prima mais importante na busca por energia limpa e sustentável. A matriz energética brasileira é destacadamente a mais renovável do mundo, na qual as fontes

renováveis representam mais de 40% (**Figura 3**). Esse dado torna-se mais impactante quando comparado com outros países em desenvolvimento ou industrializados, nos quais a contribuição das fontes renováveis é de apenas 6% e 13%, respectivamente [25]. No período entre 2002 e 2011, a participação do petróleo e seus derivados na matriz energética brasileira diminui, aproximadamente, em 5% (de 43% em 2002 para 38,6% em 2011) (**Figura 3**). Nesse mesmo período, a participação dos derivados de cana-de-açúcar aumentou, aproximadamente, em 3% (de 12,8% em 2002 para 15,7% em 2011), assumindo o segundo lugar em importância e afirmando seu papel como componente fundamental na matriz energética nacional (**Figura 3**) [25].

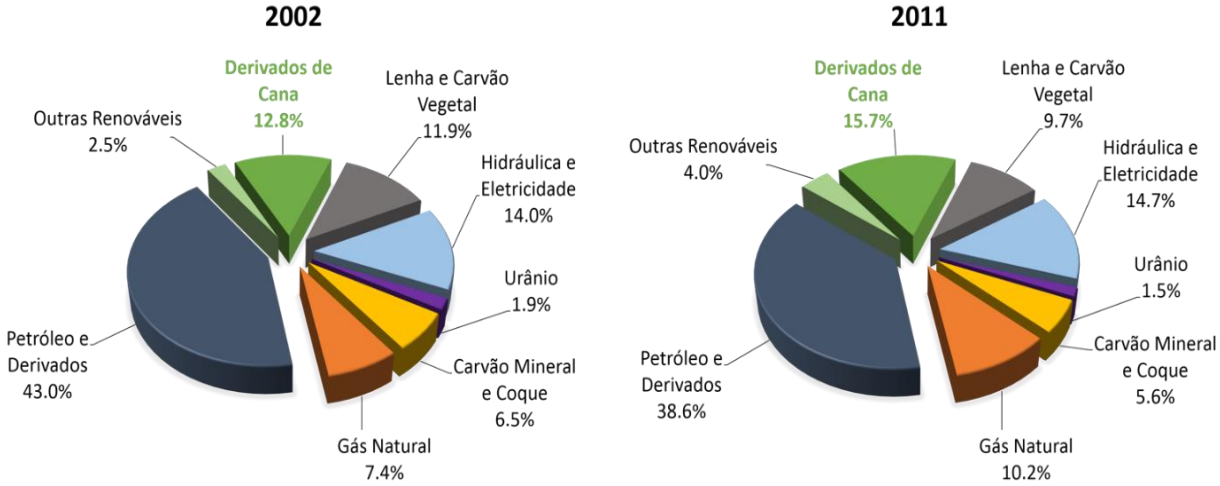


Figura 3 – Variação da oferta interna das principais fontes energéticas no Brasil no período entre 2002 e 2011.

O etanol, principal biocombustível derivado da cana-de-açúcar, atingiu na safra 2012/2013 produção superior a 23 bilhões de litros [26]. No aspecto social, a produção de etanol estimula a geração de empregos, direta ou indiretamente, apresentando impacto positivo no desenvolvimento regional. Além disso, os ganhos ambientais do uso do etanol proveniente da cana-de-açúcar são exemplificados pela menor emissão de gases de efeito estufa, cuja redução pode ultrapassar 90% quando comparado ao uso da gasolina [27]. Somado a isso, está o maior aproveitamento energético obtido com o emprego da biomassa gerada no processamento da cana-de-açúcar (e.g., bagaço, vinhaça, torta de filtro e palha da cana), de grande relevância socioeconômica para produção de bioeletricidade, biopolímeros, ração animal, aglomerados, fertilizantes, entre outros[27]. Portanto, diante da importância econômica do setor sucroalcooleiro para a economia nacional, fatores externos capazes de comprometer o rendimento produtivo devem ser rigorosamente controlados. Nesse sentido, as pesquisas devem

focar no desenvolvimento de métodos e alternativas capazes de garantir o crescimento da produção e atender ao cenário de expansão projetado para as próximas décadas sem que se faça necessário a substituição de culturas, principalmente, em áreas nas quais se cultiva alimentos.

1.1.4 Fitopatologias da cana-de-açúcar

Dentre os diversos fatores limitantes para o aumento da produção de cana-de-açúcar, destaca-se a ocorrência e a severidade de fitopatologias causadas por bactérias, fungos e vírus, além do ataque de diversas pragas [28,29]. Mais de uma centena de doenças já foram identificadas em culturas de cana-de-açúcar, entre elas destacam-se cinco como as mais importantes por causarem danos substanciais às lavouras: i. carvão – causado pelo fungo *Ustilago scitaminea* [30]; ii. ferrugem – causada pelos fungos *Puccinia melanocephala* e *Puccinia kuehnii* [30]; iii. raquitismo das soqueiras – causado pela bactéria *Leifsonia xyli* [30]; iv. mosaico – causado pelo vírus do mosaico [30]; e v. escaldadura das folhas – causada pela bactéria *Xanthomonas albilineans* [30].

1.1.5 *Xanthomonas albilineans* e a escaldadura das folhas

A *Xanthomonas albilineans* é o agente causador da escaldadura das folhas, uma das fitopatologias de maior impacto nas culturas de cana-de-açúcar. A *X. albilineans* é uma bactéria gram-negativa que apresenta formato de bastão alongado, não esporulante, aeróbica, uniflagelada, crescimento lento, aspecto amarelado e colônias não mucoidais (**Figura 4**). Essa bactéria infecta principalmente plantas monocotiledôneas da Família Poaceae, incluindo a *Saccharum* spp. [31].

A doença, disseminada em praticamente todas as regiões onde a cana-de-açúcar é cultivada, causa diminuição da produtividade, necessidade de reforma precoce dos

canaviais e queda de qualidade do caldo extraído [32,33]. O impacto econômico causado pela escaldadura das folhas é significativo, porém tem sido subestimado devido a dificuldades no diagnóstico e diferenciação dos prejuízos causados por outras fitopatologias, como o raquitismo das soqueiras [29].

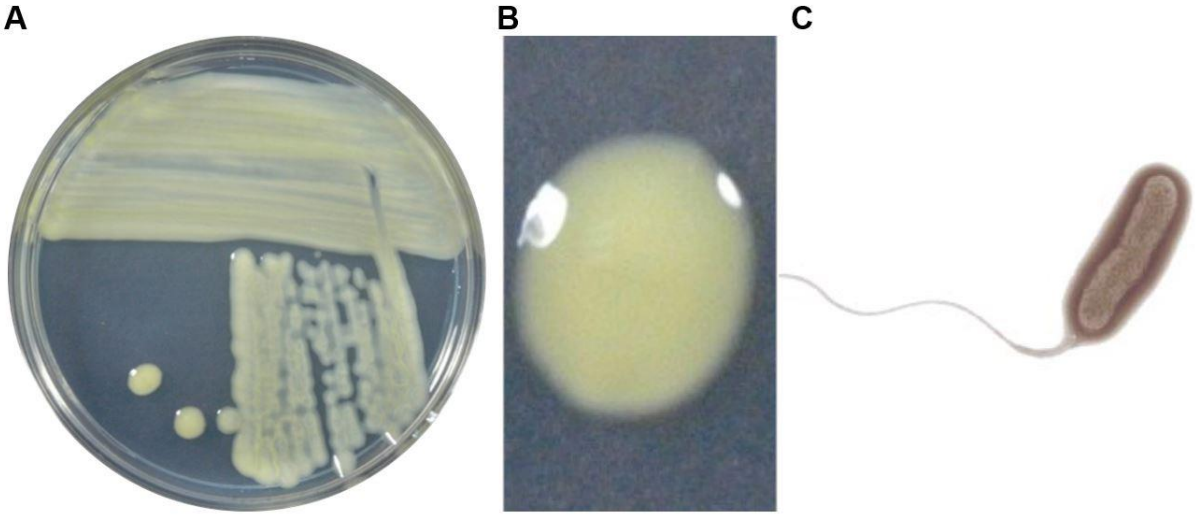


Figura 4 – (**A**) Imagem representativa de cultura de *X. albilineans* cultivada em placa. (**B**) Destaque para uma colônia isolada de *X. albilineans*. (**C**) Imagem de microscopia eletrônica de transmissão de uma única célula de *X. albilineans* [31].

A *X. albilineans* utiliza como via de entrada no hospedeiro as células do parênquima, onde há maior reserva de glicose, em seguida, a bactéria migra e se estabelece no xilema. Uma vez estabelecida, a *X. albilineans* produz um grupo de substâncias conhecidas como albicidinas, que inibem seletivamente a replicação do DNA dos cloroplastos, determinando o bloqueio da diferenciação celular [31,34]. A infecção por *X. albilineans* resulta em sintomas variáveis, cuja expressão e severidade estão relacionados ao nível de resistência da variedade de cana-de-açúcar cultivada, condições ambientais e agressividade do patógeno [35]. A manifestação da doença em variedades extremamente suscetíveis pode resultar em perdas de até 100% das lavouras. A escaldadura das folhas apresenta-se sob as formas latente, crônica e aguda. A maioria das variedades comerciais apresenta a forma latente, na qual os sintomas externos não se manifestam. Contudo, internamente, pode-se observar nos colmos uma descoloração vascular na região dos nós, assemelhando-se com os

sintomas do raquitismo das soqueiras. A forma crônica caracteriza-se pelo surgimento de estrias brancas no limbo foliar e bainha (**figura 4A–C**), enquanto o principal sintoma da forma aguda é a queima total das folhas, como se a planta tivesse sido escaldada, o que dá origem ao nome da doença (**figura 4D**). Nesse estágio, a bactéria impede a difusão de nutrientes pelo bloqueio total ou parcial do xilema que causa diminuição no conteúdo de açúcares do extrato final e determina prejuízos ao potencial energético e econômico da plantação.

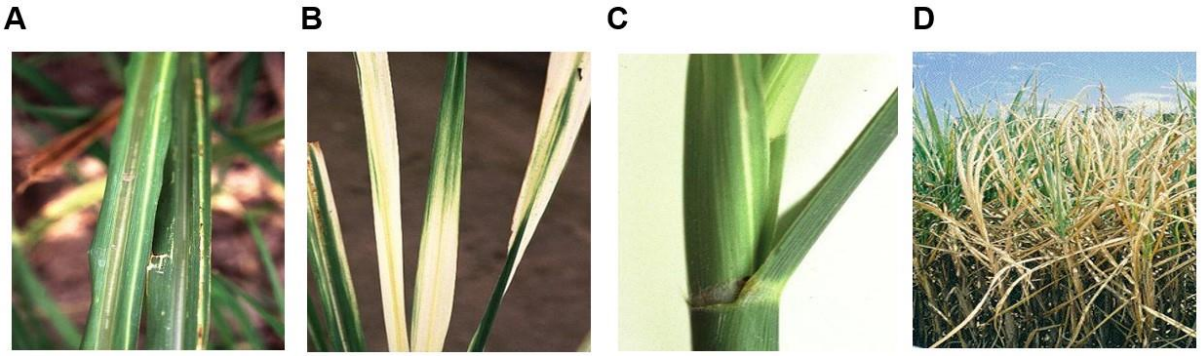


Figura 5 – Sintomas crônicos (**A–C**) e agudos (**D**) da escaldadura das folhas em cana-de-açúcar.

A *X. albilineans* presente em plantas infectadas pode ser disseminada por vento, chuva ou colheita, através de ferimentos provocados nos colmos. Com a ampliação do uso de colheitas mecanizadas, a preocupação com a escaldadura das folhas aumenta substancialmente, portanto, enfatiza-se a necessidade de manutenção do canavial sadio.

Atualmente, as estratégias de controle da doença consistem na substituição de variedades suscetíveis por resistentes. Assim, variedades de cana-de-açúcar com características agronômicas desejáveis e com alto potencial de produção, mas que são suscetíveis à doença, são frequentemente descartadas durante o processo de seleção de novas variedades comerciais. Esse cenário é ainda agravado pela propensão do patógeno em sofrer variações genéticas que favoreçam a quebra de resistência, consequentemente, transformando cultivares resistentes em altamente suscetíveis [36,37]. Além disso, não é conhecido até o momento qualquer produto químico ou biológico capaz de controlar satisfatoriamente essa fitopatologia. Esse grave cenário torna urgente o desenvolvimento de novas alternativas para o controle da escaldadura das folhas. Nesse sentido, a descoberta de alvos moleculares atrativos para o desenvolvimento de novas moléculas bioativas como defensivos

agrícolas potentes e seletivos pode contribuir significativamente para o tratamento seguro e manejo eficaz da doença.

1.1.6 Seleção de alvos moleculares para a escaldadura das folhas

O processo de planejamento de substâncias bioativas baseia-se, principalmente, na investigação de vias bioquímicas essenciais do patógeno em estudo [38]. Essa investigação tem como objetivo identificar possíveis alvos moleculares para modulação seletiva por pequenas moléculas. Nesse contexto, a elucidação do genoma de *X. albilineans* fornece informações moleculares extremamente úteis para a descoberta de rotas exclusivas e vitais do metabolismo da bactéria [39]. As biomoléculas envolvidas nessas vias bioquímicas oferecem oportunidades científicas atrativas para o desenvolvimento de novas formas de controle da doença.

A bactéria *X. albilineans* produz uma família de antibióticos e fitotoxinas conhecida como albicidinas [40]. Essas biomoléculas são fatores patogênicos extremamente potente (i.e., concentrações nanomolar) de ação bactericida rápida e amplo espectro (e.g., ativa em bactérias Gram-positivas e Gram-negativas) [41]. O mecanismo de ação molecular das albicidinas consiste na inibição da enzima DNA girase [42].

Na cana-de-açúcar, as fitotoxinas albicidinas desempenham função importante para a patogenicidade e disseminação sistêmica da *X. albilineans* [40]. Plantas de cana-de-açúcar modificadas por estratégias biotecnológicas para expressar um gene relacionado à eliminação de albicidina mostraram-se resistentes ao patógeno [43]. Além disso, estudos de mutagênese em *X. albilineans* indicaram que cepas deficientes na síntese albicidinas (mutantes Tox^-) quando inoculadas em plantas de cana-de-açúcar são incapazes de produzir os sintomas da escaldadura das folhas [44]. Portanto, as enzimas envolvidas na biossíntese de albicidinas são alvos moleculares extremamente atrativos para o planejamento de novos agroquímicos através de inibição seletiva.

1.1.7 Via de biossíntese de albicidinas e a enzima fosfopanteteinil transferase

A biossíntese de albicidina inclui vários genes importantes que se encontram reunidos em posições específicas no genoma de *X. albilineans* [45,46]. O gene *xabB* codifica uma sintetase multifuncional de 4.800 aminoácidos constituída de módulos de policetídeo sintase (PKS, do inglês, *polyketide synthase*) em fusão com uma peptídeo sintetase não-ribossomal (NRPS, do inglês, *non-ribosomal peptide synthetase*) [47]. As enzimas PKS e NRPS necessitam de modificação pós-traducional para se tornarem ativas. Esse processo, essencial para a biossíntese da albicidina [48], é realizado pelo produto do gene *xabA*, que codifica a enzima fosfopanteteína transferase (*Xa*PPT, do inglês *phosphopantetheinyl transferase* – EC 2.7.8.7) (**Figura 6A)**.

As enzimas fosfopanteteinil transferases são responsáveis pela transferência do grupo fosfopanteteína da coenzima A (CoA) (**Figura 6B**) para um resíduo de serina conservada nas PKS e NRPS (**Figura 6C**) [49,50]. Esse processo, conhecido como fosfopanteteinilação, é essencial para a biossíntese de peptídeos não ribossomais. O grupo sulfidrila da fosfopanteteína é responsável pela manutenção de intermediários da biossíntese, enquanto outras sintases promovem a modificação e extensão de cadeias peptídicas. Assim, a ausência da modificação pós-traducional torna outras proteínas não funcionais, inibindo a biossíntese de toda a via [51].

A importância da PPT para o processo bioquímico de modificação pós-traducional somado às características únicas da via de biossíntese de albicidinas em *X. albilineans* fazem dessa enzima um alvo atrativo para o planejamento de inibidores enzimáticos. Portanto, o desenvolvimento de inibidores potentes da enzima PPT de *X. albilineans* (*Xa*PPT) impedirá a produção das fitotoxinas albicidinas determinando uma condição metabólica na qual o patógeno seria incapaz de causar prejuízos as lavouras de cana-de-açúcar. Essa hipótese faz dos inibidores da enzima *Xa*PPT candidatos a novos agentes agroquímicos para o controle da escaldadura das folhas.

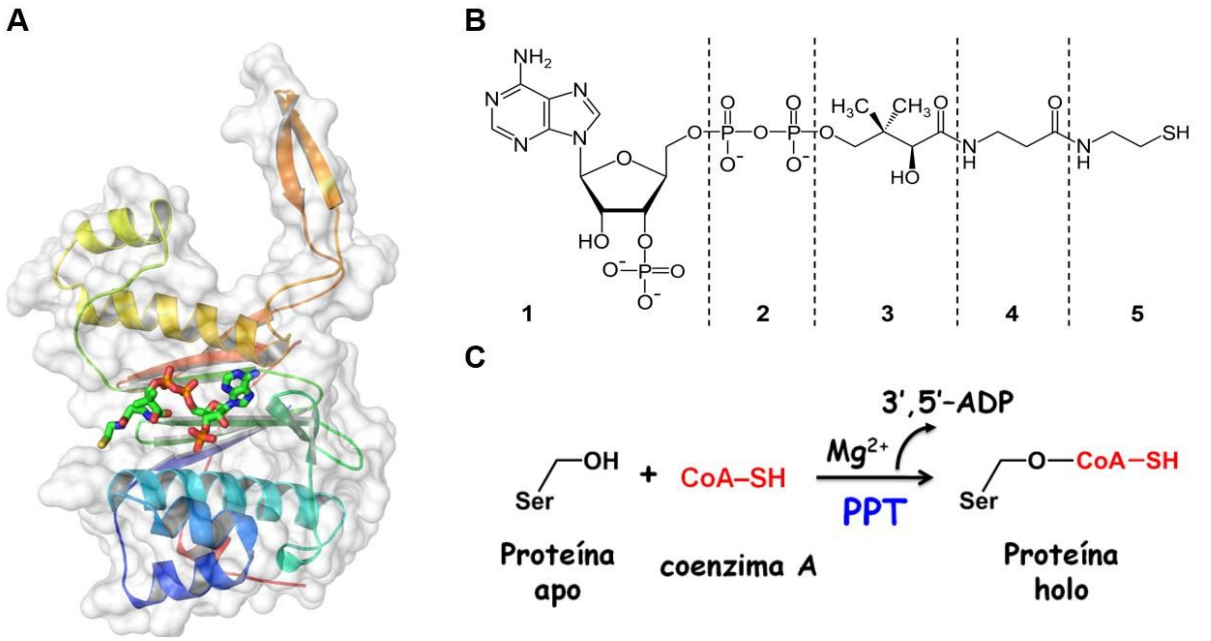


Figura 6 – (**A**) Estrutura 3D representativa do complexo PPT-coenzima A. (**B**) Estrutura química da Coenzima A (CoA). 1 = adenosina fosforilada; 2 = difosfato; 3 = ácido pantóico; 4 = β-Alanina; 5 = cisteamina. [52] Os grupos 3+4+5 formam a panteteína. (**C**) Reação bioquímica catalisada pela enzima fosfopanteteína transferase.

1.2 Métodos computacionais

1.2.1 Modelagem molecular de proteínas

Os avanços das técnicas de elucidação de estruturas 3D de proteínas fez da biologia estrutural uma ferramenta extremamente atrativa para o processo de descoberta de novos inibidores enzimáticos [53]. Contudo, a comparação entre os dados de estruturas terciária/quaternárias de macromoléculas, disponíveis no banco de dados de proteínas (PDB, do inglês, *protein databank*) [54], e estruturas primárias, disponíveis no banco de dados UniProtKB/Swiss-Prot [55], indica que existe um número significativamente maior de sequências de proteínas depositadas que estruturas de proteínas conhecidas [54,55] (**Figura 7**). Entre as diversas causas determinantes para essa diferença citam-se: i. obtenção de amostras em quantidade insuficientes para os ensaios e ii. ausência e/ou dificuldades de obtenção de cristais com qualidade adequada para os experimentos de difração de raios-X [56–58]. Diante disso, métodos alternativos para a elucidação de estruturas terciárias de proteínas foram desenvolvidos, entre eles, destaca-se a modelagem molecular comparativa de proteínas.

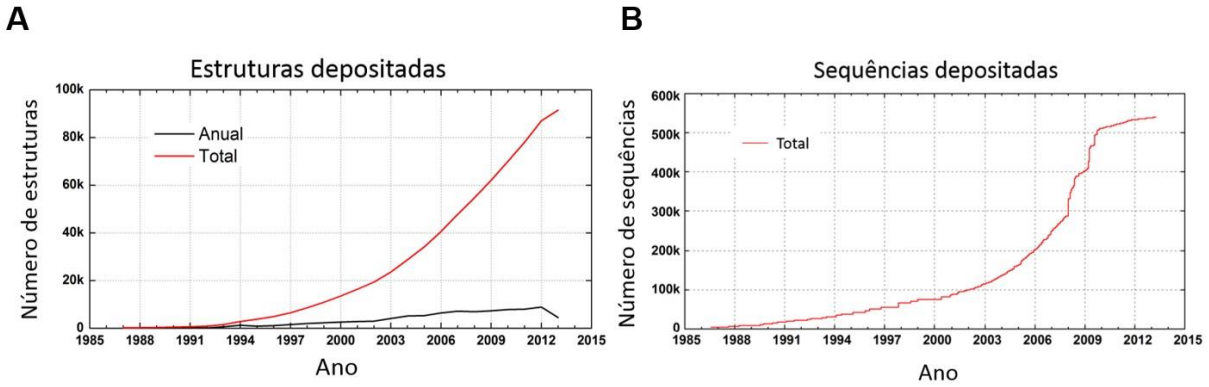


Figura 7 – (**A**) Crescimento no período 1985-2013 do número de estruturas depositas no banco de dados de proteínas (PDB) [54]. (**B**) Crescimento no período 1985-2013 do número de sequências depositas no banco de dados UniProtKB/Swiss-Prot [55].

1.2.1.1 Modelagem molecular comparativa

A modelagem comparativa de proteína (do inglês, *comparative protein modeling*), também conhecida como modelagem por homologia, é um dos métodos mais bem sucedidos de predição de estruturas 3D de proteínas [59]. A estratégia baseia-se em padrões gerais observados em nível molecular no processo de evolução biológica [60]. Brevemente, o processo de modelagem comparativa consiste em quatro etapas fundamentais: i. identificação de proteínas relacionadas com estrutura 3D conhecida (e.g., proteínas homólogas); ii. alinhamento sequencial da proteína com estrutura conhecida (estrutura molde) com a sequência de aminoácidos da proteína que se deseja modelar (estrutura alvo); iii. construção de modelos moleculares baseado na análise comparativa do alinhamento sequencial; e iv. avaliação e validação do modelo molecular.

Existem diversas estratégias e métodos disponíveis para a construção de modelos de proteínas por modelagem comparativa [61]. As primeiras abordagens utilizavam métodos de construção de corpos rígidos, nas quais as regiões conservadas entre a estrutura molde e alvo eram modeladas com base nas coordenadas atômicas dos C_α da estrutura molde [62]. Métodos mais modernos utilizam o alinhamento e identidade sequencial das proteínas molde e alvo para gerar restrições espaciais para os átomos da cadeia principal e laterais dos resíduos de aminoácidos [63,64]. Entre as ferramentas computacionais que utilizam essa abordagem, merece destaque o Modeller [61,64–67], um programa de modelagem comparativa robusto e preditivo desenvolvido pelo Andrej Sali Lab (UCSF). Modeller é capaz de construir modelos de proteínas com base na sequência de aminoácidos da proteína alvo e dados estruturais de uma ou mais estruturas 3D da proteína molde, preferencialmente homólogas à proteína alvo, com identidade sequencial superior a 25% [68–70]. Uma das características marcantes desse programa é a possibilidade de modelar alças com métodos *de novo* (e.g., *ab initio*) e a capacidade de otimização de regiões flexíveis [71]. Essas funcionalidades auxiliam significativamente a construção de modelos moleculares robustos e representativos.

1.2.2 Desenvolvimento de programas computacionais, linguagem de programação

Existem diversas estratégias e métodos para o desenvolvimento de novas ferramentas computacionais [72]. Entre as mais populares, destaca-se o modelo de desenvolvimento evolucionário, no qual versões prévias (e.g., protótipos iniciais ou versão beta) são desenvolvidas e lançadas para a comunidade em geral. Nesse processo interativo, a participação dos usuários é de fundamental importância, uma vez que são os responsáveis por indicar problemas, pontos fracos e fortes dos programas em desenvolvimento [73,74]. Essas informações são coletadas e avaliadas pelos desenvolvedores do programa que sugerem e incorporam soluções para o relançamento de versões atualizadas. O processo de atualização sequencial refina as características e funcionalidades do programa até que uma versão estável, confiável e robusta seja desenvolvida [75].

A linguagem de programação é definida como um método padronizado capaz de comunicar instruções para um computador [76]. Tais instruções podem ser executadas diretamente quando escritas em linguagem de máquina, com simples substituição quando escritas numa linguagem de montagem (do inglês, *assembly*), ou após a tradução de uma linguagem de alto nível. Apesar de existir mais de 2.000 linguagens de programação, poucas são amplamente utilizadas no desenvolvimento de programas computacionais [76]. Nesse contexto, merece destaque a linguagem Python, uma linguagem de programação interpretada de alta nível, que se destaca pela velocidade e simplicidade. Entre suas principais características, cita-se: i. clareza de sintaxe e fácil compreensão; ii. orientação a objeto intuitiva, embora não obrigatória; iii. vasta quantidade de bibliotecas disponíveis; e iv. fácil implementação de ambientes gráficos (e.g., interfaces gráficas) [77]. Além disso, a integração multiplataforma possibilita a portabilidade para a maioria dos sistemas operacionais atuais, dentre eles citam-se: i. Windows; ii. Mac OS; iii. Unix; iv. Amiga; v. OS/2; e vi. Symbian, com mínimas ou até mesmo nenhuma modificação no código. Portanto, com base em uma das mais rápidas curvas de aprendizado, extenso suporte da comunidade e distribuição livres fazem do Python uma linguagem extremamente atrativa para o desenvolvimento de programas, com destaque para aqueles com foco científico [77].

Atualmente, a linguagem é mantida em duas versões paralelas, 2.x e 3.x, que embora compartilhem muitas sintaxes, não são completamente compatíveis. A versão mais utilizada é a 2.7.3, sendo a base de muitos programas científicos de bioinformática, como PyMOL [78] e Modeller [61,64–66].

1.2.3 Ambiente gráfico e bibliotecas matemáticas

O desenvolvimento de ambientes gráficos necessita de um construtor para intermediar a interação e comunicação entre o homem e o computador [79,80]. Nesse sentido, a interface gráfica PyGTK utiliza a biblioteca GTK+ para fornecer à linguagem Python elementos e ferramentas visuais. Distribuído sob licença LGPL (*Lesser GNU Public License*), quase não existem restrições à sua modificação e redistribuição. A versão mais atual, 2.24.0, foi lançada em abril de 2011 [81].

A biblioteca Matplotlib [82] foi desenvolvida para fornecer à linguagem Python um conjunto de ferramentas capazes de construir e manipular gráficos 2D com qualidade e interatividade . Construído com o objetivo de resolver problemas simples e complexos, a Matplotlib fornece funções matemáticas predefinidas com amplo controle de parâmetros visuais, tornando as imagens geradas excelente para publicações e apresentações. Essa biblioteca fornece com sucesso ao Python as funções contidas em programas especializados como o MATLAB® e Mathematica®. Paralelamente, a biblioteca NumPy é atrativa para a computação científica por conter funções otimizadas para trabalhar com espaços N-dimensionais. Utilizado em conjunto com a Matplotlib fornece ferramentas atrativas para o tratamento de dados matemáticos.

1.2.4 Planejamento Baseado em Estrutura e Triagem Virtual

Estratégias de planejamento de candidatos a novos fármacos fundamentam-se no conhecimento prévio do processo fisiopatológico da doença e na seleção do alvo molecular adequado. O alvo selecionado (e.g., proteína, DNA ou RNA) pode ter a sua estrutura tridimensional conhecida ou não, fato que determina quais as estratégias que serão utilizadas no processo de planejamento [83]. Os grandes avanços da genômica e proteômica, bem como a evolução de métodos experimentais (e.g., cristalografia de raios X, ressonância magnética nuclear, crio-microscopia eletrônica) e computacionais (e.g., modelagem por homologia), proporcionam um aumento significativo no número de alvos moleculares que possuem suas estruturas tridimensionais disponíveis. O conhecimento da estrutura nativa do alvo macromolecular, seja uma enzima ou receptor biológico, ou da estrutura de algum complexo ligante-receptor, permite o planejamento de inibidores enzimáticos, ou agonistas/antagonistas de receptores através de estratégia de planejamento baseado na estrutura do receptor (SBDD, do inglês, *structure-based drug design*) [84].

Uma estratégia promissora para a identificação de novos ligantes é a triagem virtual de compostos (VS, do inglês, *virtual screening*). Esse método, fundamentado em estratégias computacionais, envolve a análise de grandes bases de dados virtuais de compostos através de métodos computacionais com o objetivo de identificar um pequeno subgrupo de compostos que podem ser adquiridos através de representantes comerciais ou de síntese orgânica para avaliação da atividade biológica [85]. A triagem virtual exige como pré-requisito o conhecimento dos componentes espaciais e energéticos responsáveis pela interação entre um ligante e o alvo molecular em estudo. Assim, tanto a estrutura tridimensional do alvo macromolecular como a estrutura dos ligantes devem ser previamente conhecidas e estarem disponíveis para a realização do ensaio virtual [84].

CAPÍTULO 2

OBJETIVOS

Os objetivos gerais dessa dissertação de mestrado são:

- Compreender características da *Xa*PPT passiveis de serem exploradas para o planejamento racional de agroquímicos

Os objetivos específicos dessa dissertação de mestrado são:

Estratégias Computacionais

- Desenvolver uma ferramenta robusta e eficiente para auxiliar usuários iniciantes e avançados nos estudos de modelagem comparativa.
- Construir um modelo 3D robusto e representativo da estrutur da Fosfopanteteinil transferase de *X. albilineans* (XaPPT).
- Realizar ensaio de triagem virtual com a estrutura *Xa*PPT para identificação de candidatos a novos defensivos agrícolas.

Estratégias Experimentais

- Aquisição e padronização de culturas de *Xanthomonas albilineans*
- Clonagem do gene codificante para a *Xa*PPT
- Expressão do gene codificante para a *Xa*PPT

CAPÍTULO 3

MODELAGEM MOLECULAR

3.1 Material e métodos

3.1.1 Desenvolvimento do progama ViTaMIn (*Visual Table Modeller Interface*) – uma interface gráfica para o programa Modeller

O ViTaMIn foi desenvolvido com a versão 2.7.3 do Python em ambiente de desenvolvimento integrado (IDE, do inglês, *integrated developement environment*) *Gedit 3.8.1* (The GNOME Project). Para a construção da interface gráfica foi utilizado o ambiente de desenvolvimento gráfico para GTK+ Glade 3.8.3 (The Glade Project) em sistema Linux Ubuntu 12.04 e 12.10. A interface foi construída para trabalhar com o Modeller 9.10 e 9.11.

A construção do programa foi baseada na estratégia de desenvolvimento modular, muito presente na linguagem Python. Funções de extração de informações físico-químicas a partir da sequência primária das proteínas em estudos foram importadas do projeto PDB-tools 0.1.4 [86]. As funções de modelagem foram importadas diretamente do Modeller [61,64–66]. O desenvolvimento dos menus teve como principal objetivo simplificar a construção de modelos através da orientação do usuário nos passos essenciais. Contudo, a interface é robusta e permite ao usuário experimente realizar personalizações que julgue necessárias. Uma vez carregado os arquivos com a sequência alinhada da proteína em estudo (proteína alvo) e a estrutura da proteína homóloga (proteína molde), o processo de modelagem básica pode ser todo realizado através das funções específicas na interface.

3.1.2 Busca por estruturas homólogas e alinhamento sequencial

Um estudo de caso real de modelagem comparativa foi empregado para verificar e avaliar a capacidade de construção de modelos moleculares com o auxílio da interface ViTaMIn. Para tanto, as estruturas homólogas da enzima *Xa*PPT foram

investigadas com o algoritmo *blastp*, disponível no servidor BLAST [87] e validadas no servidor HHPred [88]. Nessa busca, restringiram-se os resultados para apenas aqueles nos quais a estrutura 3D é conhecida e que podem ser úteis na construção de modelos moleculares. Nesse sentido, a busca por proteínas homólogas levou em consideração: i. a identidade sequencial (> 25%); ii. a resolução das estruturas molde (< 2.0 Å); e iii. a presença ou ausência de ligantes na estrutura, uma vez que essa informação pode ser incorporada na modelagem para a obtenção de modelos de sítio ativo mais representativo.

As estruturas homólogas de *Xa*PPT identificadas foram alinhadas por dois métodos: i. alinhamento sequencial padrão com a ferramenta ClustaW [89,90], no qual as estruturas primárias são alinhadas com relação ao grau de conservação entre os resíduos de aminoácido; e ii. alinhamento sequencial baseado na estrutura secundária com o módulo de alinhamento integrado ao programa Modeller. Nessa estratégia, intervalos (do inglês, *gaps*) são adicionados à sequência para que os enovelamentos sejam mantidos. Em ambas as estratégias, o critério utilizado para seleção do melhor alinhamento foi a representação da conservação das estruturas secundárias, preditas pelo servidor PSIPRED [91].

3.1.3 Construção, seleção e validação dos modelos moleculares da *Xa*PPT

Duas estratégias foram utilizadas para construir e comparar modelos moleculares de *Xa*PPT. A primeira estratégia empregou o método de alinhavamento (do inglês, *threading*), utilizando o servidor I-Tasser [92,93]. Nessa estratégia, foi realizada uma busca no banco de dados de proteínas por fragmentos proteicos que alinhavam com a sequência em estudo. A predição da estrutura secundária para a *Xa*PPT guiou a seleção dos fragmentos molde no banco de fragmentos. Em seguida, uma análise de significância estatística do melhor alinhamento avaliou a qualidade do alinhamento entre os fragmentos. Os fragmentos selecionados, correspondentes estruturais às regiões alinhadas, foram organizados em uma estrutura 3D e minimizados por simulações sucessivas com o método de Monte Carlo (*e.g., replica-exchange*). Esse processo foi repetido diversas vezes até que o modelo final estivesse otimizado. As

regiões do modelo que não correspondiam a nenhum fragmento estrutural identificado foram modeladas por métodos alternativos (e.g., método *ab initio*) que utilizaram restrições de campos de força molecular para definir interações e distâncias entre átomos em um determinado ambiente químico.

A modelagem comparativa utilizou o programa Modeller com o auxílio da interface ViTaMIn. Inicialmente, as proteínas homólogas a *Xa*PPT cuja estrutura 3D encontravam-se disponíveis foram obtidas do PDB. Em seguida, as características estruturais das proteínas molde foram utilizadas como parâmetros de restrição para a construção de modelos moleculares da *Xa*PPT. Os modelos foram selecionados com base no menor valor de Energia Proteica Otimizada Discreta (DOPE, do inglês, ***D**iscret **O**ptmized **P**rotein **E**nergy*), um parâmetro quantitativo que indica o potencial estatístico da energia de cada modelo gerado [94]. A validação dos modelos construídos foi realizada primeiramente no ViTaMIn e posteriormente nos servidores SAVES [95] e ProSa [96].

3.1.4 Triagem virtual

A estratégia de triagem virtual para identificação de inibidores da enzima *Xa*PPT incluiu dois métodos de busca: i) busca estocástica ou aleatória, implementada no programa Glide [97] (Schrödinger); e ii) busca sistemática, implementada no programa Surflex-Dock [98] (Tripos). O esquema da **Figura 8** ilustra as etapas utilizadas no processo de triagem virtual.

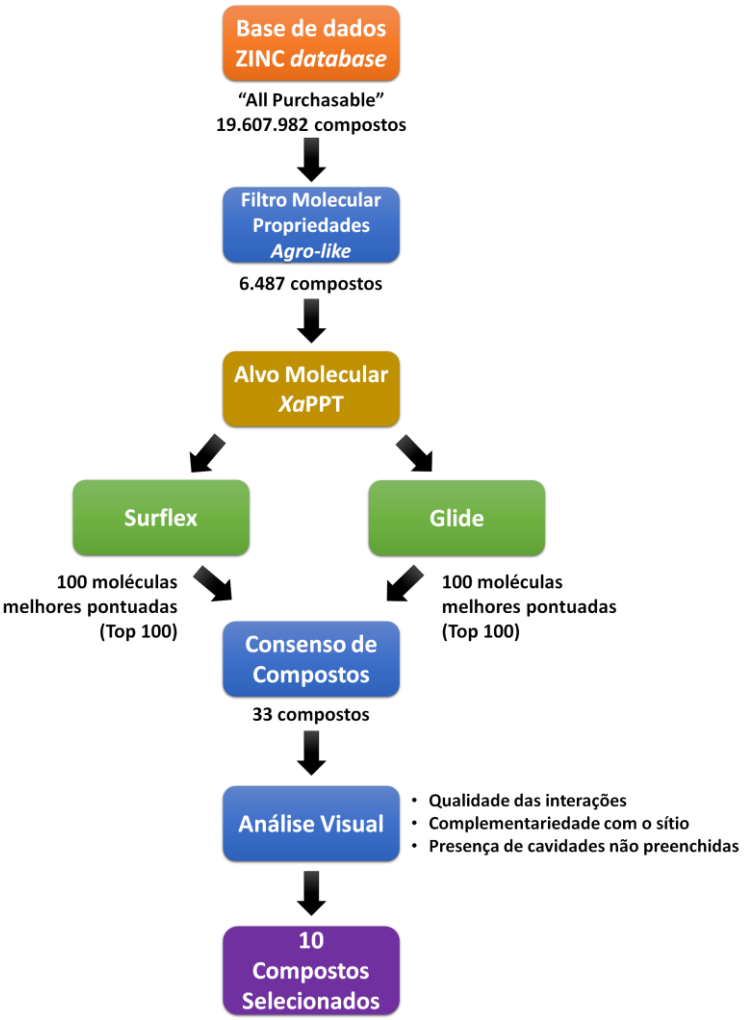


Figura 8 – Esquema das etapas envolvidas na triagem virtual de inibidores da *Xa*PPT.

3.1.5 Sítio de ligação

O modelo da *Xa*PPT, obtido por modelagem comparativa, em complexo com a coenzima A foi utilizada como alvo molecular. Os átomos de hidrogênio foram adicionados na geometria padrão utilizando o módulo Biopolymer, implementado no plataforma SYBYL X. Os resíduos de histidinas, glutaminas e asparaginas presentes na cavidade de ligação foram manualmente checados quanto à orientação, estado de protonação e tautômeros mais prováveis com o programa PyMOL 1.5.0.3. O sítio ativo para estudos de docagem molecular foi definido incorporando todos os resíduos de aminoácidos presentes no interior de uma esfera de 12 Å de raio (**Figura 9**),

posicionada ao redor da molécula do ligante. Os parâmetros padrões dos programas de docagem molecular foram utilizados para obtenção dos modelos moleculares de interação.

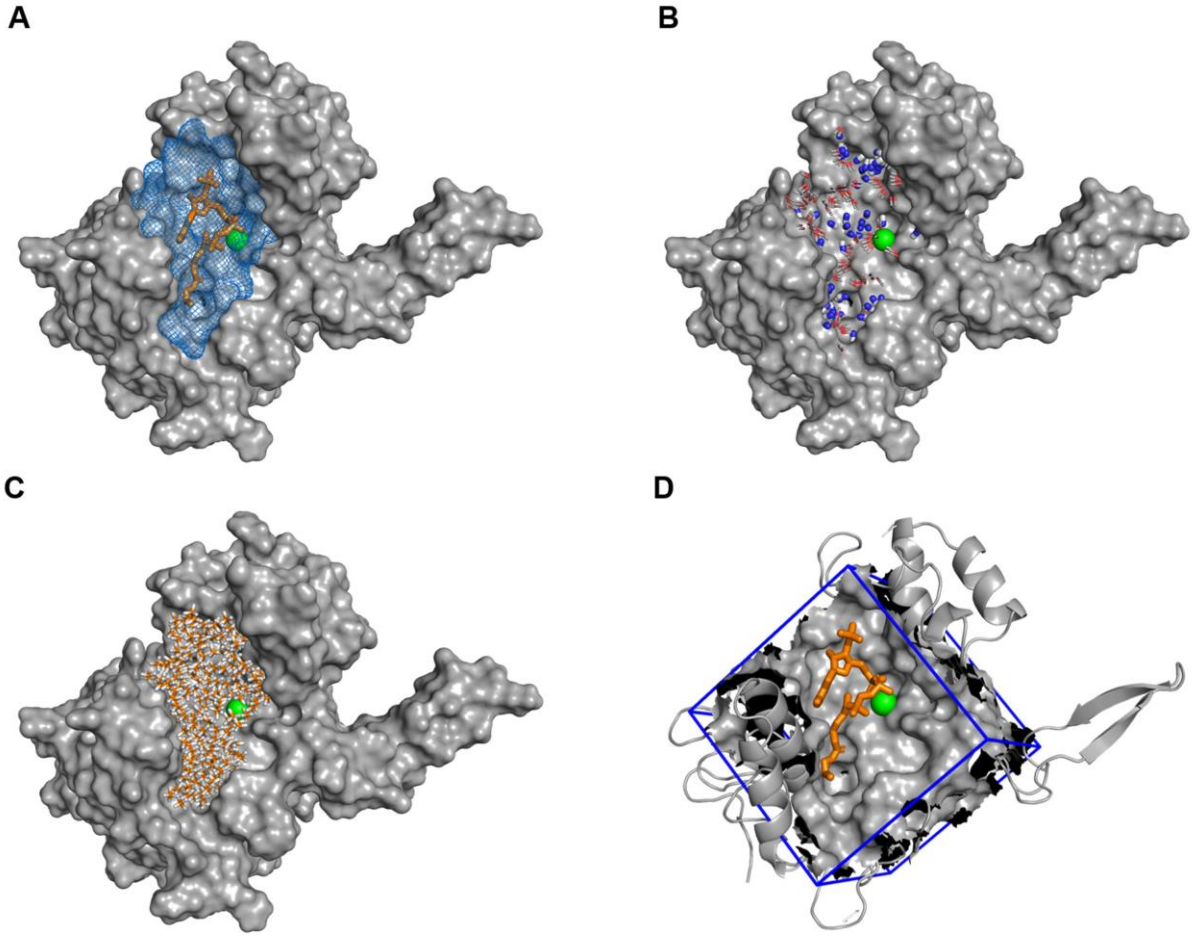


Figura 9 – Sítio de ligação utilizado na triagem virtual com a enzima *Xa*PPT. (**A**) Sítio de ligação (malha azul) explorado pelo programa Surflex. A molécula do ligante coenzima A está indicada como bastão laranja e o átomo de Mg^{2+} está indicado como esfera verde. (**B**) Mapeamento do sítio para identificação de interações favoráveis para grupos doadores (NH, bastão azul) e aceptores de ligação (C=O, bastão vermelho) de hidrogênio realizado pelo Surflex. (**C**) Mapeamento do sítio para identificação de interações favoráveis para grupos hidrofóbicos (CH_4, bastão laranja) realizado pelo Surflex. (**D**) Sítio de ligação explorado pelo programa Glide.

3.1.5.1 Base de dados de compostos

A seleção da base de dados para a realização da triagem virtual levou em consideração a facilidade para aquisição das moléculas. Diante disso, a base de dados de compostos selecionada foi um subconjunto personalizado de compostos da ZINC database [99]. As moléculas desse subconjunto tiveram as propriedades físico-químicas calculadas e aquelas que não apresentaram as características agroquímicos-similares (do inglês, *agro-like*) desejadas (i.e., doadores de ligação de hidrogênio ≤ 3, aceptores de ligação de hidrogênio ≤ 12, massa molecular ≤ 500 e coeficiente de partição log *P* ≤ 5) [100] foram excluídas da base de dados.

3.2 Resultados e discussão

3.2.1 ViTaMIn – Visual Table Modeller Interface

O programa de modelagem Modeller é uma das ferramentas computacionais mais importante para a construção de modelos moleculares de proteínas. Contudo, sua aplicação é muitas vezes comprometida devido às características intrínsecas de entrada de dados, através de linhas de comando, e necessidade de conhecimento básico da linguagem de programação Python. Diante disso, uma nova interface gráfica, denominada ViTaMIn (***Vi**sual **Ta**ble **M**odeller **In**terface*) foi desenvolvida com o objetivo de ampliar o público alvo e facilitar a construção de modelos moleculares úteis para estudos em biologia estrutural e química medicinal (**Figura 10**).

ViTaMIn é uma ferramenta autônoma programada em linguagem Python e disponível para a plataforma Linux. ViTaMIn é útil, principalmente, para usuários iniciantes no campo da modelagem molecular de proteínas, pois sua arquitetura em janelas permite a construção, análise, visualização e otimização de modelos proteicos de forma simples, fácil e direta. O programa pode ser obtido no endereço http://rafaelguido.tk/. O programa é disponibilizado em um arquivo compactado que contém o arquivo Python executável (vitaminc.py) e arquivos auxiliares. Por ser um programa disponibilizado no formato “executável” não há necessidade de instalação do pacote.

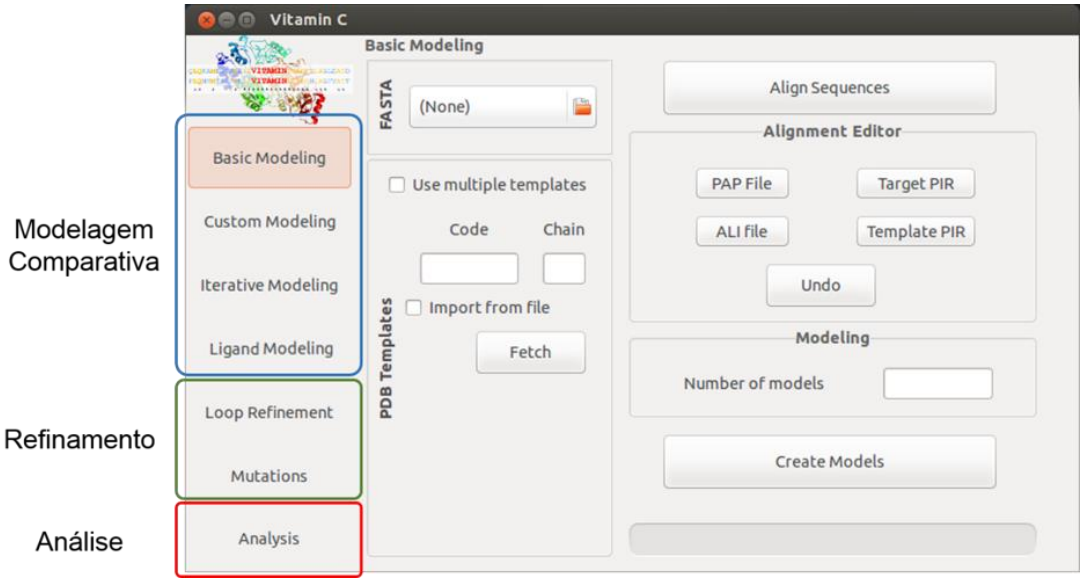


Figura 10 – Visão geral da interface gráfica ViTaMIn com destaque para as principais funções disponíveis.

Atualmente, existem outras interfaces disponíveis para o Modeller, contudo, a maioria não é mais rotineiramente atualizada e, portanto, não funcionam com as versões mais modernas do programa (i.e., 9.X). Entre as interfaces que estão disponíveis e atualizadas, destaca-se o EasyModeller [101] e o PyMOD [102]. A **Tabela 1** apresenta as principais características e funções disponíveis no ViTaMIn, EasyModeller e PyMOD.

Tabela 1 – Principais características e funções das interfaces ViTaMIn, EasyModeller e PyMOD disponíveis para o Modeller.

Função	ViTaMIn	EasyModeller	PyMOD
Integração com o PyMOL	✓		✓
Interface Amigável	✓	✓	
Personalização de *scripts*	✓	✓	
Fácil instalação	✓	✓	
Análise dos resultados	✓		
Funções avançadas do Modeller	✓		

A análise comparativa entre o ViTaMIn e o EasyModeller indica que o ViTaMIn oferece maior possibilidade de interação do usuário com o processo de modelagem,

bem como personalização de rotinas (do inglês, *scripts*), portanto, é útil para usuários iniciantes e avançados. Além disso, a arquitetura do programa permite que todo processo de construção dos modelos seja realizado em uma única janela. A navegação entre outras janelas torna-se necessária somente nas etapas subsequentes de refinamento e validação. A integração da interface com o PyMOL é um aspecto marcante que diferencia substancialmente o ViTaMIn do EasyModeller (**Figura 11**). Nesse contexto, a capacidade de avaliação e análise visual em tempo real dos modelos construídos bem como das estruturas homólogas utilizadas auxilia significativamente o usuário na tomada de decisões sobre os parâmetros utilizados na modelagem. No caso do PyMOD, essa possibilidade de integração com o PyMOL encontra-se disponível através de um *plugin*. Contudo, o PyMOD não oferece nenhuma opção para avaliação ou modificação dos parâmetros de modelagem. Além disso, a instalação do PyMOD não é simples e requer um nível intermediário de conhecimento do sistema operacional, enquanto o ViTaMIn, por ser distribuído no formato de arquivo executável, necessita apenas da execução de procedimentos triviais como a extração do conteúdo do pacote disponibilizado e a execução de um arquivo Python (disponível nas distribuições Linux). Uma característica importante incorporada no ViTaMIn consiste na aba de análise dos modelos construídos. Essa funcionalidade é inexistente no EasyModeller e PyMOD. Portanto, o ViTaMIn é uma alternativa simples, funcional e, principalmente, amigável ao usuário iniciante e experiente nos estudos de modelagem comparativa.

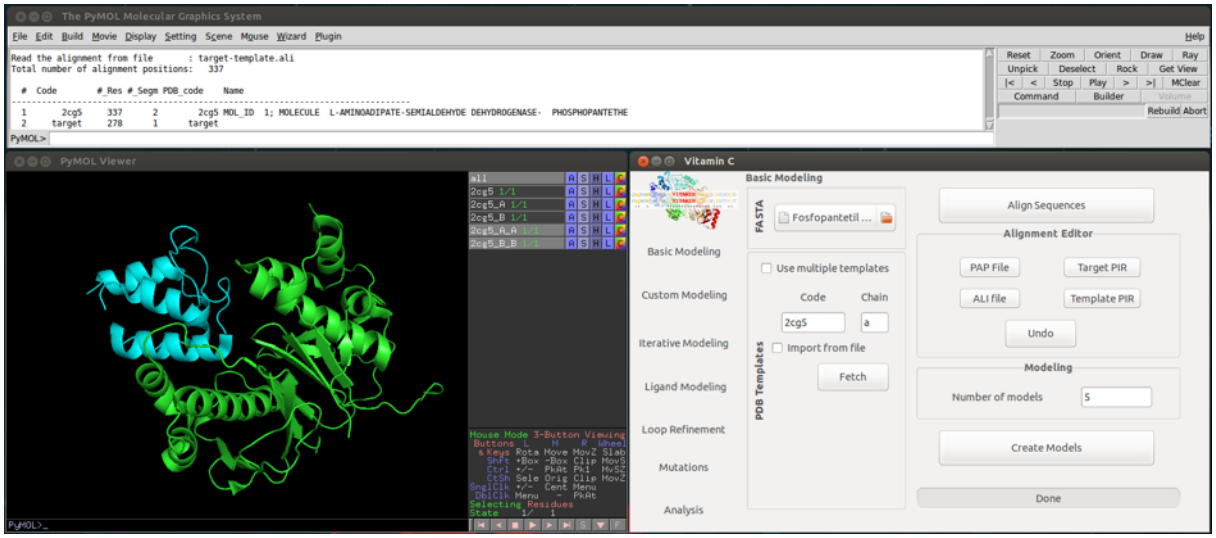


Figura 11 - Integração do ViTaMIn com o PyMOL.

ViTaMIn é divido em sete abas e três módulos: i. modelagem comparativa; ii. refinamento dos modelos; e iii. análise dos modelos gerados (**Figura 10**). Nas próximas seções são brevemente descritas as principais características de cada módulo.

3.2.1.1 Modelagem comparativa

No módulo de modelagem comparativa, há quatro opções de construção de modelos que utilizam diferentes rotinas (do inglês, *script*) do Modeller (i.e., i. *Basic Modeling*; ii. *Custom Modeling*; iii. *Iterative Modeling*; e iv. *Ligand Modeling*). A modelagem básica (*Basic Modeling*) utiliza uma rotina pré-otimizada que auxilia a construção de modelos moleculares em poucos passos e interferência opcional do usuário em relação ao alinhamento das sequências em estudo (proteína alvo e proteína molde) e a estrutura homóloga (proteína molde). Brevemente, os passos para a construção de modelos no modo de modelagem básica consistem em (**Figura 12**): i. selecionar a sequência da proteína alvo (formato FASTA); ii. identificar e selecionar a estrutura homóloga (código PDB); iii. alinhar as sequências; iv. otimizar o alinhamento (opcional); e v. selecionar o número de modelos que serão construídos.

É importante mencionar que a etapa de otimização de alinhamento é opcional, podendo ser realizada com programas específicos (e.g., JalView [103], Clustal W2 [89,90]). A estrutura homóloga pode ser indicada ao programa através da indicação do código PDB e a respectiva cadeia polipeptídica ou carregada de um arquivo local no formato PDB. A informação da cadeia é opcional, na ausência dessa informação, a construção dos modelos será realizada utilizando-se a cadeia com a melhor pontuação no alinhamento.

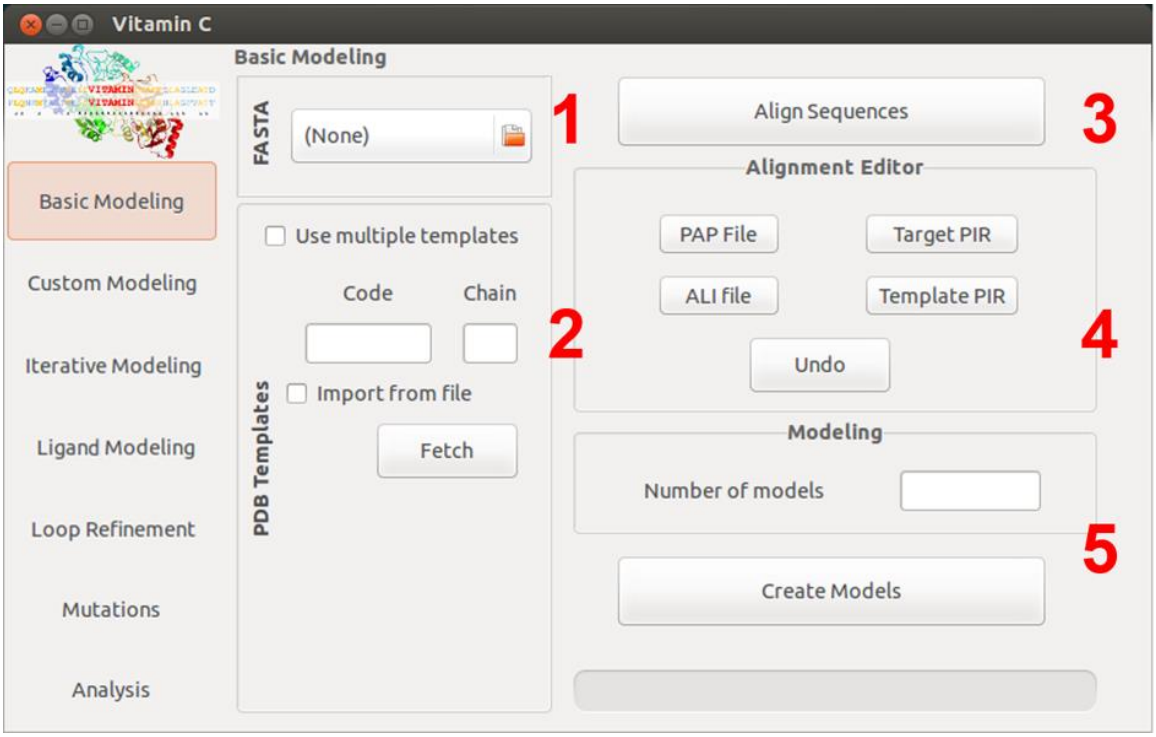


Figura 12 – Etapas para modelagem comparativa no modo *Basic Modeling*. Os números em vermelho indicam as etapas do processo de construção dos modelos. 1. seleção da sequência da proteína alvo (formato FASTA); 2. identificação e selecição da estrutura homóloga (código PDB); 3. alinhamento das sequências alvo e molde; 4. otimização do alinhamento; 5. Seleção do número de modelos que serão construídos.

O módulo de modelagem personalizada (*Custom Modeling*) permite ao usuário editar opções do algoritmo empregado no processo de construção dos modelos (e.g., método de refinamento, valor máximo de molpdf, método de construção do modelo inicial). O processo iterativo (*Iterative Modeling*) e modelagem com ajuste do ligante (*Ligand Modeling*) são realizados automaticamente, seguindo os mesmos passos da modelagem básica. Uma vez construídos, os modelos ficam disponíveis na pasta de trabalho. Nesse contexto, visando-se a melhor organização dos trabalhos devido à grande quantidade de dados gerados, é fortemente recomendado a criação de pastas específicas para cada projeto de modelagem em desenvolvimento.

3.2.1.2 Refinamento dos modelos

O refinamento das cadeias laterais é realizado automaticamente durante o processo de construção dos modelos moleculares. Além disso, para o melhor ajuste dos modelos encontra-se disponível no ViTaMIn a opção de refinamento de alças (*Loop Refinement*) e mutação sítio dirigida (*Mutations*) (**Figura 10**). Alças de tamanhos distintos e mutações pontuais são comuns em proteínas homólogas, logo, a modelagem comparativa dessas regiões pode ser desafiadora. Frequentemente, o posicionamento de algumas alças em modelos cristalográfico é incerto. Isso ocorre em função de diversos fatores, entre eles, citam-se: i. baixa resolução da estrutura; ii. empacotamento cristalino; iii. alças flexíveis e expostas ao solvente. Essa incerteza apresenta grande impacto na modelagem comparativa, pois essas alças terão baixa qualidade no modelo construído. Para minimizar esse impacto, emprega-se métodos *ab initio* para a construção de alças que apresentem conformações energeticamente favoráveis. Brevemente, o processo de refinamento das alças se divide em quatro etapas (**Figura 13**): i. seleção da região para refinamento, guiado pela análise do gráfico de pontuação DOPE para cada resíduo do modelo construído; ii. identificação do resíduo inicial e final da alça; iii. análise visual e avaliação da alça com o módulo integrado no PyMOL; e. iv seleção do número de modelos que serão construídos.

A seleção das regiões para refinamento é baseada no gráfico de pontuação DOPE para os resíduos do modelo (**Figura 13B**). A pontuação DOPE [94] é um potencial estatístico frequentemente empregado para avaliar modelos de proteínas construídos por modelagem comparativa. Essa pontuação se baseia em uma função cujo mínimo global corresponde a estrutura nativa de um conjunto não redundante de 1.472 estrutura cristalográficas. Em geral, resíduos com pontuação > -0,03 devem ser avaliados e refinados (**Figura 13B**).

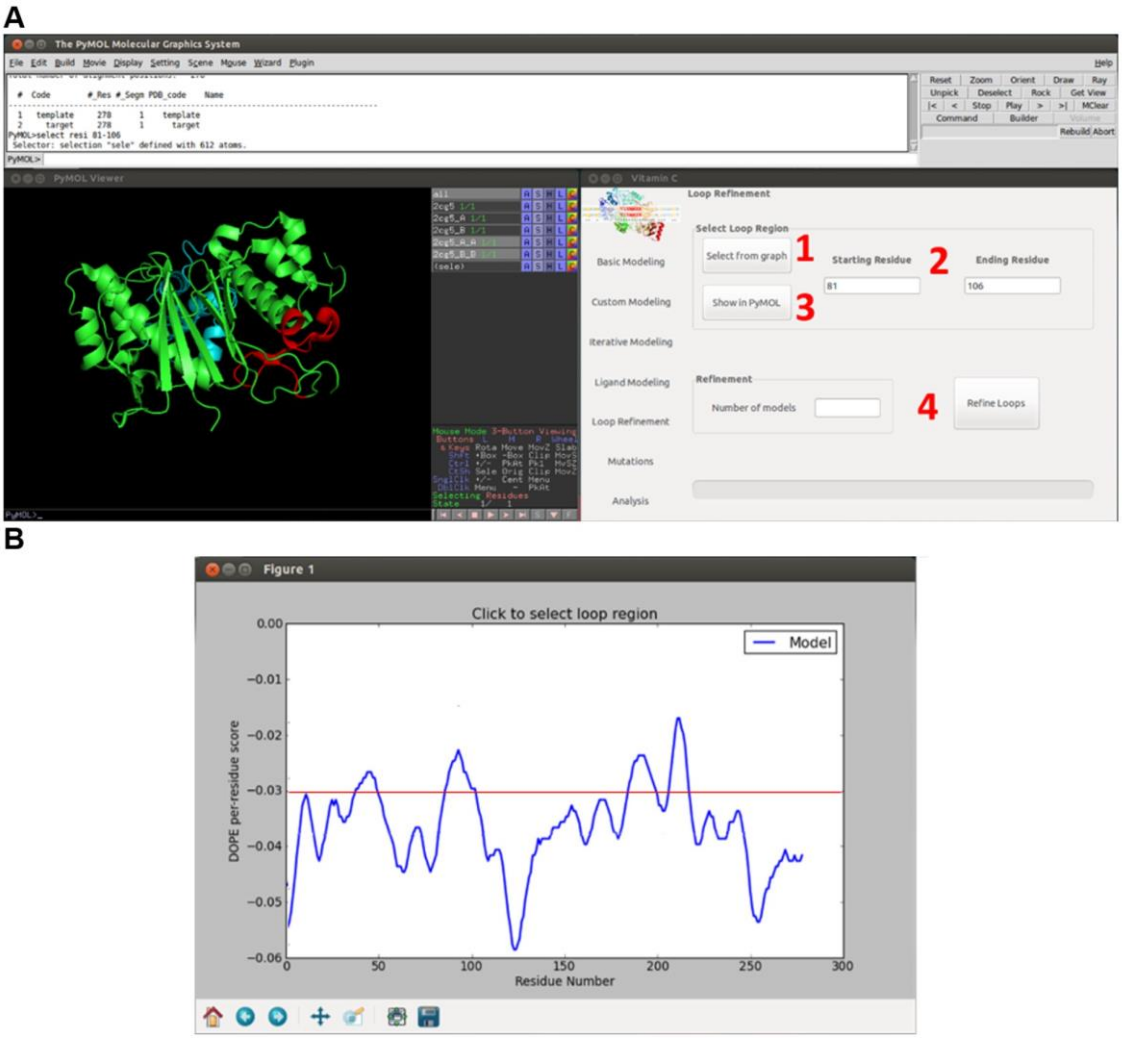


Figura 13 – (**A**) Módulo integrado ViTaMIn-PyMOL para o refinamento das alças. Região em vermelho na interface do Pymol indica a região alvo do refinamento (resíduos 81-106). Os números em vermelho indicam as etapas do processo de refinamento de alças. 1. seleção da região para refinamento; 2. identificação do resíduo inicial e final da alça; 3. visualização com o módulo integrado no PyMOL; 4. seleção do número de modelos construídos e início do processo. (**B**) Exemplo representativo de gráfico de pontuação DOPE dos resíduos do modelo construído.

A opção de refinamento de mutações (*Mutations*) é útil tanto para o refinamento de estruturas obtidas por modelagem comparativa quanto para estruturas obtidas por métodos experimentais para as quais se deseja avaliar as implicações de mutações pontuais na estrutura. A rotina para o refinamento de mutações consiste em duas etapas (**Figura 14**): i. seleção dos resíduos que serão mutados; ii. seleção do número de modelos para construção.

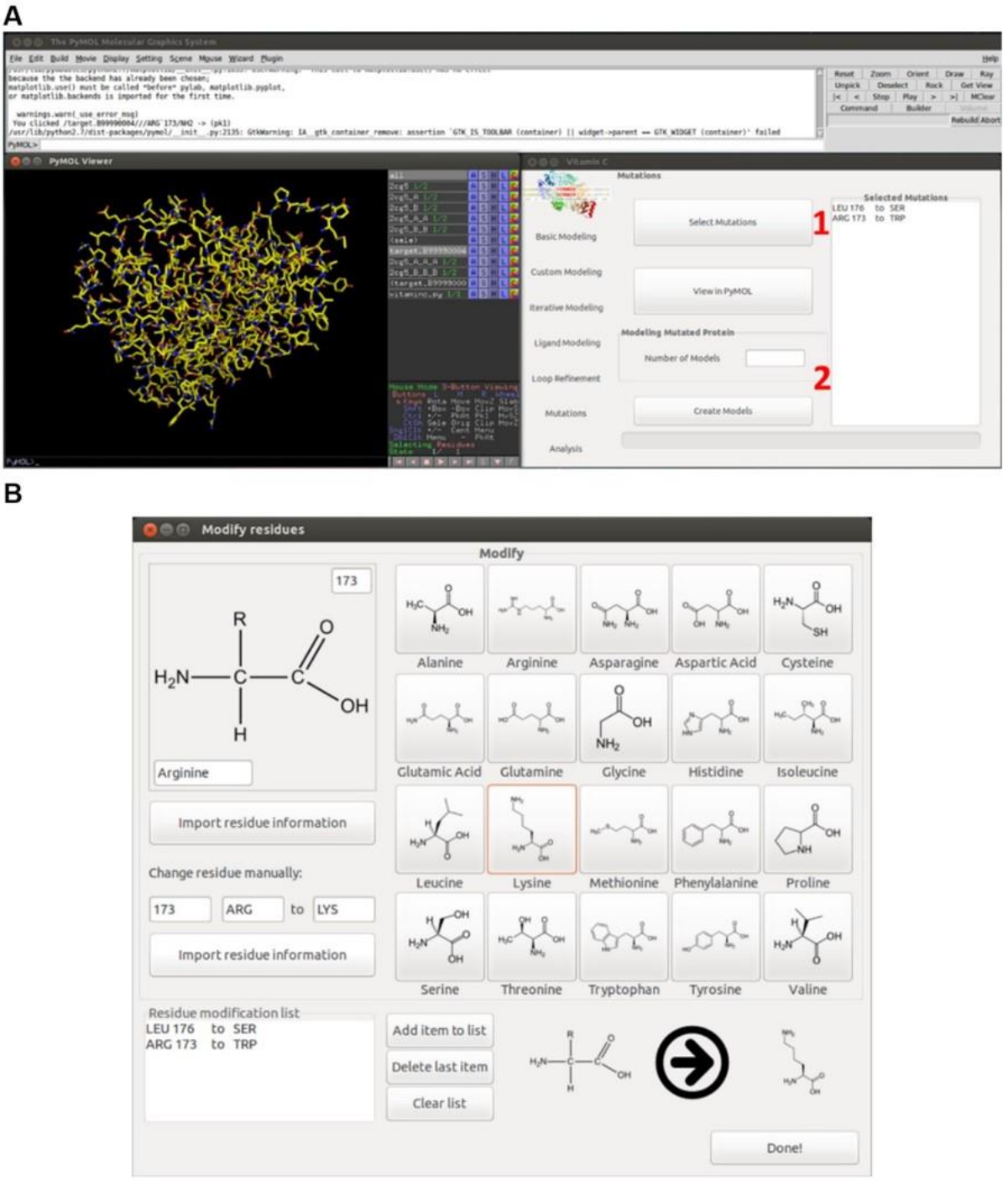


Figura 14 – (**A**) Módulo integrado ViTaMIn-Pymol para a seleção e refinamento de mutações pontuais. Os números em vermelho indicam as etapas do processo de seleção e refinamento de mutações. 1. seleção dos resíduos que serão mutados; 2. seleção do número de modelos para construção e início do processo. (**B**) Detalhe da janela de seleção de resíduos para mutação.

3.2.1.3 Análise dos modelos

A aba análise (*Analysis*) fornece informações estruturais e físico-químicas da proteína em estudo (e.g., número de aminoácidos, coordenadas do centro de massa, massa molecular, ponto isoelétrico teórico, carga total), dados para avaliação dos modelos construídos (e.g., comparação da pontuação DOPE entre os modelos gerados, comparação da pontuação DOPE entre o melhor modelo e a estrutura homóloga utilizada na modelagem) e uma ferramenta para gerar o diagrama de Ramachandran para os modelos construídos. Além disso, permite ao usuário visualizar os modelos construídos no PyMOL (**Figura 15**). É importante mencionar que a etapa de análise pode ser realizada em qualquer momento do processo de modelagem e construção dos modelos (e.g., após a modelagem básica, após o refinamento ou modelagem de mutantes).

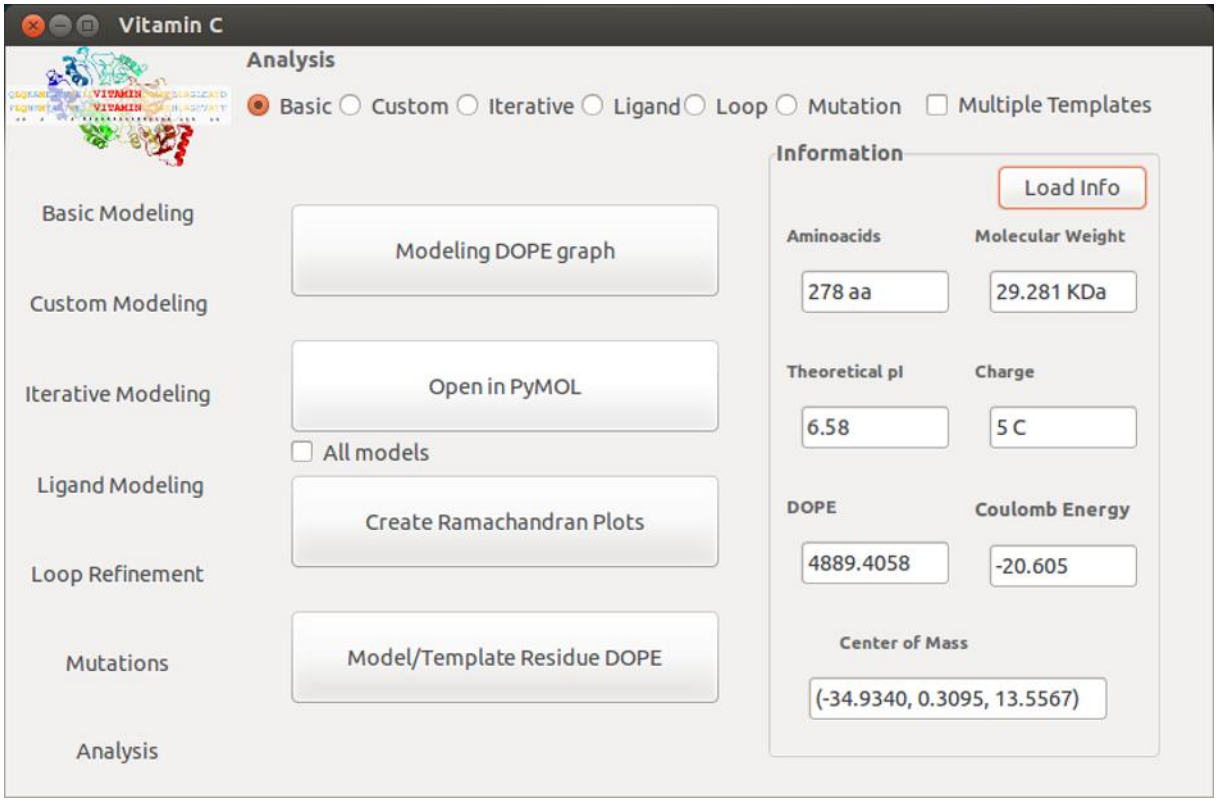


Figura 15 – Visão geral da janela de análise dos modelos moleculares.

3.2.2 Estudo de caso: Enzima fosfopanteteinil transferase de *X. albilineans*

Tendo em vista a necessidade de avaliação e aplicação da interface gráfica ViTaMIn, estudos de modelagem molecular comparativa foram conduzidos com a enzima *Xa*PPT, um alvo molecular atrativo e validado para o desenvolvimento de novos candidatos a agroquímicos para o controle da escaldadura da folhas em cana-de-açúcar. Os resultados obtidos nessa parte do trabalho validaram o ViTaMIn como uma alternativa interessante e robusta para a construção de modelos moleculares. Além disso, o melhor modelo construído indicou elementos moleculares relevantes que auxiliam a melhor compreensão dos aspectos estruturais e funcionais da enzima *Xa*PPT. As próximas seções descrevem os detalhes dos procedimentos realizados bem como as características moleculares da *Xa*PPT.

3.2.2.1 Estudo da sequência de aminoácidos da *Xa*PPT por bioinformática

A proteína *Xa*PPT é uma proteína monomérica composta por 278 resíduos de aminoácidos, massa molecular de 29 kDa e ponto isoelétrico teórico de 6,6 (calculado pela ferramenta PROTPARAM[104]). A análise dos domínios constituintes da *Xa*PPT, realizado no servidor SMART [105,106], indicou a presença de um domínio proteína carreadora de acila sintetase (ACPS, do inglês, *acyl-carrier-protein synthase* – Pfam PF01648) na região entre os resíduos de aminoácidos 160 e 226 (*e-value* = 0,0062) (**Figura 16**). O domínio ACPS está presente em proteínas da superfamília 4'-fosfopanteteinil transferase responsáveis pela transferência do grupo 4'-fosfopanteteina (4'-PP) da coenzima A (CoA) para um resíduo de serina invariante em proteínas carreadoras de acila (ACP) [107]. Além disso, realizou-se a predição de estruturas secundárias da *Xa*PPT no servidor PSIPRED [91] (**Figura 17**). O resultado da análise indicou que a *Xa*PPT pertence à classe das proteínas alfa beta (CATH 3.90.470.20), com predominância de folhas-β antiparalelas.

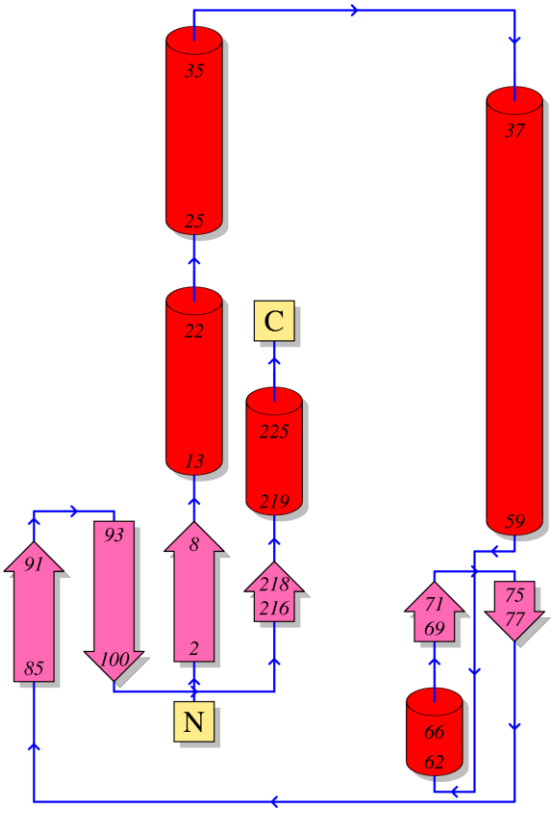


Figura 16 – Topologia representativa do domínio ACPS presente na *Xa*PPT.

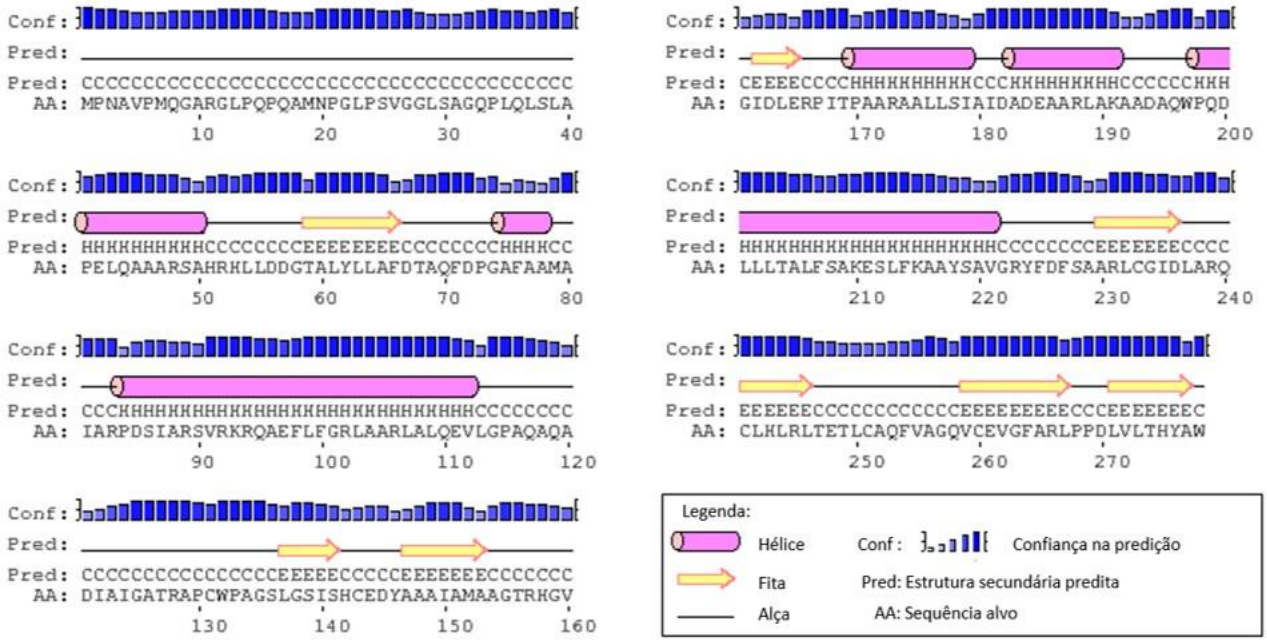


Figura 17 – Predição de estruturas secundárias da *Xa*PPT.

3.2.2.2 Identificação e análise estrutural de proteínas homologas à *Xa*PPT

A identificação de proteína homólogas a *Xa*PPT foi realizada no servidor HHPred[108], uma das ferramentas útil para alinhamento e identificação de proteínas, amplamente empregada em estudos de modelagem comparativa. A análise indicou que a *Xa*PPT é homóloga à proteína 4'-fosfopanteteinil transferase de *Bacillus subtilis* (*Bs*PPT) e à aminoadipato-semialdeído desidrogenase fosfopanteteinil transferase de humanos (*Hs*PPT) (**Figura 18**). A identidade sequencial entre *Bs*PPT e *Xa*PPT é de 44%, com cobertura de região de 53%, enquanto a identidade sequencial entre *Hs*PPT e *Xa*PPT é de 35%, com cobertura de região de 68%. Ambas as proteínas homólogas possuem estrutura 3D disponível em alta resolução (~1,9 Å) (*Bs*PPT – PDB ID, 1QR0; HsPPT – PDB ID, 2C43) [7,109]. Esses valores sugeriram que a estrutura *Bs*PPT era uma candidata adequada à proteína molde para os estudos de modelagem comparativa da *Xa*PPT.

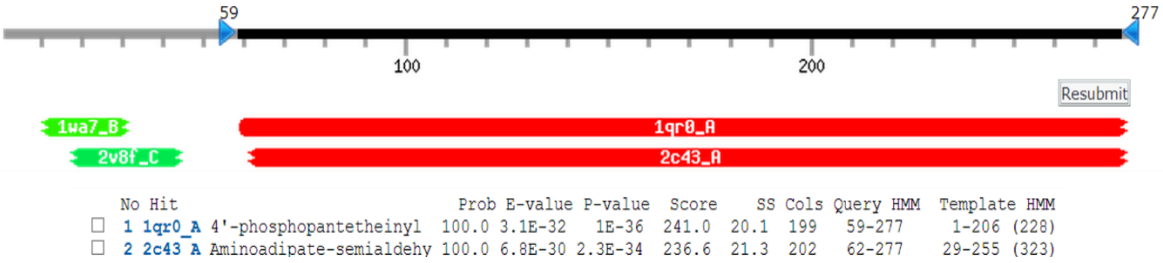


Figura 18 – Identificação de estruturas homólogas da *Xa*PPT no servidor HHPred.

Uma característica estrutural relevante das proteínas homólogas identificadas é que ambas foram determinadas na forma de complexos ternários, ou seja, proteína-cofator-íon, PPT-CoA-Mg^{2+}, respectivamente (**Figura 19A**). Tendo em vista a descoberta e desenvolvimento de novos inibidores da *Xa*PPT como candidatos a agroquímicos, essa informação estrutural torna-se extremamente interessante para o presente estudo, pois permite a identificação do sítio de ligação bem como a melhor compreensão dos determinantes estruturais responsáveis pelo processo de afinidade e reconhecimento molecular [110]. Além disso, as informações moleculares dos complexos cristalográficos fornecem dados importantes para a construção e validação do modelo da *Xa*PPT. Por exemplo, na estrutura *Bs*PPT o íon Mg^{2+} encontra-se hexacoordenado (duas coordenações axiais e quatro equatoriais) no sítio de ligação

formado pelos resíduos carregados negativamente Asp107, Glu109 e Glu151 (**Figura 19B**). Além disso, a análise estrutural do complexo ternário indica que a CoA liga-se na parte central do sítio ativo da *Bs*PPT (**Figura 19C**) estabilizado por uma extensa rede de interações polares e apolares (**Figura 19D**).

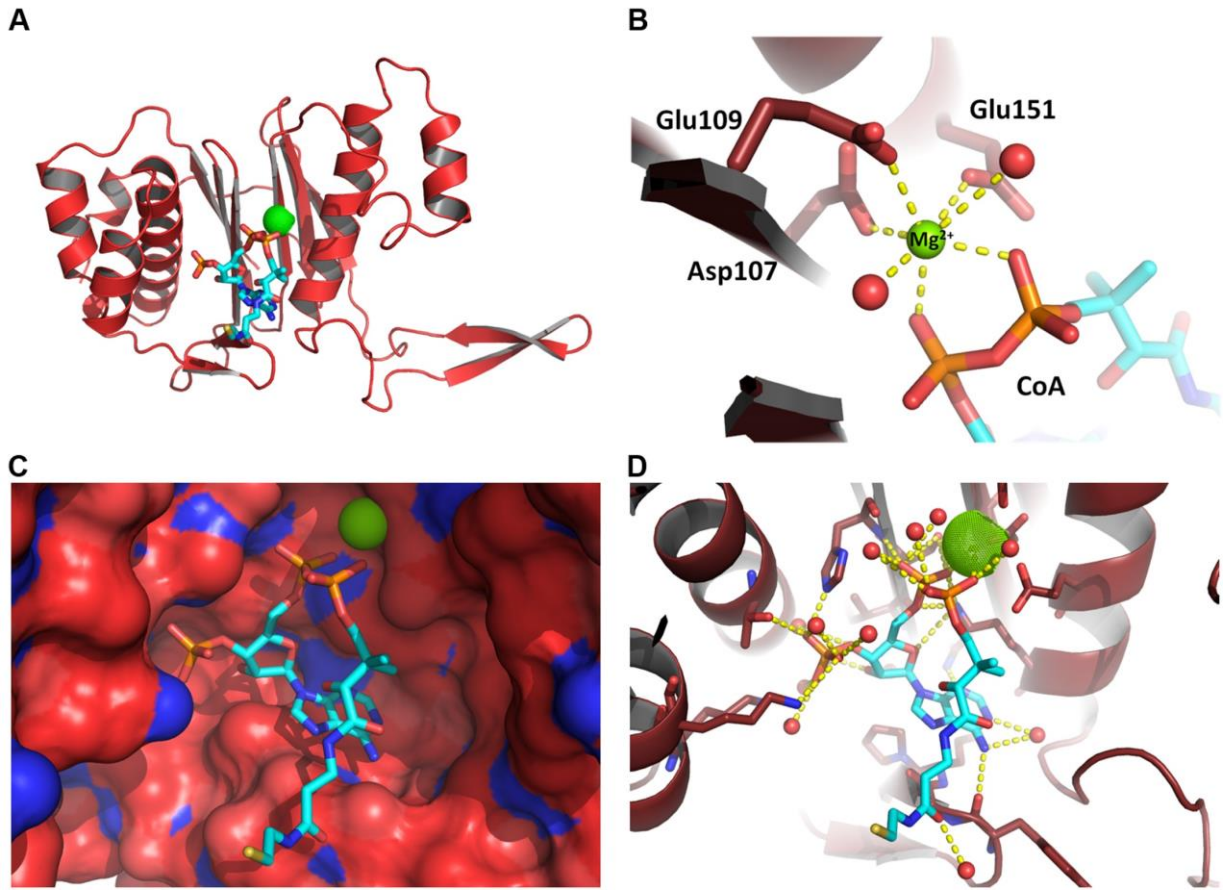


Figura 19 – (**A**) Complexo ternário da *Bs*PPT-CoA-Mg^{2+}. A *Bs*PPT está indicada como modelo de fitas, o ligante CoA está indicado como modelo de bastões, o íon Mg^{2+} está indicado como modelo de esfera e as interações polares estão indicadas como linhas tracejadas. (**B**) Sítio de ligação do Mg^{2+} hexacoordenado pelos resíduos Asp107, Glu109, Glu151, duas moléculas de água (esferas vermelhas) pelo substituinte difosfato da CoA. (**C**) Sítio de ligação do CoA. O sítio de ligação está indicado pela superfície de Connolly. (**D**) Resíduos do sítio de ligação da BsPPT e interações polares que estabilizam o modo de ligação do CoA.

3.2.2.3 Construção, seleção e validação dos modelos de *Xa*PPT

Duas estratégias foram utilizadas para a construção de modelos da *Xa*PPT: I-Tasser e ViTaMIn. A construção de modelos pelo servidor I-Tasser [92,93] utiliza a estratégia de alinhavamento e avalia os modelos obtidos de acordo com os valores de C-Score e TM-Score. No caso dos modelos construídos com o ViTaMIn, a estratégia de modelagem consiste no método de satisfação de restrições espaciais e avalição dos modelos pelos valores de pontuação DOPE.

3.2.2.3.1 Alinhavamento: I-Tasser

As coordenadas cristalográficas dos complexos ternário da *Bs*PPT e *Hs*PPT (PDB IDs, 1QR0 e 2C43, respectivamente) foram utilizadas como proteínas molde para a construção do modelo 3D da *Xa*PPT. O modelo foi construído automaticamente no servidor I-Tasser [92,93] e considerou o íon Mg^{2+} no sítio de ligação da enzima. A presença do íon Mg^{2+} no sítio de ligação é importante na estabilização da CoA [109]. Para a seleção e validação do modelo construído pelo I-Tasser foram utilizados os valores de C-score e TM-socre [111]. O valor C-score é uma pontuação que indica a confiança na estimativa da qualidade dos modelos preditos pelo I-Tasser. Os valores de C-score variam no intervalo de -5 a 2, onde valores positivos indicam maior grau de confiança no modelo construído. O parâmetro TM-Score calcula as distâncias médias entre os átomos dos resíduos equivalentes do modelo construído e das estruturas utilizadas como moldes. Contudo, o TM-Score difere do valor de raiz quadrada do desvio quadrático médio (RMSD, do inglês, *root mean square deviation*) uma vez que pondera as distâncias calculadas entre átomos equivalentes atribuindo maior peso para aqueles que apresentam menores valores distâncias. Portanto, esse parâmetro é menos susceptível a erros de modelagem locais que comumente elevam substancialmente o valor de RMSD. Valores menores de TM-Score indicam maior similaridade entre o modelo construído para a proteína

alvo e a proteína molde. Valores de TM-score > 0,5 indicam modelos construídos com topologia correta, enquanto valores de TM-score < 0,17 indicam similaridade estrutural aleatória. Os valores calculados de C-Score e TM-Score para o melhor modelo de *Xa*PPT construído pelo servidor I-Tasser foram -1,8 e 0,5, respectivamente, sugerindo que o modelo construído apresenta qualidade razoável e topologia correta (**Figura 20**).

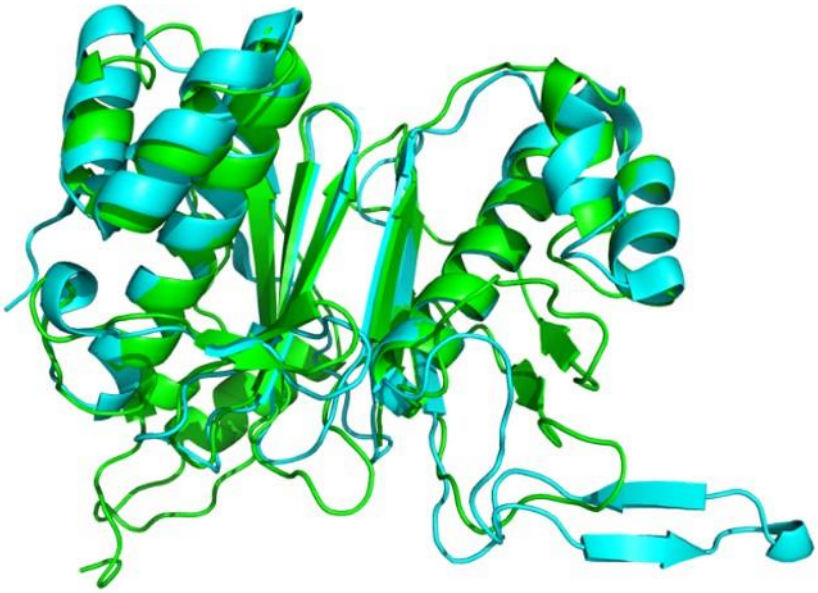


Figura 20 – Alinhamento estrutural do modelo construídos para a *Xa*PPT (verde) pelo servidor I-Tasser com a proteína molde (ciano) *Bs*PPT (PDB ID, 1QR0).

Adicionalmente, o modelo foi avaliado pelo servidor RAMPAGE [112] que fornece informações sobre a qualidade dos modelos estruturais de proteínas com base no diagrama de Ramachandran. A análise do diagrama de Ramachandran para o melhor modelo de *Xa*PPT construído pelo I-Tasser indica 87% dos resíduos em regiões favoráveis, 8% em regiões permitidas e 5 % em regiões proibidas (**Figura 21**). Os resíduos identificados em regiões proibidas são resíduos localizados em alças expostas ao solvente.

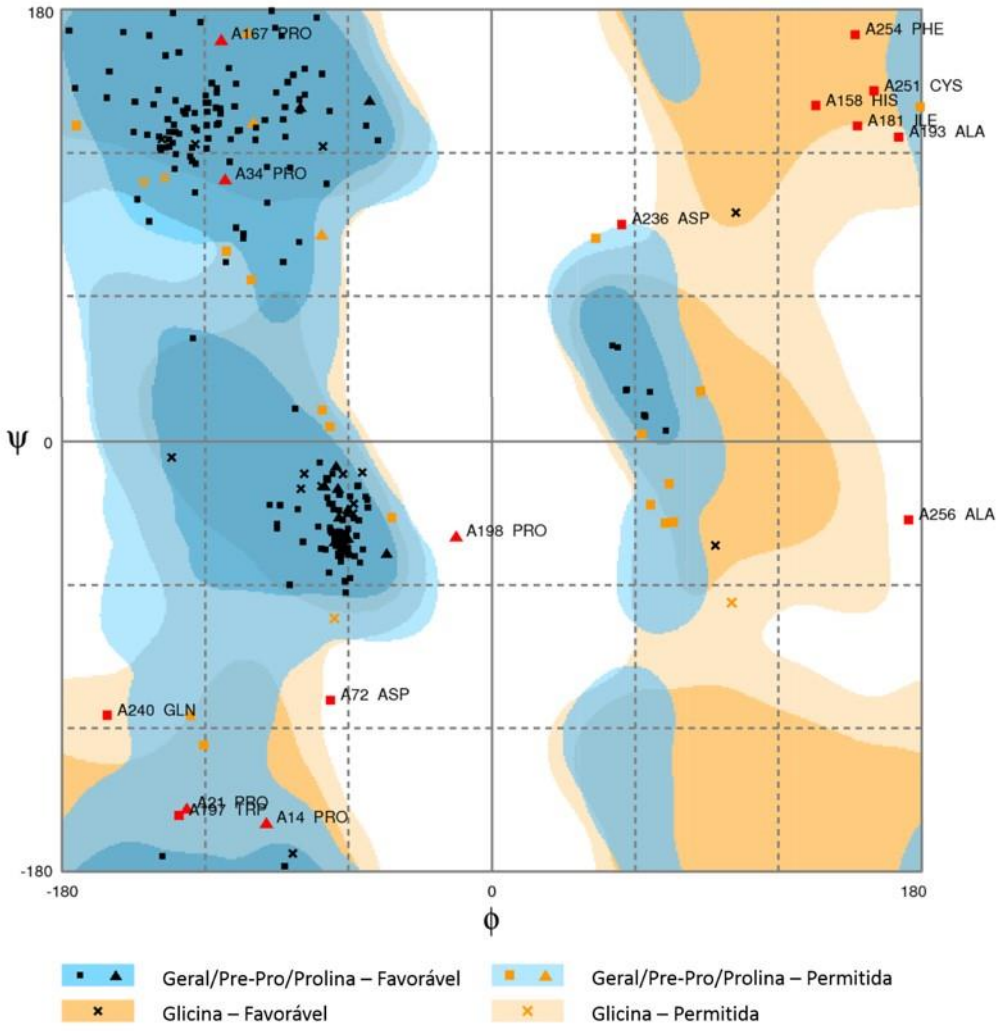


Figura 21 – Diagrama de Ramachandran para o melhor modelo construído pelo servidor I -Tasser empregando a estratégia de alinhavamento. Os pontos vermelhos correspondem aos resíduos localizados em regiões proibidas.

3.2.2.3.2 Modelagem comparativa: Modeller-ViTaMIn

Uma estratégia alternativa foi empregada para a construção de modelos 3D da *Xa*PPT. A estratégia selecionada foi a modelagem comparativa utilizando o algoritmo de restrição espacial do Modeller através da interface ViTaMIn. Nesse sentido, as coordenadas cristalográficas da *Bs*PPT foram usadas como proteínas molde. Os modelos foram construídos considerando o íon Mg^{2+} no sítio ativo. Tendo em vista a necessidade de explorar o espaço amostral para a seleção da conformação mais

representativa da *Xa*PPT, 100 modelos moleculares foram gerados e avaliados em relação à pontuação DOPE (**Figura 22**).

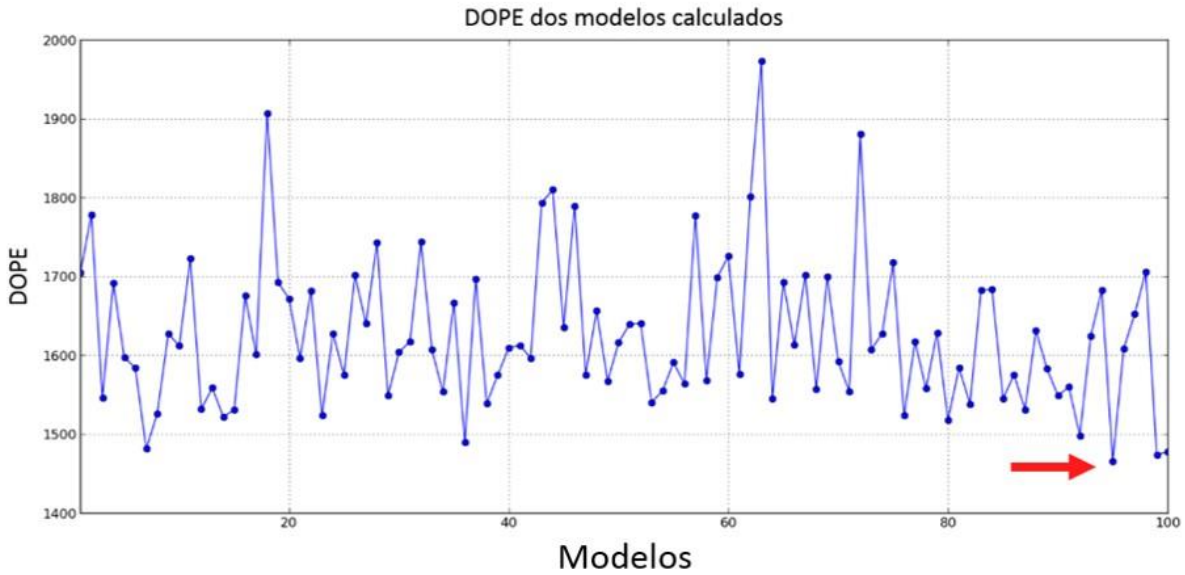


Figura 22 – Pontuação DOPE dos modelos construídos pelo ViTaMIn. A seta vermelha indica o modelo com menor pontuação DOPE.

O melhor modelo selecionado apresenta valor de pontuação DOPE = 1.463, valor de RMSD = 0,7 Å em relação à estrutura homóloga *Bs*PPT e valor de TM-Score = 0,3 (calculado pelo servidor I-Tasser) (**Figura 23**).

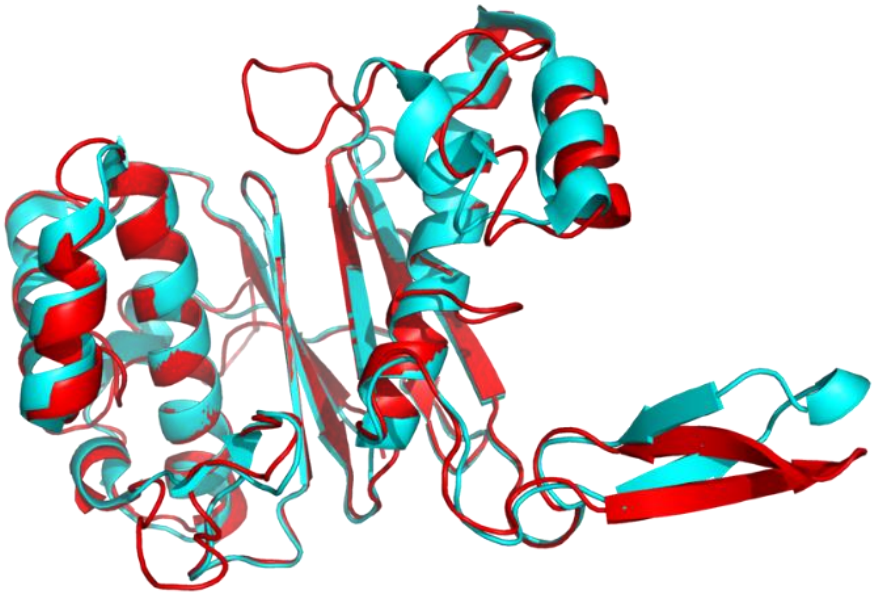


Figura 23 – Alinhamento estrutural do modelo construídos para a *Xa*PPT (vermelho) pelo ViTaMIn com a proteína molde (ciano) BsPPT (PDB ID, 1QR0).

O diagrama de Ramachandran indica 92% dos resíduos em regiões favoráveis, 7% em regiões permitidas e apenas 1% (3 resíduos) em regiões proibidas (**Figura 24**). Os resíduos observados em regiões proibidas do diagrama de Ramachandran estão localizados em região de alça exposta ao solvente.

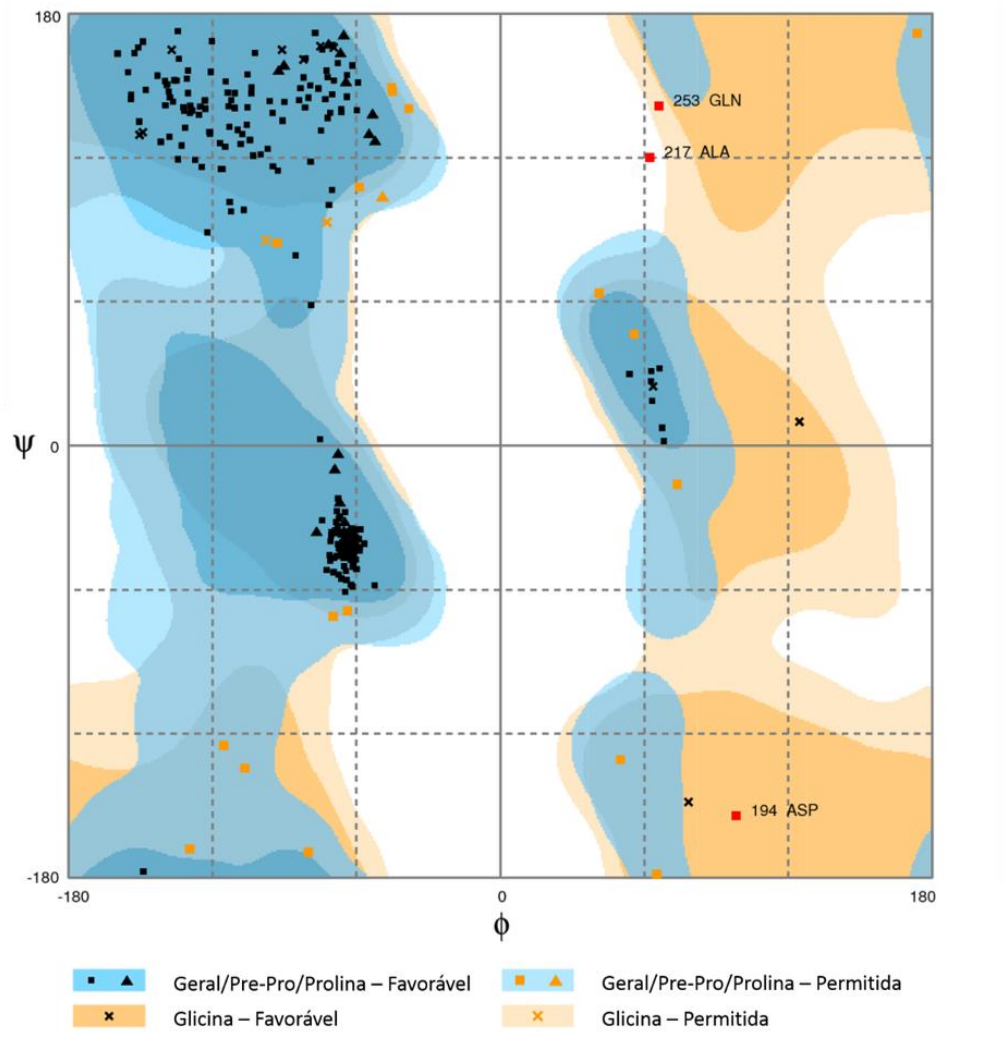


Figura 24 – Diagrama de Ramachandran para o melhor modelo construído pelo método de restrições espaciais do Modeller com a interface ViTaMIn.

3.2.2.3.3 Análise Estrutural dos modelos 3D da *Xa*PPT

O servidor Stride [113] foi utilizado para verificar o conteúdo de estruturas secundárias do modelo construído para a *Xa*PPT. A comparação entre os dados

estruturais fornecidos pelo servidor Stride e a predição das estruturas secundárias previamente preditas pelo servidor PSIPRED indicam significativa concordância entre o dado predito e o modelado (**Figura 25**).

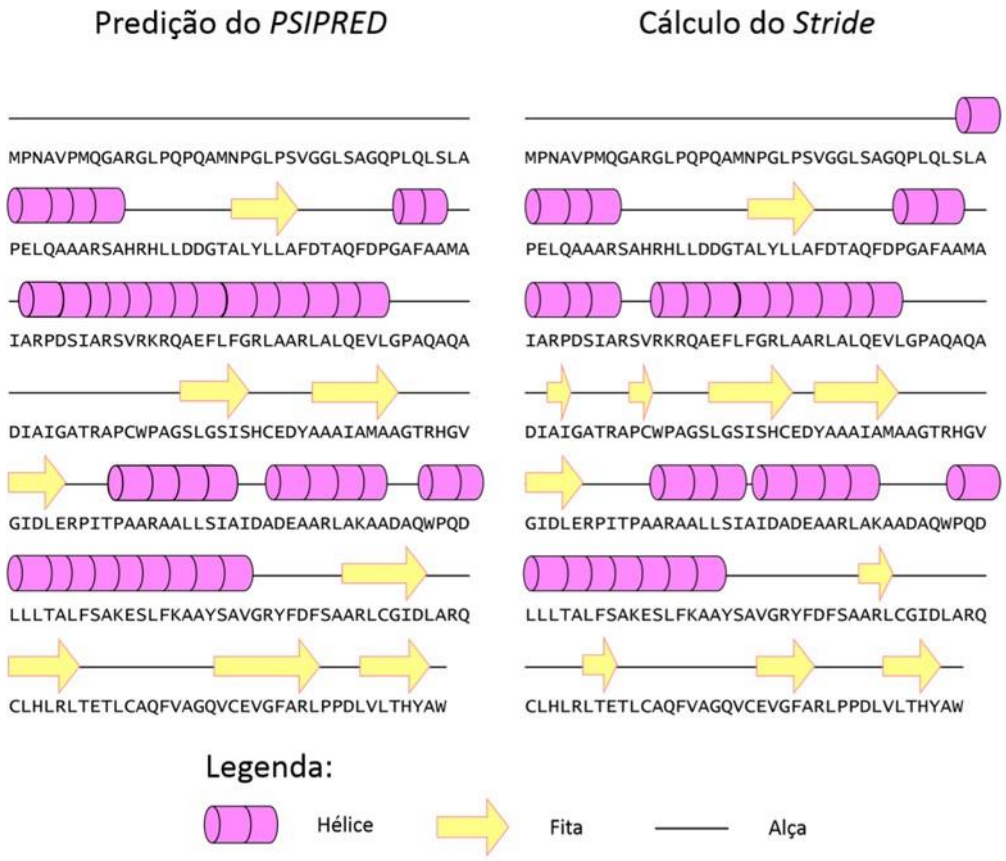


Figura 25 – Comparação entre estrutura secundária predita da *Xa*PPT pelo método PSIPRED e o verificado no modelo construído pelo método Stride.

Para auxiliar os estudos estruturais do melhor modelo construído para a *Xa*PPT foi utilizado o módulo de análise integrado com o programa PyMOL da interface ViTaMIn. O alinhamento estrutural dos modelos da *Xa*PPT construídos pelo I-Tasser e ViTaMIn sugere que a região entre os aminoácidos 200 e 220 é flexível (**Figura 26A**). Essa região apresenta um segmento de folhas-β antiparalelas conectadas por um grampo β. Esse dado é corroborado pela análise do fator B para os átomos da na região 200-220 da proteína molde *Bs*PPT. O fator B, ou fator de temperatura, é um parâmetro quantitativo de deslocamento atômico, reportado em $Å^2$, que indica a incerteza na posição atômica dos resíduos de uma macromolécula. O maior deslocamento dos resíduos 200-220 sugere desordem nessa região, fato que pode estar relacionado com a flexibilidade (**Figura 26B**). O calibre da representação e a temperatura da cor indicam a variação do deslocamento atômico para a região. Em

vermelho, as regiões com maior deslocamento e em azul as regiões com menor deslocamento.

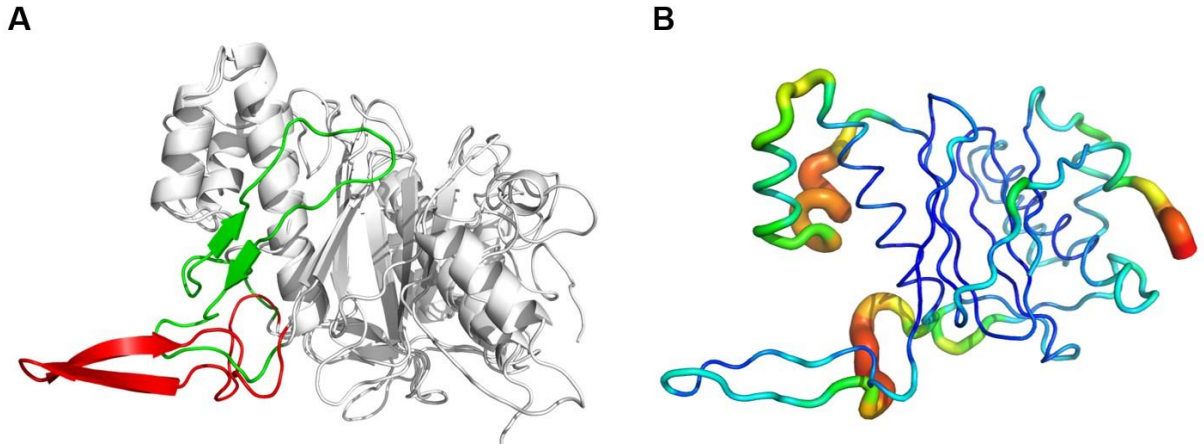


Figura 26 – (**A**) Sobreposição dos modelos construídos pelo I-Tasser (verde) e ViTaMIn (vermelho). Em destaque a região flexível dos resíduos de aminoácidos 200-220. (**B**) Representação esquemática da variação do fator B da *Bs*PPT (PDB ID, 1QR0). O calibre da representação e a temperatura da cor indicam a variação da deslocamento atômico. O espectro azul-vermelho representa o intervalo de 10-75 $Å^2$, respectivamente. Regiões em azul e vermelho apresentam menor e maior deslocamento atômico, respectivamente.

Após o refinamento das cadeias laterais, observa-se que os rotâmeros estruturais dos resíduos da *Xa*PPT são equivalente àqueles observados na estrutura da proteína molde (*Bs*PPT). Por exemplo, a adoção de um rotâmero específico do resíduo Arg92, orienta o grupo guanidina da cadeia lateral em posição favorável para doar ligações de hidrogênio para os grupos carbonilas das cadeias principais dos resíduos Leu189 e Ala190. Além disso, observa-se que o grupo eletropositivo guanidina, está orientado para o dipolo C-terminal da α-hélice central (α6). É importante mencionar que na estrutura *Bs*PPT, existe um resíduo de lisina (K75) que estabelece interações similares com resíduos equivalente (Q156 e K155) e a α-hélice (α6).

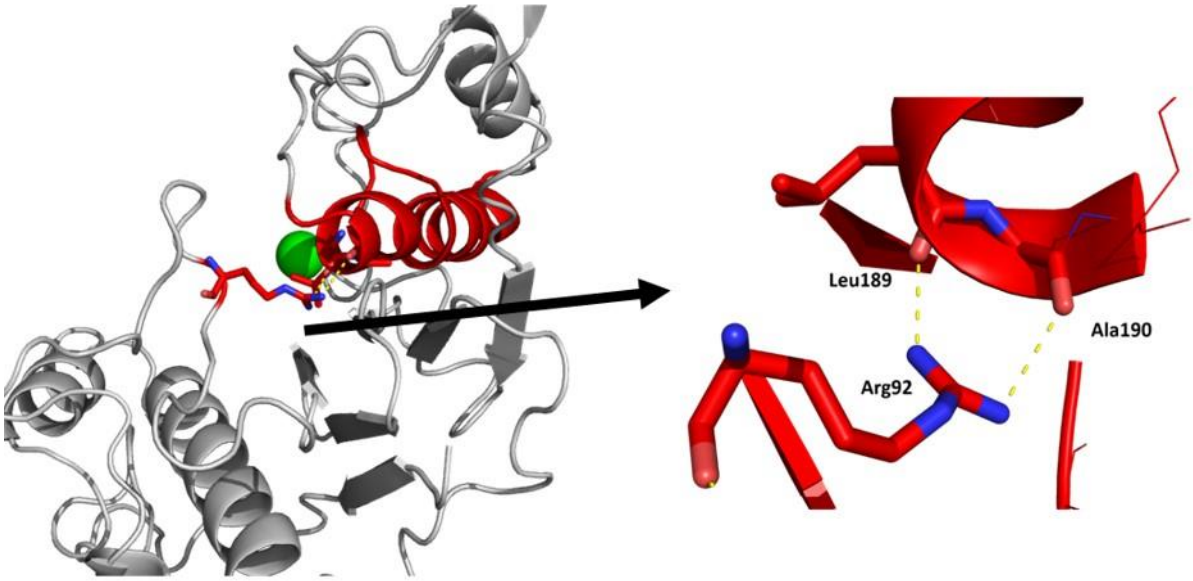


Figura 27 – Interação da Arg92 com os resíduos Leu189 e Ala190 da α-hélice

Para confirmar a representatividade conformacional do modelo de *Xa*PPT construído pelo ViTaMin foi modelado o modo de ligação da CoA no sítio ativo. O procedimento de docagem molecular foi realizado com o programa Autodock Vina, que emprega algoritmo genético como função de busca para amostragem conformacional da molécula do ligante [114]. A pontuação para o melhor modo de ligação predito do CoA foi -7,8, que equivale a um valor de afinidade na faixa nanomolar. A inspeção do modo de ligação do CoA no sítio ativo da *Xa*PPT sugere um mecanismo de reconhecimento molecular do CoA pela *Xa*PPT similar àquele observado no complexo ternário cristalográfico *Bs*PPT-CoA-Mg^{2+} (**Figura 28**), apesar da baixa identidade sequencial entre as proteínas [49].

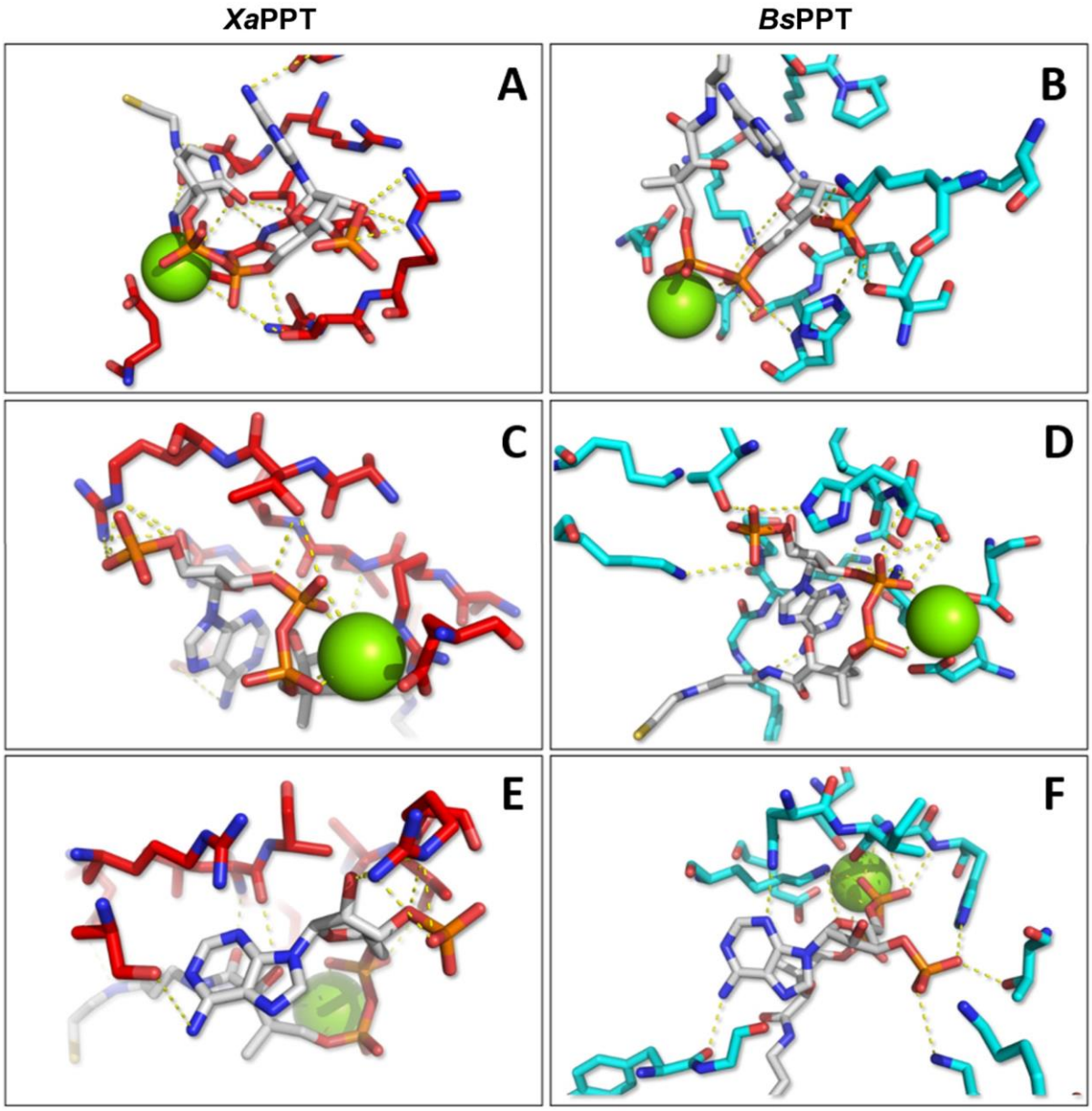


Figura 28 – Modo de ligação da CoA no sítio de ativo da *Xa*PPT e *Bs*PPT. (**A**) e (**B**) Destaque para as interações com o grupo 3-fosfato do anel ribose da adenosina. (**C**) e (**D**) Destaque para as interações entre o grupo difosfato e o íon Mg^{2+}. (**E**) e (**F**) Destaque para as interações entre o grupo adenina e os resíduos do sítio de ligação.

3.2.3 Triagem virtual de inibidores da *Xa*PPT

A estratégia de triagem virtual utilizada consistiu na aplicação de um filtro molecular para a seleção de moléculas com propriedades físico-químicas adequadas

para o desenvolvimento de novos agroquímicos. Em seguida, as características topológicas e químicas do sítio ativo da *Xa*PPT foram utilizadas para selecionar compostos capazes de se ligar na cavidade de ligação. Para tanto, dois procedimentos independentes de docagem molecular foram empregados. A última etapa consistiu na aplicação de uma análise de consenso para seleção de candidatos a inibidores da *Xa*PPT.

3.2.3.1 Alvo molecular, sítio de ligação e base de dados de compostos

O modelo estrutural da *Xa*PPT, construído e validado nessa dissertação de mestrado, foi utilizado como alvo molecular para a identificação de moléculas promissoras como candidatas a inibidores da enzima alvo. O sítio de ligação explorado na triagem virtual foi a cavidade de ligação da CoA. Nesse contexto, todos os resíduos contidos numa esfera de 12 Å de raio foram considerados como parte integrante do sítio de ligação.

A base de dados de compostos selecionada para os estudos de triagem virtual foi a *ZINC database* [99], uma base de dados de livre acesso e especializada para estudos de triagem virtual. Os compostos depositados na *ZINC database* são disponíveis comercialmente para aquisição. Atualmente, a *ZINC database* possui catalogado mais de 21 milhões de compostos disponibilizados em formato 2D e 3D (http://zinc.docking.org). Como o objetivo dessa etapa do trabalho era a identificação de compostos candidatos a novos inibidores da *Xa*PPT que apresentassem potencial para se tornarem novos agroquímicos, um filtro molecular especializado foi aplicado para a seleção de um subgrupo específico de moléculas com propriedades físico-químicas desejadas. Devido às diversas características e propriedades semelhantes entre moléculas de fármacos e agroquímicos [11], abordagens de sucesso previamente aplicadas para o estudos de propriedades de fármacos (e.g., regra dos 5 de Lipinski [115]) foram também investigadas em agroquímicos [100,116]. As conclusões obtidas nesses estudos deram origem a um conjunto de propriedades denominado “agro-similar” (do inglês, *agro-like*) ou parâmetros de Tice. A **Tabela 2** apresenta as propriedades investigadas e os limites identificados para moléculas com propriedades fármaco- e agroquímico-similar.

A análise dos dados da **Tabela 2** indica que os limites para as propriedades massa molecular (< 500) e coeficiente de partição (< 5) são os mesmos para os fármacos e agroquímicos. Esse dado é justificado uma vez que ambos os parâmetros estão associados às características de permeação em membranas celulares. Moléculas que apresentam grande massa molecular ou com elevado valor de log P tem a capacidade de atravessar membranas significativamente reduzidas devido a dificuldades de

permeação ou por apresentarem baixa solubilidade nos veículos utilizados para a administração da molécula, respectivamente.

Tabela 2 – Comparação dos limites entre as propriedades físico-químicas de fármacos (Lipinski, fármaco-similar) e agroquímicos (Tice, agroquímico-similar).

Propriedade	Lipinski	Tice
Massa Molecular	< 500	< 500
Coeficiente de partição	< 5	< 5
Nº de átomos doadores de ligação de H	< 5	< 3
Nº de átomos Aceptores de ligação de H	< 10	< 12

A **Tabela 2** indica que existem diferenças marcantes nos limites para o número máximo de doadores e aceptores de ligação de hidrogênio para os fármacos e agroquímicos. Em geral, os agroquímicos apresentam menor tolerância para grupos doadores de ligação de hidrogênio. Essa característica está provavelmente relacionada à necessidade dos agroquímicos de apresentar maior estabilidade química para resistir aos processos de biotransformação nas espécies que causam prejuízos as lavouras [11].

Ao final da aplicação do filtro molecular especializado para seleção de moléculas com propriedades agro-similar à base de dados inicialmente com 19.607.982 moléculas foi reduzida para 6.487 moléculas que apresentam características agroquímico-similar.

3.2.3.2 Docagem molecular e seleção de candidatos a inibidores da *Xa*PPT

O processo de docagem molecular pode ser dividido em duas etapas complexas: modelagem do modo de ligação (predição da conformação bioativa) e predição da afinidade do ligante pelo sítio. A predição do modo de ligação de moléculas no sítio de ligação de proteínas é considerada a etapa mais simples e robusta do processo, com

considerável nível sucesso alcançado [83]. Para realizar essa primeira etapa, os programas utilizam diferentes métodos de amostragem que atribuem flexibilidade à molécula do ligante (e.g., construção incremental, método de Monte Carlo, algoritmo genético). A etapa de predição da afinidade do ligante pela proteína é realizada baseada em funções de pontuação desenvolvidas para avaliar e classificar o modo de ligação proposto na etapa anterior. As funções de pontuação utilizadas podem ser agrupadas em três categorias básicas: i. campos de força; ii. empíricas; e iii. funções baseadas no conhecimento [84].

A etapa seguinte da triagem virtual envolveu o emprego de duas metodologias distintas de docagem molecular: i. Glide [97]; e ii. Surflex-Dock [98]. Esses dois métodos diferem, basicamente, na estratégia para amostragem conformacional das moléculas no sítio de ligação da proteína em estudo, isto é, enquanto Glide utiliza o algoritmo genético o Surflex-Dock emprega a construção incremental combinada com uma função de similaridade morfológica. A utilização de duas estratégias distintas de docagem molecular foi empregada para aumentar as chances de identificação de candidatos a inibidores da *Xa*PPT.

Brevemente, Glide (*Grid-based LIgand Docking with Energetics*) é um módulo de docagem molecular disponível na plataforma de modelagem Maestro (Schröndinger). O algoritmo de amostragem conformacional empregado no Glide (i.e., algoritmo genético) investiga interações favoráveis entre o ligante e o receptor de maneira rápida através da utilização de filtros hierárquicos que classificam os contatos polares e hidrofóbicos entre proteína e ligante. Os modos de ligação preditos dos ligantes no sítio ativo da proteína em estudo são então submetidos a um processo de minimização de energia pelo método Potencial Otimizado de Simulação de Líquidos (OPLS, do inglês, *Optimized Potentials for Liquid Simulations*) [117,118]. Finalmente, o modo de ligação predito e minimizado é avaliado pela função de pontuação empírica, derivada do ChemScore [119], denominada GlideScore.

O Surflex-Dock é um módulo de docagem molecular disponível na plataforma de modelagem Sybyl (Tripos) que utiliza a construção incremental combinada com uma função de similaridade morfológica como estratégia de amostragem conformacional. Na rotina de trabalho do Surflex-Dock, faz-se necessário gerar um arquivo denominado “protomol”. A informação contida nesse arquivo descreve as regiões de interação mais favoráveis para grupos doadores/aceptores de ligação de hidrogênio e

hidrofóbicos [120,121]. Para tanto, o método utiliza sondas moleculares (e.g., N-H, C=O e CH_4) que identificam as regiões favoráveis para interação com grupos funcionais em moléculas de ligantes. Uma vez gerado o arquivo protomol, esse é utilizado em combinação com o algoritmo de construção incremental para posicionar os respectivos grupos funcionais dos ligantes nas regiões favoráveis para interação. Essa estratégia tem-se mostrado eficiente em eliminar resultados falsos positivos [93], uma vez que somente moléculas que apresentam as características complementares ao protomol são submetidas ao processo de docagem molecular. Para avaliar a afinidade de ligação dos ligantes, o Surflex-Dock utiliza uma função de pontuação empírica, denominada Hammerhead [98,122], cujo resultado corresponde ao valor de $-\log K_D$.

A função de pontuação implementada nos programas Glide e Surflex-Dock pontuam e classificam os modos de ligação preditos durante a etapa de amostragem conformacional. Entretanto, a limitação para determinação da energia livre de ligação decorrente da falta de conhecimento específico sobre o processo de dessolvatação, constantes dielétricas do microambiente do sítio de ligação, ausência de um termo que defina corretamente a entropia do sistema, aliados a necessidade de se gerar e pontuar soluções em uma escala de tempo aceitável, faz com que as funções de pontuação não sejam robustas o suficiente para a determinação precisa da afinidade de ligação [123,124]. Diante disso, os programas de docagem molecular abordam esse problema de maneiras diferentes, adaptando a função de pontuação a uma determinada característica de interesse. Para tanto, os métodos de pontuação apresentam, por exemplo, esquemas diferentes de classificação de tipo de átomos e utilização de métodos diversos para a determinação das cargas parciais [123]. Nesse contexto, para diminuir o problema decorrente das limitações atualmente encontradas pelas funções de pontuação, a pontuação por consenso (do inglês, *consensus scoring*) vem sendo utilizada com sucesso para a identificação de moléculas bioativas [123,124].

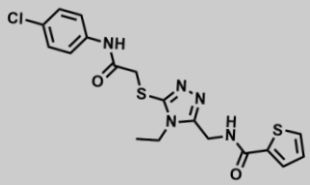
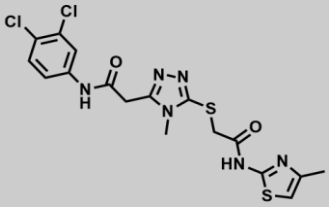
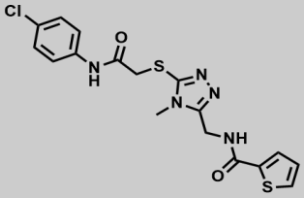
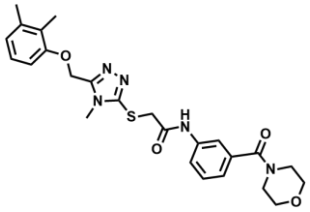
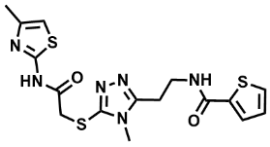
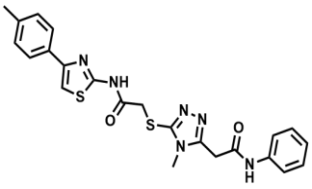
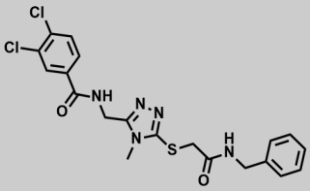
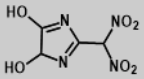
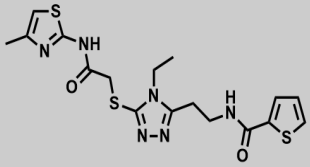
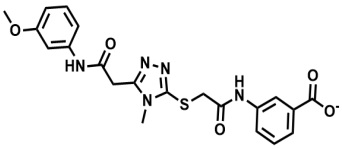
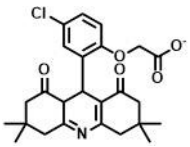
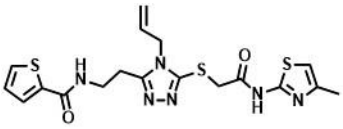
A pontuação por consenso utiliza a combinação de diferentes funções de pontuação para auxiliar na seleção de compostos para aquisição e testes biológicos[123]. Essa estratégia envolve a pontuação do modo de ligação predito dos compostos com diferentes métodos e a combinação dos *N* primeiros compostos classificados por cada função.

Nos estudos conduzidos nessa dissertação, a mesma base de dados e sítio de ligação foram explorados por dois métodos distintos de docagem molecular, os quais pontuaram o modo de ligação para os 6.487 ligantes investigados. A seleção de moléculas candidatas a inibidores da *Xa*PPT foi baseada na análise dos 100 melhores compostos pontuados (Top100) por cada método (**Figura 8**). A análise detalhada das Top100 moléculas identificadas independentemente pelo Surflex-Dock e Glide indicou que houve convergência de 33 moléculas. Isto é, embora avaliadas por metodologias distintas de amostragem conformacional e pontuação, ambos os métodos sugeriram as mesmas 33 moléculas, entre as 100 melhores pontuadas, como candidatas a inibidores da *Xa*PPT (**Tabela 3**). O conjunto de moléculas identificadas pela análise de consenso apresenta baixa diversidade química, no qual 90% dos compostos são derivados triazólicos e os demais 10% corresponde a derivados imidazólicos, tiazólico e dacahidro-9-acridinílico (**Tabela 3**).

A predição do modo de ligação de um dos derivados triazólicos (ZINC08442194) (**Figura 29A**), representativo do conjunto de moléculas selecionadas pela análise de consenso, indica que Surflex-Dock e Glide concordam na orientação do candidato a inibidor no sítio ativo da *Xa*PPT (**Figura 29B**). O valor de RMSD calculado para as soluções sugeridas pelos métodos de docagem é de 0,87 Å, indicando que a predição modo de ligação é bastante semelhante (**Figura 29B**). De acordo com os modelos, o composto ZINC08442194 ocupa o sítio de ligação do CoA próximo ao íon Mg^{2+}, (**Figura 29C** e **Figura 29D**) no qual ligações de hidrogênio estabilizam a formação do complexo. Especificamente, i. o átomo de oxigênio do grupo carbonila próximo ao anel tiazol está em posição favorável para receber ligação de hidrogênio da cadeia lateral do resíduo Arg188; ii. o átomo de nitrogênio N2 do anel triazólicos está em posição favorável para receber ligação de ligação de hidrogênio da cadeia lateral do resíduo Arg106; e iii. o átomo de nitrogênio NH da amida próxima ao anel tiofeno encontra-se em posição favorável para doar de ligação de hidrogênio para o átomo de oxigênio do grupo carbonila da cadeia principal do resíduo Leu107. Interações hidrofóbicas também desempenham função importante no modo de ligação do ZINC08442194, entre elas, citam-se: i. entre o substituinte tiazol e o resíduo Leu189; ii. entre os substituintes metilenos e os resíduos Ala108 e Ala126; e iii. entre o substituinte tiofeno e os resíduos Leu107 e His53.

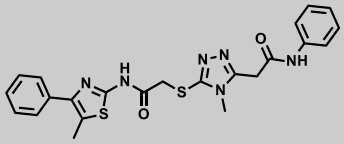
As moléculas selecionadas são encontram-se disponível no catálogo da empresa Specs (http://www.specs.net) e apresentam as seguintes características (**Tabela 4**): tamanho adequado para futuras otimizações, as massas moleculares variam entre 206 e 496, solubilidade razoável em solução aquosa, o cLog P varia entre -3,35 e 3,44, o número de átomos doadores e aceptores de ligação de hidrogênio varia entre 0 e 3 e 6 e 10, respectivamente. O pedido de compra das moléculas selecionadas foi realizado e no presente momento estas encontram-se em fase de entrega.

Tabela 3 – Conjunto de 33 moléculas entre as 100 melhores pontuadas pelo Surflex-Dock e Glide.

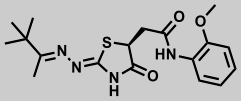
Estruturas moleculares		
 **ZINC634001**	 ZINC8442171	 ZINC633992
 **ZINC8442156**	 **ZINC8442196**	 **ZINC8442185**
 **ZINC633982**	 **ZINC19318821**	 **ZINC8442194**
 **ZINC8442189**	 **ZINC8442186**	 **ZINC8442200**

continua

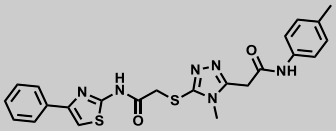
conclusão



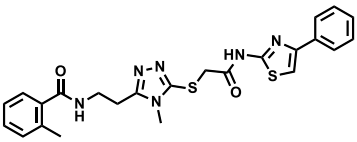
ZINC8442184



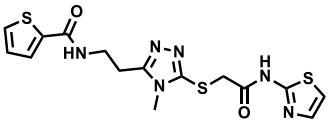
ZINC12378847



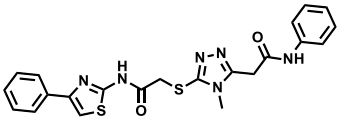
ZINC8442180



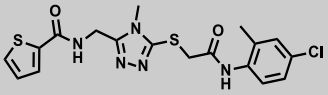
ZINC8442149



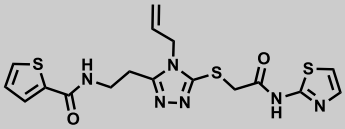
ZINC8442198



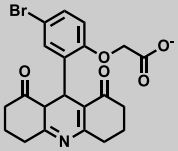
ZINC8442184



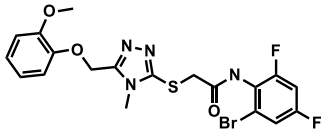
ZINC8442204



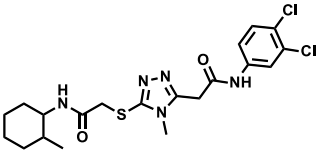
ZINC8442201



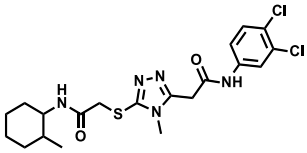
ZINC8442210



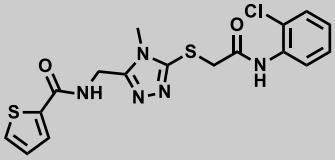
ZINC633938



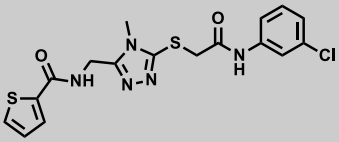
ZINC633955



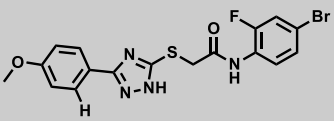
ZINC633958



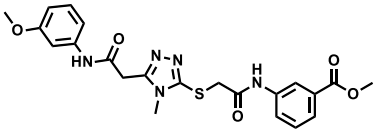
ZINC633999



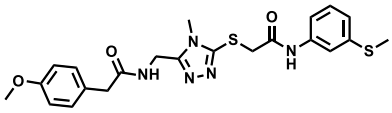
ZINC633997



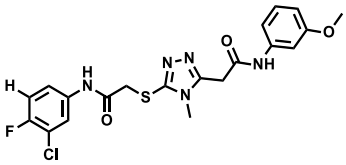
ZINC633984



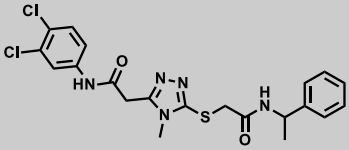
ZINC633978



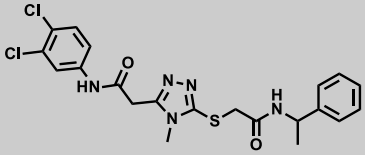
ZINC633975



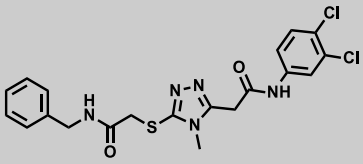
ZINC633972



ZINC633968



ZINC633966



ZINC633964

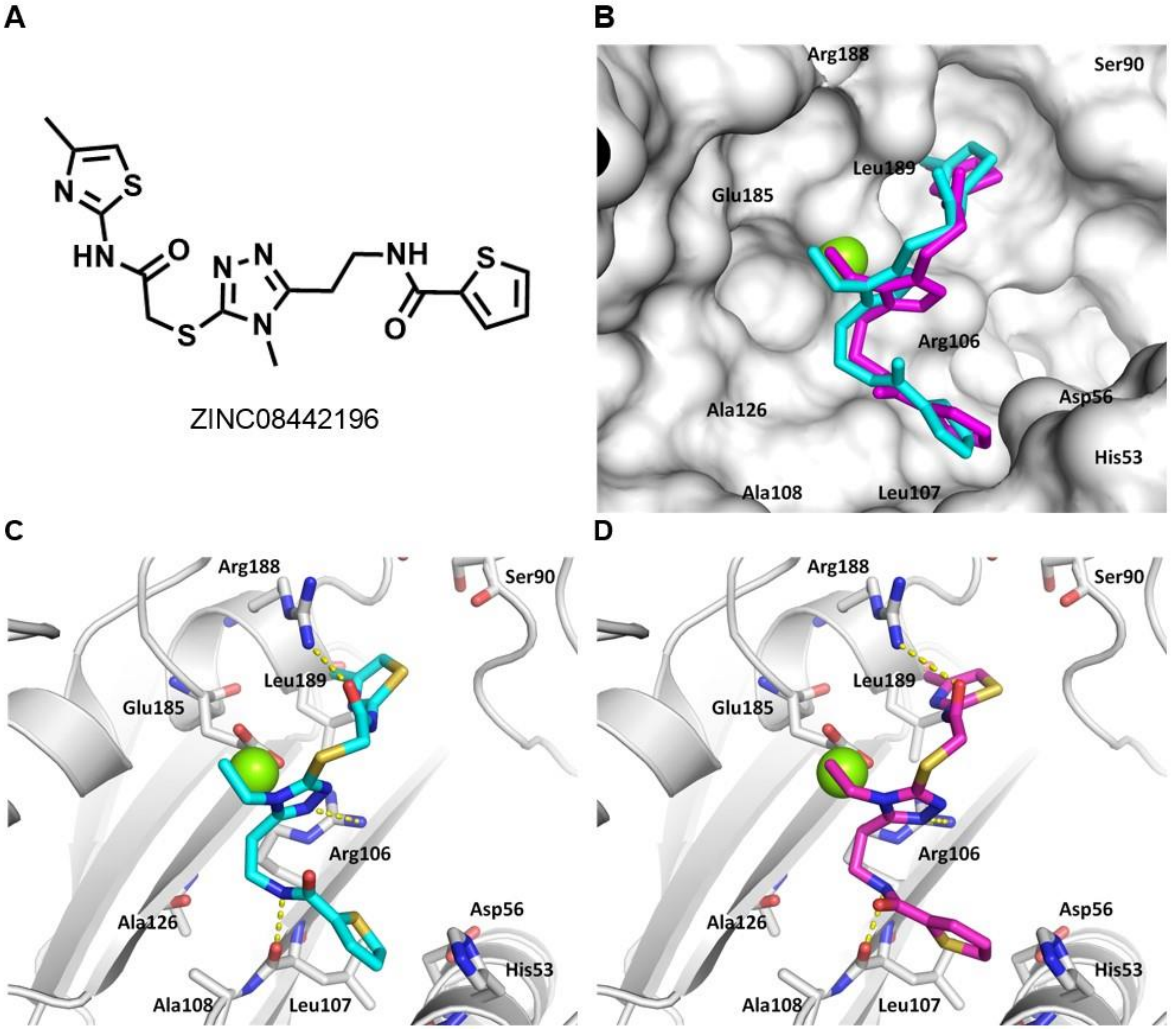
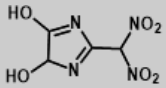
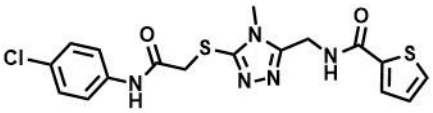
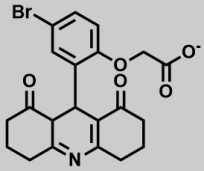
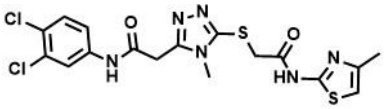
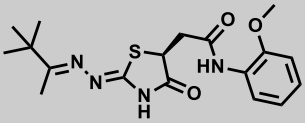
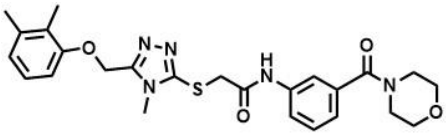
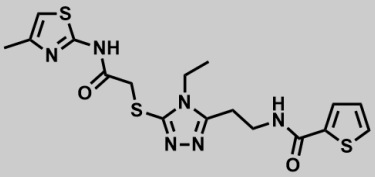
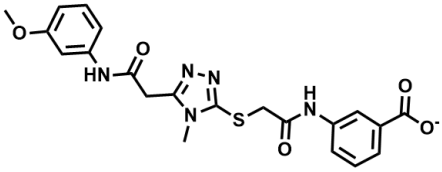
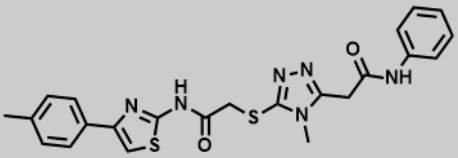


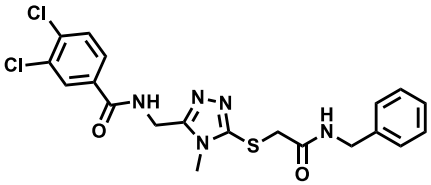
Figura 29 – (**A**) Estrutura química do candidato a inibidor ZINC08442194. (**B**) Sobreposição estrutural do modo de ligação predito para o composto ZINC08442194 pelo Glide (ciano) e Surflex-Dock (magenta). O sítio de ligação está representado pela superfície de Connolly e o átomo de Mg^{2+} está representado no modelo de esfera. (**C**) Modo de ligação do composto ZINC08442194 predito pelo Glide. (**D**) Modo de ligação do composto ZINC08442194 predito pelo Surflex-Dock. O sítio de ligação está representado em modelos de fitas e as ligações de hidrogênio estão ilustras em amarelo.

Tabela 4 – Moléculas identificadas na triagem virtual pelo consenso entre os resultados do Surflex-Dock e do Glide.

Código ZINC	Estrutura	MM	cLogP	H-Do	H-Ac	Catálogo
ZINC19318821		206	-3,35	2	10	Specs
ZINC633992		422	2,46	2	7	Specs
ZINC8442210		445	2,83	0	6	Specs
ZINC8442171		471	2,83	2	8	Specs
ZINC12378847		376	2,77	2	7	Specs
ZINC8442156		496	2,53	1	9	Specs
ZINC8442194		437	1,83	2	8	Specs
ZINC8442189		454	2,19	2	10	Specs
ZINC8442185		478	3,44	2	8	Specs

continua

conclusão

ZINC633982		464	2,86	2	7	Specs

MM = massa molecular; cLog P = coeficiente de partição octanol/água calculado; H-Do = no. de átomos doadores de ligação de hidrogênio; H-Ac = no. de átomos aceptores de ligação de hidrogênio

CAPÍTULO 4

BIOLOGIA MOLECULAR

4.1 Material e métodos

Os reagentes utilizados na parte experimental dessa dissertação de mestrado foram adquiridos, exclusivamente, da empresa Sigma-Aldrich®, sendo as exceções indicadas no corpo do texto. A utilização de *kits* comerciais seguiu estritamente as recomendações de uso de cada fabricante, sendo reportado apenas alterações pertinentes.

4.1.1 Cultura de *Xanthomonas albilineans*

O estabelecimento e padronização do cultivo da fitobactéria *X. albilineans* é essencial para a viabilização de ensaios de triagem fenotípica, além disso, permite a obtenção de material genômico, a partir da massa celular cultivada, necessário para os estudos moleculares propostos. Nesse sentido, buscou-se obter a cepa referência *X. albilineans* ATCC® 12785™, disponível no banco de microrganismos, mundialmente reconhecido, ATCC (do inglês, *American Type Culture Collection*). No entanto, devido as políticas de restrição entre os países envolvidos (Brasil e Estados Unidos da América – sede da coleção) a importação da cepa foi possível. Assim, obteve-se a cepa *X. albilineans* IBSBF1374 (comparável ATCC® 33915™) da Coleção de Culturas Fitobactérias IBSBF do Instituto Biológico de Campinas (**Tabela 5**).

O material liofilizado adquirido foi reidratado, em condições estéreis, com meio líquido YSG [125] (5 $g.L^{-1}$ extrato de levedura, 5 $g.L^{-1}$ glicose, 0,5 $g.L^{-1}$, fosfato de amônia monobásico e 0,2 $g.L^{-1}$ fosfato de potássio dibásico), meio de cultura rico em nutrientes e sem antibióticos. Para assegurar a recuperação das cepas liofilizadas, as culturas reidratadas foram semeadas em meio sólido YSG, e paralelamente, semeou-se uma alíquota do material reidratado em meio XAS [126] – meio de cultura desenvolvido para isolamento de *X. albilineans* – (1 $g.L^{-1}$ sacarose, 5 $g.L^{-1}$ bactopeptona, 500 $mg.L^{-1}$ fosfato de potássio dibásico, 250 $mg.L^{-1}$ de sulfato de

magnésio, 50 mg.L^{-1} de sulfito de sódio, 5 mg.L^{-1} de brometo de potássio, 2 mg.L^{-1} de benomil, 100 mg.L^{-1} de ciclohexamida, 25 mg.L^{-1} de cefalexina, 30 mg.L^{-1} de novobiocina, 50 mg.L^{-1} de kasugamicina e 15 g.L^{-1} de ágar), ambos cultivados em estufa bacteriológica à 28º C por 7 dias.

Colônias isoladas foram transferidas para cultivo em meio XAS líquido, à temperatura de 28º C e agitação constante de 175 rpm por 5 dias. O cultivo bacteriano foi dividido em duas alíquotas, sendo uma delas direcionada para a produção de DNA genômico (**seção 4.1.2**) e a outra utilizada para produção de um estoque glicerinado, pela adição de solução aquosa estéril de glicerol 50% na proporção 1:1, armazenado em ultrafreezer (-82 °C).

Após a clonagem da enzima *Xa*PPT (**seção 4.1.4**), a partir do DNA genômico produzido, o sequenciamento das construções estabelecidas indicou que o fragmento amplificado e clonado nos vetores propostos era proveniente da bactéria *Enterobacter cloacae,* organismo endofítico presente em tecidos de cana de açúcar [127–129]. Portanto, com o objetivo de obter culturas não contaminadas e estabelecer uma pequena biblioteca de cepas de *X. albilineans* com características fenotípicas diversas, novas cepas foram obtidas com o Instituto Biológico de Campinas (**Tabela 5**).

Tabela 5 – Culturas de *X. albilineans* adquiridas da Coleção de Culturas Fitobactérias IBSBF do Instituto Biológico de Campinas.

Identificação IBSBF	Ano da coleta	Origem	Hospedeiro
1374	1961	Fiji	*Saccharum officinarum*
326	1981	Brasil	*Saccharum officinarum*
654	1988	Brasil	*Saccharum hibridas*
740	1989	Brasil	*Saccharum officinarum*

4.1.2 Extração do DNA genômico de *X. albilineans*

A obtenção de DNA de alta qualidade é pré-requisito para a aplicação e obtenção de sucesso em técnicas de biologia molecular. Assim, empregou-se um protocolo de

extração de DNA genômico, adaptado e otimizado para o trabalho com cepas *Xanthomonas sp* [130]. Brevemente, a cepa de *X. albilineans* IBSBF 740, foi cultivada em meio líquido XAS a 28 ºC até saturação (aproximadamente 5 dias). As bactérias foram obtidas por centrifugação a 4.000 g durante 20 min. Em seguida, o meio contendo a bactéria foi ressuspenso em tampão de lise (2 % brometo de cetil trimetil amônio - CTAB, 1,4 M NaCl, 100 mM Tris-HCl pH 8.0, 20 mM EDTA) pré-aquecido e mantidas a 65 °C por 1 h. Subsequentemente, adicionou-se 1 volume de clorofórmio/álcool isoamílico (24:1) e as fases resultantes foram separadas por centrifugação (16.000g 10 min). Sobre a fase superior isolada, adicionou-se 1 volume de clorofórmio/álcool isoamílico (24:1) e separou-se novamente por centrifugação (16.000g 10 min). A precipitação dos ácidos nucléicos presentes na fase superior foi realizada pela adição de 1 volume de isopropanol (-20 ºC) seguido de incubação a -20 ºC por 30 minutos.

O pellet de DNA obtido por centrifugação a 16.000 g por 10 min foi submetido a duas etapas de lavagem consecutivas com etanol 70% e 96% e seco a temperatura ambiente. O material foi ressuspenso em água e a extração de DNA foi confirmada através de eletroforese em gel de agarose 1% (**Figura 32**). O RNA foi removido da amostra pela digestão com 100 µg/mL RNAse A (Invitrogen™) por 1 h a 37 ºC.

4.1.3 Caracterização molecular de *X. albilineans*

A caracterização molecular de *X. albilineans* foi realizada baseada na metodologia de amplificação de um fragmento gênico de 288 pares de base pertencente a região espaçadora (do inglês, *spacer*) entre as subunidades ribossomais 16S-23S [131]. A amplificação do fragmento foi realizada através da Reação em Cadeia da Polimerase – PCR (do inglês, *Polymerase Chain Reaction*) – **Tabela 6** para tanto, empregou-se o par de oligonucleotídeos específicos:

PGBL1 5’ – CTTTGGGTCTGTAGCTCAGG – 3’

PGBL2 5’ – GCCTCAAGGTCATATTCAGC – 3’

DNA polimerase de alta fidelidade (Phusion™ - New England Biolabs®), DNA genômico (**seção 4.1.2**) como molde para amplificação e ciclagem térmica com X = 42 °C e Y = 15 s. O produto da reação foi analisado através de eletroforese em gel de agarose 1% corado com 0,05 µg/mL de brometo de etídeo (**Figura 31**).

Tabela 6 – Reação de PCR utilizada nos estudos moleculares da *X. albilineans*.

Reagente		[ ] final	Ciclagem térmica
Tampão *Phusion* HF 5X		1X	
dNTP		0,2 mM	98 °C – 30 s
Oligonucleotideo senso		0,5 µM	
Oligonucleotideo anti-senso		0,5 µM	98 °C – 10 s X °C – 30 s 72 °C - Y s 35 X
Molde	DNA genômico	50-250 ng	
	DNA plasmidial	1 - 100 ng	
Phusion DNA polimerase		1 U	72 °C – 10 min
H_2O		qsp. 50 µL	

4.1.4 Clonagem da enzima *Xa*PPT

Para maximizar as chances de obtenção da proteína em sua forma solúvel, necessária para os estudos funcionais e estruturais, explorou-se o aumento de solubilidade, tipicamente, promovido pela inclusão de diferentes proteínas de fusão (e.g., SUMO, tiorredoxina, NusA). Para tanto, adotou-se dois sistemas de clonagem em paralelo: i. clássico, empregando-se os vetores de expressão pET28a e pETSUMO; ii. clonagem independente de ligação – LIC (do inglês, *Ligation Independent Cloning*) [132], empregando-se os vetores pETM-11, pETTrx-1a, pETNus-1a, adaptados (**Figura 30**) e gentilmente cedidos pelo Dr. Arie Geerlof do Laboratório de Biologia Molecular Europeu (EMBL – Hamburgo – Alemanha). As características das construções proposta são descritas esquematicamente na **Tabela 7**.

Tabela 7 – Características das construções gênicas da enzima *Xa*PPT nos vetores de expressão empregados.

Vetor de Expressão	Proteína de Fusão	Massa Molecular da Construção	Enzima Empregada para Remoção da Proteína de Fusão
pET28a	Cauda de Hexahistina (3,6 KDa)	33 KDa	Trombina
pETSUMO	SUMO (11 KDa)	40 KDa	SUMO protease
pETM11	Cauda de Hexahistina (3 KDa)	32 KDa	TEV protease
pETTrx-1a	Tiorredoxina (14 KDa)	43 KDa	TEV protease
pETNus-1a	NusA (57 KDa)	86 KDa	TEV protease

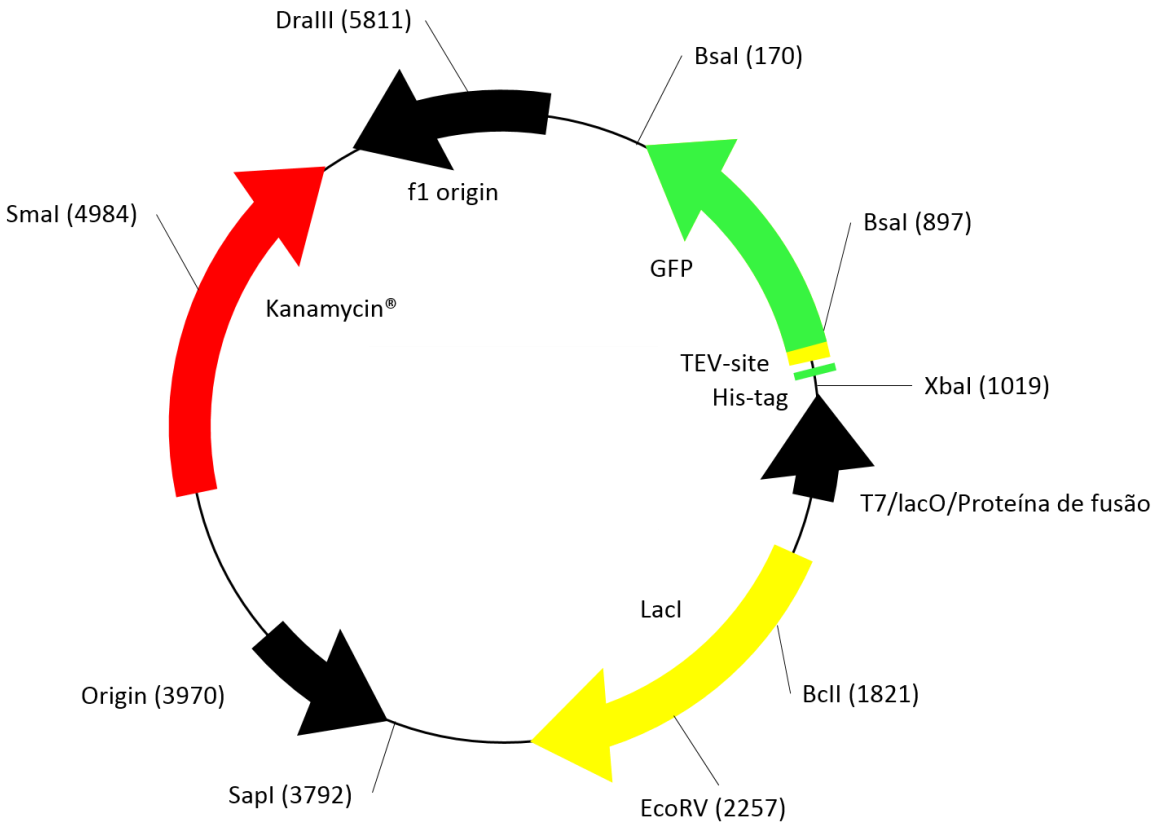


Figura 30 – Mapa geral dos vetores LIC. As proteínas de fusão utilizadas estão descritas na Tabela 7.

4.1.4.1 Clonagem clássica nos vetores de expressão pET28a e pETSUMO

No sistema de clonagem clássico, tipicamente, o fragmento de interesse é amplificado por reação de PCR a partir de um DNA molde e um conjunto de oligonucleotídeos iniciadores (senso e anti-senso), os quais contêm uma sequência de reconhecimento para enzimas endonucleases. Em seguida, o fragmento é submetido a uma digestão com as respectivas enzimas e ligado em vetores, previamente tratados com as mesmas enzimas, em reação catalisada pela enzima DNA ligase. Assim, baseado na sequência do gene que codifica a *Xa*PPT (*AlbXXI*), obtida no banco de dados GenBanK (AAG28384.1), um par de oligonucleotideos iniciadores:

senso: 5' – **GAATTC**ATGCCCAATGCCGTACC – 3'

anti-senso: 5' – **CTCGAG**TTATCACCAGGCGTAGTG – 3'

O par de oligonucleotídeos iniciadores foi planejado e sintetizado (Sigma® – Life Science), contendo sítios de reconhecimento (em destaque) para as enzimas endonucleases **EcoRI** e **XhoI**, respectivamente.

A reação de amplificação do gene *AlbXXI* foi conduzida com o sistema Phusion™ (*New England Biolabs®),* **Tabela 6**, com X = 55 °C, Y = 30 s e 200 ng de DNA genômico (**seção 4.1.2**) como molde para amplificação. O produto da reação foi analisado através de eletroforese em gel de agarose 1% corado com 0,05 µg/mL de brometo de etídeo (**Figura 34**). O fragmento amplificado foi recuperado com o auxílio do kit *Wizard® SV Gel and PCR Clean-Up System* (Promega). O fragmento purificado foi quantificado (NanoDrop 2000 – Thermo Scientific) e submetido a reação de ligação (50 ng de vetor, 70 ng de fragmento purificado) no vetor pJET 1.2 (Thermo Scientific) catalisada pela enzima T4 DNA ligase.

O produto da reação de ligação foi utilizado para transformação de bactérias competentes *E. coli* DH10B, produzidas pelo método Inoue [133]. Brevemente, 100 µL de células foram incubados com 100 µL de mistura reacional (10 µL do produto da

reação de ligação + tampão de transformação - 100 mM KCl, 30mM $CaCl_2$, 50 mM $MgCl_2$, 1,5% PEG 4.000) em gelo por 30 min. Após o período de incubação, as células foram submetidas a choque térmico na temperatura ambiente por 10 min. Em seguida, adicionou-se 1 mL de meio líquido SOC (20 $g.L^{-1}$ triptona, 5 $g.L^{-1}$ extrato de levedura, 0,5 $g.L^{-1}$ NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glicose) e incubou-se por 1 h a 37 °C sob agitação constante 175 rpm. Subsequentemente, semeou-se o cultivo em placas de Petri contendo meio sólido LB (10 $g.L^{-1}$ triptona, 5 $g.L^{-1}$ extrato de levedura, 10 $g.L^{-1}$ NaCl) suplementado com ampilicilina (100 µg/mL).

Colônias isoladas foram cultivadas em 5 mL de meio líquido LB, suplementados com ampicilina (100 µg/mL), por 16 h a 37 °C. O DNA plasmidial foi obtido com o auxílio do kit *Wizard® Plus Minipreps DNA Purification System* (*Promega*). A obtenção do fragmento e vetores (pET28a e pETSUMO) com extremidades coesivas compatíveis foi realizada através da reação de dupla digestão com as enzimas EcoRI e XhoI (*Fermentas*) de 1 µg de pJET1.2-*Xa*PPT e 5 µg de cada vetor, analisada por eletroforese em gel de agarose 1%. O fragmento e o vetor digerido, excisados e recuperados do gel com auxílio do kit *Wizard® SV Gel and PCR Clean-Up System* (*Promega*), foram submetidos a reação de ligação (5:1 - 50 ng de vetor/37 ng de fragmento) promovida pela T4 DNA polimerase (*Fermentas*).

O produto da reação foi utilizado para a transformação de bactérias competentes DH10B, conforme previamente descrito, e semeadas em placas de Petri contendo meio sólido LB suplementado com canamicina (30 µg/mL) e incubadas em estufa bacteriológica por 16 h a 37° C. A presença do plasmídeo nos clones recombinantes foi verificada através de reação de PCR de colônia (**Tabela 6**), com X = 55 °C, Y = 30 s e 1 µL do cultivo em meio líquido LB como molde para reação de amplificação. As colônias positivas foram cultivadas em meio líquido LB por 16 h a 37° C e o DNA plasmidial obtido com o auxílio do kit *Wizard® Plus Minipreps DNA Purification System* (*Promega*). A integridade das construções foi verificada através de sequenciamento em um sequenciador automático ABI-Prism 377 (Perkin Elmer), utilizando a técnica de didesoxinucleotídeo marcado.

O resultado do sequenciamento das construções revelou que o fragmento amplificado e clonado pertencia a uma bactéria contaminante (*Enterobacter cloacae*) proveniente da cultura de *X. albilineans* adquirida. Devido às dificuldades para a obtenção de novas cepas, adquiriu-se o gene sintético (GenScript) *AlbXXI* inserido no

vetor pUC57 o qual foi utilizado como molde para reconstrução dos sistemas de expressão propostos.

4.1.4.2 Clonagem independente de ligação (LIC) nos vetores de expressão pETM11, pETTrx-1a, pETNus-1a

O sistema de clonagem independente de ligação (LIC) difere do método clássico por não empregar uma etapa dedicada a formação de extremidades coesivas pela digestão com enzimas endonucleases, além disso, fundamentalmente, dispensa-se a etapa de ligação com a enzima DNA ligase. No entanto, a metodologia empregada requer o planejamento diferenciado para os oligonucleotideos iniciadores. Portanto, os seguintes oligonucleotideos:

senso: 5' – **CAGGGCGCCAT**GCCCAATGCCGTAC – 3'
anti-senso: 5' – **GACCCGACGCGGT**TACCAGGCGTAGTGG – 3'

foram planejado e sintetizado (Sigma® – Life Science), contendo as caudas salientes (em destaque) correspondentes ao sistema LIC empregado.
O gene *AlbXXI* foi amplificado a partir do pUC57, através da técnica de PCR (**Tabela 6**), utilizando o sistema Phusion™ (*New England Biolabs®)* com X = 55 °C e Y= 30 s. O produto de reação foi analisado através de eletroforese em gel de agarose 1% (**Figura 33**) e purificado com o auxílio do kit *Wizard® SV Gel and PCR Clean-Up System* (*Promega*). Após quantificação espectrofotométrica (*NanoDrop 2000 – Thermo Scientific*), o fragmento (150 ng) foi tratado com T4 DNA polimerase (Fermentas) em reação (1X Tampão T4 DNA polimerase, 2,5 mM dATP, 5 mM DTT, 1U T4 DNA polimerase) por 30 min a 22 °C, para a formação das extremidades coesivas de ligação ao sistema LIC empregado.

O fragmento tratado foi anelado aos vetores de expressão (proporção 3:1), previamente tratados com T4 DNA polimerase em condições reacionais semelhantes ao tratamento do fragmento (exceto pela utilização de dTTP em substituição a dATP), em temperatura ambiente por 30 min. A reação de anelamento foi utilizada para a transformação de células competentes *E. coli* DH10B (conforme descrito na seção

4.1.4.1) semeadas e cultivadas em placas de Petri contendo meio sólido LB suplementado com canamicina (30 µg/mL) por 16 h a 37 °C em estufa bacteriológica. A presença do plasmídeo nos clones recombinantes foi avaliada através de PCR de colônia (**Tabela 6**), empregando 1µL do cultivo em meio LB líquido como molde para reação de amplificação.

As colônias positivas foram cultivadas em meio líquido LB por 16h a 37° C, suplementado com canamicina (30 µg/mL), e DNA plasmidial obtido com o auxílio do kit *Wizard® Plus Minipreps DNA Purification System* (*Promega*). Todas as construções tiveram sua integridade verificada por sequenciamento.

4.1.4.3 Testes de expressão da enzima *Xa*PPT

Com o objetivo de identificar o melhor sistema de expressão proteica, ou seja, o sistema de expressão com a maior capacidade de produzir proteína em sua forma solúvel, testes de expressão para todas construções obtidas foram conduzidos. Para tanto, DNA plasmidial (50 ng) de cada construção foi utilizado para transformação de células competentes *E. coli* BL21(DE3) (conforme descrito na **seção 4.1.4.1**), cepa geneticamente modificada para propiciar altos níveis de expressão proteica.

Colônias isoladas foram cultivadas aerobicamente a 37° C em meio líquido LB até atingirem a densidade ótica (DO, em 600 nm) entre 0,6-0,8. Em seguida, induziu-se a expressão proteica com a adição de 1 mM de isopropil-β-D-galactosídeo (IPTG). Dois protocolos distintos foram empregados para avaliação da condição de expressão mais apropriada: i. expressão por 3 h a 37° C e ii. expressão por 16 h a 18° C. Após o período de expressão, as células foram obtidas por centrifugação a 4.000 g por 20 min e ressuspensas em tampão de lise (50 mM Tris pH 8.0, 500 mM NaCl, 1 mM PMSF - fluoreto de fenilmetanosulfonil, 4 mM DTT - Ditiotreitol) em um décimo do volume de cultura inicial. Em seguida, as células foram submetidas a lise celular por sonicação em gelo. As frações solúvel e insolúvel do clarificado produzido foram separadas por centrifugação a 16.000 g por 20 min e submetidas à análise de eletroforese em gel de poliacrilamida 15% (SDS-PAGE), corado com 0,01% de *Comassie Brillant Blue* R-250, para avaliação do perfil de expressão.

4.2 Resultados

4.2.1 Identificação molecular de *X. albilineans*

A identificação molecular de fitobactérias através de técnicas moleculares (*e.g.,* amplificação de regiões gênicas específicas) é amplamente empregada no processo de detecção e diagnóstico de fitopatologias [134–137]. Além da elevada especificidade, o emprego de métodos de identificação molecular apresenta grandes vantagens, entre elas, destaca-se: i. sensibilidade; ii. rapidez; e iii. não obrigatoriedade de cultivo dos microrganismos para sua detecção.

A limitação dos métodos tradicionais de identificação de *X. albilineans* em culturas de cana-de-açúcar permite o desenvolvimento e alastramento das cepas bacterianas. Portanto, o diagnóstico precoce da doença é essencial para reduzir o impacto causado sobre a lavoura. Contudo, esse tipo de diagnóstico é possível somente através de métodos altamente sensíveis. Nesse sentido, foi desenvolvido um método molecular extremamente sensível, baseado em reação de PCR e oligonucleotideos iniciadores altamente específicos, que é capaz de identificar a presença de *X. albilineans* em culturas de cana-de-açúcar, [131]. O método identifica um fragmento gênico específico de *X. albilineans* de 288 pares de base localizado na região espaçadora entre as subunidades ribossomais 16S-23S.

As quatro cepas de *X. albilineans* (i.e., IBSBF 1374, 326, 654, 740), adquiridas do Instituto Biológico de Campinas, foram validadas pela método molecular. A amplificação do fragmento gênico de 288 bp (**Figura 31**) confirmou a identidade das cepas adquiridas. No entanto, é importante mencionar que a positividade nesse teste indica que a cepa *X. albilineans* está presente na amostra, porém não exclui a existência de demais bactérias.

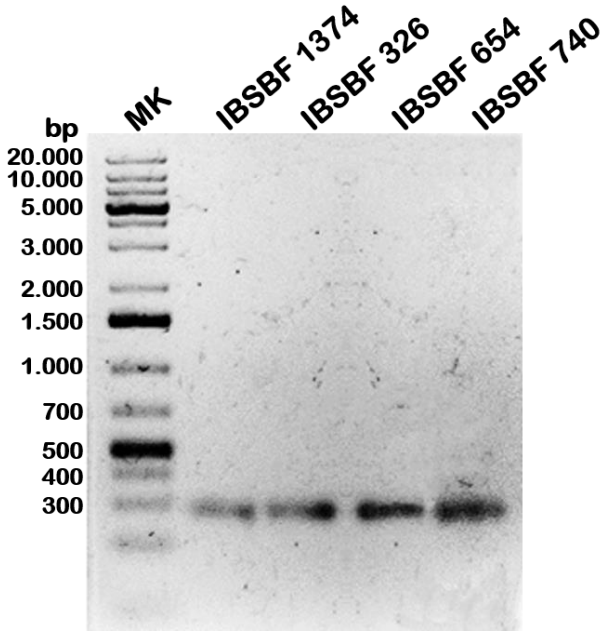


Figura 31 – Gel de eletroforese em agarose 1%. Reação de identificação molecular de *X. albilineans* através da amplificação de região gênica específica (288 bp). MK = Marcador de massa molecular

4.2.2 Clonagem da enzima *Xa*PPT

O processo de clonagem da enzima *Xa*PPT iniciou-se pela obtenção do molde empregado na reação de PCR para amplificação do gene *AlbXXI*. As bactérias desse gênero produzem e secretam grandes quantidades de polissacarídeos extracelulares que dificultam a obtenção do material genômico. Portanto, foi empregado o protocolo baseado no emprego do detergente catiônico brometo de cetil trimetil amônio (CTAB), cujo resultado consiste em uma mistura de ácidos nucléicos (DNA+RNA) (**Figura 32**). O RNA foi removido da amostra pela digestão com a enzima RNAse A.

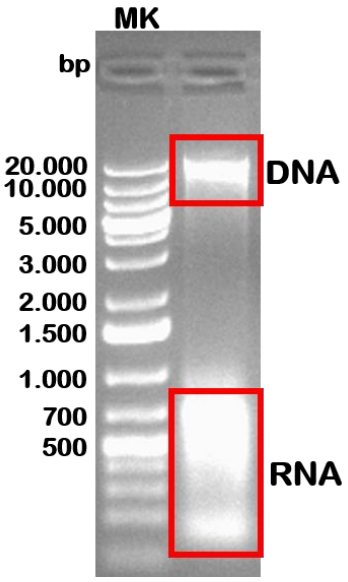


Figura 32 – Gel de eletroforese em agarose 1%. Extração de DNA genômico de *X. albilineans*.

O emprego do sistema moderno de clonagem independente de ligação (LIC) [132] apresenta grandes vantagens em relação ao sistema clássico, destacando-se a possibilidade do planejamento e utilização de um único par de oligonucleotídeos para clonagem em diversos vetores. Além disso, os vetores LIC empregados possuem algumas características otimizadas, como um sítio para TEV protease (uma protease produzida de maneira recombinante em nosso laboratório) para remoção das proteínas de fusão, além da possibilidade inclusão de proteínas de fusão na região N-terminal. Essas proteínas são reconhecidas pela notável capacidade de aumento de solubilidade. Paralelamente, empregou-se o sistema clássico em virtude de alguns vetores estarem em fase de adaptação ao sistema LIC.

O gene *AlbXXI* foi amplificado a partir do DNA genômico gerando os fragmentos necessários para o sistema clássico e sistema LIC (**Figura 33**). A clonagem de *Xa*PPT nos vetores propostos (i.e., pET28a, pETSUMO, pETM11, pETTrx-1a e pETNus-1a) foi realizada com sucesso (**Figura 34**). As construções tiveram sua integridade verificada por sequenciamento, portanto, o DNA plasmidial obtido foi empregado na transformação de células de expressão.

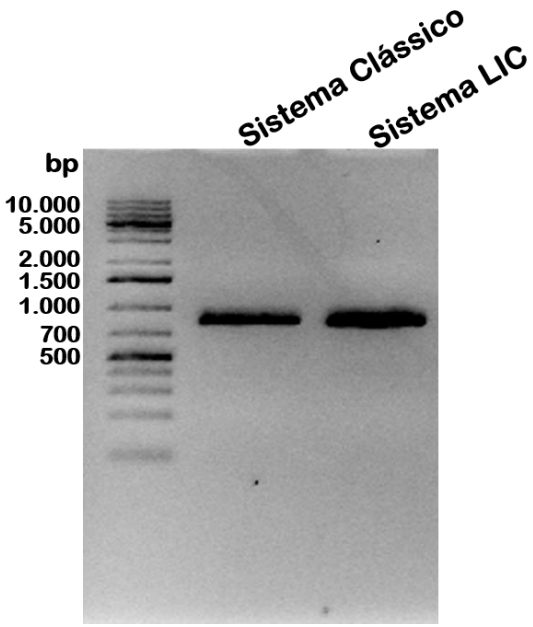


Figura 33 – Gel de eletroforese em agarose 1%. Reação de PCR para amplificação do gene *AlbXXI*.

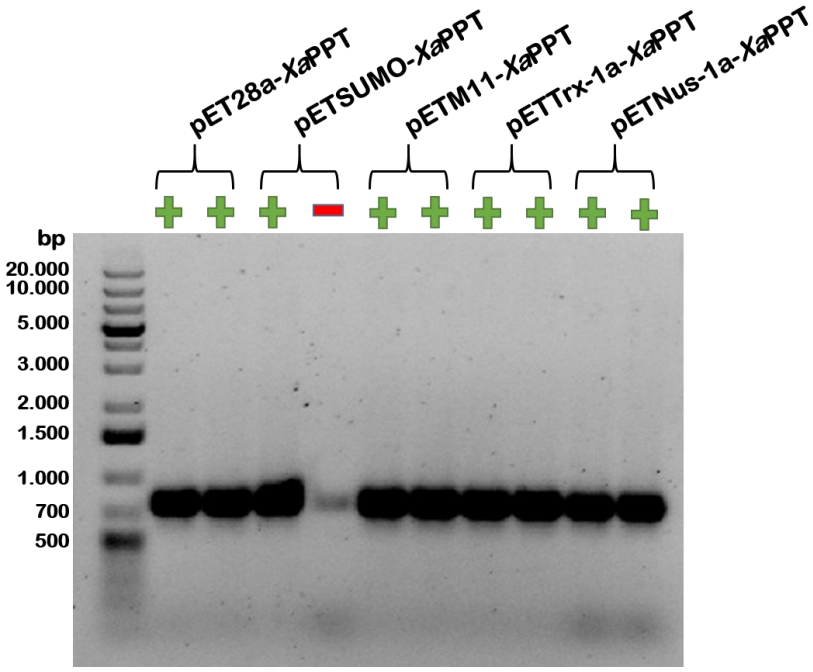


Figura 34 – Gel de eletroforese em agarose 1%. Reação de PCR dos clones recombinantes obtidos para as construções propostas.

4.2.3 Testes preliminares de expressão de *Xa*PPT

A avaliação do perfil de expressão proteica (i.e., distribuição da proteína alvo nas frações solúvel e insolúvel) foi, durante muito tempo, conduzida primordialmente buscando adaptar-se as necessidades e disponibilidades dos grupos de pesquisa (i.e., sistema de expressão, alvos moleculares e custos). Dessa maneira, as estratégias e os sucessos obtidos por determinado grupo dificilmente podiam ser adaptados as realidades de outro. No entanto, com o estabelecimento de centros genômicos e estruturais uma quantidade substancial de dados e experiências foram geradas permitindo, através da avaliação estatística, que as estratégias com os maiores níveis de sucesso fossem identificadas.

Atualmente, essas estratégias encontram-se compiladas em protocolos otimizados, capazes de guiar os experimentos iniciais e reduzir tanto o tempo quanto os custos para identificação de condições de expressão promissoras. Esses protocolos empregam avaliações do sistema de expressão em pequena escala, fornecendo dados quantitativos e qualitativos valiosos que guiarão a escolha do sistema de expressão para os experimentos em maior escala.

O número de parâmetros que afetam a expressão proteica, mesmo em um sistema de expressão bacteriano, são extremamente elevados (e.g., cepa bacteriana, temperatura, agente indutor, co-expressão, etc). Portanto, a avaliação sistemática de todos os fatores é desafiadora. Nesse trabalho, empregou-se a cepa bacteriana a *E. coli* BL21 (DE3), uma cepa adaptada para expressão proteica pela ausência de específicas proteases. Os testes de expressão avaliaram a influência da temperatura (*i.e.*, 20 °C e 37 °C) e a inclusão de proteínas de fusão (i.e., cauda de hexahistidina, SUMO, tiorredoxina, NusA) na região N-terminal [138,139].

Os resultados preliminares dos testes de expressão indicaram que a *Xa*PPT clonada no vetor de expressão pET28a apresentou melhor perfil de expressão proteica quando conduzida a 20 °C (**Figura 35**). Embora, a distribuição entre as frações solúvel e insolúvel não é ideal, o perfil obtido é suficientemente adequado para os ensaios em maior escala, permitindo a continuidade dos estudos funcionais e estruturais. Além disso, a avaliação das demais construções permitirá a escolha do sistema de expressão melhor adaptado as necessidades para condução do projeto.

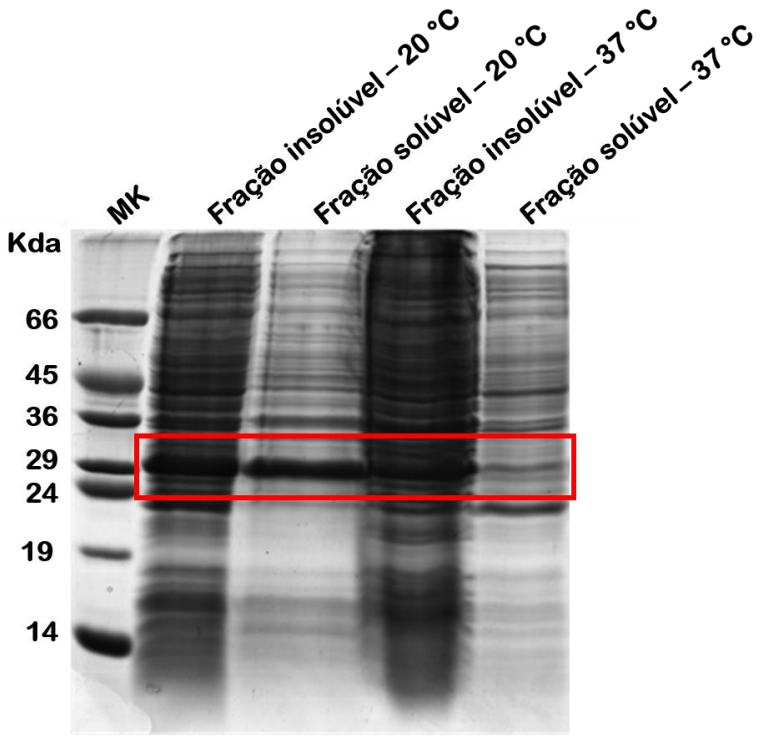


Figura 35 – Gel de eletroforese SDS-PAGE 15%. Teste de expressão para a construção *Xa*PPT-pET28a. MK = Marcador de massa molecular

CAPÍTULO 5

CONCLUSÕES E PERSPECTIVAS

A cultura de cana-de-açúcar destaca-se na economia nacional e tem impulsionado as exportações na última década com expectativa extremamente favorável de crescimento nos próximos anos. A necessidade de se aumentar a produtividade das lavouras para que haja um aumento na produção dos derivados da cana-de-açúcar sem que se necessite expandir as áreas de cultivo tem um forte aspecto econômico-social, visto que a substituição de áreas plantadas com alimentos por plantações cana-de-açúcar agravaria o crescente aumento do preço dos alimentos, dificultando ainda mais o acesso das populações.

O combate às pragas e doenças que afetam a produtividade das lavouras é fundamental para o crescimento sustentável do setor sucroalcooleiro. Entre as doenças mais severas que comprometem os canaviais, destaca-se a escaldadura das folhas, uma fitopatologia causada pela bactéria *X. albilineans* que não possui formas de controle. Nesse contexto, devido à importância da via de produção de albicidinas para o desenvolvimento da *X. albilineans*, enzimas dessa via são alvos moleculares em potencial para o tratamento da escaldadura das folhas. Entre elas, destaca-se a enzima fosfopanteteinil transferase (*Xa*PPT), uma enzima chave da via de albicidinas.

Nesta dissertação de mestrado foram apresentados os resultados obtidos no decorrer de um ano e meio de trabalho. O desenvolvimento dessa dissertação baseou-se em vários métodos e estratégias computacionais e experimentais que foram empregados para investigar as propriedades estruturais e biológicas da enzima *Xa*PPT. Os trabalhos computacionais conduzidos incluíram: i. o desenvolvimento de uma interface amigável e robusta denominada VitaMIn para o programa Modeller. VitaMIn auxilia, pesquisadores iniciantes e experimentes na modelagem molecular, a construir e avaliar os modelos moleculares de proteínas. A integração do VitaMIn com o PyMOL permite a visualização dos modelos gerados em tempo real. Tutoriais em forma de texto e vídeo encontram-se disponíveis no endereço http://rafaelguido.tk. A distribuição sob a licença Apache e GPL permite que a comunidade científica possa fazer alterações e incrementos específicos. ii. um estudo de caso conduzido com a enzima *Xa*PPT validou o VitaMIn como uma opção atrativa para a modelagem molecular comparativa de proteínas. iii. o modelo 3D da enzima *Xa*PPT foi utilizado para a triagem virtual de inibidores. Uma estratégia moderna baseada no conceito de consenso incluiu dois métodos distintos de docagem molecular para a identificação

de 10 candidatos a inibidores que serão avaliados frente a enzima em testes enzimáticos padronizados.

Na parte experimental foram conduzidos trabalhos que focaram em métodos de biologia molecular para a clonagem e expressão da enzima *Xa*PPT. Nesse sentido, foi realizada a clonagem do gene *AlbXXI* nos vetores pET-SUMO, pETM-11 LIC, pET-NUSA LIC e pET-Trx LIC, visando-se a obtenção da *Xa*PPT expressa na forma solúvel. Resultados promissores foram obtidos com o sistema pET28a.

Perspectivas

Esta dissertação de mestrado faz parte de uma das principais linhas de pesquisa estabelecidas no Laboratório de Química Medicinal e Computacional (LQMC), que visa o desenvolvimento de novos agentes agroquímicos. Diante disso, obtenção da enzima XaPPT pura será extremamente útil para o desenvolvimento de ensaios cinéticos padronizados para a identificação de inibidores através de triagens biológicas, bem como para os estudos estruturais que visam a determinação experimental da estrutura 3D da *Xa*PPT. Além disso, a padronização do cultivo de culturas de *X. albilineans* será útil para o desenvolvimento de ensaios *in vitro* de inibição do crescimento da bactéria. Esse ensaio será importante para o desenvolvimento de candidatos a agroquímicos com potente ação frente a *X. albilineans.* A disponibilidade de diferentes cepas em nosso laboratório também será útil para a avaliação da capacidade inibitória dos melhores candidatos a defensivos agrícolas para o controle da escaldadura das folhas. O desenvolvimento do ViTaMIn será continuado com objetivo de integrar métodos de dinâmica e docagem molecular, com o GTKDynamo [140] e o Autodock Vina [114], respectivamente. Além disso, a integração com o PyRosetta [141] será interessante para a construção de regiões de baixa identidade através de modelagem *ab initio*. Esses trabalhos serão conduzidos no doutorado.

REFERÊNCIAS

1 KISLEV, M.E.; WEISS, E.; et al. Impetus for sowing and the beginning of agriculture: Ground collecting of wild cereals. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 9, p. 2692–2695, 2004.

2 OERKE, E.-C. Crop losses to pests. **The Journal of Agricultural Science**, v. 144, n. 01, p. 31, 2005.

3 International code of conduct on the distribution and use of pesticides. Disponível em <http://www.who.int/whopes/recommendations/FAO_WHO_Guidelines_Pesticide_Advertising.pdf >. Acesso em 22 jun. 2013.

4 MICHAEL L. MCKINNEY, ROBERT M. SCHOCH, L.Y. Environmental Science: Systems And Solutions. p. 642, 2007.

5 NOBELPRIZE.ORG Paul Müller - Biography. 2013. Disponível em: <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1948/muller-bio.html>. Acesso em: 18 jun. 2013.

6 LIU, W.; LI, C.; et al. Distribution of DDT in a typical DDT waste contaminated site. **Bulletin of environmental contamination and toxicology**, v. 80, n. 3, p. 280–2, 2008.

7 BUNKOCZI, G.; PASTA, S.; et al. Mechanism and substrate recognition of human holo ACP synthase. **Chemistry & biology**, v. 14, n. 11, p. 1243–53, 2007.

8 LI, Y.F.; CAI, D.J.; et al. Gridded usage inventories of technical hexachlorocyclohexane and lindane for China with 1/6 degrees latitude by 1/4 degrees longitude resolution. **Archives of environmental contamination and toxicology**, v. 41, n. 3, p. 261–6, 2001.

9 TREN, R.; KAMWI, R.N.; et al. The UN is premature in trying to ban DDT for malaria control. **BMJ (Clinical research ed.)**, v. 345, n. oct10_8, p. e6801, 2012.

10 ENVIRONMENTAL PROTECTION AGENCY Pesticides: regulating pesticides. 2012. Disponível em: < http://www.epa.gov/opp00001/regulating/>. Acesso em 18 jun. 2013.

11 DELANEY, J.; CLARKE, E.; et al. Modern agrochemical research: a missed opportunity for drug discovery? **Drug discovery today**, v. 11, n. 17-18, p. 839–45, 2006.

12 KRIEGER, R. Reviewing Some Origins of Pesticide Perceptions. **Outlooks on Pest Management**, v. 16, n. 6, p. 244–248, 2005.

13 COOKSON, C. Global competition sparks spending spree. Disponível em: <http://www.ft.com/intl/cms/s/2/2e966372-6771-11db-8ea5-0000779e2340.html>. Acesso em: 22 jun. 2013.

14 MCDOUGALL, P. R&D trends in crop protection. 2012. Disponível em: <http://www.ecpa.eu/files/attachments/R&D%20Trends.pdf>. Acesso em: 22 jun. 2013.

15 MCDOUGALL, P. The Cost of New Agrochemical Product Discovery , Development and Registration in 1995 , 2000 and R & D expenditure in 2007 and expectations for 2012. n. January 2010, 2012.

16 Syngenta Financial Report 2011. 2011. Disponível em: <http://annualreport2011.syngenta.com/downloads/pdf/Syngenta_FinancialReport2011.pdf>. Acesso em: 22 jun. 2013.

17 Syngenta Full Year Results - 2011. 2011. Disponível em: <http://www.syngenta.com/global/corporate/en/news-center/news-releases/pages/120208.aspx>. Acesso em: 22 jun. 2013.

18Bayer Annual Report 2011. 2011. Disponível em: <http://www.annualreport2011.bayer.com/en/bayer-annual-report-2011.pdfx>. Acesso em: 22 jun. 2013.

19 SANTOS, G.R. DOS Características, sistema de registros de produtos econcorrência no mercado de agrotóxicos no Brasil. Disponível em: <https://www.ipea.gov.br/agencia/images/stories/PDFs/radar/120704_radar20_apresentacao_gesmar.pdf>. Acesso em: 22 jun. 2013.

20 Centro de Estudos Avançados em Economia Aplicada. **PIB do Agronegócio**. Disponível em: <http://www.cepea.esalq.usp.br/pib/>. Acesso em: 22 jun. 2013.

21 MIRANDA, J.R. History of sugarcane, 2008. 167 p.

22 Centro de Gestão e Estudos Energéticos. **Bioetanol combustível: uma oportunidade para o Brasil**. 2009. Disponível em: <http://www.cogen.com.br/paper/2010/Livro_Bioetanol_Cana_Acucar_2009.pdf>. Acesso em: 22 jun. 2013.

23 Food and Agriculture Organization of the United Nations. 2011. Disponível em: <http://faostat.fao.org/site/339/default.aspx>. Acesso em: 22 jun. 2013.

24 BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Relações Internacionais do Agronegócio. **Balança comercial do agronegócio** - dez/2012. 2013. Disponivel em: <http://www.agricultura.gov.br/arq_editor/file/0arquivos/12_%20Dezembro%202012.doc. Acesso em: 14 jun. 2013.

25 BRASIL. Ministério de Minas e Energia. Empresa de Pesquisa Energética. **Balanço energético nacional 2012**: ano base 2011. 2012. Disponivel em: <https://ben.epe.gov.br/downloads/Resultados_Pre_BEN_2012.pdf>. Acesso em: 14 jun. 2013.

26 BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Produção e Agroenergia. Departamento da Cana-de-açúcar e Agroenergia. **Análise comparativa do desempenho das safras 2012/2013 e 2011/2012** - dados acumulados. 2013. Disponivel em: <http://www.agricultura.gov.br/arq_editor/file/Desenvolvimento_Sustentavel/Agroenergia/estatisticas/producao/ABRIL_2013/analise%20comparativa%2029_04_13%20%20acumulada.pdf>. Acesso em: 12 jun. 2013.

27 SOUSA, E.L. DE Etanol e bioeletricidade : a cana-de-açúcar no futuro da matriz energética. 2009.

28 BLANCH, M.; LEGAZ, M.-E.; et al. Xanthan production by Xanthomonas albilineans infecting sugarcane stalks. **Journal of plant physiology**, v. 165, n. 4, p. 366–74, 2008.

29 TOKESHI, H. Doenças da cana-de-açúcar (híbridos de Saccharum spp.). Volume 2: Doenças das Plantas Cultivadas. **Manual de Fitopatologia**, p. 207–225, 1997.

30 Jackson, R.W. **Plant Pathogenic Bateria: Genomics and Molecular Biology**. Horizon Scientific Press, 2009.

31 BIRCH, G.; KONG, H. Xanthomonas albilineans and the antipathogenesis approach. **Molecular Plant Pathology**, v. 2, n. 1, p. 1–11, 2001.

32 DAVIS, M.J.; ROTT, P.; et al. Intraspecific Genomic Variation Within Xanthomonas albilineans, the Sugarcane Leaf Scald Pathogen. **Phytopathology**, v. 87, n. 3, p. 316–24, 1997.

33 JAUFEERALLY-FAKIM, Y.; AUTREY, J.C.; et al. Genetic polymorphism in Xanthomonas albilineans strains originating from 11 geographical locations, revealed by two DNA probes. **Letters in applied microbiology**, v. 30, n. 4, p. 287–93, 2000.

34 ARMAS, R. DE; MARTINEZ, M.; et al. Free and Conjugated Polyamines and Phenols in Raw and Alkaline-Clarified Sugarcane Juices. **Journal of Agricultural and Food Chemistry**, v. 47, n. 8, p. 3086–3092, 1999.

35 LOPES, S.A.; DAMANN, K.E.; et al. Infectivity Titration for Assessing Resistance to Leaf Scald Among Sugarcane Cultivars. **Plant Disease**, v. 85, n. 6, p. 592–596, 2001.

36 DABBAS, K.M.; FERRO, M.I.T.; et al. Genes diferencialmente expressos em cana-de-açúcar inoculada com Xanthomonas albilineans, o agente causal da escaldadura da folha. **Summa Phytopathologica**, v. 32, n. 4, p. 328–338, 2006.

37 SILVA, M. DE S. E; BEDENDO, I.P.; et al. Caracterização molecular e patogênica de isolados de Xanthomonas albilineans (Ashby) Dowson, agente causal da escaldadura das folhas da cana-de-açúcar. **Summa Phytopathologica**, v. 33, n. 4, p. 341–347, 2007.

38 GUIDO, R.V.C.; ANDRICOPULO, A.D.; et al. Planejamento de fármacos, biotecnologia e química medicinal: aplicações em doenças infecciosas. **Estudos Avançados**, v. 24, n. 70, p. 81–98, 2010.

39 PIERETTI, I.; ROYER, M.; 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.

40 BIRCH, R.G. Xanthomonas albilineans and the antipathogenesis approach to disease control. **Molecular plant pathology**, v. 2, n. 1, p. 1–11, 2001.

41 BIRCH, R.G., PATIL, S.S. Antibiotic and process for the production thereof. 1985.

42 HASHIMI, S.M.; WALL, M.K.; et al. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. **Antimicrobial agents and chemotherapy**, v. 51, n. 1, p. 181–7, 2007.

43 ZHANG, L.; XU, J.; et al. Engineered detoxification confers resistance against a pathogenic bacterium. **Nature biotechnology**, v. 17, n. 10, p. 1021–4, 1999.

44 BIRCH, R.G.; PATIL, S.S. Correlation between albicidin production and chlorosis induction by Xanthomonas albilineans, the sugarcane leaf scald pathogen. **Physiological and Molecular Plant Pathology**, v. 30, n. 2, p. 199–206, 1987.

45 ROTT, P.C.P.C.; COSTET, L.; et al. At least two separate gene clusters are involved in albicidin production by Xanthomonas albilineans. **J. Bacteriol.**, v. 178, n. 15, p. 4590–4596, 1996.

46 WALL, M.K.; BIRCH, R.G. Genes for albicidin biosynthesis and resistance span at least 69 kb in the genome of Xanthomonas albilineans. **Letters in Applied Microbiology**, v. 24, n. 4, p. 256–260, 1997.

47 HUANG, G.; ZHANG, L.; et al. A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in Xanthomonas albilineans. **Microbiology (Reading, England)**, v. 147, n. Pt 3, p. 631–42, 2001.

48 HUANG, G.; ZHANG, L.; et al. Albicidin antibiotic and phytotoxin biosynthesis in Xanthomonas albilineans requires a phosphopantetheinyl transferase gene. **Gene**, v. 258, n. 1-2, p. 193–9, 2000.

49 COPP, J.N.; NEILAN, B.A. The Phosphopantetheinyl Transferase Superfamily : Phylogenetic Analysis and Functional Implications in Cyanobacteria †. v. 72, n. 4, p. 2298–2305, 2006.

50 WALSH, C.T.; GEHRING, A M.; et al. Post-translational modification of polyketide and nonribosomal peptide synthases. **Current opinion in chemical biology**, v. 1, n. 3, p. 309–15, 1997.

51 CANE, D.E.; WALSH, C.T.; et al. Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. **Science**, v. 282, n. 5386, p. 63–68, 1998.

52 NEUROTIKER . Structure of coenzyme A (labeled) 2007. Disponível em: < http://en.m.wikipedia.org/wiki/File:Coenzym_A_beschriftet.svg>. Acesso em 18 jun. 2013.

53 SCAPIN, G. Structural biology and drug discovery. **Current pharmaceutical design**, v. 12, n. 17, p. 2087–97, 2006.

54 BERNSTEIN, F.C.; KOETZLE, T.F.; et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. **Journal of molecular biology**, v. 112, n. 3, p. 535–42, 1977.

55 THE UNIPROT CONSORTIUM Reorganizing the protein space at the Universal Protein Resource (UniProt). **Nucleic acids research**, v. 40, n. Database issue, p. D71–5, 2012.

56 YEATES, T.O.; FAM, B.C. Protein crystals and their evil twins. **Structure**, v. 7, n. 2, p. R25–R29, 1999.

57 BURLEY, S.K.; ALMO, S.C.; et al. Structural genomics: beyond the human genome project. **Nature genetics**, v. 23, n. 2, p. 151–7, 1999.

58 BLOW, D.M.; CHAYEN, N.E.; et al. Control of nucleation of protein crystals. **Protein science : a publication of the Protein Society**, v. 3, n. 10, p. 1638–43, 1994.

59 BAKER, D.; SALI, A. Protein structure prediction and structural genomics. **Science (New York, N.Y.)**, v. 294, n. 5540, p. 93–6, 2001.

60 LESK, A.M.; CHOTHIA, C.H. The Response of Protein Structures to Amino-Acid Sequence Changes. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 317, n. 1540, p. 345–356, 1986.

61 ESWAR, N.; WEBB, B.; et al. Comparative protein structure modeling using MODELLER. **Current protocols in protein science / editorial board, John E. Coligan ... [et al.]**, v. Chapter 2, p. Unit 2.9, 2007.

62 JOHNSON, M.S.; SRINIVASAN, N.; et al. Knowledge-based protein modeling. **Critical reviews in biochemistry and molecular biology**, v. 29, n. 1, p. 1–68, 1994.

63 HAVEL, T.F.; SNOW, M.E. A new method for building protein conformations from sequence alignments with homologues of known structure. **Journal of molecular biology**, v. 217, n. 1, p. 1–7, 1991.

64 SALI, A.; BLUNDELL, T.L. Comparative protein modelling by satisfaction of spatial restraints. **Journal of molecular biology**, v. 234, n. 3, p. 779–815, 1993.

65 MARTÍ-RENOM, M.A.; STUART, A.C.; et al. Comparative protein structure modeling of genes and genomes. **Annual review of biophysics and biomolecular structure**, v. 29, p. 291–325, 2000.

66 FISER, A.; DO, R.K.; et al. Modeling of loops in protein structures. **Protein science : a publication of the Protein Society**, v. 9, n. 9, p. 1753–73, 2000.

65 Karadaghi, S. Al Introduction to Protein Crystallization: Basics ideas. Disponível em: <http://www.proteinstructures.com/Experimental/Experimental/protein-crystallization.html>. Acesso em: 22 jun. 2013.

68 ORENGO, C.A.; JONES, D.T.; et al. Protein superfamilies and domain superfolds. **Nature**, v. 372, n. 6507, p. 631–4, 1994.

69 WANG, Z.X. How many fold types of protein are there in nature? **Proteins**, v. 26, n. 2, p. 186–91, 1996.

70 ZHANG, C.T. Relations of the numbers of protein sequences, families and folds. **Protein engineering**, v. 10, n. 7, p. 757–61, 1997.

71 XIANG, Z. Advances in homology protein structure modeling. **Current protein & peptide science**, v. 7, n. 3, p. 217–27, 2006.

72 A spiral model of software development and enhancement. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=59&tag=1>. Acesso em: 22 jun. 2013.

73 CRINNION, J. Evolutionary Systems Development: A Practical Guide to the Use of Prototyping within a Structured Systems Methodology. 1992.

74 DAVIS, A.M. Operational prototyping: a new development approach. **IEEE Software**, v. 9, n. 5, p. 70–78, 1992.

75 SOMMERVILLE, I. Software engineering. Boston: Addison Wesley, 2007. 840 p.

76 HEMMENDINGER, D. Computer programming language. **Encyclopedia Britannica**, 2012. Disponível em: < http://global.britannica.com/EBchecked/topic/130670/computer-programming-language>. Acesso em: 18 jun. 2013.

77 ROSSUM, G. VAN; DRAKE, F.L. Python Reference Manual. 2006. Disponível em: < http://docs.python.org/2/reference/index>. Acesso em: 18 jun. 2013.

78 SCHRÖEDINGER. The PyMOL Molecular Graphics System. Disponível em: <http://www.pymol.org/pymol>. Acesso em: 22 jun. 2013.

79 JANSEN, B.J. The graphical user interface: an introduction. 1998. Disponível em: <http://faculty.ist.psu.edu/jjansen/academic/pubs/chi.html>. Acesso em: 14 jun. 2013.

80 MARTINEZ, W.L. Graphical user interfaces. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 3, n. 2, p. 119–133, 2011.

81 PYGTK TEAM PyGTK: GTK+ for Python. 2009. Disponível em: <http://www.pygtk.org/downloads.html>. Acesso em: 22 jun. 2013.

82 HUNTER, J.D. Matplotlib: A 2D graphics environment. **Computing In Science & Engineering**, v. 9, n. 3, p. 90–95, 2007.

83 KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. **Drug discovery today**, v. 11, n. 13-14, p. 580–94, 2006.

84 GUIDO, R.V.C.; OLIVA, G.; et al. Virtual screening and its integration with modern drug design technologies. **Current medicinal chemistry**, v. 15, n. 1, p. 37–46, 2008.

85 GUIDO, R.V.C.; OLIVA, G.; et al. Modern drug discovery technologies: opportunities and challenges in lead discovery. **Combinatorial chemistry & high throughput screening**, v. 14, n. 10, p. 830–9, 2011.

86 HARMS, M.J. PDB-tools 0.1.4. 2012. Disponível em: <http://code.google.com/p/pdb-tools/>. Acesso em: 22 jun. 2013.

87 ALTSCHUL, S.F.; MADDEN, T.L.; et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–402, 1997.

88 HILDEBRAND, A.; REMMERT, M.; et al. Fast and accurate automatic structure prediction with HHpred. **Proteins**, v. 77 Suppl 9, p. 128–32, 2009.

89 LARKIN, M.A.; BLACKSHIELDS, G.; et al. Clustal W and Clustal X version 2.0. **Bioinformatics (Oxford, England)**, v. 23, n. 21, p. 2947–8, 2007.

90 GOUJON, M.; MCWILLIAM, H.; et al. A new bioinformatics analysis tools framework at EMBL-EBI. **Nucleic acids research**, v. 38, n. Web Server issue, p. W695–9, 2010.

91 JONES, D.T. Protein secondary structure prediction based on position-specific scoring matrices. **Journal of molecular biology**, v. 292, n. 2, p. 195–202, 1999.

92 ROY, A.; KUCUKURAL, A.; et al. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature protocols**, v. 5, n. 4, p. 725–38, 2010.

93 ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC bioinformatics**, v. 9, n. 1, p. 40, 2008.

94 SHEN, M.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein science : a publication of the Protein Society**, v. 15, n. 11, p. 2507–24, 2006.

95 UCLA The Structure Analysis and Verification Server - SAVES. 2012. Disponível em: <http://nihserver.mbi.ucla.edu/SAVES/>. Acesso em: 22 jun. 2013.

96 WIEDERSTEIN, M.; SIPPL, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic acids research**, v. 35, n. Web Server issue, p. W407–10, 2007.

97 SCHRÖDINGER, I. Glide. 2009. Disponível em: <http://www.schrodinger.com/productpage/14/5/>. Acesso em: 22 jun. 2013.

98 JAIN, A.N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. **Journal of medicinal chemistry**, v. 46, n. 4, p. 499–511, 2003.

99 IRWIN, J.J.; STERLING, T.; et al. ZINC: A Free Tool to Discover Chemistry for Biology. **Journal of chemical information and modeling**, 2012.

100 TICE, C.M. Selecting the right compounds for screening: does Lipinski's Rule of 5 for pharmaceuticals apply to agrochemicals? **Pest management science**, v. 57, n. 1, p. 3–16, 2001.

101 KUNTAL, B.K.; APAROY, P.; et al. EasyModeller: A graphical interface to MODELLER. **BMC research notes**, v. 3, p. 226, 2010.

102 BRAMUCCI, E.; PAIARDINI, A.; et al. PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. **BMC bioinformatics**, v. 13 Suppl 4, n. Suppl 4, p. S2, 2012.

103 WATERHOUSE, A.M.; PROCTER, J.B.; et al. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. **Bioinformatics (Oxford, England)**, v. 25, n. 9, p. 1189–91, 2009.

104 WILKINS, M.R.; GASTEIGER, E.; et al. Protein identification and analysis tools in the ExPASy server. **Methods in molecular biology (Clifton, N.J.)**, v. 112, p. 531–52, 1999.

105 SCHULTZ, J.; MILPETZ, F.; et al. Colloquium Paper: SMART, a simple modular architecture research tool: Identification of signaling domains. **Proceedings of the National Academy of Sciences**, v. 95, n. 11, p. 5857–5864, 1998.

106 LETUNIC, I.; DOERKS, T.; et al. SMART 7: recent updates to the protein domain annotation resource. **Nucleic acids research**, v. 40, n. Database issue, p. D302–5, 2012.

107 LAMBALOT, R.H.; GEHRING, A.M.; et al. A new enzyme superfamily - the phosphopantetheinyl transferases. **Chemistry & biology**, v. 3, n. 11, p. 923–36, 1996.

108 MARIANI, V.; KIEFER, F.; et al. Assessment of template based protein structure predictions in CASP9. **Proteins**, v. 79 Suppl 1, p. 37–58, 2011.

109 REUTER, K.; MOFID, M.R.; et al. Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily. **The EMBO journal**, v. 18, n. 23, p. 6823–31, 1999.

110 ANDRICOPULO, A.D.; SALUM, L.B.; et al. Structure-based drug design strategies in medicinal chemistry. **Current topics in medicinal chemistry**, v. 9, n. 9, p. 771–90, 2009.

111 ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. **Proteins**, v. 57, n. 4, p. 702–10, 2004.

112 LOVELL, S.C.; DAVIS, I.W.; et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. **Proteins**, v. 50, n. 3, p. 437–50, 2003.

113 HEINIG, M.; FRISHMAN, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. **Nucleic acids research**, v. 32, n. Web Server issue, p. W500–2, 2004.

114 TROTT, O.; OLSON, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of computational chemistry**, v. 31, n. 2, p. 455–61, 2010.

115 LIPINSKI, C.A.; LOMBARDO, F.; et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced drug delivery reviews**, v. 46, n. 1-3, p. 3–26, 2001.

116 CLARKE, E.D.; DELANEY, J.S. Physical and Molecular Properties of Agrochemicals: An Analysis of Screen Inputs, Hits, Leads, and Products. **CHIMIA International Journal for Chemistry**, v. 57, n. 11, p. 731–734, 2003.

117 JORGENSEN, W.L.; MAXWELL, D.S.; et al. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. **Journal of the American Chemical Society**, v. 118, n. 45, p. 11225–11236, 1996.

118 KAMINSKI, G.A.; FRIESNER, R.A.; et al. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides †. **The Journal of Physical Chemistry B**, v. 105, n. 28, p. 6474–6487, 2001.

119 ELDRIDGE, M.D.; MURRAY, C.W.; et al. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. **Journal of computer-aided molecular design**, v. 11, n. 5, p. 425–45, 1997.

120 KITCHEN, D.B.; DECORNEZ, H.; et al. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature reviews. Drug discovery**, v. 3, n. 11, p. 935–49, 2004.

121 TAYLOR, R.D.; JEWSBURY, P.J.; et al. A review of protein-small molecule docking methods. **Journal of computer-aided molecular design**, v. 16, n. 3, p. 151–66, 2002.

122 WELCH, W.; RUPPERT, J.; et al. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. **Chemistry & biology**, v. 3, n. 6, p. 449–62, 1996.

123 CLARK, R.D.; STRIZHEV, A.; et al. Consensus scoring for ligand/protein interactions. **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 281–295, 2002.

124 CHARIFSON, P.S.; CORKERY, J.J.; et al. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. **Journal of Medicinal Chemistry**, v. 42, n. 25, p. 5100–5109, 1999.

125 DYE, D.W. The Inadequacy of the Usual Determinative Tests for the Identification of Xanthomonas Spp. **New Zealand Journal of Science**, v. 5, p. 393–416, 1962.

126 DAVIS, M.J. Evaluation of Selective Media and Immunoassays for Detection of Xanthomonas albilineans, Causal Agent of Sugarcane Leaf Scald Disease. **Plant Disease**, v. 78, n. 1, p. 78, 1994.

127 BODDEY, R.M.; URQUIAGA, S.; et al. Biological nitrogen fixation associated with sugar cane. **Plant and Soil**, v. 137, n. 1, p. 111–117, 1991.

128 BODDEY, R.M.; URQUIAGA, S.; et al. Endophytic nitrogen fixation in sugarcane: present knowledge and future applications. **Plant and Soil**, v. 252, n. 1, p. 139–149, 2003.

129 MIRZA, M.; AHMAD, W.; et al. Isolation, partial characterization, and the effect of plant growth-promoting bacteria (PGPB) on micro-propagated sugarcane in vitro. **Plant and Soil**, p. 47–54, 2001.

130 JAUFEERALLY-FAKIM, Y.; DOOKUN, A. Extraction of high quality DNA from polysaccharides-secreting xanthomonads. **Sci Technol**, v. 6, p. 33–40, 2000.

131 PAN, Y.-B.; GRISHAM, M.P.; et al. Development of Polymerase Chain Reaction Primers Highly Specific for Xanthomonas albilineans , the Causal Bacterium of Sugarcane Leaf Scald Disease. **Plant Disease**, v. 83, n. 3, p. 218–222, 1999.

132 ASLANIDIS, C.; JONG, P.J. DE Ligation-independent cloning of PCR products (LIC-PCR). **Nucleic acids research**, v. 18, n. 20, p. 6069–74, 1990.

133 INOUE, H.; NOJIMA, H.; et al. High efficiency transformation of Escherichia coli with plasmids. **Gene**, v. 96, n. 1, p. 23–8, 1990.

134 GAGNEVIN, L.; LEACH, J.E.; et al. Genomic Variability of the Xanthomonas Pathovar mangiferaeindicae, Agent of Mango Bacterial Black Spot. **Applied and environmental microbiology**, v. 63, n. 1, p. 246–53, 1997.

135 KERKOUD, M.; MANCEAU, C.; et al. Rapid Diagnosis of Pseudomonas syringae pv. papulans, the Causal Agent of Blister Spot of Apple, by Polymerase Chain Reaction Using Specifically Designed hrpL Gene Primers. **Phytopathology**, v. 92, n. 10, p. 1077–83, 2002.

136 LOUWS, F.; RADEMAKER, J.; et al. THE THREE DS OF PCR-BASED GENOMIC ANALYSIS OF PHYTOBACTERIA: Diversity, Detection, and Disease Diagnosis. **Annual review of phytopathology**, v. 37, p. 81–125, 1999.

137 HENSON, J.M.; FRENCH, R.C. **The polymerase chain reaction and plant disease diagnosis**. 1993. Disponível em: <http://digitalcommons.unl.edu/plantpathpapers/2/ >. Acesso em: 22 jun. 2013.

138 VINCENTELLI, R.; CIMINO, A.; et al. High-throughput protein expression screening and purification in Escherichia coli. **Methods (San Diego, Calif.)**, v. 55, n. 1, p. 65–72, 2011.

139 BIRD, L.E. High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. **Methods (San Diego, Calif.)**, v. 55, n. 1, p. 29–37, 2011.

140 BACHEGA, J.F.R.; TIMMERS, L.F.S.M.; et al. GTKDynamo: A PyMOL plug-in for QC/MM hybrid potential simulations. **Journal of Computational Chemistry**, v. 00000, p. n/a–n/a, 2013.

141 CHAUDHURY, S.; LYSKOV, S.; et al. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. **Bioinformatics (Oxford, England)**, v. 26, n. 5, p. 689–91, 2010.